



## RESEARCH ARTICLE

# Surface wind over Europe: Data and variability

Cristina Rojas-Labanda<sup>1</sup>  | Fidel González-Rouco<sup>1</sup> |  
 Elena García-Bustamante<sup>2</sup>  | Jorge Navarro<sup>2</sup> | Etor E. Lucio-Eceiza<sup>3,4</sup> |  
 Gerard Van der Schrier<sup>5</sup> | Frank Kaspar<sup>6</sup> 

<sup>1</sup>Universidad Complutense de Madrid, and Instituto de Geociencias (UCM-CSIC), Madrid, Spain

<sup>2</sup>Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

<sup>3</sup>Free University Berlin, Institute of Meteorology, Berlin, Germany

<sup>4</sup>DKRZ German Climate Computing Center, Hamburg, Germany

<sup>5</sup>Royal Netherlands Meteorological Institute, the Netherlands

<sup>6</sup>Deutscher Wetterdienst, Germany

**Correspondence**

Cristina Rojas-Labanda, Universidad Complutense de Madrid, and Instituto de Geociencias (UCM-CSIC), Madrid, Spain.  
 Email: [crojaslabanda@gmail.com](mailto:crojaslabanda@gmail.com)

**Abstract**

This work improves the characterization and knowledge of the surface wind climatology over Europe with the development of an observational database with unprecedented quality control (QC), the European Surface Wind Observational database (EuSWiO). EuSWiO includes more than 3,829 stations with sub-daily resolution for wind speed and direction, with a number of sites spanning the period of 1880–2017, a few hundred time series starting in the 1930s and relatively good spatial coverage since the 1970s. The creation of EuSWiO entails the merging of eight different data sets and its submission to a common QC. About 5% of the total observations were flagged, correcting a great part of the extreme and unrealistic values, which have a discernible impact on the statistics of the database. The daily wind variability was characterized by means of a classification technique, identifying 11 independent subregions with distinct temporal wind variability over the 2000–2015 period. Significant decreases in the wind speed during this period are found in five regions, whereas two regions show increases. Most regions allow for extending the analysis to earlier decades. Caution in interpreting long-term trends is needed as wind speed data have not been homogenized. Nevertheless, decreases in the wind speed since the 1980s can be noticed in most of the regions. This work contributes to a deeper understanding of the temporal and spatial surface wind variability in Europe. It will allow from meteorological to climate and climate change studies, including potential applications to the analyses of extreme events, wind power assessments or the evaluation of reanalysis or model-data comparison exercises at continental scales.

**KEYWORDS**

observational database, quality control, regionalization, surface wind

## 1 | INTRODUCTION

Near-surface wind (simply referred to as wind or surface wind hereafter) has countless sociological, political,

economic and environmental implications. It is a dominant factor affecting, among some others, the transport and dispersion of pollutants and pests, evapotranspiration processes and crop growth (Cleugh *et al.*, 1998;

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

Marsac and Le Blanc, 1998; Darby, 2005; Farquhar and Roderick, 2007; McVicar *et al.*, 2012). Therefore, an improved understanding of wind variability at different timescales is key, not only because it offers important scientific challenges but also because it is policy relevant for better planning and adaptation. Such progress in understanding wind variability is gained from the analysis of model simulations and observational products.

In recent years, modelling capabilities have allowed the analyses of wind speed and wind power projections and the impact of climate change on these variables have increased (Tobin *et al.*, 2015, 2016, 2018; Reyers *et al.*, 2016; Carvalho *et al.*, 2017; Wohland *et al.*, 2017; Karnauskas *et al.*, 2018). However, while historical and scenario simulations are informative about the responses of atmosphere circulation to external forcings (Eyring *et al.*, 2016), they will not grasp the actual trajectory of the real climate system which is dependent, to a large extent, on internal variability. To assess the actual evolution of wind fields and to evaluate climate models, observational products are needed. Reanalysis are hybrid products that combine a so-called frozen state-of-the-art numerical model with the assimilation of past observations from several sources (Fujiwara *et al.*, 2017) to produce uniform in time and space data sets of multidecadal (e.g., ERA5; Hersbach *et al.*, 2020) and even centennial (e.g., 20CRc2c; Compo *et al.*, 2011) length. The assimilation of observations drives the model close to the actual variability of the variable of interest. In the case of wind-related variables, this enables producing wind resource assessments, typically after downscaling the large-scale information to regional or local-scale products through using meso- and micro-scale models (Dörenkämper *et al.*, 2020; Hahmann *et al.*, 2020), or directly using regional reanalyses (Kaspar *et al.*, 2020). Reanalysis products provide large inter-model consistency and virtually an observational quality in areas of larger observational density, and still offer useful information as a means of physical interpolation in areas of data scarcity (Brands *et al.*, 2012; Ramón *et al.*, 2019). However, for the case of surface wind, owed to the different model physics, assimilation networks and observational uncertainties, different reanalysis products show large differences in regional multidecadal trends (Chen *et al.*, 2013; Hartmann *et al.*, 2013; Olauson, 2018; Ramón *et al.*, 2019; Lucio-Eceiza *et al.*, 2020).

An alternative to reanalyses are satellite based and near-surface observations. Satellite observations (Zhang *et al.*, 2006; Atlas *et al.*, 2011) allow for global coverage over ocean basins. They offer high temporal and reasonably high spatial resolution, although the time span is limited and they are sensitive to the use of different sensors, calibration and measurement methods as with other satellite observations (Hartmann *et al.*, 2013; Kent *et al.*, 2013). In turn, onsite near-surface observations are provided at wind

tower and meteorological station data. Such observations offer the most valuable and trustful information of near-surface local wind speed, as the measurements reflect all the cascade of interactions from the large-scale down to the particularities of a given site (Dörenkämper *et al.*, 2020). Wind towers are particularly relevant for the energy sector as they provide information about the wind resource at hub height. However, their time and spatial coverage is limited, particularly owed to restrictions from the private sector (Standen *et al.*, 2017; Ramón *et al.*, 2019). Surface site observations are comparatively more abundant (Staffell and Pfenninger, 2016; Harris *et al.*, 2020), but still, unlike temperature or precipitation, long time series of surface wind observations are difficult to find (Valero *et al.*, 1996; Molina *et al.*, 2021) and the limited spatial density of stations in some areas, particularly in those of complex orography, hampers a realistic representation of wind variability (Jiménez *et al.*, 2008).

Additionally, as in the case of reanalysis products, both satellite and surface based data sets present large discrepancies in different regions of the globe in the representation of multidecadal trends (Hartmann *et al.*, 2013). Discrepancies occur at regional and continental scales both within products of the same nature, like for different satellite (Zhang *et al.*, 2006; Atlas *et al.*, 2011) or surface wind (Pryor *et al.*, 2009; Smith *et al.*, 2011; Minola *et al.*, 2016) data sets and also resulting in differences among different reanalysis, satellite or surface wind data sets. Indeed, with some discrepancies across data sets (McVicar *et al.*, 2012; Azorin-Molina *et al.*, 2014, 2016) positive trends are found over some ocean basins (Wentz *et al.*, 2007; Tokinaga and Xie, 2011; Young *et al.*, 2011) some coastal areas (Pinard, 2007) and high-latitude regions (McVicar *et al.*, 2012; Minola *et al.*, 2016). However, multidecadal trends over land areas have been a matter of debate, often related to the detection of diminishing wind speed (Roderick *et al.*, 2007; Vautard *et al.*, 2010; McVicar *et al.*, 2012), although recently, global trends have been detected to reverse to positive after 2010 (Zeng *et al.*, 2019). This so-called stilling effect has been discussed to be potentially related to different problems or mechanisms: either internal variability and thus related to large-scale circulation changes (Jiang *et al.*, 2010; Torralba *et al.*, 2017; Wu *et al.*, 2018; Lucio-Eceiza *et al.*, 2019, 2020), changes in the land use or land cover (Vautard *et al.*, 2010; Guo *et al.*, 2011; Bichet *et al.*, 2012; Wever, 2012) or problems in data quality (Zahradníček *et al.*, 2019).

Therefore, the availability of either observational or hybrid model-observational products that consistently describe inter-annual, decadal and multi-decadal timescales is still a challenge for many regions. Specifically, the European region is an interesting case that gathers some of the longest time series of surface wind. Many different national services in Europe have been monitoring

surface wind, as other meteorological variables, with relatively good density, incorporating often different measurement procedures, data management or quality control (QC) approaches (Jiménez *et al.*, 2008; Smith *et al.*, 2011; Kaspar *et al.*, 2013; Lorente-Plazas *et al.*, 2015; Dunn *et al.*, 2016).

This study is specifically oriented to the development of the European Surface Wind Observational data set (EuSWiO) after integrating information of wind speed and direction from several global-scale data sets, as well as from a few national meteorological services (eight different sources in total). The resulting data set has benefited from the spatial and temporal coverage and resolution of the contributing sources.

The information provided by the various data sources may have been tested for quality in different ways (Lott, 2004; Dunn *et al.*, 2012, 2016; DWD, CDC, 2018). A number of QC routines has been applied here to account for errors derived from data management and measurement issues based on the routines developed in Lucio-Eceiza *et al.* (2018, 2018), who evaluated 526 sites over northeastern North America. Here those routines are adapted to about 7.5 times more data spanning mostly since the 1970s to present, including a smaller group (ca. 2%) reaching back to the first decades of the century.

The results of the QC will provide information about typical problems in surface wind over Europe and the spatial and temporal distribution of their occurrence. The QC analyses included herein do not produce an error free data set because QC tests can be subjected to future improvements or the routines can be improved or expanded with new tests in the future. These routines address different types of errors related to data management and other measurement-related pathologies, both

in wind speed and direction. Additionally, tests included target long-term changes for wind direction, but not for wind speed. Therefore, any evaluation of long-term trends in wind velocity must be done with care and future versions of the data set will consider implementing homogenization corrections.

Section 2 provides an account of the QC methods, diagnoses on the impacts of the QC and description of the data set climatology used hereafter. The results of the QC are described in Section 3, including the information about the spatial and temporal dependencies of the errors detected, as well as illustrative examples of their behaviour and the amount of records affected. Finally, a basic assessment of the climatology before and after the application of the QC and a first description of multidecadal trends is also provided, allowing for an initial assessment of multidecadal variability.

## 2 | DATA AND METHODS

### 2.1 | Compilation and merging

EuSWiO blends the information of global and pan-European wide ( $-30^{\circ}$ – $45^{\circ}$ E  $\times$   $25^{\circ}$ – $74^{\circ}$ N) data sets that are publicly available (see Table 1): the National Center for Atmospheric Research (NCAR), the Global Telecommunications System (GTS) and the European Climate Assessment and Dataset (ECA&D). It integrates additionally data, provided on a basis of opportunity, from some national meteorological services: the Deutscher Wetterdienst (DWD), the Royal Netherlands Meteorological Institute (KNMI) and the Swedish Meteorological and Hydrological Institute (SMHI). Only records

**TABLE 1** Original data sources composing EuSWiO (acronyms in column 1); Number of time series available from each source (column 2), temporal resolution (column 3), time span (column 4), and additional bibliographic and web information (column 5). NCAR1, NCAR2 and NCAR3 refer to the NCAR data sets included: ds461.0, ds464.0, and ds463.3, respectively

Data set Acronym	Time Series	Temporal Resolution	Time Span	Additional information
NCAR1	3,693	1, 3, 6 h	1999–2017	ds461.0 (NCEP, 2004); <a href="https://doi.org/10.5065/4F4P-E398">https://doi.org/10.5065/4F4P-E398</a>
NCAR2	3,987	1, 3, 6 h	1975–2007	ds464.0 (NCEP, 1980); <a href="http://rda.ucar.edu/datasets/ds464.0">http://rda.ucar.edu/datasets/ds464.0</a>
NCAR3	185	1, 3, 6 h	1929–2011	ds463.3 (NCEP, 2005); <a href="http://rda.ucar.edu/datasets/ds463.3">http://rda.ucar.edu/datasets/ds463.3</a>
				See also Smith <i>et al.</i> (2011)
GTS	3,694	6 h	1980–2017	WMO (2015)
ECA&D	405	1 d	1877–2017	Klein Tank <i>et al.</i> (2002)
DWD	337	1 h	1893–2017	DWD, CDC (2018)
KNMI	58	1 h	1950–2017	<a href="https://www.knmidata.nl/data-services/knmi-producten-overzicht">https://www.knmidata.nl/data-services/knmi-producten-overzicht</a>
SMHI	521	1, 3, 6 h	1940–2017	<a href="https://www.smhi.se/data/oppna-data">https://www.smhi.se/data/oppna-data</a>
EuSWiO	12,880	1, 3, 6 h, 1 d	1929–2017	

with information for both (nominally) 10 m wind speed and direction were collected, adding up to a total of 15,970 time series for each wind variable.

The incorporation of data from different sources made it necessary to perform a standardization of formats, dates, missing data and other codes. One example is the criteria to define calm and true north wind directions. A value of  $0^\circ$  was assigned for calm cases, differentiating it from the true north case, when wind speed is identified with  $360^\circ$ . Typographical, chronological errors and repeated dates were also addressed while compiling each individual data set. Hourly time series were created for the entire data set, using a missing code when the resolution was lower. Recording time was in UTC and units for wind speed (direction) in  $\text{m}\cdot\text{s}^{-1}$  (degrees). Data from February 2001 to July 2002 from the ds461.0 data set from NCAR were transformed from knots to  $\text{m}\cdot\text{s}^{-1}$  following Schuster (2003).

An initial screening of time series was made discarding cases in which the total time span of the records was  $<1$  year or for which the metadata indicated changes in their position larger than 1 km. The resulting data set comprised 12,880 time series for each wind speed and direction. Table 1 and Figure 1 describe the availability of information at this stage of the QC. Table 1 indicates the source of data, the amount of time series collected, their resolution and time span. The largest contribution is from the NCAR data sets, followed by the GTS. Except for the daily resolution of ECA&D, all incoming data had intra-daily resolution. The data cover not only the European domain but also beyond this region, for example, to the east of Europe and south of the Mediterranean. Figure 1a illustrates the heterogeneous density of time series, larger over central Europe (mostly a contribution of DWD and KNMI; Figure 1b) and lower in eastern Europe and northern Africa. Time series from different sources may have a very different time span (Figure 1c), overall showing many cases reaching back to the 1970s and some back to the first half of the 20th century.

As shown in Figure 1b, several sources can provide information from coincident sites. This information may not be identical, with versions of the same site having different temporal resolution, time span or different previous quality assessments may have been applied to each. We follow a similar approach to Rennie et al. (2014) to merge time series corresponding to the same site. To identify records that belonged to the same site, each time series was compared to all those within a distance of 10 km using root mean squared differences (rmsd) and considering the data at the lowest common time resolution. Minimal rmsd values helped to identify time series of the same site. Comparisons were also made, regardless of their position, if available metadata showed common

codes (e.g., WMO). For the merging process, the time segments provided by different institutions followed a hierarchy. For example, the segments coming from national institutes were considered over any other source, as they offered the highest resolution (Table 1) and were expected to have a better quality as well. Data that offered no improvement in a resolution or time extent over other existing records were discarded. Because of its daily resolution (Table 1), ECA&D received the lowest priority, keeping its information when no other source provided data for a given time interval.

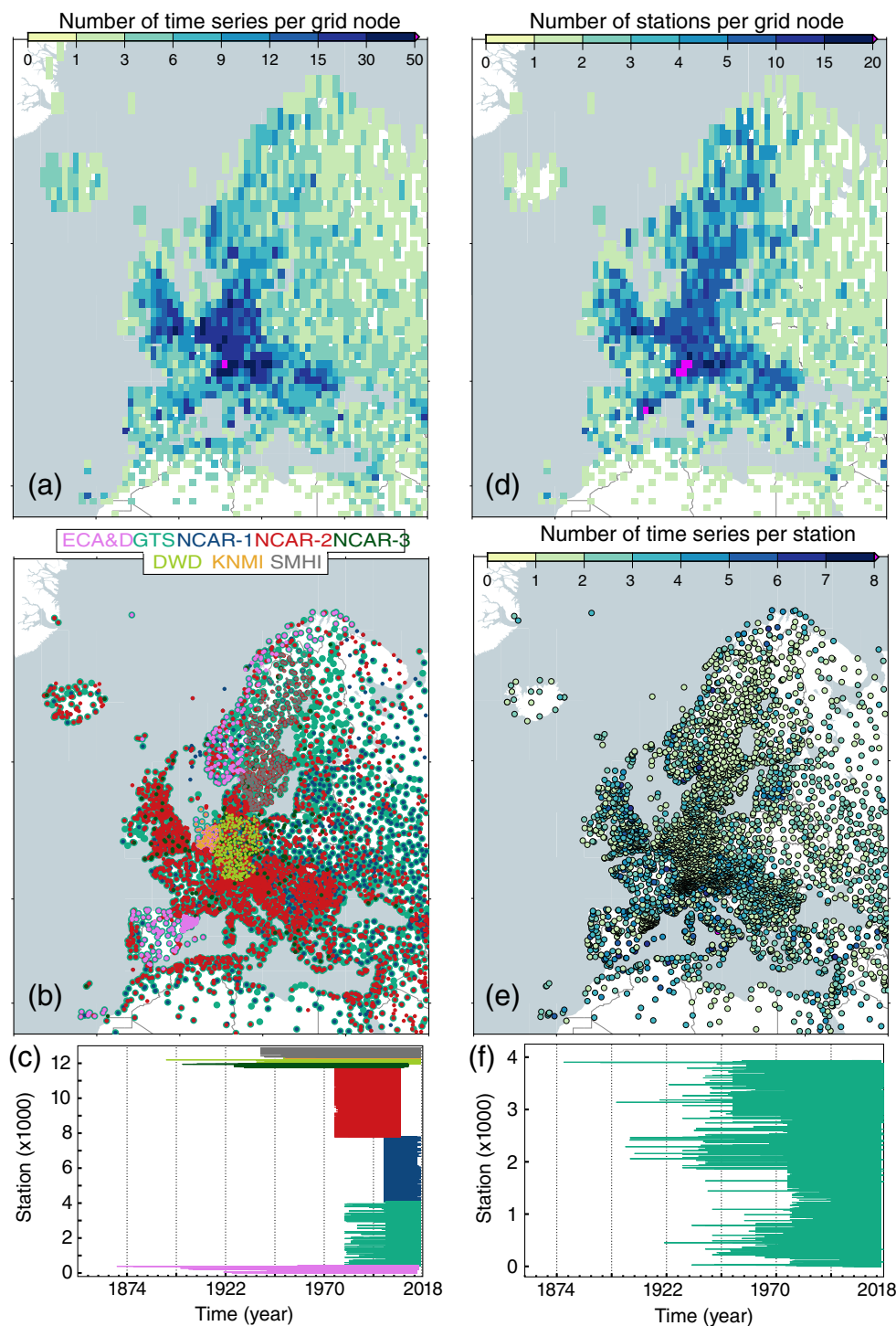
The merging procedure led eventually to 3,936 final time series corresponding each to a different site. Such station information was typically the result of integrating between 1 and 11 time series (Figure 1d,e), with higher numbers for central and northern Europe, where more sources of data overlap (Figure 1b). The resulting coverage extends over Europe and borderlands, although the density of station sampling is, as discussed earlier, heterogeneous. The time span of data focuses mostly on the last two decades with 70.1% of the records starting after 1990, 21.4% of the records start within the period of 1970 to 1990, 6.4% between 1950 and 1970, and with 2.1% of the records reaching back the first half of the 20th century.

It is likely that some of the original data had been previously quality controlled to some extent. However, *metadata* reporting on this process were available to a limited extent. Therefore, the same QC routines have been uniformly applied to all time series.

## 2.2 | EuSWiO: Quality control

The QC applied follows the procedures in Lucio-Eceiza et al. (2018, 2018), grouped into five steps. The first steps deal with issues often related to data storage and management (Lucio-Eceiza et al., 2018), whereas the latter ones deal with measurement or instrumental problems, like unreliable performance, calibration, siting and changes in exposure of the surrounding environment, among others (Lucio-Eceiza et al., 2018). For the sake of conciseness only a brief synthesis of each step of the methodology is included here. Section 3.1 shows the results to the corresponding steps presented below.

*Step 1. Intra- and inter-site duplications* are duplicated periods of data within the same station or between two different station reports, respectively, which may appear during any step of data transcription (Kunkel et al., 1998; Durre et al., 2010; Lawrimore et al., 2011; Dunn et al., 2016; Lucio-Eceiza et al., 2018). Thus, chains of repeated non-missing observations were inspected. Transcription errors are prone to happen at identical or similar dates, helping as an additional condition to discern



**FIGURE 1** Distribution of time series taking part in the compilation of EuSWiO: (a) spatial density in a  $1^\circ \times 1^\circ$  lat  $\times$  lon grid; (b) distribution of time series according to the data providing institution (see labels); and (c) the time span of the series provided by each institution (colours correspond to the colour of labels in b). (d-e) Description of the time series after the merging procedure to achieve one time series per site: (d) as (a) but for the final number of stations; (e) number of time series contributing to each of the sites; and (f) time span of each merged site time series [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

erroneous periods. Chains of constant periods were targeted in step 3 and thus, not considered herein. Neither were chains of repeated data shorter than 15 observations as they could be due to natural causes (Jiménez *et al.*, 2010; Lucio-Eceiza *et al.*, 2018).

When a duplication was found, the evaluation of whether one of the two time segments was correctly placed was done via correlations with neighbour stations, expecting high significant correlations with the correct

time intervals. If the use of neighbouring information from EuSWiO was not conclusive, correlations with ERA-Interim reanalysis (Dee *et al.*, 2011) were used. If these comparisons did not help to identify the correct time interval, both periods were erased.

*Step 2. Consistency in values* tests address the detection of impossible observations. While minimum wind speed is obviously zero, different approaches can be used to set a physical maximum wind speed (Chávez-Arroyo and

Probst, 2015; Lucio-Eceiza *et al.*, 2018). Due to the lack of instrumental metadata, the upper limit was set to the maximum wind speed observed ( $113.2 \text{ m}\cdot\text{s}^{-1}$ ), updated from the official WMO climate extreme repositories following Dunn *et al.* (2012). Thus, all wind direction (speed) values falling outside  $0^\circ\text{--}360^\circ$  ( $0\text{--}113.2 \text{ m}\cdot\text{s}^{-1}$ ) were erased. Additionally, some stations that reported directions between  $0^\circ$  and  $36^\circ$  were transformed to  $0^\circ\text{--}360^\circ$ .

*Step 3. Abnormally low and high variability* (ALV and AHV) time intervals are targeted to assess the consistency of temporal variability at each site. ALV focuses on the detection of long chains of constant data, that are mostly due to malfunctioning of the instruments, like iced or blocked sensors (WMO, 2008). Erroneous constant periods were analysed separately for calm ( $<1 \text{ m}\cdot\text{s}^{-1}$ ) and non-calm ( $\geq 1 \text{ m}\cdot\text{s}^{-1}$ ) situations (e.g., Jiménez *et al.*, 2010; Chávez-Arroyo and Probst, 2015; Lucio-Eceiza *et al.*, 2018). Only sequences of repeated observations longer than five records were analysed, as shorter chains can occur naturally at any site (e.g., Jiménez *et al.*, 2010; Chávez-Arroyo and Probst, 2015; Lucio-Eceiza *et al.*, 2018).

Three different tests were applied to detect unreliably long chains of repeated data. The first one, searches for calm and non-calm constant periods surrounded by a large proportion of missing data as part of a long interval of potential malfunction. This test evaluates the percentage of missing data during the constant period, and the preceding and following 24 h. When any two of these percentages exceed the 90% of missing data, the period is flagged. The second test adapts the approach of Lucio-Eceiza *et al.* (2018) by calculating the likelihood of sequences of non-calm ( $\geq 1 \text{ m}\cdot\text{s}^{-1}$ ) constant values of any given length, deriving one threshold per site and time resolution. Finally, the third test concerns only erroneous wind speed calm periods ( $<1 \text{ m}\cdot\text{s}^{-1}$ ). These periods can be identified using a regional series built from averaging the closest five neighbour stations significantly correlated with the target site, as realistic calms should be reflected on both target and regional series. For each selected site the period spanning 24 h before and after the suspect sequence was standardized to 0 mean and unit standard deviation.

The resulting average regional time series,  $ra(t)$ , was evaluated at each time step to obtain a ratio between the range of wind speeds during the calm ( $\Delta_{in}(t)$ ) and the periods immediately before and after it ( $\Delta_{out}$ ):  $Ratio(t) = \Delta_{in}(t)/\Delta_{out}$ , where  $\Delta_{in}(t)$  is the difference  $ra(t) - V_{min}$  at each time step, and  $\Delta_{out}$  the range  $V_{max} - V_{min}$ ; with  $V_{min}$  and  $V_{max}$  being the minimum and maximum wind speed values during the candidate calm and the immediate 24 h before and after the calm, respectively.

A statistical threshold was set to flag low speed values that were not supported by high ratios in the regional series (Lucio-Eceiza *et al.*, 2018).

AHV are often caused by electronic failures, data manipulation or transmission errors (WMO, 2008), leading to three different error types: spikes, dips and long period. These errors were targeted using a combination of a blip test (Fiebrich *et al.*, 2010), which looks for successive increases and decreases in values, with a spatial check. Spikes and dips were detected by a blip temporal test. Frequency distributions of consecutive positive and negative differences were obtained for each station time series to set thresholds for erroneous positive (spikes) or negative (dips) intervals, following Lucio-Eceiza *et al.* (2018). For the detection of steps, records from neighbour stations were considered. Long periods were identified when the blip test detected a suspect positive step, followed by a period that was spatially flagged as anomalously high.

*Step 4. The bias detection* phase is aimed at detecting systematic errors, biases, in wind speed and direction. These biases might be resulting from changes in the position, height or exposure of the station, instrumentation changes or malfunction, changes in measurement or averaging methods (Pryor *et al.*, 2009; Wan *et al.*, 2010; Azorin-Molina *et al.*, 2014). Only one source of data (DWD; Table 1) provides this information in its metadata. Such lack of supporting information hampered the correction of long-term wind speed biases from documented changes in anemometer heights. Thus, only undocumented biases that affect wind speed records for periods of several weeks to months are addressed with this test (Lucio-Eceiza *et al.*, 2018). Long-term changes in the wind direction due to anemometer rotation can also be detected.

The flagging of wind speed biases was done using daily averages and based on three statistical parameters: the 15-day estimation in a moving window of the mean, standard deviation and the coefficient of variation of the daily time series. Typical ranges of variability for each parameter and thresholds for suspect values at each site were obtained following the same criteria as in Lucio-Eceiza *et al.* (2018). When the threshold of at least one of the three parameters was exceeded during 15 days or more, the period was flagged. Shorter chains down to 6 days were also flagged if at least two thresholds were exceeded. A manual inspection of the flagged periods based on expert criteria was required, ignoring long-term inhomogeneities which were not targeted herein.

Shifts in direction, which may affect inter-annual to multidecadal timescales are targeted by considering consecutive annual wind roses and calculating the angle in which their rmsd are at a minimum (Lucio-Eceiza *et al.*, 2018). Only years with relative shifts greater than  $10^\circ$  were flagged. Subsequently, the comparison was expanded to the complete time intervals between the flagged years. As in the case of wind speed, the flagged

cases had to be checked manually to discard rotations of wind roses with equitable wind directions. The correction of the flagged data was done by rotating the wind rose of the prior period/s to match the most recent one. Directional statistics (Mardia and Jupp, 2009) have been applied at all instances dealing with wind direction.

*Step 5.* Short chains of *isolated data* remaining between longer intervals of missing data erased during the application of the previous routines are unreliable (Lawrimore *et al.*, 2011), and thus targeted at this final phase for their elimination. Any sequence of observations of up to 24 h of length that is surrounded by intervals of missing observations was flagged if the sum of the lengths of the preceding and posterior missing periods is at least eight times the length of the target isolated period, that is, the ratio of a period of 3 h versus a whole day. Eventually, as the number of available observations of a given site might have diminished after applying the QC, following the same criteria used in Section 2.1, time series with few remaining observations were excluded.

## 2.3 | Quality control impacts and wind climatology

The impact of the QC on the wind distribution of the data set is assessed based on the changes in the spatial distribution of the mean, standard deviation, skewness and kurtosis, obtained from the first four order moments of the distribution for the entire time span of each record before and after applying the QC. Additionally, the temporal variability of surface wind during the last decades is also described. Because an average over the whole European domain would smooth out the variability of specific regions and associated wind regimes (Jiménez *et al.*, 2009; Dafka *et al.*, 2016; Minola *et al.*, 2016) we show inter-annual to multidecadal changes in wind variability over regions of homogeneous wind variability. For this purpose, rotated empirical orthogonal functions (REOFs; Richman, 1986; von Storch and Zwiers, 1999) are used to identify a distribution of regions that covers the spatial extent of the data set and within which wind variability is uncorrelated from region to region.

The analysis was applied to a subset of EuSWiO spanning the time interval 2000–2015, the spatially better represented period in the whole data set, selecting only the time series from each  $1^\circ \times 1^\circ$  lat  $\times$  lon grid cell in Figure 1d, which had the lowest percentage of missing data in the 2000–2015 interval. Daily data were considered and an estimation of the mean annual cycle is obtained by averaging all values of each calendar day over all years with available measurements at the site. The resulting 366-day estimate is filtered with a 15-day running mean, thus providing a

smooth estimate of the annual cycle with daily resolution (see Jiménez *et al.*, 2008; Lucio-Eceiza *et al.*, 2018); missing days in series of short span are interpolated. Subsequently, anomalies were calculated with respect to the mean annual cycle at each site and finally a principal component analysis (Preisendorfer, 1988; von Storch and Zwiers, 1999; Hannachi *et al.*, 2007). was performed based on the diagonalization of the inter-site correlation matrix. The homogeneous spatial sampling, the reduced presence of missing data and the use of the correlation instead of the covariance matrix are intended to mitigate potential spatial sampling and variance biases in the calculation of the maps loadings, that is, the eigenvectors or empirical orthogonal functions (EOFs). A number of EOFs were kept for the subsequent *varimax* rotation, selected herein because of its simplicity and because it preserves the orthogonality of the eigenvectors (Jiménez *et al.*, 2008). In a perfect simple structure, each site corresponds to a high load for one REOF loading map and a null load for the rest. In practice, several non null loads are obtained for the same site (Jiménez *et al.*, 2008), although, in general, the loadings of a given site tend to be highest within one or two of the resulting REOFs. Thus, the classification of regions was done considering the REOFs that produced the highest load at each site.

As this methodology assigns one principal mode to each region, it is possible to analyse the temporal wind variability of each region by calculating the time series of the corresponding scores, that is, the principal component (PC). Once the regions were defined, the inter-annual to multidecadal trends within each region were compared with the PC obtained during the 2000–2015 and the analysis was also extended back in time using the time series available within each region. This allowed both for having insights into longer timescales and also evidencing potential limitations related to still uncorrected homogeneity issues.

## 3 | RESULTS

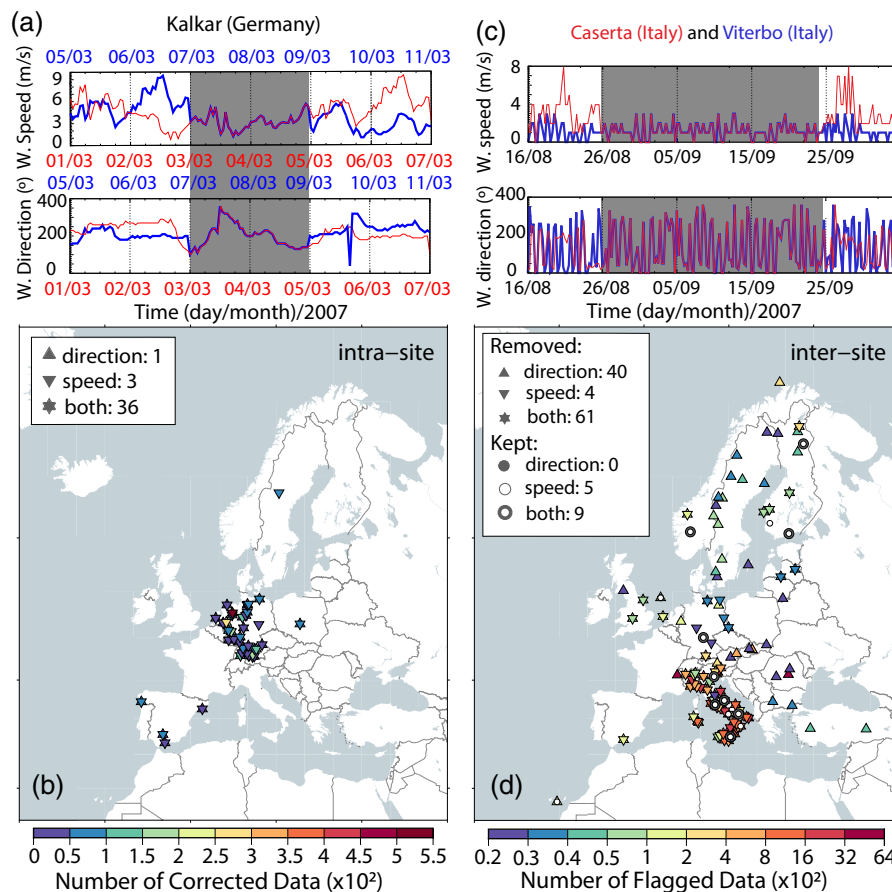
### 3.1 | Diagnosis of the quality control

Results of each step of the QC are summarized in Table 2, indicating the number of stations affected, the number of records corrected for each variable, the total accumulated values, as well as the corresponding percentages relative to the initial records prior the QC.

*Step 1.* Figure 2a illustrates an example of an *intra-site duplication* detected in a station in Kalkar (Germany), in which 2 days of hourly data were duplicated at two different periods for both variables. *Intra-site duplications* are the least frequently detected error (Table 2), affecting around 40 stations concentrated mainly in Germany, as

**TABLE 2** Number of affected stations and data during each step (columns 1 and 2) of the QC for wind speed (columns 3 and 4), wind direction (columns 5 and 6) and the total of both (column 7). Percentages, in parentheses, are given with reference to the initial number of wind speed, wind direction or total records, in each case

Test	QC step	Wind speed		Wind direction		Total
		Stations	Observations (%)	Stations	Observations (%)	Observations (%)
Intra-site	1	39	3,561 (<0.01)	37	3,093 (<0.01)	6,654 (<0.01)
Inter-site	1	65	41,505 (0.01)	101	60,753 (0.02)	102,258 (0.01)
Consistency in values	2	1,305	725,498 (0.21)	1,529	2,087,506 (0.61)	2,813,004 (0.41)
0°–36°	2	—	—	52	116,401 (0.03)	116,401 (0.02)
ALV (<1 m·s <sup>-1</sup> , –)	3	2,950	3,129,220 (0.91)	—	—	3,129,220 (0.45)
ALV (≥1 m·s <sup>-1</sup> , ≠0°)	3	1,426	496,580 (0.14)	1,661	1,973,970 (0.57)	2,470,550 (0.36)
AHV	3	3,804	1,527,203 (0.44)	—	—	1,527,203 (0.22)
Bias	4	61	262,521 (0.08)	337	24,432,950 (7.10)	24,695,471 (3.59)
Calms after AHV	4	879	520,775 (0.15)	—	–	520,775 (0.08)
Isolated records	5	2,055	35,516 (0.01)	1,973	42,246 (0.01)	77,762 (0.01)
Total			6,742,379 (1.96)		28,716,919 (8.34)	35,459,298 (5.15)



**FIGURE 2** Step1 of quality control: Examples of flagged periods (shaded area) for wind speed (top panel) and wind direction (bottom panel) with an (a) intra-site duplication in Kalkar, Germany, and an (c) inter-site duplication between Caserta and Viterbo, Italy. In (a), red and blue timelines match exactly, indicating each the two different periods of the time series. In (c) the red (blue) line indicates the period of the time series from Caserta (Viterbo) with the duplication. Spatial distribution of stations affected with (b) intra-site duplications and (d) inter-site duplications. In (c,d) the regular (inverted) triangle indicates duplications in wind direction (speed), and the colour represents the amount of duplicated data. Note the change in scale in the x axis. In (d) the grey (white) circle indicates inter-site duplications in wind direction (wind speed) which are correct and, therefore, preserved [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



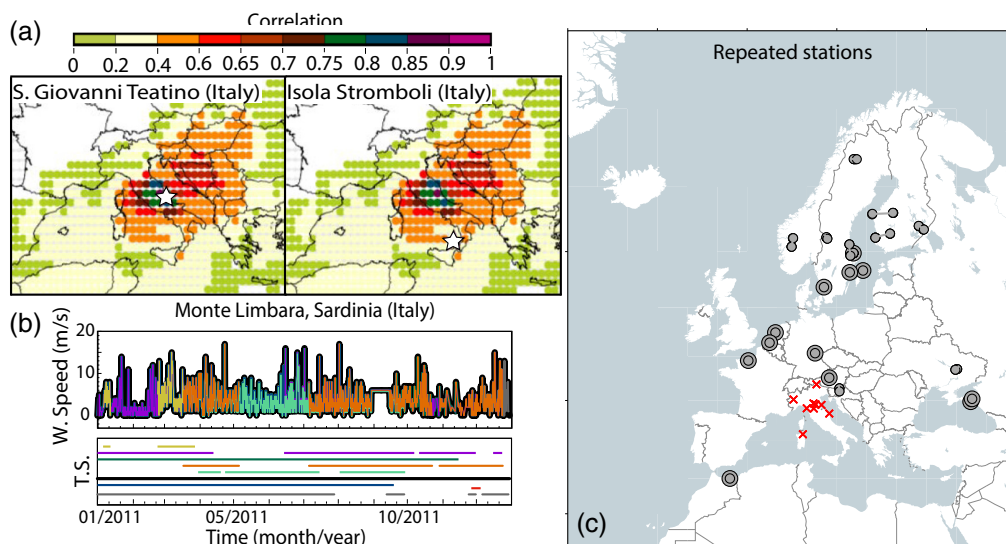
shown in Figure 2b. Most of these data were received via the DWD and occurred simultaneously in both variables, reinforcing the idea of the erroneous nature of such cases.

As an example of an *inter-site duplication*, an identical period of almost 1 month of data for both variables detected in a station in Caserta and in another one from Viterbo (both from Italy) is represented in Figure 2c. Figure 2d illustrates the distribution of the total of 110 (79) stations with wind direction (speed) inter-site duplications. Most of the sites over Italy are GTS stations showing the effects of this issue more than other regions and sources. The correct site could only be identified in 14 (9) cases for wind speed (direction) by comparing correlations in the vicinity of both sites involved in the duplication with the reanalysis. Figure 3a illustrates an example in which the area of maximum correlation with reanalysis allowed to identify the correct site (the one in S. Giovanni Teatino). Periods of the correct site, identified with a circle in Figure 2d, were kept. In contrast, periods of either erroneous sites, or from cases in which correlations gave no evidence of the correct site, were erased and are identified with a triangle in Figure 2d.

Italian GTS records were particularly problematic, with a remarkable example of inter-site duplication that occurred among nine stations and during their entire time span. One year of a station in Sardinia involved in this case is plotted in Figure 3b, showing that it is

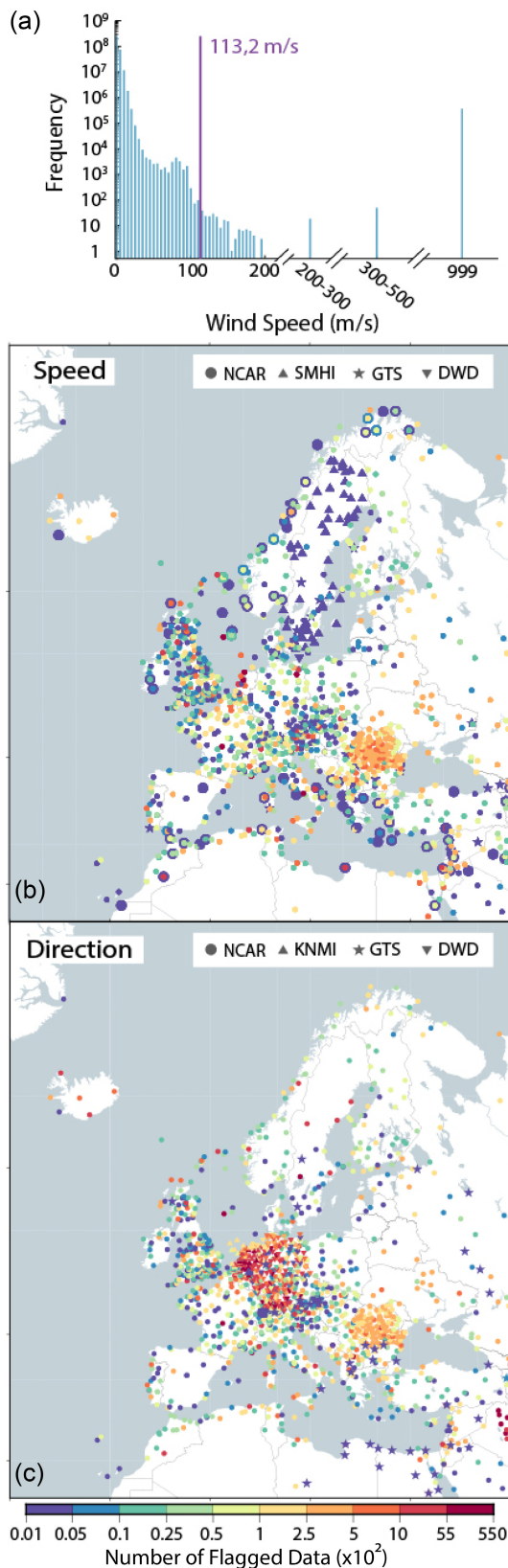
composed by the overlapping of the other eight stations. All nine stations, represented with a red cross in Figure 3c, were withdrawn. Moreover, 19 pairs of time series were found to be almost identical. Each pair of time series might belong to the same site, but they were not identified as equal during the merging step. The distribution of these pairs of stations is shown in Figure 3c with circles. For duplications with time series at the same location, the correction applied consisted on merging them if this provided additional information to the final time series, or keeping only the most complete time series having higher temporal resolution or more meta-data. If locations in the metadata were different for identical time series, those records were withdrawn as there is no confidence in assigning the observed wind variability to a specific site.

*Step 2.* The histogram of wind speed observations of the entire data set (Figure 4a) gives insights of the impact of the *consistency in values* test. One would expect a convergence of the wind speed histogram to values well before the maximum limit admitted, as it was a very conservative threshold. However, an abnormally large amount of observations are close to  $113.2 \text{ m}\cdot\text{s}^{-1}$ . As mentioned, the lack of metadata hampered the selection of a more restrictive threshold, but some of these unrealistic values were expected to be flagged in subsequent steps. Observations above the global maximum wind speed observed (purple vertical line in the histogram) were



**FIGURE 3** (a) Maps of correlations with reanalysis of two stations (stars) in Italy affected by an inter-site duplication: S. Giovanni Teatino (left panel) and Isola Stromboli (right panel). (b) Example of a station in Monte Limbara, Cerdeña, for which the wind speed time series (top panel) mixes parts of other eight stations also in Italy. The panel below shows the overlap and time span covered by each time series (T. S.), represented in a different colour. (c) Spatial distribution of EuSWiO repeated station time series. Pairs of repeated stations in the same position or within a short distance are represented with two concentric circles of different size. Identical time series with different assigned position are identified with grey circles of the same size connected with a line. Red crosses show the sites affected by the duplication in (b) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

considered unrealistic and thus erased. It is noteworthy the large amount of values above  $113.2 \text{ m}\cdot\text{s}^{-1}$ , and specifically the high number of observations equal to  $999 \text{ m}\cdot\text{s}^{-1}$ ,



that were probably due to missing data transcribed erroneously. Such large values distributed over Europe in time series coming from NCAR data sets: more than the 99% from the ds464.0, and the rest from the ds461.0 (Figure 4b). Some specific regions are notably more affected, such as Romania, which was mainly fed by the ds464.0 in the absence of other sources. High values different from  $999 \text{ m}\cdot\text{s}^{-1}$  do not seem to respond to any specific spatial pattern and stem from various sources: mostly NCAR, SMHI and GTS.

Regarding wind direction, values exceeding realistic limits were mainly equal to  $999^\circ$  or  $990^\circ$ , and thus, also probably due to an erroneous transcription of missing data (Figure 4c). NCAR sites from the ds464.0 show a similar pattern as for wind speed, highlighting as well the area of Romania with a high density of stations affected. Alternatively, a large amount of observations with missing codes defined as  $990^\circ$  were detected in the DWD and KNMI: 34% and 19%, respectively, of the total number of cases. Finally, 59 observations of values different to  $999^\circ$  or  $990^\circ$  were flagged, coming all of them from the GTS except for two cases from the ds464.0 and the KNMI. Corrections affected 33% (39%) of the wind speed (direction) time series (Table 2).

Several records were also found to report only within the range of  $0^\circ$  to  $36^\circ$  during some period of time. This issue, related to data transcription, was detected in 52 stations (Table 2) from Spain, originally provided by the ECA&D.

*Step 3.* Examples of the detected *abnormally low variability errors* for a non-calm and a calm situation are highlighted in the shaded area of Figure 5a,b. Comparisons with neighbour stations made it possible to identify reliable calms as illustrated in the examples from Figure 5b,c. On one hand, Figure 5b shows a case in which the target site differs from the regional time series, created from its neighbour sites, during the calm period (shaded area), leading to a high ratio, illustrated schematically for the wind speed  $v(t)$  at a given time step in Figure 5b,c. In this case the period is considered erroneous and erased. On the other hand, in Figure 5c the target site and the regional time series behaves similarly

**FIGURE 4** Step2 of quality control (QC): (a) wind speed histogram before the QC. The vertical purple line indicates the threshold set for maximum wind speed. Spatial distribution of the stations with unphysical wind speed (b) and wind direction (c) observations. The colour scale in the map indicates the number of affected data. Sources of data are identified with symbols (see legend). Smaller symbol size refers to observations equal to  $999 \text{ m}\cdot\text{s}^{-1}$ ,  $999^\circ$  or  $990^\circ$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

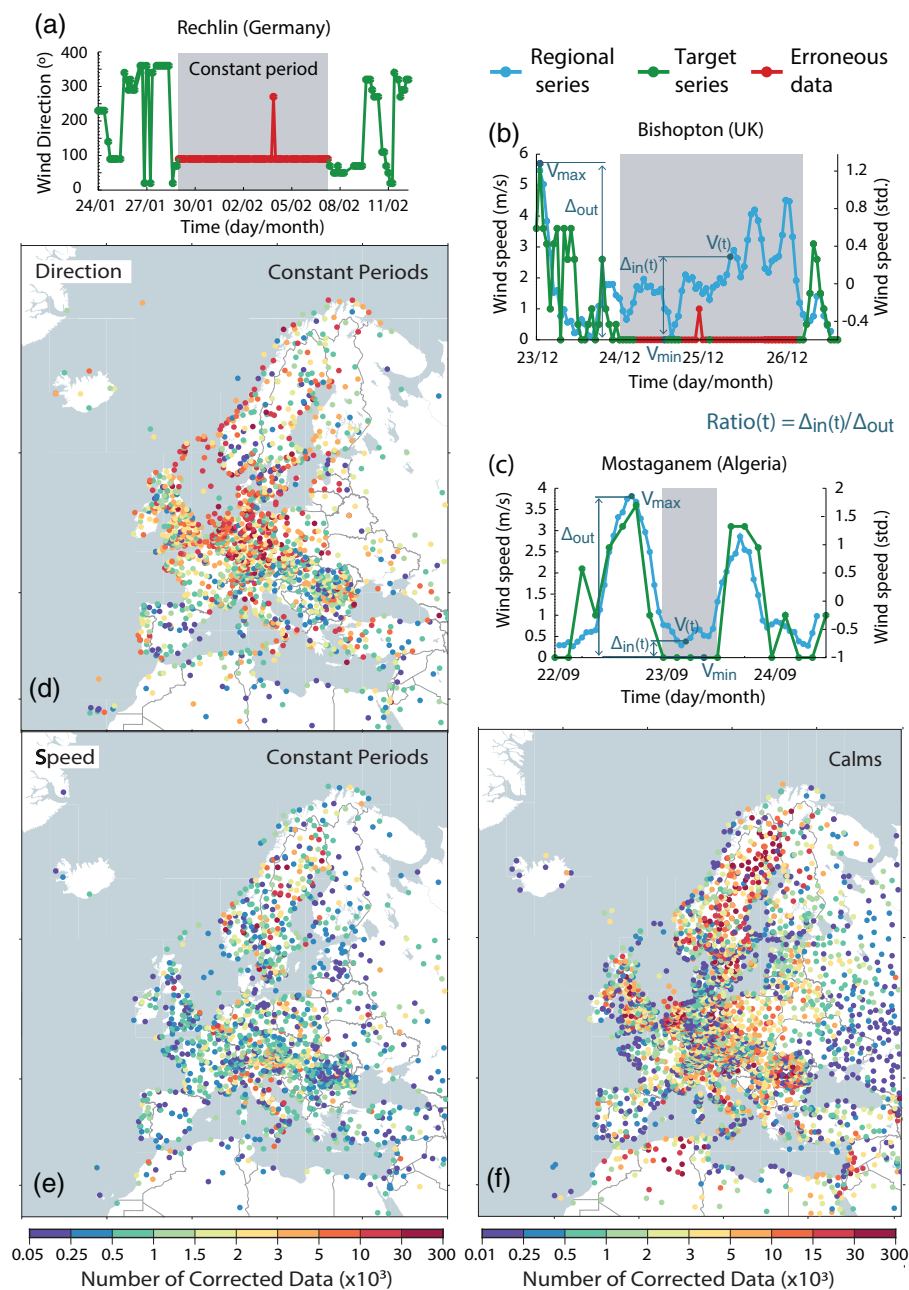


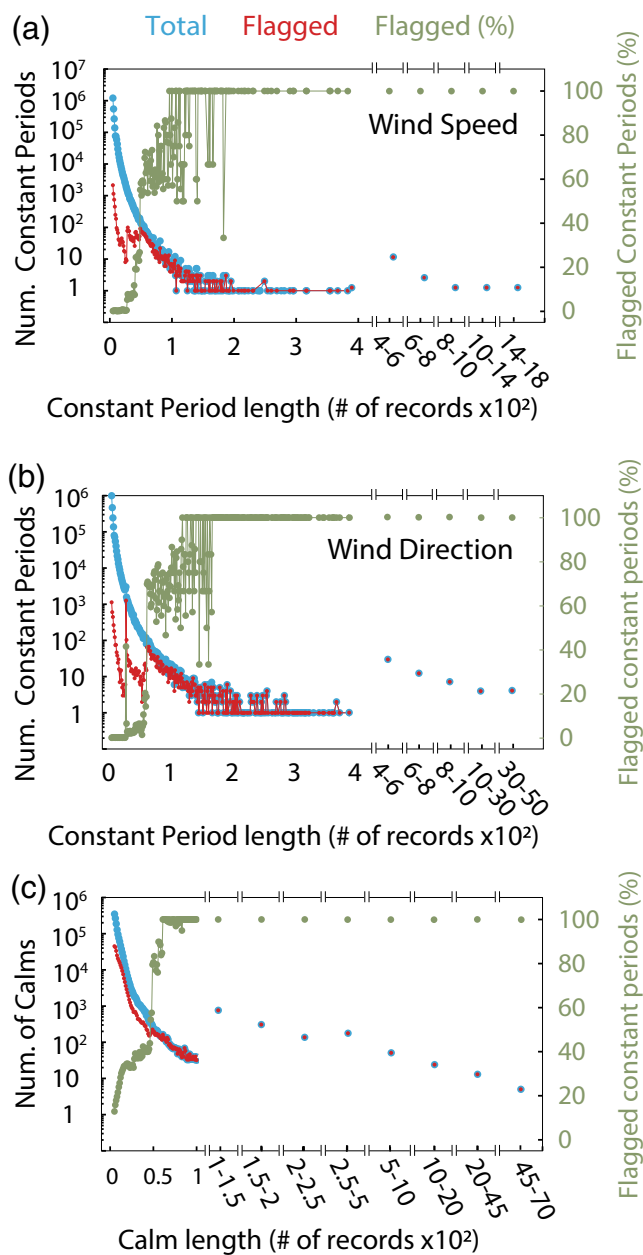
FIGURE 5 Step 3 of quality control:

Examples of low variability erroneous data (red dots) for (a) a long constant period in wind direction in Rechlin, Germany, and (b) a calm situation for wind speed in Bishopton, United Kingdom. (c) Example of realistic calm situation in Mostaganem, Algeria. Shaded areas represent suspect periods. In (b) and (c), the target and the regional series are plotted using the right axis, standardized and thus dimensionless; the left axis shows the absolute range for the target series (see Section 2.2, step 3, for details about the definition of ratios and related parameters). Spatial distribution of stations with long constant non-calm sequences, for wind direction (d) and wind speed (e), and (f) for wind speed calms. The same colour scale indicates the number of corrected data for both variables [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

during the calm period. Such low ratio indicates a non erroneous behaviour, and thus, it was not corrected. Abnormally low variability errors were very frequent for both variables and distributed all over the domain as shown in Figure 5d,e,f. As expected (Section 2.2), the frequency of wind speed calm periods detected (Figure 5f) is more than six times higher than for constant periods (Figure 5e), with almost the 75%, against the 36%, of the stations affected in each case (Table 2). The percentage of flagged sequences increases rapidly with its length until around the 70% for sequences of constant periods longer than 56 (62) for wind speed (direction) records (Figure 6). Percentages above this length oscillate highly because they encompass a small number of cases, up to the

absolute threshold set of 168 records (7 days in hourly resolution), over which the 100% of the periods were flagged. Regarding calms, the rise of percentage is even sharper, achieving an 80% of flagged periods for a length of 49 records, and over 61 records it remains virtually constant.

Spikes, dips and steps were the three different typologies of *high variability errors* targeted. Examples of each are illustrated in red in Figures 7a–c. Figure 7d shows the total distribution of *high variability errors*, which were also very frequent and spread all over the domain affecting almost the 97% of the stations. The largest values are attained over central Europe and the Scandinavian Peninsula. Note that less data were affected than in the case



**FIGURE 6** Absolute frequency distribution of the number of (a) wind speed constant periods, (b) wind direction constant periods and (c) calms (left y axis) vs. their length in number of records excluding missing data and not considering differences in resolution. The total number of periods detected (blue) and the flagged ones (red) are represented in the left y axis, while the percentage of flagged periods over the total evaluated (green) is shown in the right y axis [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

of low variability errors (Table 2). After applying this test, new long low variability intervals were detected in 879 stations (Table 2). They became detectable after the correction of high variability errors that in all these cases interrupted constant sequences of data hampering their correction. Therefore, the ALV test was run again to

account for these errors, prior to Step4. The statistics resulting from that are included in those of ALV in Step 3.

*Step 4. Bias errors* for wind speed (direction), as the example in Figure 8a (Figure 8b), were detected only in about 10% of the stations, with a higher percentage for wind direction than for wind speed as seen in Table 2 and Figures 8c,d. However, a high amount of data was removed (corrected) for wind speed (direction), as a result of the long-term span of the errors.

*Step 5. Isolated records* were eventually detected and removed from a high amount of stations (Table 2). The final clean-up removed additional 76 stations from the database that did not meet the requirements of minimum length imposed. A total of 3,829 stations remained in the final version of EuSWiO.

### 3.2 | Impact of the quality control and wind climatology

The first four moments of the wind speed distribution (mean, standard deviation, skewness and kurtosis) were obtained before and after the QC for the entire time span of each record, to evaluate its impacts on EuSWiO and provide a first assessment of the wind climate over Europe.

Changes in the mean and standard deviation of wind speed before and after the QC (Figure 9), are related because wind speed is a positive zero-limited variable, with increases (decreases) in the mean producing increases (decreases) in the width of the distribution. High values in the mean and variance tend to be observed in the coast lands, in particular in the Atlantic countries. However, very high wind speeds occur locally and without a specific geographic pattern from the north-west to the southeast side of the domain, with a noticeable outlier region in Romania. Such sites tend to show also very high standard deviations. After the corrections, mostly the influence of Step 2 addressing the detection of physically unrealistic observations, wind speed average and variability still show the contrast between coastlines and the interior, but within plausible ranges of values. Although some local impacts on wind direction are visible, changes tend to be minor.

Skewness is expected to be positive for wind speed (Figure 10a). However, before the QC, a large number of sites showed very high unreliable values mostly over the eastern half of the domain and also negative values at random locations. After the QC, mainly owed to the correction of high variability cases, the pattern is spatially more homogeneous with values ranging mostly between zero and three.

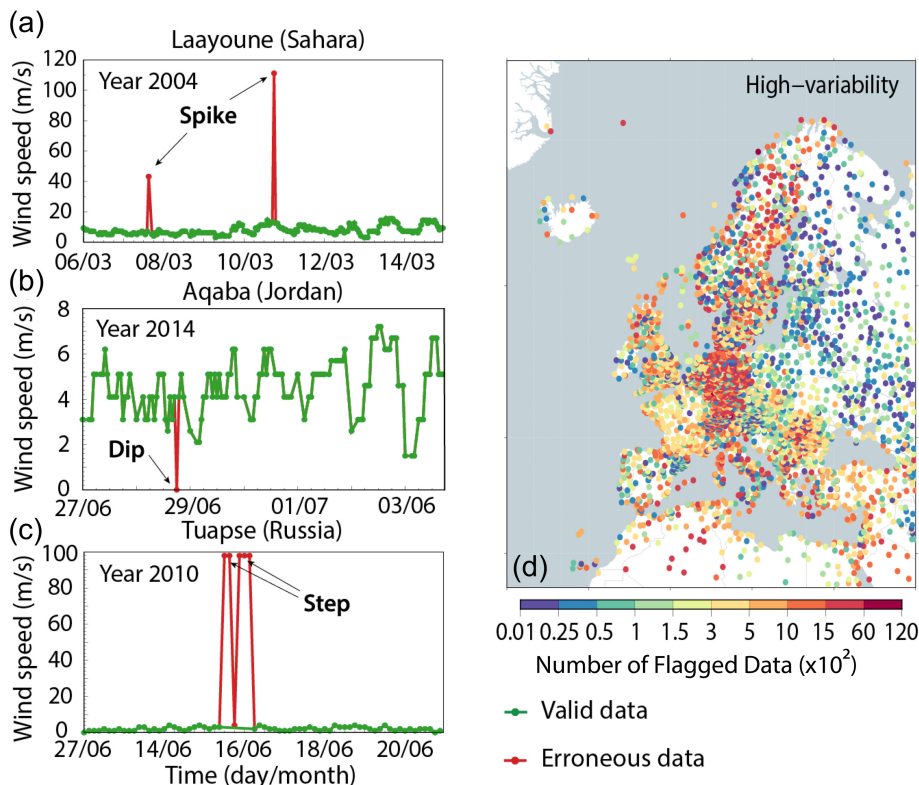


FIGURE 7 Step 3 of quality control: Example of: (a) two spikes at Laayoune, Sahara; (b) a dip at Aqaba, Jordan; and (c) two steps at Tuapse, Russia. (d) Spatial distribution of stations with high variability errors. The colour scale indicates the number of affected data [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

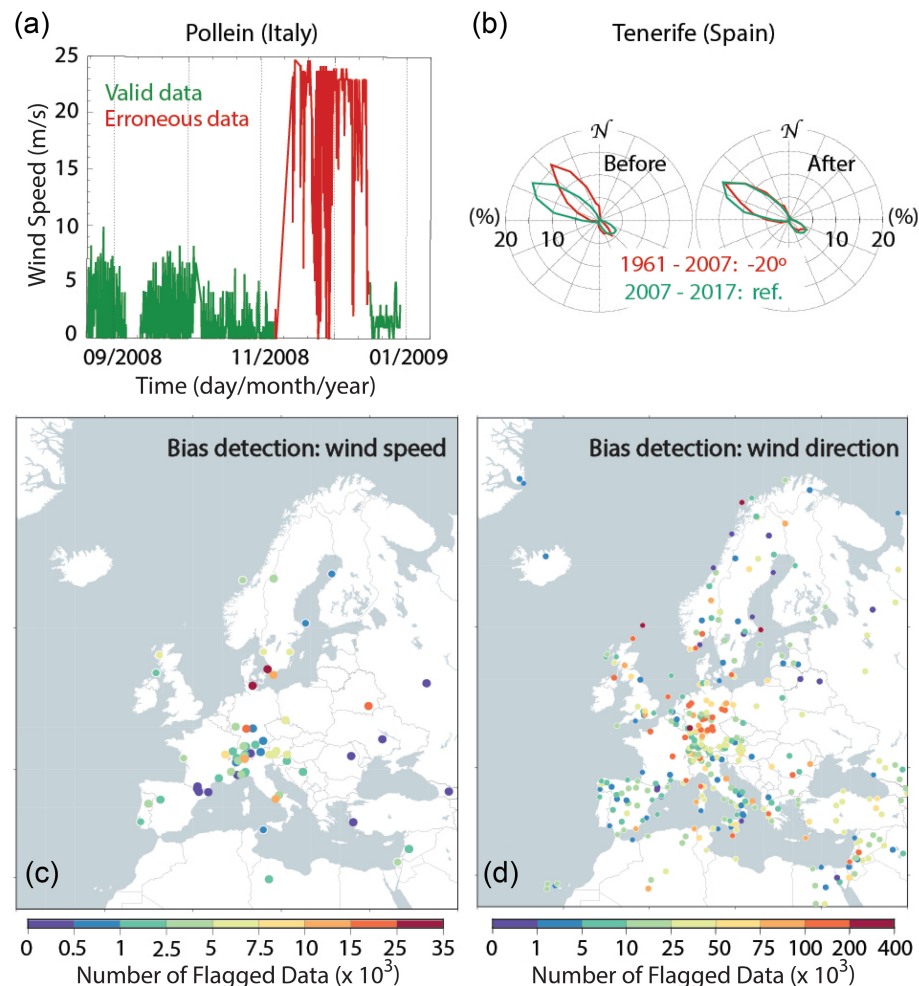
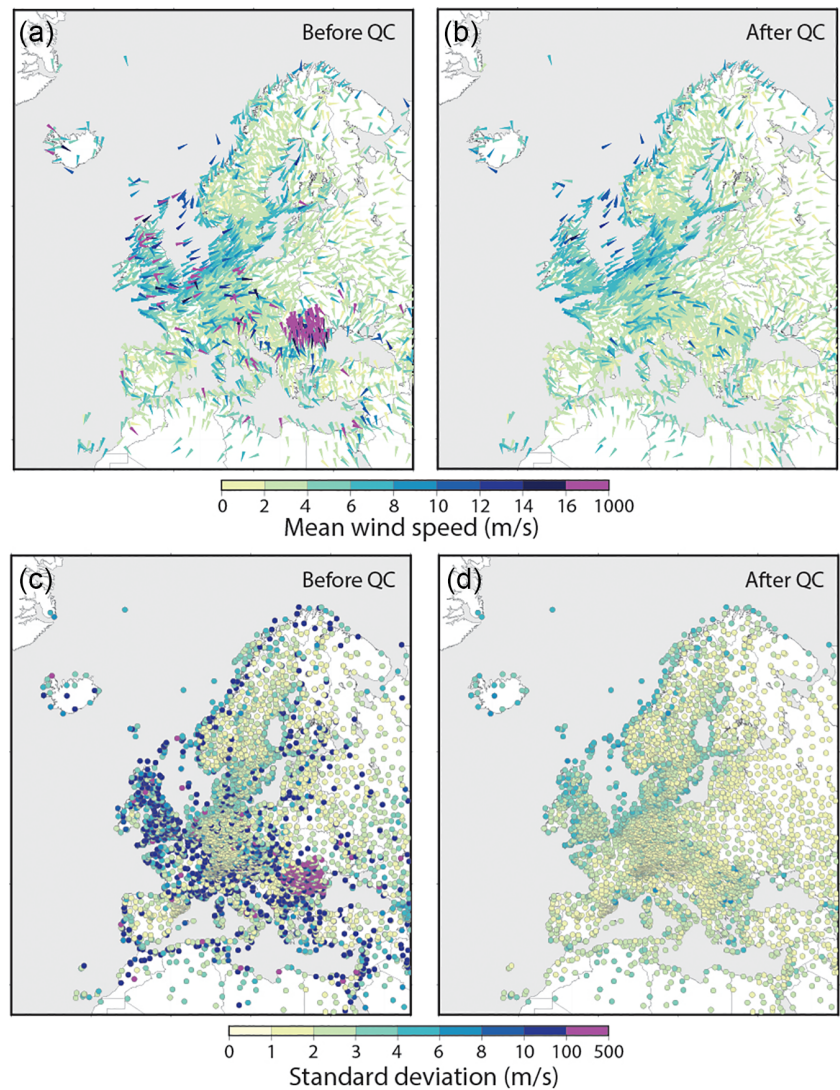


FIGURE 8 Step 4 of quality control: (a) example of a temporal wind speed bias detected at Pollein, Italy. The red part indicates the flagged period of 33 days. (b) Wind rose showing a wind direction bias of 20° at Tenerife, Spain in 2007. Colours stand for values before and after correction. The rose is rotated by correcting the earlier data and keeping the most recent distribution. Spatial distribution of stations with (c) detected wind speed and (d) direction biases. Note the colour scales indicating the number of affected data are different for each variable [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**FIGURE 9** Comparison of (a,b) mean wind speed and direction, and (c,d) wind speed standard deviation before and after the quality control, respectively. Directional statistics from Mardia and Jupp (2009) are used for the calculation of mean wind direction [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



Kurtosis (Figure 10c,d) shows also a wide range of values with respect to zero, after subtracting the reference value of three of the normal distribution. High negative values distribute primarily over the western half and highly positive over the eastern half. After the QC the pattern is spatially more homogeneous, with most values in the positive range indicating a tendency towards leptokurtic (i.e., peaky) sites. Some local negative values tend to coincide with sites of low skewness whereas in general high kurtosis occurs with the high positive skewnees.

The maps of the mean and standard deviation (Figure 9) provide a simple description of the wind behaviour in Europe. As observed in previous studies (e.g., Troen and Lundtang Petersen, 1989; Bett *et al.*, 2013) the highest winds cover the region around the North Sea, such as The United Kingdom, The Netherlands, Belgium and Denmark, corresponding also to the ones of highest variability. This area is characterized by well defined mean geostrophic westerlies over

this large region. However, the influence of the major orographic barriers such as the Alps, the Pyrenees, the Scandinavian mountains and the Mediterranean Sea is also remarkable giving rise to many local winds, like the Cierzo in northeastern Spain (Jiménez *et al.*, 2009), the mistral in southern France (Jiang *et al.*, 2003), the Etesians winds along the Aegean Sea (Dafka *et al.*, 2016) and the Bora over the Black Sea (Smith, 1987).

A first assessment of the temporal variability is provided in Figures 11 and 12. As performing an average of the broader domain would smooth out different regional behaviours as those indicated above, the focus is placed on objectively defining regions that are independent in their time variability. The first step for the regionalization was to perform an EOF analysis over a subset of EuSWiO where the presence of missing values and spatial heterogeneity was minimized (Section 2.3). A subset of 1,334 stations spanning the entire period of 2000–2015 and covering the whole European domain was selected (see

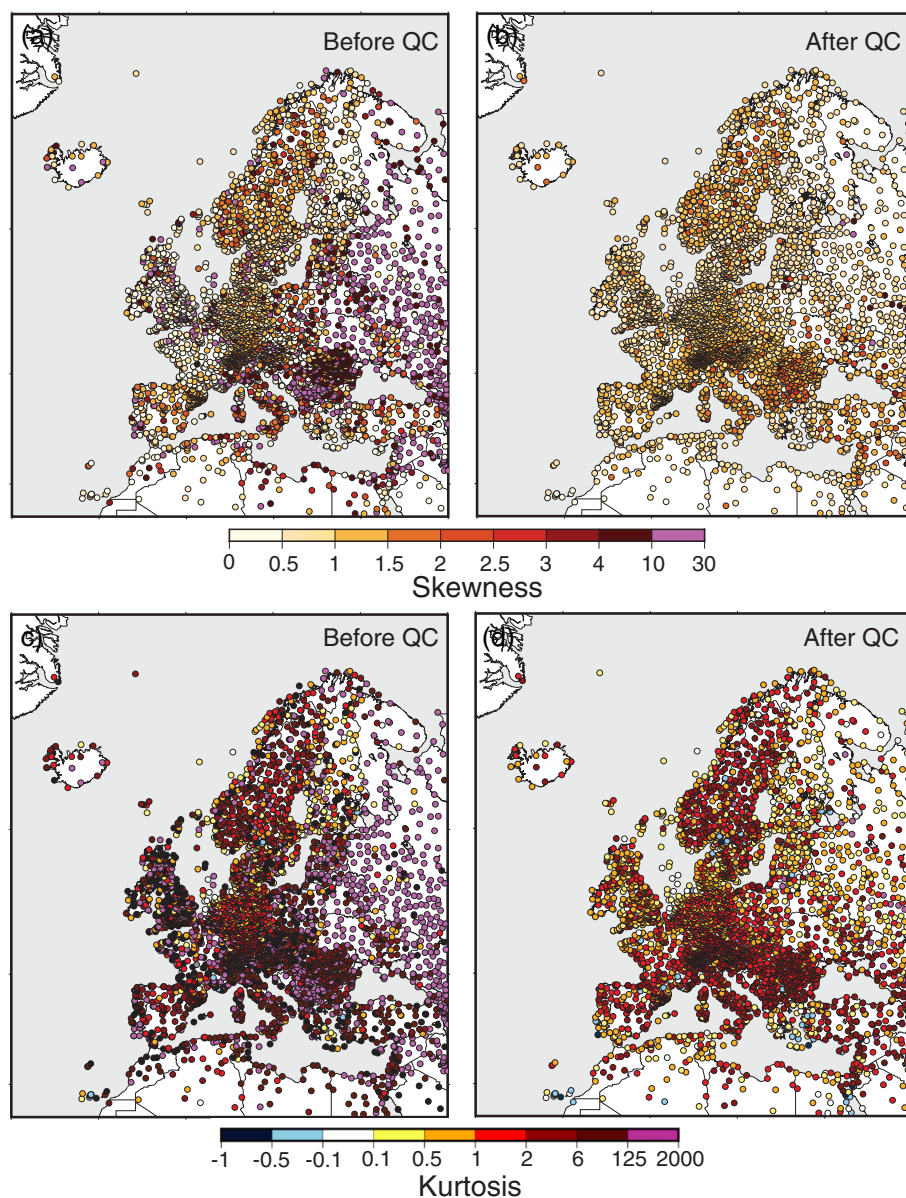


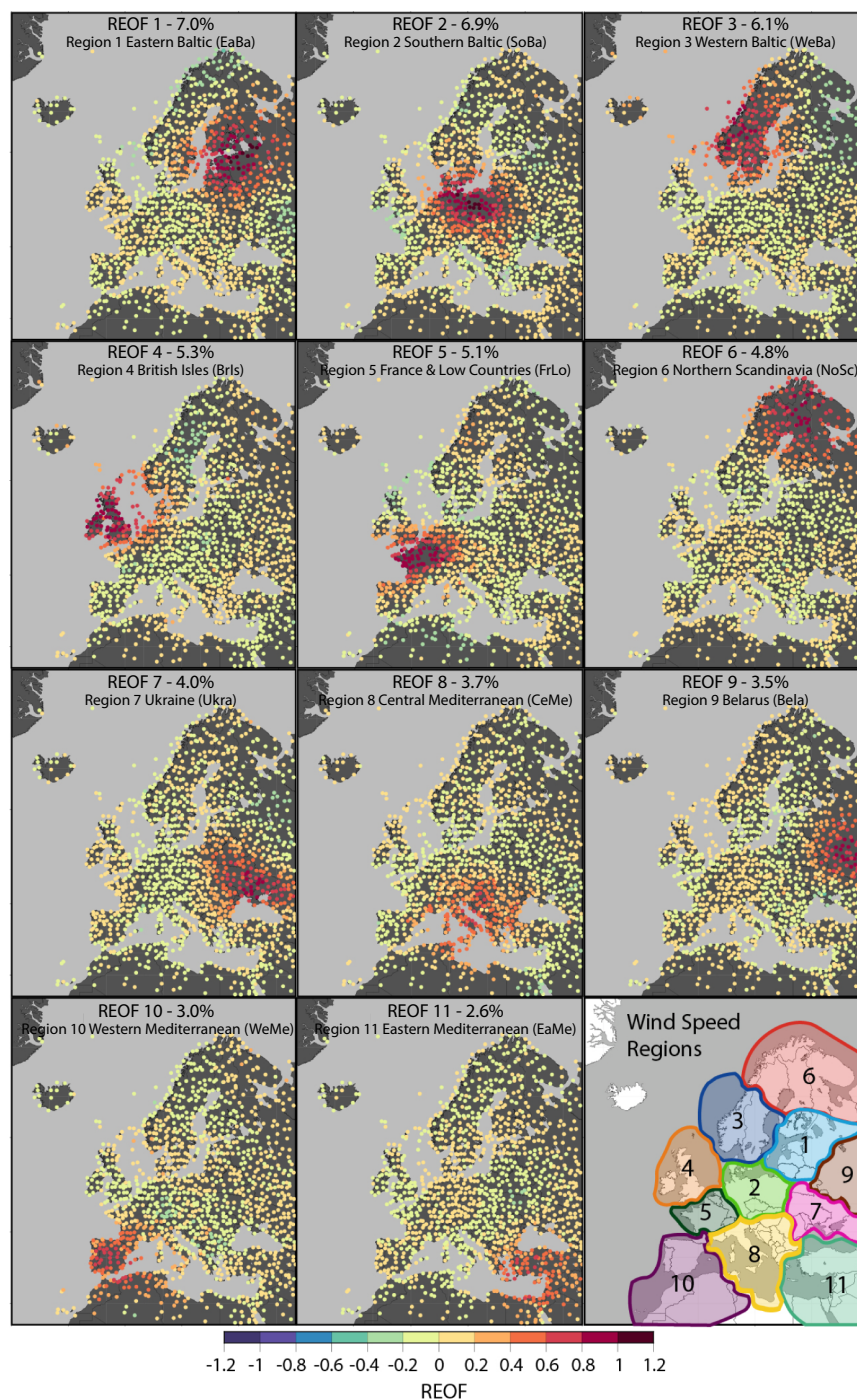
FIGURE 10 Comparison of (a,b) skewness and (c,d) kurtosis for wind speed before and after the quality control, respectively [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7739)]

Figure 11 for the distribution of locations) and daily averages calculated. Site dependent annual cycles were subtracted and EOFs were obtained from the correlation matrix and the first 11 modes, accounting for a total of 34% of variance were kept. The retained EOFs were subsequently rotated for a better characterization of the regions (Richman, 1986). The REOFs explained the 52% of the total variance with individual explained variances ranging between 26% and 7%. Eleven modes were determined to be enough to describe the whole domain, avoiding, at the same time, a high overlapping of the regions. Attempts to increase the number of rotated EOF modes lead to increasing inter-region overlap and to overemphasizing some small domains. The spatial patterns of the resulting 11 REOFs are displayed in Figure 11. Note that each REOF shows higher loadings

over a specific region (indicative region names and acronyms are provided in each panel) and that the regionalization (Figure 11 bottom right panel) covers the broad European domain.

The temporal wind variability can be now analysed independently for each region by considering all the sites belonging to the defined regions. Average regional time series were obtained considering sites within the limits of each region which span for the entire period chosen for the calculation of REOFs, and represented along with the corresponding time series of scores in Figure 12 (black and red solid lines, respectively). The average regional and the scores time series are highly correlated. The lowest value obtained is of 0.91 ( $p < 0.05$ ) for Region 8 (Central Mediterranean, CeMe), thus indicating that the corresponding principal mode of each region

**FIGURE 11** Spatial patterns and percentage of the explained variance of the 11 retained rotated empirical orthogonal functions (REOFs) (dimensionless) calculated in the period 2000–2015 with a selected subset of 1334 stations. Wind regions inferred from each REOF mode are indicated in the bottom right map. Indicative region names and acronyms are provided in each panel [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

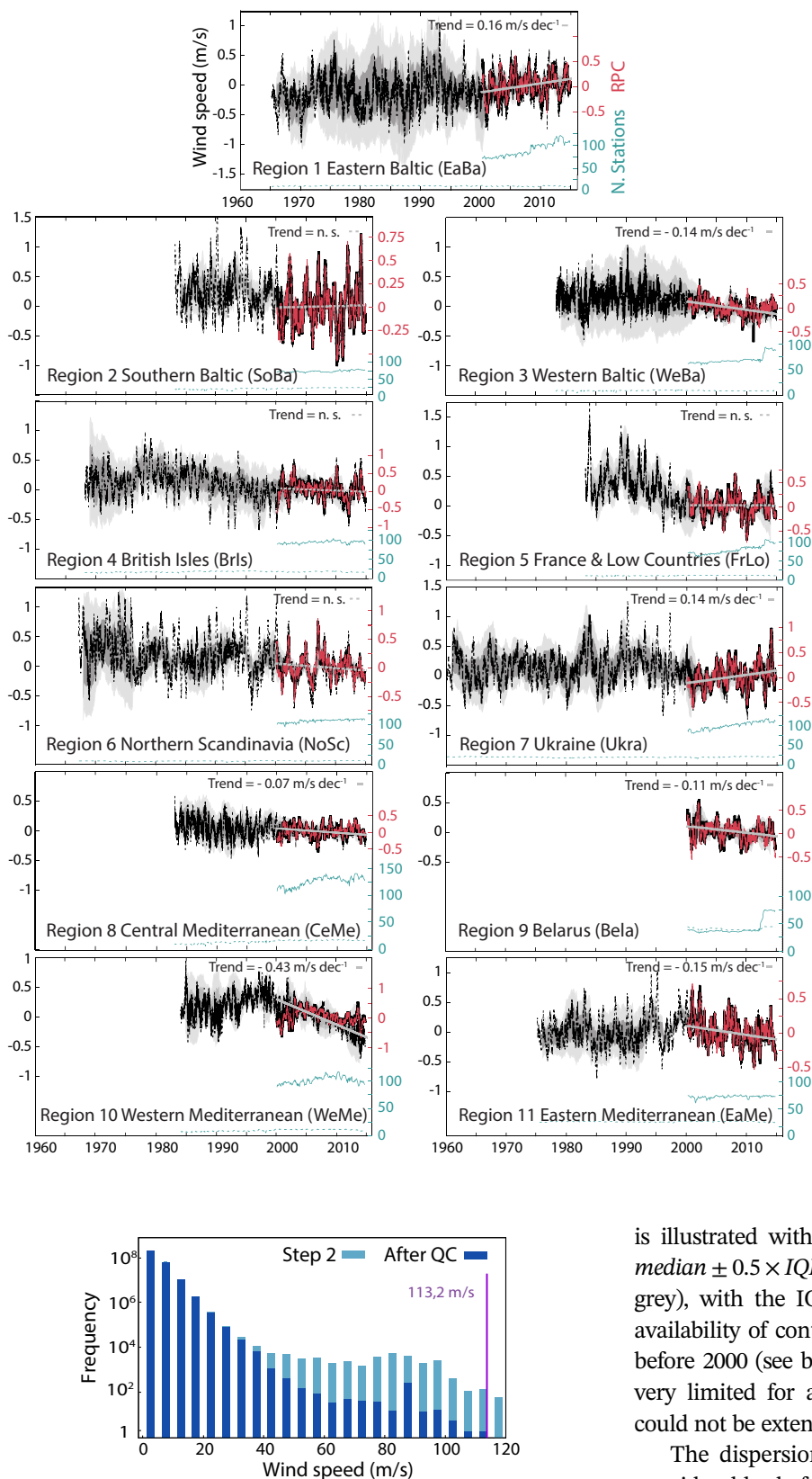


characterize well the mean wind speed behaviour during that period. Note also the very different inter-regional temporal behaviour, where some regions show very high frequency variability (e.g., British Isles and Central Mediterranean; regions 4 and 8) while others show more variance at low frequencies (e.g. France & Low Countries and Northern Scandinavia; regions 5 and 6) during the 2000–2015 interval. Inter-regional correlations are not significant ( $p < 0.05$ ) and range 0.01 and 0.30 for most regions with a maximum value of 0.39 between Regions 5 and 7 (France & Low Countries and Ukraine). Trends

during the 2000–2015 interval vary also between regions with significantly ( $p < 0.05$ ) increasing wind in some of them (e.g., Eastern Baltic and Ukraine; regions 1 and 7) and decreasing in others (e.g., regions 3-WeBa, 8-CeMe, 9-Bela, 10-WeMe, 11-EaMe). Decadal trends in wind speed regional averages and their significance are also shown for each region in Figure 12.

To gain more insight into the long-term trends in wind speed for timescales longer than the 2000–2015 period, only time series with a continuous time coverage previous to 2000 were selected. The range of observational uncertainty





**FIGURE 13** Histogram of the wind speed distribution of the EuSwiO before applying corrections to physical limits (step 2) and after the quality control [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**FIGURE 12** Rotated principal components filtered by a 61-day moving average (red line and red right y axis labels) of the first 11 rotated empirical orthogonal functions (Figure 11). The left y axis and black lines represent the median filtered by a 61-day moving average from each region obtained with all the stations with data during the period 2000–2015 (solid) and only with the stations that can be extended to previous years (dashed). For the 2000–2015 interval a linear fit is shown with grey solid (dashed) line for significant (non significant) trends; significance is evaluated for at  $p < 0.05$  and accounting for autocorrelation. The rate of change is also indicated in the top right of each panel; n.s. stands for non significant trends. Shaded grey colours denote the  $median \pm 0.5 \times IQR$  (dark grey) and  $median \pm IQR$  (light grey) ranges. Blue lines show the number of available sites to calculate the regional average during the 2000–2015 period (solid blue) and in the extended period of availability for each region (dashed blue). Region names and acronyms correspond to those of Figure 11 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

is illustrated with the shaded areas corresponding to the  $median \pm 0.5 \times IQR$  (dark grey) and  $median \pm IQR$  (light grey), with the IQR being the Inter Quartile Range. The availability of continuous series of data varies considerably before 2000 (see blue lines in Figure 12). For instance, it is very limited for areas of Russia, and Region 9 (Belarus) could not be extended to the past.

The dispersion of data within each region increases considerably before 2000. Recall that homogenization corrections of long-term changes in wind speed are not applied here, which does not allow for a confident assessment of long-term trends. Nevertheless, some robust features arise. Some regions widen their uncertainty range

around the median before 2000 (e.g., regions 1-EaBa, 3-WeBa, 8-CeMe, 10-WeMe) hampering any confidence on long-term mean speed changes prior to 2000. Nevertheless, some periods show robust windier conditions like the mid 1970s and 1990s in Region 1 (EaBa) or the late 1990s in Region 10 (WeMe). Most of the other regions show consistent increases of mean wind speed to the past with respect to the 2000–2015 average (e.g., regions 2-SoBa, 4-BrIs, 5-FrLo, 6-NoSc). It is noteworthy that in Region 4 (BrIs) and Region 10 (WeMe), 2000–2015 trends are the continuation of changes at longer timescales. These inspection underlines the need for a more insightful analysis involving an assessment of inhomogeneities but also exhibits potential for the understanding of wind-related applications to relevant fields (e.g., wind energy, agriculture; Cleugh *et al.*, 1998; Garcia-Bustamante *et al.*, 2013; Devis *et al.*, 2018).

#### 4 | SUMMARY AND CONCLUSIONS

This work presents a first version of a European Surface Wind quality controlled observational database, EuSWiO, which compiles wind speed and direction observations from heterogeneous origins.

The semi-automatic QC applied, adapted from Lucio-Eceiza *et al.* (2018, 2018), is briefly described herein, providing a useful framework of similar QC techniques that may be extended to other surface wind data sets. Results of the QC techniques are discussed on a case-by-case basis, providing examples and the distribution of detected unreliable data over the domain. A summary of the overall results of the QC applied is shown in Table 2. Around 5% of the observations were flagged, a high proportion of them corresponding to wind direction observations (80% of both variables), mainly due to bias (7%). In turn, for the case of wind speed, calms were the most abundant errors. However, some of them, like low variability intervals were mixed with other, mostly high variability, error types and were only detected with subsequent tests or during the manual assessment required in the bias phase. Another example is the histogram of the wind speed distribution at Step 2 (Figure 4a), that evidences the limitations of this test to correct a high amount of unreliable values because of the very conservative threshold imposed. The sharp cut-off in Figure 4a, evidences that many of those values are likely excessively high to be realistic. In Figure 13, this histogram is compared with the wind speed distribution of the final version of EuSWiO, showing that multiple observations over  $30 \text{ m}\cdot\text{s}^{-1}$  were corrected afterwards. Nevertheless, observations between  $80$  and  $90 \text{ m}\cdot\text{s}^{-1}$  are still suspicious

and indicates that future versions of the data set should evaluate this further. The identification of other issues such as wind direction time series transcribed from  $0^\circ$  to  $36^\circ$ , that are not the specific target of any of the tests, also exhibits the ability of the procedure to detect an erroneous performance of wind times series, which were not expected. These examples suggest a level of complementarity between the tests, that allow for detecting additional data of a suspicious character, different from that being targeted.

An improvement of the quality in the general behaviour of the database is shown with the representation of the first four moments of the wind that show more realistic values after the QC. In this regard, the impact of the QC is considerable as the resulting patterns are spatially more homogeneous, owed to the reduction of outlier behaviour of many sites, particularly after the correction of high variability errors.

The REOF-based regionalization allows for defining 11 regions that cover the broader European domain and that show independent temporal variability during the 2000–2015 period. The inner regions (e.g., 1-EaBa, 2-SoBa and 7-Ukra) show increasing wind speed during this period while the outer regions show diminished speeds (e.g., 3-WeBa, 4-BrIs, 10-WeMe, and 11-EaMe). For earlier decades, uncertainties increase, thus confirming the need of homogeneity tests for long-term trend applications. Nevertheless, robust conclusions can still be drawn as several regions present consistency in the information available. For instance, most regions suggest decreases in wind speed since the early 1980s.

This work issues a first version of an unprecedented quality-controlled European surface wind speed and direction observational database, EuSWiO. The data set expands over the broader European domain and southern Mediterranean lands. The data, with mixed hourly to daily resolutions, span over the last four decades with relatively good spatial coverage, including some regions that extend back to the 1960s and some sites that reach back to the beginning of the 20th century.

EuSWiO bears potential for a range of different analysis from meteorological to climate and climate change oriented studies, including potential applications such as analyses of extreme events (Usbeck *et al.*, 2010; Cheng *et al.*, 2012; Cheng, 2014; García Bustamante *et al.*, 2021), wind power assessments (García-Bustamante *et al.*, 2009; Tian *et al.*, 2019; Zeng *et al.*, 2019) or the evaluation of reanalysis or model-data comparison exercises at continental scales (Staffell and Pfenninger, 2016; Molina *et al.*, 2021).

#### ACKNOWLEDGEMENTS

CRL was supported by the NEWA ERANET project (PCIN-2014-015-C07-06, PCIN-2016-009) and

GReatModelS RTI2018-102305-B-C21 from MINECO. This research has been conducted under the Joint Research Unit between UCM and CIEMAT, by the Collaboration Agreement 7158/ 2016. We acknowledge the data providers in the ECA&D project, the Royal Netherlands Meteorological Institute, the Swedish Meteorological and Hydrological Institute, the Deutscher Wetterdienst and the National Center for Atmospheric Research.

Initial and future versions of the EuSWiO will be made available under the European Climate Assessment & Dataset ([www.ecad.eu](http://www.ecad.eu)).

## FUNDING INFORMATION

Ministerio de Ciencia, Innovación y Universidades, NEWA ERANET project (PCIN-2014-015-C07-06, PCIN-2016-009) and GReatModelS (RTI2018-102305-B-C21).

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## ORCID

Cristina Rojas-Labanda  <https://orcid.org/0000-0002-7191-0947>

Elena García-Bustamante  <https://orcid.org/0000-0002-2677-0252>

Frank Kaspar  <https://orcid.org/0000-0001-8819-8450>

## REFERENCES

- Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, S.M., Jusem, J.C., Smith, D.K. and Gombos, D. (2011) A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bulletin of the American Meteorological Society*, 92, 157–174. <https://doi.org/10.1175/2010BAMS2946.1>.
- Azarin-Molina, C., Vicente-Serrano, S.M., Mcvicar, T.R., Jerez, S., Sanchez-Lorenzo, A., López-Moreno, J.I., Revuelto, J., Trigo, R. M., Lopez-Bustins, J.A. and Espirito-Santo, F. (2014) Homogenization and assessment of observed near-surface wind speed trends over Spain and Portugal, 1961–2011. *Journal of Climate*, 27, 3692–3712. <https://doi.org/10.1175/JCLI-D-13-00652.1>.
- Azarin-Molina, C., Vicente-Serrano, S.M., Mcvicar, T.R., Revuelto, J., Jerez, S. and López-Moreno, J.I. (2016) Assessing the impact of measurement time interval when calculating wind speed means and trends under the stilling phenomenon. *International Journal of Climatology*, 37, 480–492. <https://doi.org/10.1002/joc.4720>.
- Bett, P.E., Thornton, H.E. and Clark, R.T. (2013) European wind variability over 140 yr. *Advances in Science and Research*, 10, 51–58. <https://doi.org/10.5194/asr-10-51-2013>.
- Bichet, A., Wild, M., Folini, D. and Schr, C. (2012) Causes for decadal variations of wind speed over land: sensitivity studies with a global climate model. *Geophysical Research Letters*, 39, 4–9. <https://doi.org/10.1029/2012GL051685>.
- Brands, S., Gutiérrez, J.M., Herrera, S. and Cofiño, A.S. (2012) On the use of reanalysis data for downscaling. *Journal of Climate*, 25, 2517–2526. <https://doi.org/10.1175/JCLI-D-11-00251.1>.
- Carvalho, D., Rocha, A., Gómez-Gesteira, M. and Silva Santos, C. (2017) Potential impacts of climate change on european wind energy resource under the CMIP5 future climate projections. *Renewable Energy*, 101, 29–40. <https://doi.org/10.1016/j.renene.2016.08.036>.
- Chávez-Arroyo, R. and Probst, O. (2015) Quality assurance of near-surface wind velocity measurements in Mexico. *Meteorological Applications*, 22, 165–177. <https://doi.org/10.1002/met.1432>.
- Chen, L., Li, D. and Pryor, S.C. (2013) Wind speed trends over China: quantifying the magnitude and assessing causality. *International Journal of Climatology*, 33, 2579–2590. <https://doi.org/10.1002/joc.3613>.
- Cheng, C.S. (2014) Evidence from the historical record to support projection of future wind regimes: an application to Canada. *Atmosphere - Ocean*, 52, 232–241. <https://doi.org/10.1080/07055900.2014.902803>.
- Cheng, C.S., Li, G., Li, Q., Auld, H. and Fu, C. (2012) Possible impacts of climate change on wind gusts under downscaled future climate conditions over Ontario, Canada. *Journal of Climate*, 25, 3390–3408. <https://doi.org/10.1175/JCLI-D-11-00198.1>.
- Cleugh, H.A., Miller, J.M. and Böhm, M. (1998) Direct mechanical effects of wind on crops. *Agroforestry Systems*, 41, 85–112. <https://doi.org/10.1023/A:1006067721039>.
- Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Jr., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D. and Worley, S.J. (2011) The twentieth century reanalysis project. *Royal Meteorological Society*, 137, 1–28. <https://doi.org/10.1002/qj.776>.
- Dafka, S., Xoplaki, E., Toreti, A., Zanis, P., Tyrlis, E., Zerefos, C. and Luterbacher, J. (2016) The Etesians: from observations to reanalysis. *Climate Dynamics*, 47, 1569–1585. <https://doi.org/10.1007/s00382-015-2920-7>.
- Darby, L.S. (2005) Cluster analysis of surface winds in Houston, Texas, and the impact of wind patterns on ozone. *Journal of Applied Meteorology*, 44, 1788–1806. <https://doi.org/10.1175/JAM2320.1>.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>.
- Devis, A., Van Lipzig, N.P. and Demuzere, M. (2018) Should future wind speed changes be taken into account in wind farm

- development? *Environmental Research Letters*, 13, 064012. <https://doi.org/10.1088/1748-9326/aabff7>.
- Dörenkämper, M., Olsen, B., Witha, B., Hahmann, A., Davis, N., Barcons, J., Ezber, Y., García-Bustamante, E., González Rouco, J.F., Navarro, J., Sastre-Marugá, M., Sile, T., Trei, W., Žagar, M., Badger, J., Gottschall, J., Sanz Rodrigo, J. and Mann, J. (2020) The making of the new European wind Atlas, part 2: production and evaluation. *Geoscientific Model Development Discussions*, 13, 5079–5102. <https://doi.org/10.5194/gmd-13-5079-2020>.
- Dunn, J.H.R., Willett, M.K., Parker, E.D. and Mitchell, L. (2016) Expanding HadISD: quality-controlled, sub-daily station data from 1931. *Geoscientific Instrumentation, Methods and Data Systems*, 5, 473–491. <https://doi.org/10.5194/gi-5-473-2016>.
- Dunn, R.J.H., Willett, K.M., Thorne, P.W., Woolley, E.V., Durre, I., Dai, A., Parker, D.E. and Vose, R.S. (2012) HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Climate of the Past*, 8, 1649–1679. <https://doi.org/10.5194/cp-8-1649-2012>.
- Durre, I., Menne, M.J., Gleason, B.E., Houston, T.G. and Vose, R.S. (2010) Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, 49, 1615–1633. <https://doi.org/10.1175/2010JAMC2375.1>.
- DWD, CDC (2018) Historical hourly station observations of wind speed and wind direction for Germany.
- Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the coupled model Intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>.
- Farquhar, G. and Roderick, M. (2007) Worldwide changes in evaporative demand. *Water Environment*, 2, 81–103.
- Fiebrich, C.A., Morgan, Y.R., McCombs, A.G., Hall, P.K. and McPherson, R.A. (2010) Quality assurance procedures for mesoscale meteorological data. *Journal of Atmospheric and Oceanic Technology*, 27, 1565–1582. <https://doi.org/10.1175/2010JTECHA1433.1>.
- Fujiwara, M., Wright, J.S., Manney, G.L., Gray, L.J., Anstey, J., Birner, T., Davis, S., Gerber, E.P., Harvey, V.L., Hegglin, M.I., Homeyer, C.R., Knox, J.A., Krüger, K., Lambert, A., Long, C.S., Martineau, P., Molod, A., Monge-Sanz, B.M., Santee, M.L., Tegtmeier, S., Chabrilat, S., Tan, D.G.H., Jackson, D.R., Polavarapu, S., Compo, G.P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J.S. and Zou, C.-Z. (2017) Introduction to the SPARC reanalysis Intercomparison project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, 17, 1417–1452. <https://doi.org/10.5194/acp-17-1417-2017>.
- García Bustamante, E., González Rouco, J.F., Navarro, J., Lucio Eceiza, E.E. and Rojas Labanda, C. (2021) Expected recurrence of extreme winds in northwestern sahara and associated uncertainties. *Energies*, 14(21), 6913. <https://doi.org/10.3390/en14216913>.
- García-Bustamante, E., González-Rouco, J.F., Jiménez, P.A., Navarro, J. and Montávez, J.P. (2009) A comparison of methodologies for monthly wind energy estimation. *Wind Energy*, 12, 640–659. <https://doi.org/10.1002/we.315>.
- García-Bustamante, E., Gonzalez-Rouco, J.F., Navarro, J., Xoplaki, E., Luterbacher, J., Jimenez, P.A., Montalvez, J.P., Hidalgo, A. and Lucio-Eceiza, E.E. (2013) Relationship between wind power production and North Atlantic atmospheric circulation over the northeastern Iberian Peninsula. *Climate Dynamics*, 40, 935–949. <https://doi.org/10.1007/s00382-012-1451-8>.
- Guo, H., Xu, M. and Hu, Q. (2011) Changes in near-surface wind speed in China: 1969–2005. *International Journal of Climatology*, 31, 349–358. <https://doi.org/10.1002/joc.2091>.
- Hahmann, A., Sile, T., Witha, B., Davis, N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González Rouco, J.F., Navarro, J., Olsen, B. and Söderberg, S. (2020) The making of the new European wind Atlas, part 1: model sensitivity. *Geoscientific Model Development Discussions*, 13, 5053–5078. <https://doi.org/10.5194/gmd-13-5053-2020>.
- Hannachi, A., Jolliffe, I.T. and Stephenson, D.B. (2007) Empirical orthogonal functions and related techniques in atmospheric science: a review. *International Journal of Climatology*, 27, 1119–1152. <https://doi.org/10.1002/joc.1499>.
- Harris, I., Osborn, T.J., Jones, P. and Lister, D. (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7, 109. <https://doi.org/10.1038/s41597-020-0453-3>.
- Hartmann, D.L., Klein Tank, A.M.G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R., Kaplan, A., Soden, B.J., Thorne, P.W., Wild, M., and Zhai, P.M. (2013) Observations: atmosphere and surface. In: Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. pp. 159–254. <https://doi.org/10.1017/CBO9781107415324.008>.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Jiang, Q., Smith, R.B. and Doyle, J. (2003) The nature of the mistral: observations and modelling of two MAP events. *Quarterly Journal of the Royal Meteorological Society*, 129, 857–875. <https://doi.org/10.1256/qj.02.21>.
- Jiang, Y., Luo, Y., Zhao, Z. and Tao, S. (2010) Changes in wind speed over China during 1956–2004. *Theoretical and Applied Climatology*, 99, 421–430. <https://doi.org/10.1007/s00704-009-0152-7>.
- Jiménez, P.A., González-Rouco, J.F., Montávez, J.P., García-Bustamante, E. and Navarro, J. (2009) Climatology of wind patterns in the northeast of the Iberian Peninsula. *International Journal of Climatology*, 29, 501–525. <https://doi.org/10.1002/joc.1705>.

- Jiménez, P.A., García-Bustamante, E., González-Rouco, J.F., Valero, F., Montávez, J.P. and Navarro, J. (2008) Surface wind regionalization in complex terrain. *Journal of Applied Meteorology and Climatology*, 47, 308–325. <https://doi.org/10.1175/2007JAMC1483.1>.
- Jiménez, P.A., González-Rouco, J.F., Navarro, J., Montávez, J.P. and García-Bustamante, E. (2010) Quality assurance of surface wind observations from automated weather stations. *Journal of Atmospheric and Oceanic Technology*, 27, 1101–1122. <https://doi.org/10.1175/2010JTECHA1404.1>.
- Karnauskas, K.B., Lundquist, J.K. and Zhang, L. (2018) Southward shift of the global wind energy resource under high carbon dioxide emissions. *Nature Geoscience*, 11, 38–43. <https://doi.org/10.1038/s41561-017-0029-9>.
- Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A. and Deutschländer, T. (2013) Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre. *Advances in Science and Research*, 10, 99–106. <https://doi.org/10.5194/asr-10-99-2013>.
- Kaspar, F., Niermann, D., Borsche, M., Fiedler, S., Keller, J., Potthast, R., Rösch, T., Spanghel, T. and Tinz, B. (2020) Regional atmospheric reanalysis activities at Deutscher Wetterdienst: review of evaluation results and application examples with a focus on renewable energy. *Advances in Science and Research*, 17, 115–128. <https://doi.org/10.5194/asr-17-115-2020>.
- Kent, E.C., Fangohr, S. and Berry, D.I. (2013) A comparative assessment of monthly mean wind speed products over the global ocean. *International Journal of Climatology*, 33, 2520–2541. <https://doi.org/10.1002/joc.3606>.
- Klein Tank, A.M., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A.F., Forland, E., Miletus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V. and Petrovic, P. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology*, 22, 1441–1453. <https://doi.org/10.1002/joc.773>.
- Kunkel, K.E., Andsager, K., Conner, G., Decker, W.L., Hillaker, H. J., Knox, P.N., Nurnberger, F.V., Rogers, J.C., Scheeringa, K., Wendland, W.M., Zandlo, J. and Angel, J.R. (1998) An expanded digital daily database for climatic resources applications in the Midwestern United States. *Bulletin of the American Meteorological Society*, 79, 1357–1366. [https://doi.org/10.1175/1520-0477\(1998\)079<1357:AEDDDF>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<1357:AEDDDF>2.0.CO;2).
- Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. and Rennie, J. (2011) An overview of the global historical climatology network monthly mean temperature data set, version 3. *Journal of Geophysical Research Atmospheres*, 116, 1–18. <https://doi.org/10.1029/2011JD016187>.
- Lorente-Plazas, R., Montávez, J.P., Jimenez, P.A., Jerez, S., Gómez-Navarro, J.J., García-Valero, J.A. and Jimenez-Guerrero, P. (2015) Characterization of surface winds over the Iberian Peninsula. *International Journal of Climatology*, 35, 1007–1026. <https://doi.org/10.1002/joc.4034>.
- Lott, J.N. (2004) The quality control of the integrated surface hourly database. *Conference proceedings (extended abstract). 84th AMS Annual Meeting*, pp. 5039–5045. Seattle, WA. <https://ams.confex.com/ams/pdfpapers/71929.pdf>.
- Lucio-Eceiza, E.E., González-Rouco, J.F., García-Bustamante, E., Navarro, J. and Beltrami, H. (2019) Multidecadal to centennial surface wintertime wind variability over northeastern North America via statistical downscaling. *Climate Dynamics*, 53, 41–66. <https://doi.org/10.1007/s00382-018-4569-5>.
- Lucio-Eceiza, E.E., González-Rouco, J.F., García-Bustamante, E., Navarro, J., Rojas-Labanda, C. and Beltrami, H. (2020) Summertime surface wind variability over northeastern North America at multidecadal to centennial time scales via statistical downscaling. *Journal of Climate*, 33, 1969–1990. <https://doi.org/10.1175/JCLI-D-19-0331.1>.
- Lucio-Eceiza, E.E., González-Rouco, J.F., Navarro, J. and Beltrami, H. (2018) Quality control of surface wind observations in northeastern North America. Part I: data management issues. *Journal of Atmospheric and Oceanic Technology*, 35, 163–182. <https://doi.org/10.1175/JTECH-D-16-0204.1>.
- Lucio-Eceiza, E.E., González-Rouco, J.F., Navarro, J., Beltrami, H. and Conte, J. (2018) Quality control of surface wind observations in northeastern North America. Part II: measurement errors. *Journal of Atmospheric and Oceanic Technology*, 35, 183–205. <https://doi.org/10.1175/JTECH-D-16-0205.1>.
- Mardia, K.V., and P.E. Jupp. (2009) *Directional statistics. Wiley Series in probability and statistics*. 494, 456. London, United Kingdom: John Wiley & Sons. <https://doi.org/10.1002/9780470316979>
- Marsac, F. and Le Blanc, J.-L. (1998) Interannual and enso-associated variability of the coupled ocean-atmosphere system with possible impacts on the yellowfin tuna fisheries of the Indian and Atlantic oceans. *Collective Volume of Scientific Papers - International Commission for the Atlantic Tunas*, 50, 345–377.
- McVicar, T.R., Roderick, M.L., Donohue, R.J., Li, L.T., Van Niel, T.G., Thomas, A., Grieser, J., Jhajharia, D., Himri, Y., Mahowald, N.M., Mescherskaya, A.V., Kruger, A.C., Rehman, S. and Dinpashoh, Y. (2012) Global review and synthesis of trends in observed terrestrial near-surface wind speeds: implications for evaporation. *Journal of Hydrology*, 416–417, 182–205. <https://doi.org/10.1016/j.jhydrol.2011.10.024>.
- Minola, L., Azorin-Molina, C. and Chen, D. (2016) Homogenization and assessment of observed near-surface wind speed trends across Sweden, 1956–2013. *Journal of Climate*, 29, 7397–7415. <https://doi.org/10.1175/JCLI-D-15-0636.1>.
- Molina, M.O., Gutiérrez, C. and Sánchez, E. (2021) Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *International Journal of Climatology*, 41, 1–15. <https://doi.org/10.1002/joc.7103>.
- NCEP (1980) NCEP ADP Operational Global Surface Observations, February 1975–February 2007. Retrieved from <http://rda.ucar.edu/datasets/ds464.0/>. [Accessed January 13, 2017].
- NCEP (2004) NCEP ADP global surface observational weather data, October 1999 - continuing. Retrieved from <https://doi.org/10.5065/4F4P-E398>. [Accessed 13-January-2017].

- NCEP (2005) NCDC TD3505 Integrated Surface Hourly Data. Retrieved from <http://rda.ucar.edu/datasets/ds463.3/>. [Accessed 13-January-2017].
- Olauson, J. (2018) ERA5: the new champion of wind power modelling? *Renewable Energy*, 126, 322–331. <https://doi.org/10.1016/j.renene.2018.03.056>.
- Pinard, J.P. (2007) Wind climate of the Whitehorse area. *Arctic*, 60, 227–237. <https://doi.org/10.14430/arctic215>.
- Preisendorfer, R. W. (1988) Principal component analysis in meteorology and oceanography.
- Pryor, S.C., Barthelmie, R.J. and Takle, G.S. (2009) Wind speed trends over the contiguous USA. *IOP Conference Series: Earth and Environmental Science*, 6, 092023. <https://doi.org/10.1088/1755-1307/6/9/092023>.
- Ramón, J., Lledó, L., Torralba, V., Soret, A. and Doblas-Reyes, F. (2019) What global reanalysis best represents near surface winds? *Quarterly Journal of the Royal Meteorological Society*, 145, 3236–3251. <https://doi.org/10.1002/qj.3616>.
- Reyers, M., Moemken, J. and Pinto, J.G. (2016) Future changes of wind energy potentials over Europe in a large CMIP5 multi-model ensemble. *International Journal of Climatology*, 36, 783–796. <https://doi.org/10.1002/joc.4382>.
- Rennie, J.J., Lawrimore, J.H., Gleason, B.E., Thorne, P.W., Morice, C. P., Menne, M.J., Williams, C.N., de Almeida, W.G., Christy, J.R., Flannery, M., Ishihara, M., Kamiguchi, K., Klein-Tank, A. M.G., Mhanda, A., Lister, D.H., Razuvaev, V., Renom, M., Rusticucci, M., Tandy, J., Worley, S.J., Venema, V., Angel, W., Brunet, M., Dattore, B., Diamond, H., Lazzara, M.A., le Blancq, F., Luterbacher, J., Mächel, H., Revadekar, J., Vose, R.S. and Yin, X. (2014) The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geoscience Data Journal*, 1, 75–102. <https://doi.org/10.1002/gdj3.8>.
- Richman, M.B. (1986) Review article: rotation of principal components. *Journal of Climatology*, 6, 293–335. <https://doi.org/10.1002/JOC.3370060305>.
- Roderick, M.L., Rotstayn, L.D., Farquhar, G.D. and Hobbins, M.T. (2007) On the attribution of changing pan evaporation. *Geophysical Research Letters*, 34, 1–6. <https://doi.org/10.1029/2007GL031166>.
- Schuster, D. (2003) 2 x wind speed problem. Retrieved from [http://dss.ucar.edu/datasets/ds464.0/docs/2xwind\\_final.pdf](http://dss.ucar.edu/datasets/ds464.0/docs/2xwind_final.pdf).
- Smith, A., Lott, N. and Vose, R. (2011) The integrated surface database: recent developments and partnerships. *Bulletin of the American Meteorological Society*, 92, 704–708. <https://doi.org/10.1175/2011BAMS3015.1>.
- Smith, R.B. (1987) Aerial observations of the Yugoslavian bora. *Journal of the Atmospheric Sciences*, 44, 269–297.
- Staffell, I. and Pfenninger, S. (2016) Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, 114, 1224–1239. <https://doi.org/10.1016/j.energy.2016.08.068>.
- Standen, J., Wilson, C., Vosper, S. and Clark, P. (2017) Prediction of local wind climatology from met Office models: virtual met mast techniques. *Wind Energy*, 20, 411–430. <https://doi.org/10.1002/we.2013>.
- von Storch, H. and Zwiers, F.W. (1999) Misuses of statistical analysis in climate research. In: *Statistical Analysis in Climate Research*. New York: Cambridge University Press.
- Tian, Q., Huang, G., Hu, K. and Niyogi, D. (2019) Observed and global climate model based changes in wind power potential over the northern hemisphere during 1979–2016. *Energy*, 167, 1224–1235. <https://doi.org/10.1016/j.energy.2018.11.027>.
- Tobin, I., Greuell, W., Jerez, S., Ludwig, F., Vautard, R., Van Vliet, M.T. and Breón, F.M. (2018) Vulnerabilities and resilience of European power generation to 1.5°C, 2°C and 3°C warming. *Environmental Research Letters*, 13, 044024. <https://doi.org/10.1088/1748-9326/aab211>.
- Tobin, I., Jerez, S., Vautard, R., Thais, F., Van Meijgaard, E., Prein, A., Déqué, M., Kotlarski, S., Maule, C.F., Nikulin, G., Noël, T. and Teichmann, C. (2016) Climate change impacts on the power generation potential of a European mid-century wind farms scenario. *Environmental Research Letters*, 11, 034013. <https://doi.org/10.1088/1748-9326/11/3/034013>.
- Tobin, I., Vautard, R., Balog, I., Bréon, F.M., Jerez, S., Ruti, P.M., Thais, F., Vrac, M. and Yiou, P. (2015) Assessing climate change impacts on European wind energy from ENSEMBLES high-resolution climate projections. *Climatic Change*, 128, 99–112. <https://doi.org/10.1007/s10584-014-1291-0>.
- Tokinaga, H. and Xie, S.P. (2011) Wave- and anemometer-based sea surface wind (WASWind) for climate change analysis. *Journal of Climate*, 24, 267–285. <https://doi.org/10.1175/2010JCLI3789.1>.
- Torralba, V., Doblas-Reyes, F., MacLeod, D., Christel, I. and Davis, M. (2017) Seasonal climate prediction: a new source of information for the management of wind energy resources. *Journal of Applied Meteorology and Climatology*, 56, 1231–1247. <https://doi.org/10.1175/JAMC-D-16-0204.1>.
- Troen, I. and Lundtang Petersen, E. (1989) *European wind atlas*. Roskilde: Risø National Laboratory.
- Usbeck, T., Wohlgemuth, T., Pfister, C., Volz, R., Beniston, M. and Dobbertin, M. (2010) Wind speed measurements and forest damage in Canton Zurich (Central Europe) from 1891 to winter 2007. *International Journal of Climatology*, 30, 347–358. <https://doi.org/10.1002/joc.1895>.
- Valero, F., Gonzalez, J.F., Doblas, F.J. and García-Miguel, J.A. (1996) A method for the reconstruction and temporal extension of climatological time series. *International Journal of Climatology*, 16, 213–227. [https://doi.org/10.1002/\(SICI\)1097-0088\(199602\)16:2<213::AID-JOC996>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1097-0088(199602)16:2<213::AID-JOC996>3.0.CO;2-7).
- Vautard, R., Cattiaux, J., Yiou, P., Thépaut, J.-N. and Ciais, P. (2010) Northern hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nature Geoscience*, 3, 756–761. <https://doi.org/10.1038/ngeo979>.
- Wan, H., Wang, X.L. and Swail, V.R. (2010) Homogenization and trend analysis of Canadian near-surface wind speeds. *Journal of Climate*, 23, 1209–1225. <https://doi.org/10.1175/2009JCLI3200.1>.
- Wentz, F.J., Ricciardulli, L., Hilburn, K. and Mears, C. (2007) How much more rain will global warming bring? *Science*, 317, 233–235. <https://doi.org/10.1126/science.1140746>.
- Wever, N. (2012) Quantifying trends in surface roughness and the effect on surface wind speed observations. *Journal of Geophysical Research Atmospheres*, 117, 1–14. <https://doi.org/10.1029/2011JD017118>.

- WMO (2008) Guide to meteorological instruments and methods of observation. *Tech. Rep. WMO-8*, World Meteorological Organization.
- WMO (2015) Manual on the global telecommunication system. Annex III to the WMO technical regulations. *Tech. Rep. WMO-386*, World Meteorological Organization.
- Wohland, J., Reyers, M., Weber, J. and Witthaut, D. (2017) More homogeneous wind conditions under strong climate change decrease the potential for inter-state balancing of electricity in Europe. *Earth system Dynamics (Pembroke, Ont.)*, 8, 1047–1060. <https://doi.org/10.5194/esd-8-1047-2017>.
- Wu, J., Zha, J., Zhao, D. and Yang, Q. (2018) Changes in terrestrial near-surface wind speed and their possible causes: an overview. *Climate Dynamics*, 51, 2039–2078. <https://doi.org/10.1007/s00382-017-3997-y>.
- Young, I.R., Zieger, S. and Babanin, A.V. (2011) Global trends in wind speed and wave height. *Science*, 332, 451–455. <https://doi.org/10.1126/science.1197219>.
- Zahradníček, P., Brázdil, R., Štěpánek, P. and Řezníčková, L. (2019) Differences in wind speeds according to measured and homogenized series in The Czech Republic, 1961–2015. *International Journal of Climatology*, 39, 235–250. <https://doi.org/10.1002/joc.5800>.
- Zeng, Z., Ziegler, A.D., Searchinger, T., Yang, L., Chen, A., Ju, K., Piao, S., Li, L.Z.X., Ciais, P., Chen, D., Liu, J., Azorin-molina, C., Chappell, A., Medvigy, D. and Wood, E.F. (2019) A reversal in global terrestrial stilling and its implications for wind energy production. *Nature Climate Change*, 9, 979–985. <https://doi.org/10.1038/s41558-019-0622-6>.
- Zhang, H.M., Bates, J.J. and Reynolds, R.W. (2006) Assessment of composite global sampling: sea surface wind speed. *Geophysical Research Letters*, 33, L17714. <https://doi.org/10.1029/2006GL027086>.

**How to cite this article:** Rojas-Labanda, C., González-Rouco, F., García-Bustamante, E., Navarro, J., Lucio-Eceiza, E. E., Van der Schrier, G., & Kaspar, F. (2023). Surface wind over Europe: Data and variability. *International Journal of Climatology*, 43(1), 134–156. <https://doi.org/10.1002/joc.7739>