Energy and AI 13 (2023) 100269

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Ensemble Machine Learning greatly improves ERA5 skills for wind energy applications

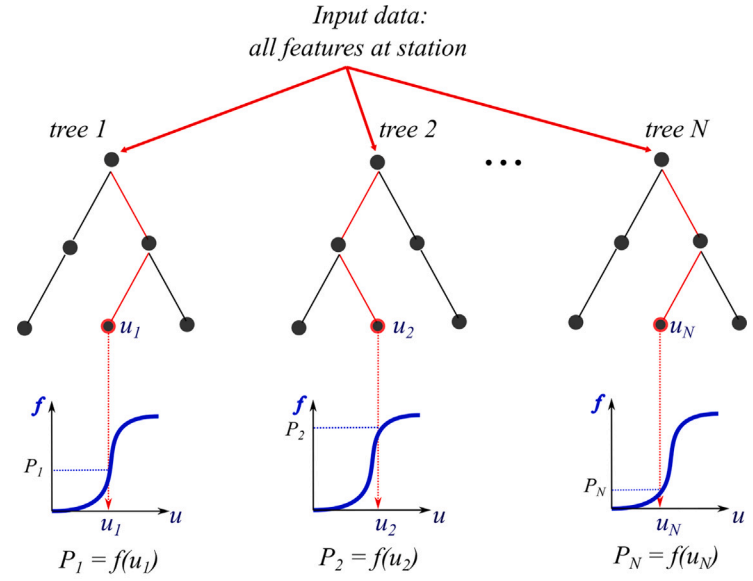
Mattia Cavaiola [a,b,*], Peter Enos Tuju [b], Francesco Ferrari [a], Gabriele Casciaro [c], Andrea Mazzino [a,b]

[a] DICCA, Department of Civil, Chemical and Environmental Engineering. University of Genoa, Via Montallegro 1, Genoa, 16145, Genoa, Italy
[b] INFN, National Institute of Nuclear Physics, Genoa section, Via Dodecaneso 33, Genoa, 16146, Genoa, Italy
[c] EGO Data S.r.l., Via Felice Romani 9/5, Genoa, 16122, Genoa, Italy

HIGHLIGHTS

- ERA5 skills for wind energy predictions are assessed in a time span of 20 years.
- Climatology-based wind-energy predictions systematically outperform raw ERA5.
- Our ensemble ML-based calibration brings out the ERA5 value.
- Our best ensemble ML-based based calibration outperforms a 3-km ERA5 downscaling.
- Costly high-resolution downscaling can be avoided for wind energy applications.

GRAPHICAL ABSTRACT



$$P = \frac{1}{N}\sum_{i=1}^{N} P_i \neq f\left(\frac{1}{N}\sum_{i=1}^{N} u_i\right)$$

ARTICLE INFO

Keywords:
Wind energy
ERA5 reanalysis
Quantile random forest

ABSTRACT

The skill of ERA5 has been assessed in relation to the prediction of the wind energy associated with 28 SYNOP stations located in Italy for a time span of 20 years (2001–2020). For comparison, a WRF-based high-resolution downscaling (3 km horizontally) was also produced for the same period. We found that simple predictions based on materialized past wind measures outperform the wind energy predictions from ERA5.

* Corresponding author.
E-mail addresses: mattia.cavaiola@edu.unige.it (M. Cavaiola), peter.enos.tuju@edu.unige.it (P.E. Tuju), francesco.ferrari@unige.it (F. Ferrari), gabriele.casciaro.r@gmail.com (G. Casciaro), andrea.mazzino@unige.it (A. Mazzino).

https://doi.org/10.1016/j.egyai.2023.100269
Received 16 January 2023; Received in revised form 28 April 2023; Accepted 1 May 2023
Available online 9 May 2023

Machine learning calibration

This result can be ascribed to the particularly complex characteristics of the Italian territory. Motivated by this expected behavior, we have implemented a Quantile Random Forest (QRF) calibration which greatly alleviates the problems encountered in the ERA5 reanalysis dataset. This technique provides a calibrated ensemble prediction system for the wind speed at the station. Surprisingly, the calibrated ERA5 outperforms wind energy estimations from the high-resolution 3-km downscaling. Once properly calibrated, the high-resolution downscaling provides predictions very similar to the calibrated ERA5. Limiting our conclusions to the estimation of wind energy over a long time span (here 20 years), having at disposal a high-resolution wind-field dataset does not necessarily mean greater accuracy. A careful calibration of the original coarser wind-field dataset produces better results than the raw high-resolution dataset.

1. Introduction

In recent years, wind energy production from the wind industry has proven a strong and positive growth trend. It is expected that before the year 2025 the production of wind energy from the wind industry will exceed 1 TW in global cumulative installations of onshore and offshore wind energy [1]. The wind sector has thus become a primary energy policy among several European countries for decreasing carbon dioxide emissions thus attenuating climate change impacts. Its contribution to 2030 and 2050 EU greenhouse emission reduction targets will be thus crucial.

The wind is however highly intermittent in space and in time thus its prediction presents the most challenging task, even for the shortest look-ahead forecast horizons of interest for both the energy industry and the power system regulators [1]. The issue of having accurate predictions for wind power is however even more dramatic. The relationship between wind speed and wind power is indeed cubic implying that a given relative error in the wind speed prediction becomes a factor 3 larger for the wind power. The intermittent nature of the wind source also causes difficulties/inaccuracies to assess the wind potential of a given area. Indeed, observed data are rarely available for a sufficiently long time in the past in regions having a potential interest in wind energy exploitation. One thus has to resort to reanalysis models [2–5] to estimate such a potential [6].

From a very general point of view, a reanalysis is a state-of-the-art strategy to produce datasets for climate monitoring. Reanalyses are created via a suitable data assimilation scheme and numerical weather prediction (NWP) models which ingest all available observations (e.g., data from radiosondes, satellites, SYNOP stations, buoys, air crafts, and ships) every 6–12 h over the period being analyzed. Observations distributed irregularly in space and time are in this way synthesized by NWP models into global regularly gridded meteorological datasets at high frequencies.

Although reanalysis datasets possess several key strengths (e.g., they are global datasets, they are built with a consistent spatial and temporal resolution covering several decades, they include hundreds of variables, their spatial resolution continues to increase, they incorporate millions of observations into a stable data assimilation system) some limitations are also present. The most important is probably due to changes in the observing system, affecting its reliability. Observations can indeed considerably vary depending on the location, time period, and variable considered. Spurious variability and non-physical trends may be thus present in the record. Despite the fact that a reanalysis resolution has been continuously increased over the years, it is still too coarse for some applications [7], in particular for applications related to the assessment of wind potential in regions of complex terrain [8]. Focusing on the Swiss Alps, the authors of [8] concluded that both reanalyses considered in their study, with resolution of 2 and 6 km, are insufficient to reproduce site-specific wind speeds in Switzerland's complex terrain.

A similar conclusion has been drawn by [9] where authors concluded that using the ERA5 reanalysis from ECMWF [10] in place of higher-resolution regional reanalysis products should be avoided when addressing sites with high variation of topography and, in particular, land use. A comprehensive recent review on those aspects has been provided by [6]. On the other hand, studies performed in northern Europe showed good ERA5 performances also in relation to extreme events [11–14]. Moreover, comparisons between ERA5 and other reanalysis products show as ERA5 is the most suitable data source for wind energy applications [15,16] in regions with gentle orography. There is thus a clear evidence that ERA5 skills do strongly depend on the complexity of the terrain, a fact which is not surprising in view of the coarse resolution of the dataset which necessarily ignores topographic variations having wavelengths smaller than about 30 km.

Our paper offers a change of paradigm with respect to the current scientific knowledge on the supposed uselessness of coarse-resolution reanalysis as ERA5, for the assessment of wind potential in regions of complex terrain. To quantitatively ascertain this fact, a time span of 20 years (from 2001 to 2020) will be considered for a set of 28 different locations where we will provide different estimations for the wind energy accumulated over the 20 years and the corresponding degree of variability (i.e. the wind-energy variance). Being able to capture the correct variance of the wind energy signal is important when a given location is a potential candidate for wind energy exploitation. Indeed, among two sites having the same accumulated wind energy, the one associated with the smallest variance is surely preferable. Our estimates will be quantitatively compared against wind power obtained from ground measurements of the 10-m wind speed available from official SYNOP stations. The study area selected for the present work is a subset of the Italian territory (see Fig. 1 where pink crosses refer to the SYNOP stations). Italy indeed presents a huge variety in its orographic shape, ranging from alpine regions, with station elevations up to 3500 m, to flat areas as, e.g., the Padana valley, the wind regime of which is however strongly affected by the surrounding Alps. Moreover, the mutual interaction between land and sea circulations makes wind prediction even more challenging [17]. To estimate the 20-year accumulated wind energy and its variance, different strategies will be considered. Two of them belong to the realm of differential-equation-based strategies: the coarse-in-space world-class global atmospheric reanalysis ERA5 from ECMWF ([10], resolution of 30 km horizontally) and its high-resolution (3 km horizontally) downscaling based on the Weather Research and Forecasting (WRF) model [18]. While ERA5 is freely available and easy to handle, its high-resolution downscaling is demanding in terms of computational costs, especially for a 20-year horizon. The downscaling strategy is nowadays considered as the most accurate strategy to deal with small-scale (e.g. orographically-induced) effects. The main question addressed here is: can a robust AI-based strategy, easy to train and economic in terms of computational cost, provide a relevant added value with respect to the costly high-resolution downscaling, when used in synergy with ERA5? Said in equivalent terms, can the features from ERA5 be used to train a suitable AI-based strategy to obtain a prediction with higher skills than the WRF-based (costly) strategy? To answer these questions, the so-called Quantile Random Forest (QRF) [19,20], an evolution of the more popular Random Forest (RF) algorithm [19], is exploited here. Several reasons motivated our choice. (i) QRF is robust to overfitting, easy to train and computationally cheap; (ii) QRF being and ensemble methods, it is particularly suitable to pass from the wind speed prediction to the wind energy prediction: each ensemble member must be separately mapped from the wind speed to the wind power, member by member, and successively processed to compute the relevant wind power/energy statistics; (iii) for the main paper message, it seems relevant that even a relatively simple AI strategy is

Fig. 1. The SYNOP stations considered in the present study (pink crosses) from 2001 to 2020. The background shows (part) of the Italian peninsula with colors denoting different orography elevations. Blue: sea areas, green: flat regions; brown: regions with orography, the elevation of which increases as it becomes darker and darker. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

enough to transform a prediction with low skills to a prediction greatly overcoming a costly high-resolution downscaling.

To anticipate the main paper conclusions, a clear evidence of the disruptive role of AI in capturing local effects (i) otherwise lost by coarse dynamical models; (ii) otherwise captured only via high-resolution dynamical downscaling, but at a steep price, will clearly emerge.

The paper is organized as follows. Section 2 describes the collected wind data from SYNOP stations over Italy and the ERA5 dataset, and Section 3 presents the motivation of our work. Section 4 compares simple predictions based on materialized past wind measures against the raw reanalysis-based predictions. In Section 5 we introduce the concept of calibration via EMOS and QRF, and in Section 6 we show the results of the calibrated reanalysis. Section 7 analyzes whether a high-resolution downscaling brings out an added value in term of accuracy.

Finally, Section 8 draws the conclusions.

2. Wind data

2.1. Observed data from SYNOP stations

The observation data used in this study are surface-based synoptic data from 28 different stations scattered in central and northern Italy. Synoptic observation stations are representative of an area up to 100 km around the station, but for local applications, the area can be considered to have a dimension of 10 km or less [21]. The station synoptic data used are retrieved for a period of 20 years from 1st January 2001 to 31st December 2020. The temporal frequency of the data is 3 h. The 3-h time frequency corresponds to synoptic hours at 00, 03, 06, 09, 12, 15, 18, and 21 UTC. Details on the synoptic station data used in the present study are provided as supplementary material.

2.2. The ERA5 reanalysis dataset

ERA5 is the fifth-generation atmospheric reanalysis of the global climate covering the period from January 1950 to the present. ERA5 is produced by the Copernicus Climate Change Service (C3S) at the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 provides hourly estimates of a large number of atmospheric, land, and oceanic climate variables. The data cover the Earth on a 30 km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80 km [22].

In the present study, in addition to the 10 m wind speed, we collected in the period 2001–2020 other atmospheric variables listed in the supplementary material, that will be used as meteorological features to train our ML algorithm.

3. Motivation of the work and method

Let us consider the issue of siting selection for planning wind energy development, and suppose to have at disposal two main ingredients: a sufficiently long record, say 1 year, of measured wind speed in a given site, and a long-term say 20 years, state-of-the-art reanalysis dataset, ERA5 in the present study. Also, imagine using the one-year measurement to extrapolate the wind energy accumulated over the entire period of 20 years. Let us define this parsimonious way to infer the wind energy in the 20-year time span, based on information collected in a sole year, as the climatology-based prediction (climatology in short, even if we are well aware that climatology focuses on the processes that create climate patterns and multi-decade variability in the atmospheric science jargon).

Our question now is whether the longer reanalysis may provide more accurate information than climatology on the wind energy associated to the site under consideration accumulated in a period of 20 years from 2001 to 2020. Imagine focusing the attention just on a specific site to which the ERA5 wind speed is associated via a standard bilinear interpolation; the same considerations will be trivially extended to all sites in Fig. 1 where wind energy must by evaluated. To evaluate the wind power necessary to compute the wind energy either from the measured wind speed or from the reanalysis dataset, we used in way of example a typical transfer function corresponding to a turbine having a capacity of 2 MW.

Provided the observed wind, and thus the associated wind energy accumulated in the 20 years from 2001 to 2020 (the truth), any prediction can be compared against the truth. Let E denote the true accumulated wind energy in the 20 years, E_c the predicted climatology based on one observed reference year, and E_h the prediction based on ERA5. While E and E_h are two single values, one may perform $N = 20$ different climatological predictions depending on the specific selected reference year at disposal. Let us thus replace E_c by E_c^i, $i = 1, \cdots, N$ to account for all possible climatologies.

4. Comparing the climatology-based predictions against the raw reanalysis-based predictions

Fig. 2 tells us which among climatology E_c and E_h better represents E. For each SYNOP station (along the x-axis, from north to south in Fig. 1) we have reported the following error indices:

$$\Delta_h = \frac{|E - E_h|}{\mathcal{E}} \qquad \Delta_c = \frac{1}{N}\sum_{i=1}^{N}\frac{|E_c^i - E|}{\mathcal{E}} \tag{1}$$

$\mathcal{E}$ being the normalization factor corresponding to the maximum energy one can extract from the site in the N years. Plus symbols (blue line) are relative to Δ_h, the error associated with the reanalysis; open circles (green line) represent Δ_c, the error associated with climatology. For the majority of sites, the climatology outperforms the reanalysis-based prediction. The same conclusion holds (not shown) if the year from which the climatology-based prediction is built is removed from the test set. The skill of the climatology-based estimation is highlighted in the inset of Fig. 2 where the skill score $((\Delta_c - \Delta_h)/\Delta_c) \times 100$ (see the appendix of [23]) is reported for all sites using as a reference error the one obtained from climatology. The fact that it is negative for 25 out of 28 sites confirms what we have already concluded from the main frame of the figure. The inset also shows that the error associated with

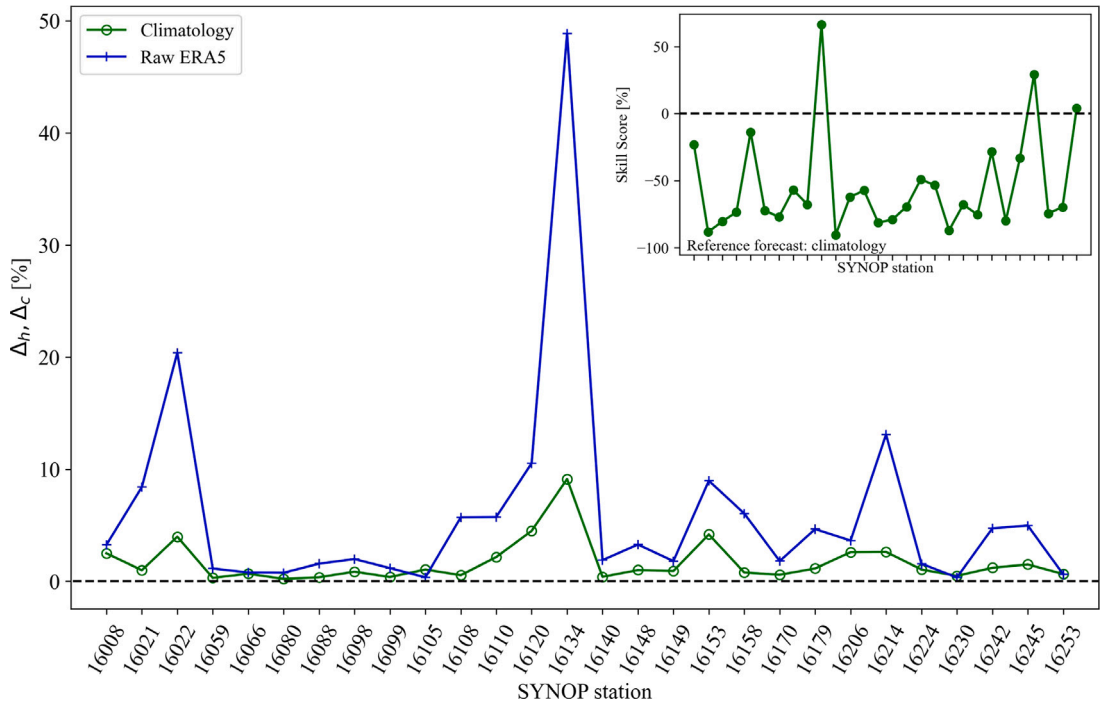


Fig. 2. Main frame: behavior of the two dimensionless error indices Δ_c and Δ_h (expressed as a percentage) measuring how much climatology (green line, open circles) and the reanalysis (blue line, crosses) are successful to predict the available wind energy accumulated in the 20-year time span for all sites reported in Fig. 1. Inset: the skill score of Δ_h using Δ_c as a reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

climatology is, by averaging over all stations, about 50% smaller than the error associated with the reanalysis-based prediction.

Let us now analyze the role of the length of the time period where observed data are supposed to be known (1 year in Fig. 2) and from which the climatology-based predictions are formulated. In Fig. 3 we address this issue by showing the skill score of Δ_h, averaged over all sites, using climatology as a reference. The abscissa reports different lengths (measured in months) where measurements are supposed to be available. The case of Fig. 2 corresponds to the abscissa taking the value 12 months. As expected, the added value of climatology with respect to the reanalysis-based predictions tends to reduce as the number of months where measures are available reduces. Quite surprisingly, climatology however still overcomes reanalysis-based predictions also for time intervals smaller than 1 month, even if they tend to be practically indistinguishable from each other.

Do our results suggest that the reanalysis has no value for wind energy site selection in regions of complex meteorological conditions as the one considered in the present study? As we will see in the next sections, a suitable calibration strategy can bring out the best from the reanalysis.

5. Calibration strategies: Quantile Random Forest (QRF), Random Forest (RF), and Ensemble Model Output Statistics (EMOS)

Although at first glance the results of the previous section might appear surprising, actually a modeling strategy with a spatial resolution necessarily coarse, because of the need to cover the entire globe, can hardly capture local effects. These effects are surely important in the Italian territory characterized by the interplay of orographic effects and sea-induced flows. Our aim here is to show that those small-scale contributions can be incorporated into the reanalysis by means of suitable Machine-Learning (ML) algorithms with a resulting added value. Our choice here falls on three well-known strategies: the Quantile Random Forest (QRF) [19,20] a supervised ML algorithm still in its infancy in relation to applications in the field of wind energy, the Random Forest (RF) [19], and a widely used statistical calibration strategy, the Ensemble Model Output Statistics (EMOS) [24–27]. Here, RF and EMOS are used as a benchmark, because of their simplicity, affordability, and widely-used character. The general ideas of the three methods are reported in the supplementary material.

6. Results on the calibrated reanalysis

Let us proceed in the same spirit of Section 3 and suppose to consider a year (among the 20) where wind speed measures are available and use these measures to train both the QRF and the RF, using the features reported in the supplementary material, and the standard EMOS algorithm. For the same reason for which different (20 in this specific example) climatology-based predictions can be constructed, we similarly have 20 possible QRF calibrations depending on the year one selects for the training. Let us denote $E^i_{QRF}, i = 1, 20$ such 20 predictions for the wind energy accumulated over the entire period of 20 years. The value of E^i_{QRF}, for each training year labeled by the index i is obtained from the predicted QRF wind speed ensemble upon transforming each wind speed member into a wind power member and finally a wind energy member. Having the entire distribution of members of the wind energy for each year, it is an easy task to obtain the mean of the distribution and then sum up to obtain the accumulated energy in the 20 years. The resulting total energy when using the ith year as a training set is just E^i_{QRF} as defined above. Given the predictions E^i_{QRF} the error with respect to the truth can be defined in analogy to (1):

$$\Delta_{QRF} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|E^i_{QRF} - E\right|}{\mathcal{E}} \tag{2}$$

The added value of the QRF-based reanalysis calibration can be detected from Fig. 4 where the skill score $((\Delta_c - \Delta_{QRF})/\Delta_c) \times 100$ is reported (filled squares, red line) for all sites using a reference error obtained from climatology. For 23 sites out of 28 this index is positive signaling now a clear added value with respect to the climatology-based prediction. For comparison, the blue line (plus symbols) refers to the same skill score as before but is now computed using the EMOS-based prediction. QRF clearly outperforms the standard EMOS as expected.

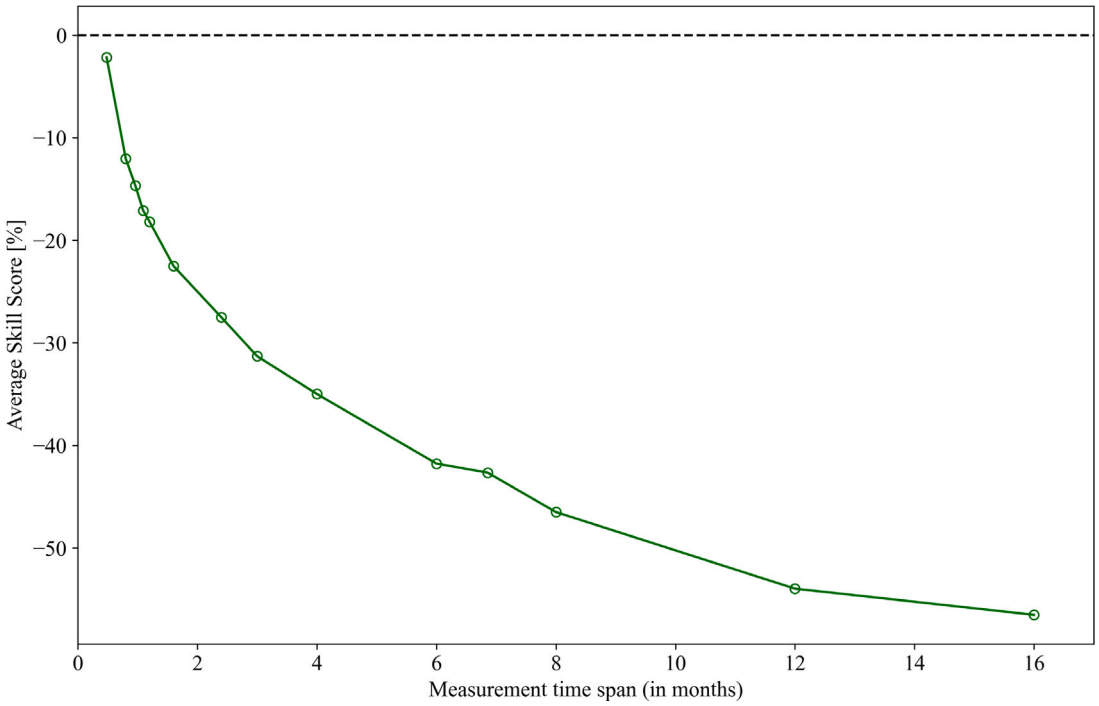


Fig. 3. Skill score of Δ_h, averaged over all SYNOP stations, using climatology as reference. The abscissa reports different durations of the hypothetical, measurement campaign (in months) where wind speed measurements are supposed to be available.

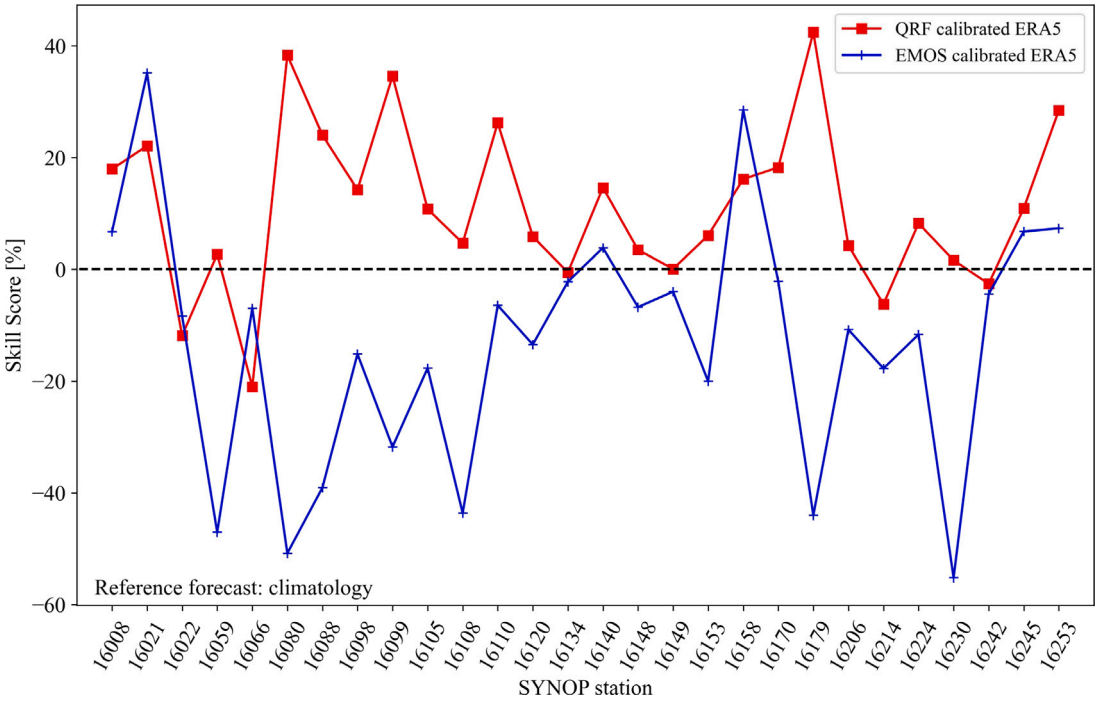


Fig. 4. Skill score of Δ_{QRF} using Δ_c as a reference (red line, filled squares) for all considered SYNOP stations (along the abscissa). The blue line (crosses) is analog but it refers to the EMOS-based calibration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It is now interesting to investigate the role of the measurement interval length (one year in the previous analysis) on the resulting skill of the calibrated reanalysis. This interval coincides with the period of training of EMOS, RF, and QRF strategies and also serves to build the climatology-based prediction. It is not a priori obvious the final outcome. Indeed, if on the one hand increasing the length of the training set would correspond to a more robust calibration, on the other hand for a longer measure interval the prediction based on climatology is also expected to achieve larger skills.

The results reported in Fig. 5 show an optimal length at about 7 months of training at which the skill score of the QRF error with respect to the error of the climatology-based prediction $((\Delta_c - \Delta_{QRF})/\Delta_c) \times 100$ averaged over all analyzed sites, reaches the peak (red line, filled

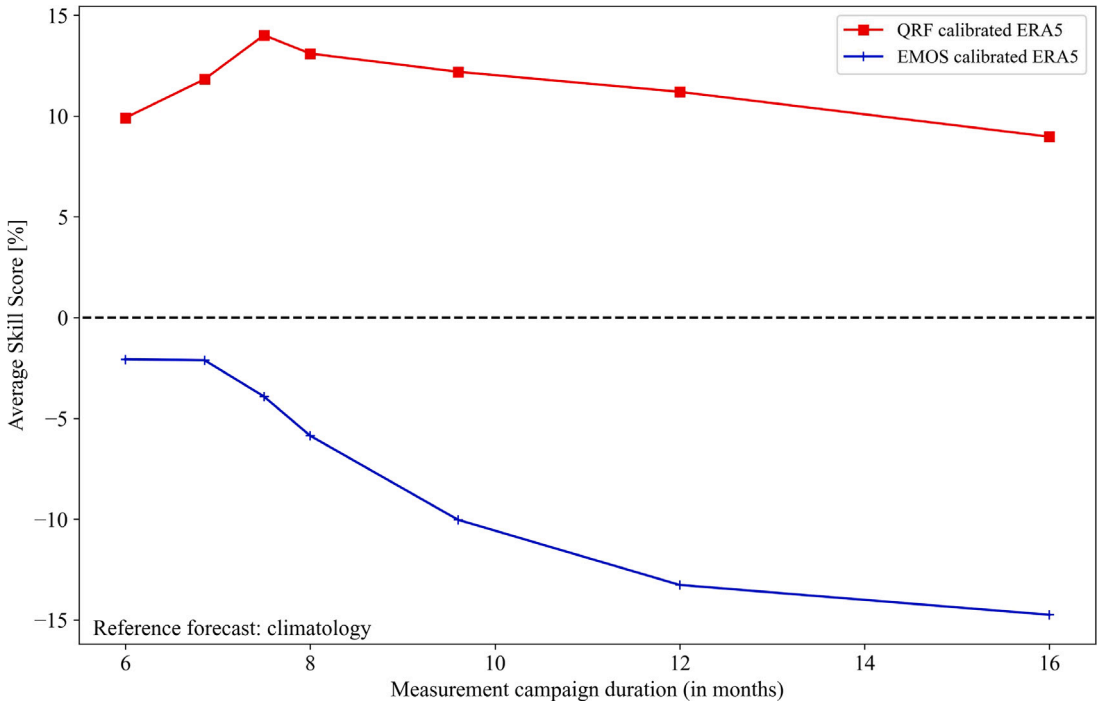


Fig. 5. Skill score of Δ_{QRF}, averaged over all SYNOP stations, using climatology as reference (red line, filled square). The abscissa reports different durations of the, hypothetical, measurement campaign (in months) where wind speed measurements are supposed to be available. The blue line (crosses) is as the red line but refers to the EMOS calibration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

squares). For longer training, climatology increases its importance remaining, however, less effective than the calibrated reanalysis. The EMOS strategy does not show a peak, with its skill score line being always negative and monotonically decreasing as the length of the training set increases. No added value is thus brought out by the EMOS strategy for all considered lengths of the training set.

We are now ready to show the superiority of the QRF strategy with respect to the RF strategy in predicting the accumulated wind energy in the 20-year interval. Fig. 6, analogous to Fig. 5, shows the skill score of Δ_{QRF}, averaged over all SYNOP stations, using the RF-based predictions as reference (main frame). The abscissa reports the different duration of the measurement campaign where wind speed measurements are supposed to be available.

How the skill score behaves for all SYNOP stations is shown in the inset of Fig. 6. The resulting skill scores are remarkably larger, up to 40%, signaling a consistently added value brought by the QRF strategy against the simpler, but more widely used RF strategy.

Having shown that the QRF-calibrated reanalysis outperforms both the climatology-based predictions and the RF-based predictions in estimating the wind energy accumulated in a time interval of 20 years, we now pass to address the question on whether the QRF calibration also brings an added value with respect to the raw reanalysis in relation to the description of the inter-annual variability of the wind energy in the N = 20 years. In doing that, let us fix to 1 year the length of the time span on which we want to assess the wind energy variability, while allowing to vary the length of the training period which coincides with the time span in which measurements are supposed to be available. Having defined E (see Section 3) the total wind energy in the 20-year time span considered, let us now define by E^i, $i = 1, \cdots, 20$ the wind energy in the ith year of the 20 years considered. In terms of E^i, we can describe the inter-annual variability of wind energy through the standard deviation;

$$\sigma_{tr} = \sqrt{\frac{\sum_{i=1}^{N}\left(E^i - \langle E^i\rangle\right)^2}{N}} \qquad \langle E^i\rangle = \frac{\sum_{i=1}^{N} E^i}{N} \tag{3}$$

Our aim is now to compare against the truth the analogous standard deviations from the calibrated ERA5 (via QRF and EMOS). Climatology based on 1-year measurements trivially gives no inter-annual variability and will thus not be considered. For the raw reanalysis, we define σ_h in strict analogy with (3) by simply replacing (i) E^i with the reanalysis-based prediction for the ith year, (ii) $\langle E^i\rangle$ with the average reanalysis-based annual wind energy. For the calibrated reanalysis we have M possible standard deviations, depending on the position of the selected time span (of length 20 years/M) corresponding to the training period. Namely, we define

$$\sigma^{(j)}_{QRF} = \sqrt{\frac{\sum_{i=1}^{N}\left(E^{(j)}_{QRF\,i} - \langle E^{(j)}_{QRF\,i}\rangle\right)^2}{N}} \qquad \langle E^{(j)}_{QRF\,i}\rangle = \frac{\sum_{i=1}^{N} E^{(j)}_{QRF\,i}}{N} \tag{4}$$

$E^{(j)}_{QRF\,i}$ denoting the QRF-based prediction of the wind energy in the year ith (among the 20) relative to the jth training set (among all possible different M training sets). In a similar way, we can define $\sigma^{(j)}_{EMOS}$, which differs from (4) simply because the predictions now come from the EMOS-based calibration. Having defined σ_h, $\sigma^{(j)}_{QRF}, \sigma^{(j)}_{EMOS}, j = 1, \cdots, M$ we can quantify the errors with respect to the truth σ_{tr}. This can be easily done, e.g., in terms of the following normalized error index:

$$\Delta_{\sigma_{QRF}} = \frac{\sum_{j=1}^{M}\left|\sigma^{(j)}_{QRF} - \sigma_{tr}\right|}{M\sigma_{tr}} \tag{5}$$

$$\Delta_{\sigma_{EMOS}} = \frac{\sum_{j=1}^{M}\left|\sigma^{(j)}_{EMOS} - \sigma_{tr}\right|}{M\sigma_{tr}} \tag{6}$$

$$\Delta_{\sigma_h} = \frac{\left|\sigma_h - \sigma_{tr}\right|}{\sigma_{tr}} \tag{7}$$

Fig. 7 (panel (a)) shows the behavior of $\Delta_{\sigma_{QRF}}$, $\Delta_{\sigma_{EMOS}}$, and Δ_{σ_h} for all considered SYNOP stations (abscissa). The superiority of the QRF calibration (red line, filled squares) in reproducing the wind energy inter-annual variability (for the majority of stations, 22 out of 28) clearly emerges from this figure both with respect to the raw reanalysis (red line, open circles) and with respect to the EMOS-based calibration (blue line, crosses). Panel (b) shows the skill score of the inter-annual variability quantified from the (i) QRF $((\Delta_{\sigma_h} - \Delta_{\sigma_{QRF}})/\Delta_{\sigma_h})$ using the raw

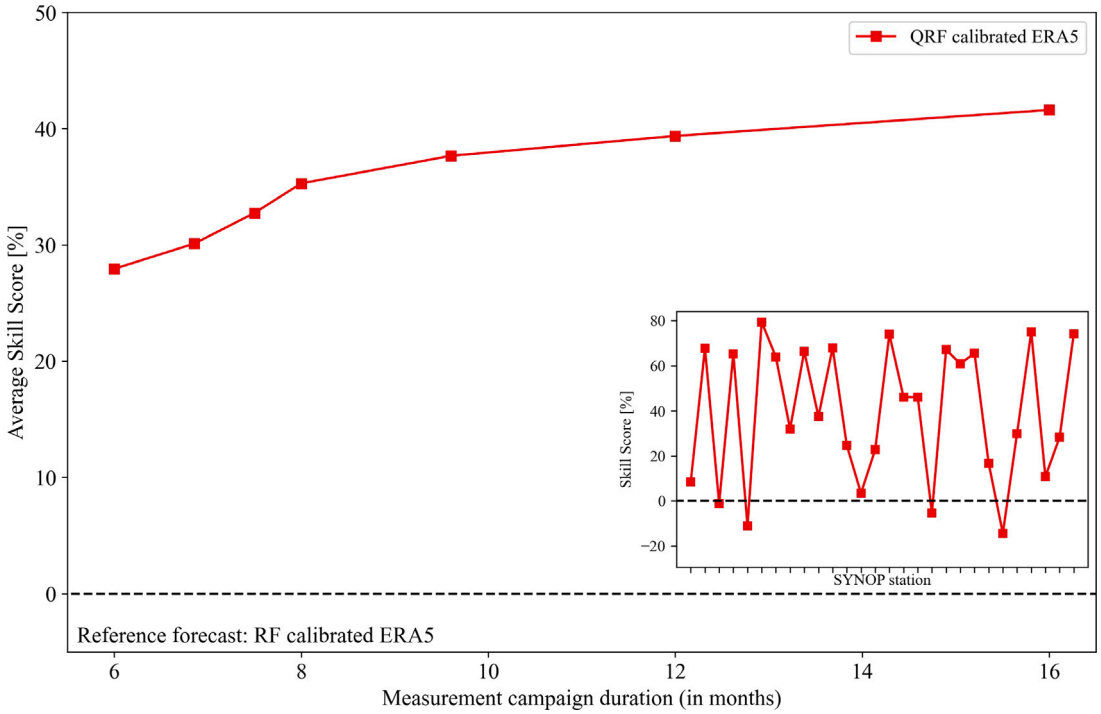


Fig. 6. Main frame: skill score of Δ_{QRF} averaged over all SYNOP stations, using the RF-based predictions as reference. The abscissa reports different durations of the hypothetical, measurement campaign (in months) where wind speed measurements are supposed to be available. Inset: for a training period of 12 months, the skill score of Δ_{QRF} using the RF-based predictions as a reference, is shown for all considered SYNOP stations.

reanalysis prediction as a reference (red line, filled squares), and (ii) EMOS $((\Delta_{\sigma_h} - \Delta_{\sigma_{EMOS}})/\Delta_{\sigma_h})$ again using the raw reanalysis prediction as a reference (blue line, crosses). The added value of the QRF-based calibration over the EMOS-based calibration is evident. For the QRF we have 22 stations out of 28 for which the skill score is positive (i.e. signaling an added value of the calibration with respect to the raw prediction in accounting for the inter-annual variability of the wind energy) but only 11 stations out of 28 manifest an added value when exploiting the EMOS strategy. We conclude by analyzing how the skill scores of Fig. 7, averaged over all SYNOP stations, depending on the length of the training set. From Fig. 8, it is possible to see that both considered calibrations only show minor variations in their performances, a fact indicating their robustness in assessing the inter-annual variability of the wind energy, for the analyzed range of lengths of the training set.

A quantitative evaluation of the skills of the ERA5 QRF-based calibration against the raw ERA5 estimations for the 10-m wind speed is provided in the supplementary material.

7. Does high-resolution downscaling always mean high accuracy?

In this section, we ask the question on whether or not a high-resolution downscaling may be competitive with respect to our best ML-based calibration. The question raises from a practical perspective related to wind energy assessment where high-resolution hindcasts are typically built from a dynamical downscaling where a coarser hindcast, as for example ERA5, provides the necessary initial/boundary conditions for the high-resolution simulations. In order to answer the question we performed a downscaling starting from ERA5 in the same period already considered in previous sections. In doing that, the Weather Research and Forecasting (WRF) model [18] has been used. The model configuration is as in [28] a part for the resolution that here is of 3 km for the area considered in Fig. 1 and 10 km for the remaining area covering the European region. Results are reported in Fig. 9. This figure is analogous to Fig. 4 and shows the skill score of Δ_{QRF} from the calibrated ERA5 now using Δ from WRF as a reference, both in its raw version (dashed line) and after calibration (continuous line). Calibrations have been performed via the same QRF used to calibrate ERA5 (also in relation to the used features). From this figure a clear answer to our question arises. The calibrated ERA5 is superior with respect to the raw WRF even if a high-resolution (3 km) is used. Differences between the calibrated WRF predictions and those based on the calibrated ERA5 are small, and negligible in practice.

8. Conclusion

The possible added value to estimate the 20-year time span wind energy has been explored by using the ERA5 dataset vs. its high-resolution downscaling based on WRF at a resolution of 3 km. 28 SYNOP stations located in Italy in a time span from 2001 to 2020 are considered to define the truth against which different predictive models are considered, including two machine learning algorithms known as Random Forest (RF) and Quantile Random Forest (QRF), a more parsimonious and well-known parametric calibration strategy referred to as EMOS, and a climatology-based prediction. This latter strategy determines the cumulative wind energy in a 20-year time span by simply multiplying by 20 the wind energy resulting from one year of observations. The results suggest that the climatology-based prediction, despite its simplicity, outperforms greatly the raw ERA5-based prediction. A possible reason for such a failure can be due to the complex characteristics of the Italian territory which is well-known to cause nontrivial dynamical sub-synoptical effects not properly captured by models too coarse in space.

Motivated by this result, we implemented a machine learning algorithm capable of providing not just a mean wind speed in a given station but rather, the whole probability density function (PDF) of the wind speed. Once this PDF is obtained one can easily determine the resulting wind power PDF from which the mean wind power, or the mean wind energy, can be computed. The selected ensemble calibration strategy has been the so-called Quantile Random Forest. As a result of our analysis, the QRF-based calibration brings out the power of the ERA5 reanalysis, showing a significant improvement with respect to

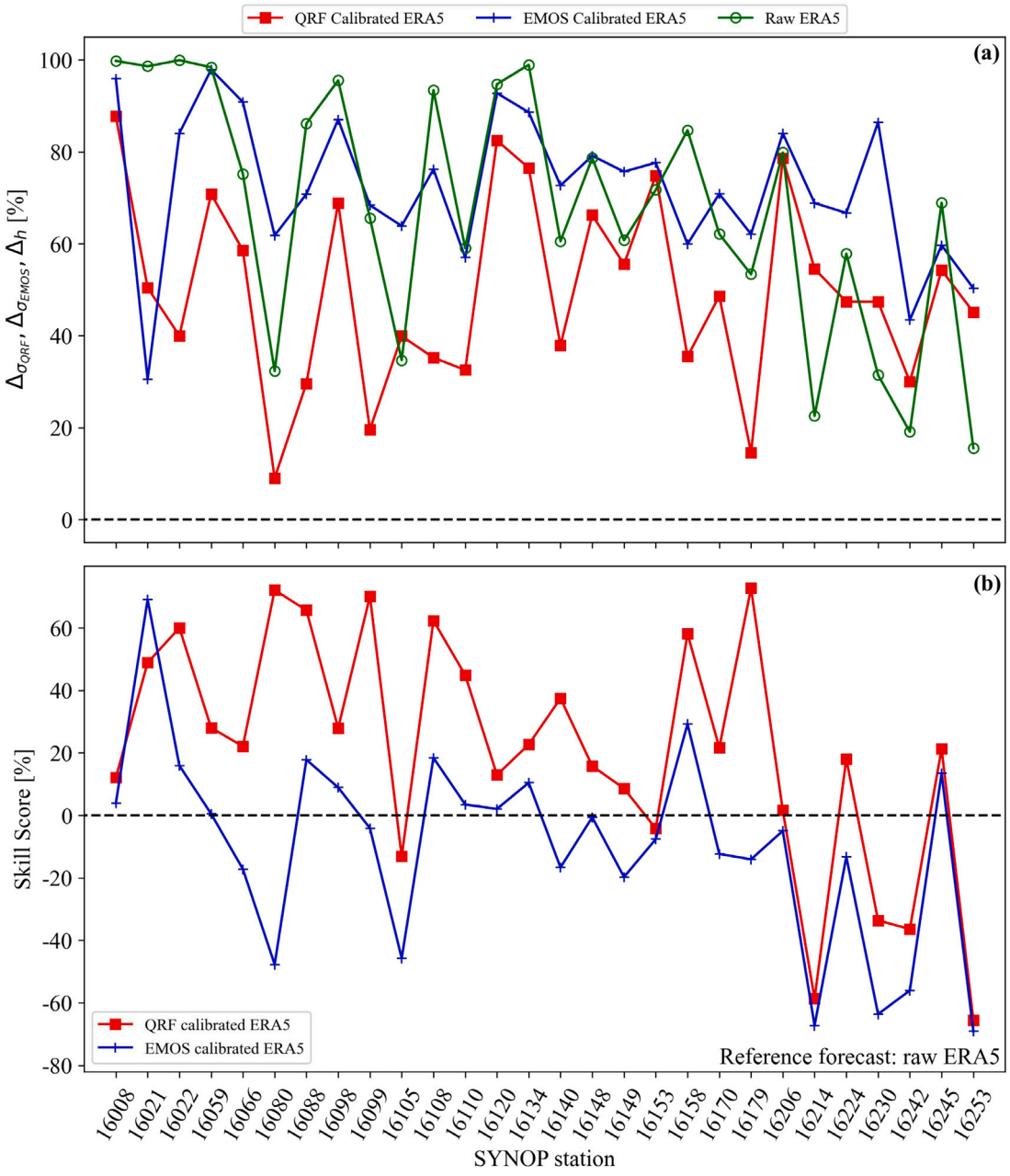


Fig. 7. Panel (a): Δ_{σ_h}, $\Delta_{\sigma_{QRF}}$, and $\Delta_{\sigma_{EMOS}}$ are shown for all analyzed SYNOP stations. Panel (b): Skill score of $\Delta_{\sigma_{QRF}}$ using Δ_{σ_h} as a reference (red line, filled squares) for all considered SYNOP stations (along the abscissa). The blue line (crosses) is analog but it refers to the EMOS-based calibration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

both the RF-based prediction, the climatology-based prediction, and the raw reanalysis. The improvement brought by the QRF calibration turned out to be not limited to the sole wind energy in the 20-year time span. It manifests also when assessing the inter-annual variability of the wind energy in the time span of 1 year.

The results of the QRF-based predictions have also been compared against the more parsimonious EMOS-based and RF-based predictions finding a significant added value of the first strategy with respect to the other two strategies.

Finally, we compared our best ERA5 calibration against its 3-km dynamic downscaling done with the WRF model, both in its raw version and with the same calibration used for ERA5. The results show that the calibrated ERA5 provides more accurate estimations of the 20-year time span wind energy than the raw WRF-based high-resolution downscaling. Differences become negligible when the raw WRF simulations are replaced by their calibrated counterpart done in the same way as for ERA5. In view of the fact that producing high-resolution hindcasts is expensive, both for the time it takes and for the cost of the computational resources, using a properly calibrated coarser hindcast appears a more convenient option.

Let us now provide some motivations on why our results can be of interest for the wind energy industry despite the fact that, in a wind park, wind turbines operate at heights higher than 10 m a.g.l. A first possible objection in this respect is that SYNOP stations can be heavily affected by the terrain and land-use around them. Actually, also assuming that local phenomena affect the measured wind velocity from SYNOP stations, this seems an important strength from wind energy applications. A wind farm is indeed an environment strongly perturbed by wake effects. Being able to capture this kind of nonlocal

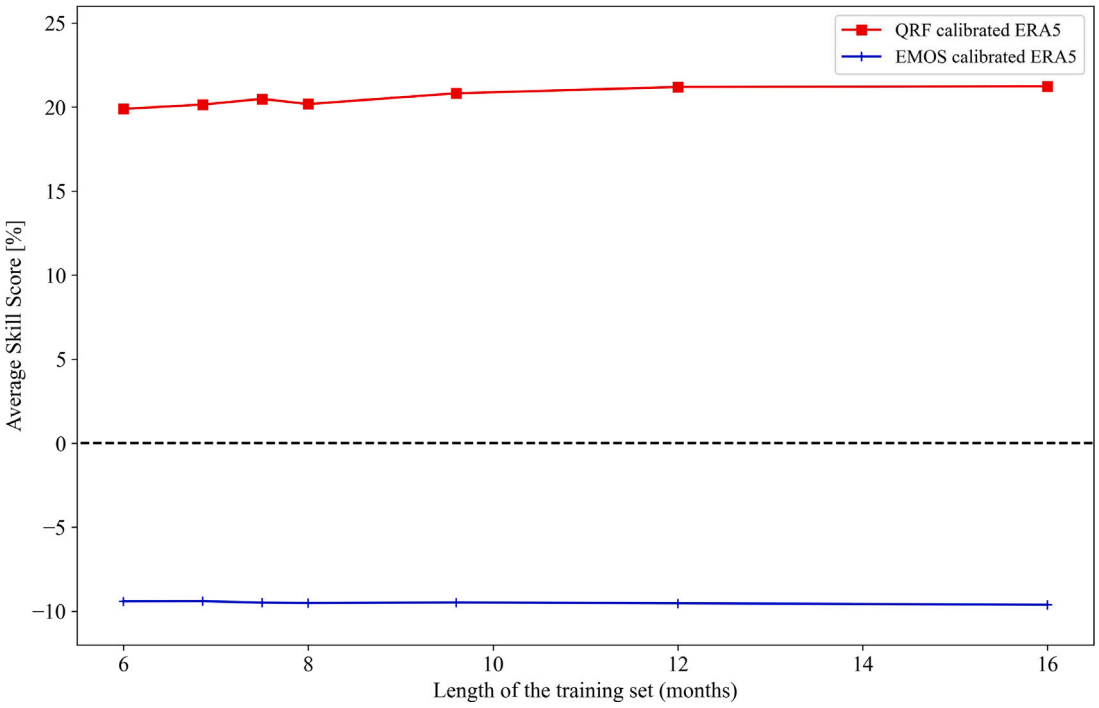


Fig. 8. Skill scores of Fig. 7, averaged over all SYNOP stations, as a function of the length of the training set.

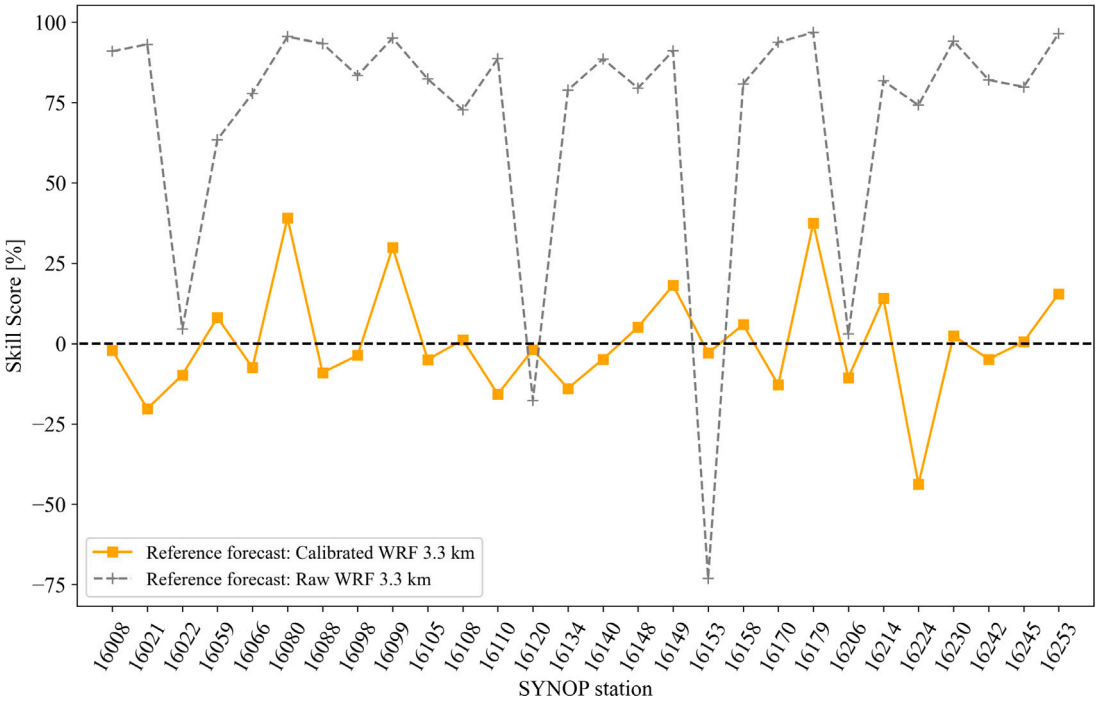


Fig. 9. Skill score of Δ_{QRF} using Δ from the raw WRF simulations as a reference (gray dashed-line, crosses) and from the calibrated WRF simulations (orange line, filled squares). The skill scores have been calculated for all considered SYNOP stations (along the abscissa). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

phenomena via suitable AI methods is thus a key fact proving how our conclusions can be of direct interest to the wind energy industry. A second possible objection is that the sampling time from the wind tower nacelle anemometer is of 10 minutes and thus shorter than the one of SYNOP stations. This fact however does not seem a problem, the AI-based tool we have implemented is indeed totally agnostic with respect to the wind speed sampling frequency. We conclude by noticing that our results are expected to bring advantages in other energy-related realms beyond the one of wind industry. Having accurate calibrations of high-resolution downscaling capable to bring out their possible superiority with respect to coarser hindcasts, especially for the 10 m wind speed, is an issue of paramount importance in met-ocean applications where the 10-m wind velocity forces wave models through which the wave potential in a given sea region can be estimated [29–33] also for

issues related to climate change [34] or to study coupling mechanisms between atmosphere and sea [35].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Andrea Mazzino reports financial support was provided by Compagnia di San Paolo.

Data availability

Data will be made available on request

Acknowledgments

A.M. thanks the financial support from the Compagnia di San Paolo, Italy, Project MINIERA No. I34I20000380007. We thank the Aeronautica Militare – Servizio Meteorologico – for providing us with the SYNOP data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.egyai.2023.100269.

References

[1] Impram S, Nese SV, Oral B. Challenges of renewable energy penetration on power system flexibility: A survey. Energy Strategy Rev 2020;31:100539.
[2] Kalnay E. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 1996;77:431–7.
[3] Uppala S. The ERA-40 Re-analysis. Q J R Meteorol Soc 2005;131:2961–3012.
[4] Onogi K. JRA-25: Japanese 25-year re-analysis project–progress and status. Q J R Meteorol Soc 2005;131:3259–68.
[5] Dee DP. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q J R Meteorol Soc 2011;137:553–97.
[6] Kaspar F, Niermann D, Borsche M, Fiedler S, Keller J, Potthast R, Rösch T, Spangehl T, Tinz B. Regional atmospheric reanalysis activities at Deutscher Wetterdienst: review of evaluation results and application examples with a focus on renewable energy. Adv Sci Res 2020;17:115–28. http://dx.doi.org/10.5194/asr-17-115-2020.
[7] Gregow H, Jylhä K, Mäkelä H, Aalto J, Manninen T, Karlsson P, Kaiser-Weiss A, Kaspar F, Poli P, Tan D, et al. Worldwide survey of awareness and needs concerning reanalyses and respondents views on climate services. Bull Am Meteorol Soc 2016;97(8):1461–73.
[8] Pickering B, Grams C, Pfenninger S. Sub-national variability of wind power generation in complex terrain and its correlation with large-scale meteorology. Environ Res Lett 2020;15:044025.
[9] Gualtieri G. Reliability of era5 reanalysis data for wind resource assessment: A comparison against tall towers. Energies 2021;14(14):4169.
[10] Hersbach H. The ERA5 Atmospheric Reanalysis.. In: AGU fall meeting abstracts, 2016. 2016, NG33D–01.
[11] Luzia G, Hahmann AN, Koivisto MJ. Evaluating the mesoscale spatio-temporal variability in simulated wind speed time series over northern Europe. Wind Energy Sci 2022;7(6):2255–70.
[12] Kalverla PC, Holtslag AA, Ronda RJ, Steeneveld G-J. Quality of wind characteristics in recent wind atlases over the North Sea. Q J R Meteorol Soc 2020;146(728):1498–515.
[13] Jurasz J, Mikulik J, Dąbek PB, Guezgouz M, Kaźmierczak B. Complementarity and 'Resource Droughts' of solar and wind energy in Poland: An ERA5-based analysis. Energies 2021;14(4). http://dx.doi.org/10.3390/en14041118, URL https://www.mdpi.com/1996-1073/14/4/1118.
[14] Kalverla PC, Duncan Jr. JB, Steeneveld G-J, Holtslag AA. Low-level jets over the North Sea based on ERA5 and observations: together they do better. Wind Energy Sci 2019;4(2):193–209.
[15] Olauson J. ERA5: The new champion of wind power modelling? Renew Energy 2018;126:322–31.
[16] Araveti S, Quintana CA, Kairisa E, Mutule A, Adriazola JPS, Sweeney C, Carroll P. Wind energy assessment for renewable energy communities. Wind 2022;2(2):325–47.
[17] Cassola F, Ferrari F, Mazzino A, Miglietta MM. The role of the sea on the flash floods events over Liguria (northwestern Italy). Geophys Res Lett 2016;43(7):3534–42. http://dx.doi.org/10.1002/2016GL068265.
[18] Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG. A description of the advanced research WRF version 2. Tech. rep., National Center For Atmospheric Research Boulder Co Mesoscale and Microscale ...; 2005.
[19] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[20] Meinshausen N, Ridgeway G. Quantile regression forests. J Mach Learn Res 2006;7(6).
[21] World Meteorological Organization. Guide to instruments and methods of observation. 1, 2018,
[22] 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 RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J-N. The ERA5 global reanalysis. Q J R Meteorol Soc 2020;146(730):1999–2049. http://dx.doi.org/10.1002/qj.3803.
[23] Casciaro G, Cavaiola M, Mazzino A. Calibrating the CAMS European multi-model air quality forecasts for regional air pollution monitoring. Atmos Environ 2022;287:119259. http://dx.doi.org/10.1016/j.atmosenv.2022.119259, URL https://www.sciencedirect.com/science/article/pii/S1352231022003247.
[24] Gneiting T, Raftery AE, Westveld III AH, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Rev 2005;133(5):1098–118.
[25] Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E. Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method. J Amer Statist Assoc 2006;101(475):968–79.
[26] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Ser B (Statistical Methodology) 2007;69(2):243–68.
[27] Casciaro G, Ferrari F, Cavaiola M, Mazzino A. Novel strategies of ensemble model output statistics (EMOS) for calibrating wind speed/power forecasts. Energy Convers. Manage. 2022;271:116297.
[28] Ferrari F, Cassola F, Tuju P, Mazzino A. RANS and LES face to face for forecasting extreme precipitation events in the liguria region (northwestern Italy). Atmos Res 2021;259:105654.
[29] Mentaschi L, Besio G, Cassola F, Mazzino A. Developing and validating a forecast/hindcast system for the mediterranean sea. J Coast Res 2013;(65 (10065)):1551–6.
[30] Mentaschi L, Besio G, Cassola F, Mazzino A. Problems in RMSE-based wave model validations. Ocean Model 2013;72:53–8.
[31] Mentaschi L, Besio G, Cassola F, Mazzino A. Performance evaluation of wavewatch III in the Mediterranean Sea. Ocean Model 2015;90:82–94.
[32] Besio G, Mentaschi L, Mazzino A. Wave energy resource assessment in the Mediterranean Sea on the basis of a 35-year hindcast. Energy 2016;94:50–63.
[33] Ferrari F, Besio G, Cassola F, Mazzino A. Optimized wind and wave energy resource assessment and offshore exploitability in the Mediterranean Sea. Energy 2020;190:116447.
[34] Lira-Loarca A, Ferrari F, Mazzino A, Besio G. Future wind and wave energy resources and exploitability in the Mediterranean Sea by 2100. Appl Energy 2021;302:117492.
[35] Rizza U, Canepa E, Miglietta MM, Passerini G, Morichetti M, Mancinelli E, Virgili S, Besio G, De Leo F, Mazzino A. Evaluation of drag coefficients under medicane conditions: Coupling waves, sea spray and surface friction. Atmos Res 2021;247:105207.