ORIGINAL RESEARCH **Open access**

Enhancing Wind Energy Estimation from ERA5 through Ensemble Machine Learning Calibration

Conor O'Brien[1*], Aisling Kelly[1], Niamh Ryan[2], Eoin Byrne[3]

Abstract

The performance of ERA5 has been evaluated in predicting wind energy across 28 SYNOP stations throughout Italy over a 20-year timeframe (2001–2020). For comparison, a high-resolution regional downscaling (3 km horizontal grid spacing) using the WRF model was also generated for the same duration. The findings indicate that basic prediction methods relying on actual historical wind observations yield better wind energy forecasts than raw ERA5 data. This outcome can be linked to the highly intricate topographical traits of the Italian region. Motivated by this anticipated performance, a Quantile Random Forest (QRF) calibration was implemented to substantially mitigate the errors found within the ERA5 reanalysis dataset substantially. This method yields a calibrated ensemble system for predicting wind speed at specific station sites. Remarkably, the calibrated ERA5 data outperforms wind energy assessments derived from the costly 3-km high-resolution downscaling. When the high-resolution downscaling is itself properly calibrated, it offers predictions that are highly comparable to the calibrated ERA5. Restricting the assessments to wind energy estimations over an extended period (20 years in this case), the availability of a high-resolution wind-field dataset does not automatically guarantee superior precision. Implementing precise calibrations on the original, lower-resolution wind-field dataset yields superior outcomes compared to using an uncalibrated high-resolution dataset.

Keywords Wind energy, ERA5 reanalysis, Quantile random forest, Machine learning calibration

*Correspondence:
Conor O'Brien
conor.obrien@ucd.ie
[1] Department of AI for Electricity Market Analytics, School of Engineering, University College Dublin, Dublin, Ireland
[2] Department of Intelligent Power Distribution, Faculty of Engineering, Dublin City University, Dublin, Ireland
[3] Department of Smart Grid Technologies, Faculty of Engineering, University of Galway, Galway, Ireland

Introduction

Over the last few years, the generation of wind energy by the commercial wind sector has shown robust upward growth. It is anticipated that before 2025, global cumulative installations for both onshore and offshore wind applications will drive wind industry energy production beyond 1 TW [1]. Consequently, the wind sector has turned into a cornerstone of energy strategies for many European nations aiming to curb carbon dioxide emissions and mitigate the consequences of climate change. Its role in meeting the 2030 and 2050 European Union greenhouse gas mitigation targets is therefore of paramount importance.

However, wind resources exhibit high spatial and temporal intermittency, making their prediction an incredibly difficult endeavor, even across the briefest look-ahead forecasting windows relevant to energy producers and electrical grid supervisors [1]. The physical characteristics of the resource amplify the challenge of securing precise wind power forecasts. Specifically, the relationship linking wind speed to wind power is cubic, meaning that any relative error introduced during wind speed estimation is magnified by a factor of 3 when translating to wind power. The variable nature

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

of wind also introduces complications and errors when attempting to quantify the wind potential of a particular geographic area. In reality, actual observed records covering a sufficiently long historical duration are seldom accessible in territories that hold potential value for wind power development. Analysts must therefore utilize reanalysis models [2-5] to calculate these resource potentials [6].

Broadly speaking, a reanalysis represents a cutting-edge approach for generating datasets intended for climate monitoring. These products are assembled by combining advanced data assimilation frameworks with numerical weather prediction (NWP) models, which incorporate all accessible observational records (such as measurements from weather balloons, satellites, SYNOP surface stations, marine buoys, commercial aircraft, and ships) at 6- to 12-hour intervals over the historical target timeframe. Spatially and temporally irregular observations are thereby consolidated by NWP systems into uniform, high-frequency, globally gridded meteorological products.

Even though reanalysis products possess multiple major advantages (such as global coverage, uniform spatial and temporal resolutions spanning multiple decades, inclusion of hundreds of distinct variables, regularly improving spatial resolution, and the assimilation of millions of data points into a standardized framework), certain constraints persist. The most significant constraint likely stems from historical alterations in the observing network, which impact data consistency. Observations can vary markedly depending on geographic position, historical era, and the specific parameter under review. As a result, artifact variations and non-physical trends can infiltrate the historical records. Furthermore, even though the spatial resolution of reanalysis products has progressively improved over time, it remains too coarse for specific local applications [7], particularly for assessing wind resources across topographically complex areas [8]. In an investigation centered on the Swiss Alps, Pickering *et al*. [8] determined that both reanalysis models evaluated in their research, featuring grid spacings of 2 and 6 km, were unable to capture site-specific wind speeds within Switzerland's intricate terrain accurately.

An identical inference was reached by Gualtieri [9], where the investigators stated that utilizing the ECMWF ERA5 reanalysis [10] instead of higher-resolution regional reanalysis alternatives should be avoided when evaluating locations characterized by sharp gradients in topography and, more specifically, land use. A thorough and recent synthesis of these challenges was presented by Kaspar *et al*. [6]. Conversely, investigations conducted across northern Europe demonstrated strong ERA5 performance, even when evaluating extreme weather phenomena [11-14]. Furthermore, cross-examinations between ERA5 and alternative reanalysis products demonstrate that ERA5 serves as the most appropriate data resource for wind energy use cases [15, 16] in territories characterized by smooth terrain. Consequently, clear evidence indicates that ERA5's performance relies heavily on topographical complexity. This finding is expected given the dataset's coarse grid spacing, which naturally glosses over terrain features with wavelengths shorter than roughly 30 km.

Our study introduces a paradigm shift relative to the prevailing scientific view that lower-resolution reanalyses like ERA5 are impractical for evaluating wind resources in topographically complex areas. To verify this quantitatively, a 20-year timeframe (spanning 2001 to 2020) is examined across a network of 28 distinct sites, where we analyze varied estimates of wind energy accumulated over the two decades along with its associated level of fluctuation (i.e., wind-energy variance). Successfully capturing the precise variance of the wind energy signal is critical when a specific site is being evaluated for commercial wind power development. In fact, if two locations exhibit identical total accumulated wind energy, the site tied to the lower variance is undeniably more favorable. Our calculated estimates are rigorously validated against wind power values derived from actual surface measurements of 10-m wind speed collected from official SYNOP stations. The geographic region chosen for this research is a portion of the Italian territory (as depicted in **Figure 1**, where pink crosses indicate the SYNOP station positions). Italy features an incredibly diverse terrain matrix, stretching from alpine zones with stations positioned up to 3500 meters high, down to low-lying plains like the Po Valley, where wind behavior is nevertheless strongly shaped by the flanking Alps. Additionally, the complex interplay between land and sea breezes renders wind forecasting even more difficult in this region [17]. To calculate the 20-year cumulative wind energy and its variance, a variety of methodologies are evaluated.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Two of these fit within the category of physics-based differential equation methodologies: the spatially coarse, premium global atmospheric reanalysis ERA5 from ECMWF ([10], featuring a 30-km horizontal grid spacing) and its high-resolution (3-km grid spacing) downscaled counterpart generated via the Weather Research and Forecasting (WRF) model [18]. While ERA5 data is openly accessible and straightforward to manage, its high-resolution downscaled equivalent demands intensive computational resources, particularly for a 20-year simulation window. Dynamic downscaling is currently viewed as the most precise approach to resolve fine-scale local phenomena (such as terrain-induced behaviors). The principal inquiry explored in this study is: can a dependable AI-driven methodology, which is straightforward to train and computationally economical, offer meaningful improvement over expensive high-resolution dynamic downscaling when coupled with ERA5? Framed differently, can ERA5 predictors be utilized to train an appropriate AI model to achieve a forecast that outperforms the computationally demanding WRF-based setup? To address these questions, the Quantile Random Forest (QRF) method [19, 20], an extension of the widely used Random Forest (RF) framework [19], is implemented here. Several factors guided this choice: (i) QRF resists overfitting, is straightforward to optimize, and demands minimal computing power; (ii) because QRF is an ensemble technique, it is exceptionally well-suited for transitioning from wind speed forecasts to wind energy estimates, as every single ensemble member can be converted individually from wind speed to wind power before aggregating the data to extract the necessary wind power and energy statistics; (iii) regarding the core takeaway of this study, it is highly notable that a basic AI approach is sufficient to convert a poorly performing prediction into an output that significantly outshines an expensive, high-resolution dynamic downscaling.

To preview the primary outcomes of this study, clear evidence will reveal the transformative capacity of AI in resolving localized phenomena that are either missed entirely by coarse-grid physics models or captured only via high-resolution dynamic downscaling at a very high computational cost.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press



Figure 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).

Wind Data

Observed data from SYNOP stations

The observational measurements utilized in this research consist of surface synoptic data gathered from 28 distinct meteorological stations distributed across northern and central Italy. While synoptic stations generally reflect meteorological conditions within a 100 km radius, their representativeness can be scaled down to 10 km or less for highly localized applications [21]. The synoptic records employed here span a 20-year timeframe, beginning on 1st January 2001 and ending on 31st December 2020. These measurements possess a 3-h temporal resolution, aligning with the standard synoptic hours of 00, 03, 06, 09, 12, 15, 18, and 21 UTC.

The ERA5 reanalysis dataset

ERA5 represents the fifth generation of global atmospheric climate reanalysis, encompassing data from January 1950 to the current day. Developed by the Copernicus Climate Change Service (C3S) at the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5 delivers hourly approximations for an extensive array of oceanic, land, and

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

atmospheric climate parameters. This dataset spans the globe on a 30 km spatial grid and maps the atmosphere vertically across 137 levels, extending from the Earth's surface up to an altitude of 80 km [22].

For this investigation, alongside the 10 m wind speed measurements, we gathered additional atmospheric variables spanning the 2001–2020 timeframe. These variables serve as meteorological input features for training our ML algorithm.

Motivation of The Work and Method

Consider the challenge of selecting suitable locations for wind energy infrastructure development. Assume we have access to two primary data sources: a reasonably long observational record—such as 1 year of measured wind speeds at a specific site—and a comprehensive, long-term state-of-the-art reanalysis dataset covering 20 years, which in this study is ERA5. We aim to utilize the single year of observational data to project the total accumulated wind energy over the entire 20-year duration. We refer to this conservative approach of estimating a 20-year wind energy yield based on a single year of collected data as a climatology-based prediction (or simply climatology, recognizing that in atmospheric science, climatology typically refers to the mechanisms driving climate patterns and multidecadal variations).

The primary question is whether the more extensive reanalysis dataset offers a more precise estimation of the total accumulated wind energy at a specific site from 2001 to 2020 than the climatological approach. Suppose we focus on a single location where the ERA5 wind speed is estimated through conventional bilinear interpolation; these same principles apply trivially to all the locations shown in **Figure 1** where wind energy must be assessed. To determine the necessary wind power for calculating wind energy—from either the reanalysis data or measured wind speeds—we apply a standard transfer function representative of a 2 MW capacity wind turbine as an illustrative example.

Given the observed wind data, and consequently the true accumulated wind energy over the 20 years from 2001 to 2020 (the ground truth), we can evaluate any predictive model against this baseline. Let E represent the actual accumulated wind energy over the 20 years, Ec the climatological projection based on one reference year of observation, and Eh the prediction derived from ERA5. While E and Eh yield single aggregated values, one could generate N = 20 distinct climatological forecasts depending on which specific reference year is chosen. Therefore, we substitute Ec by , i = 1, ⋯, N to encompass all potential climatologies.

Comparing the Climatology-Based Predictions Against the Raw Reanalysis-Based Predictions

Figure 2 illustrates whether climatology Ec or reanalysis Eh serves as a better approximation of E. For each SYNOP station (arranged along the x-axis from north to south as depicted in **Figure 1**), we calculated the following error indices:

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

$\mathcal{E}$ being the normalization factor representing the maximum potential energy extractable from the location over the N years. The plus symbols (blue line) denote Δh, which is the error linked to the reanalysis; the open circles (green line) indicate Δc, the error tied to the climatology. For most locations, the climatology-based approach yields superior results compared to the reanalysis prediction. This finding remains consistent (though not depicted) even when the specific year used to generate the climatological prediction is excluded from the testing dataset. The effectiveness of the climatology-based approximation is further underscored in the inset of **Figure 2**, which plots the skill score $((\Delta c - \Delta h)/\Delta c) \times 100$ [23] for all locations, utilizing the climatological error as the baseline reference. The observation that this value is negative for 25 out of 28 sites corroborates the conclusion drawn from the main frame of the figure. Furthermore, the inset indicates

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

that, on average across all stations, the climatology-related error is roughly 50% lower than the error corresponding to the raw reanalysis-based forecast.

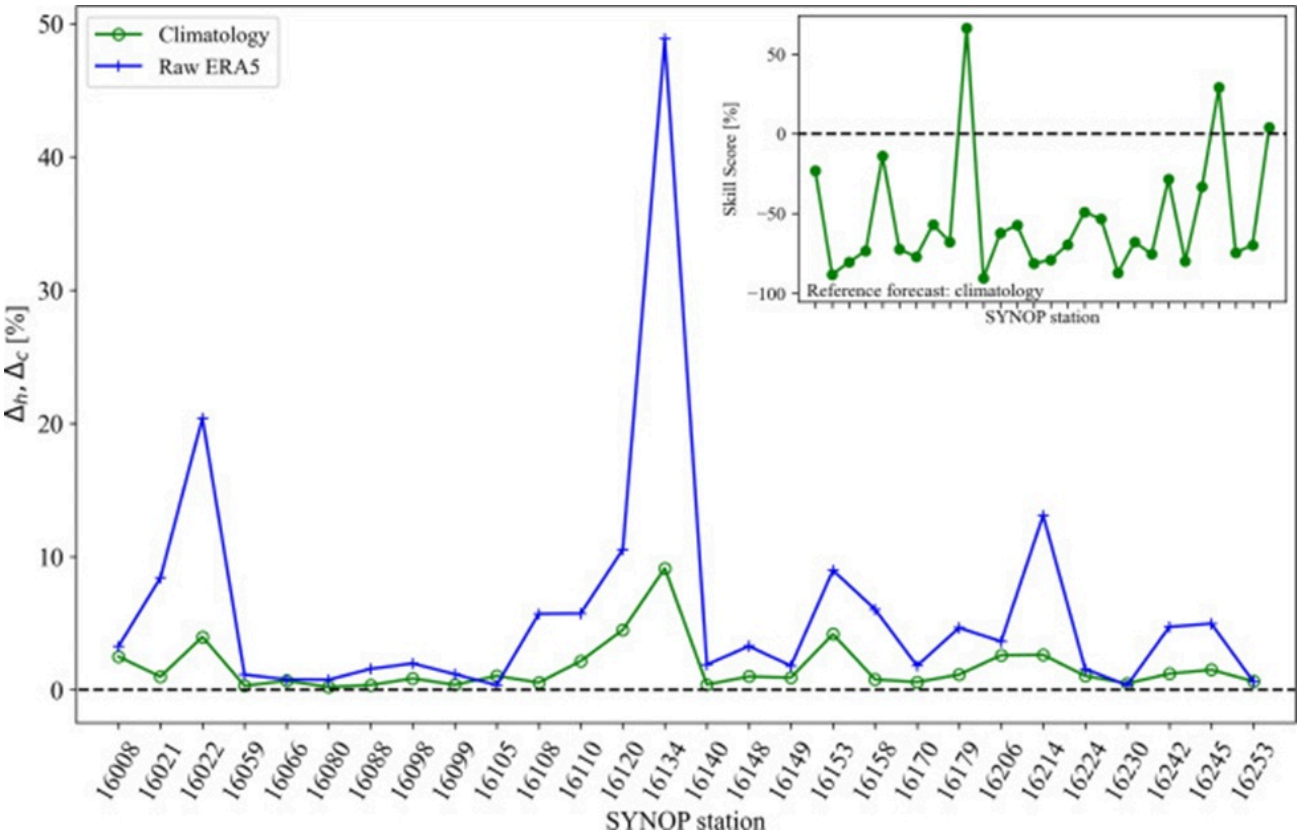


Figure 2. Main frame: behavior of the two dimensionless error indices Δc and Δh (expressed as a percentage) measuring how much climatology (green line, open circles) and the reanalysis (blue line, crosses) are successful in predicting the available wind energy accumulated in the 20 years for all sites reported in **Figure 1**. Inset: the skill score of Δh using Δc 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).

We now explore how the duration of the observational period—from which the climatology predictions are derived (1 year in **Figure 2**)—affects the outcomes. In **Figure 3**, we investigate this by displaying the skill score of Δh, averaged across all SYNOP stations, with climatology as the baseline reference. The abscissa represents various durations (in months) of available measurement data. The scenario presented in **Figure 2** aligns with the abscissa taking the value 12 months. Naturally, as the availability of measured data decreases, the relative advantage of climatology over reanalysis-based predictions diminishes. Surprisingly, however, climatology continues to outperform reanalysis predictions even for measurement periods shorter than 1 month, although they tend to become practically indistinguishable at that scale.

Do these findings imply that reanalysis data holds no utility for selecting wind energy sites in areas with complex meteorological dynamics, such as the region analyzed in this study? As demonstrated in the subsequent sections, an appropriate calibration methodology can effectively extract maximum value from the reanalysis dataset.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

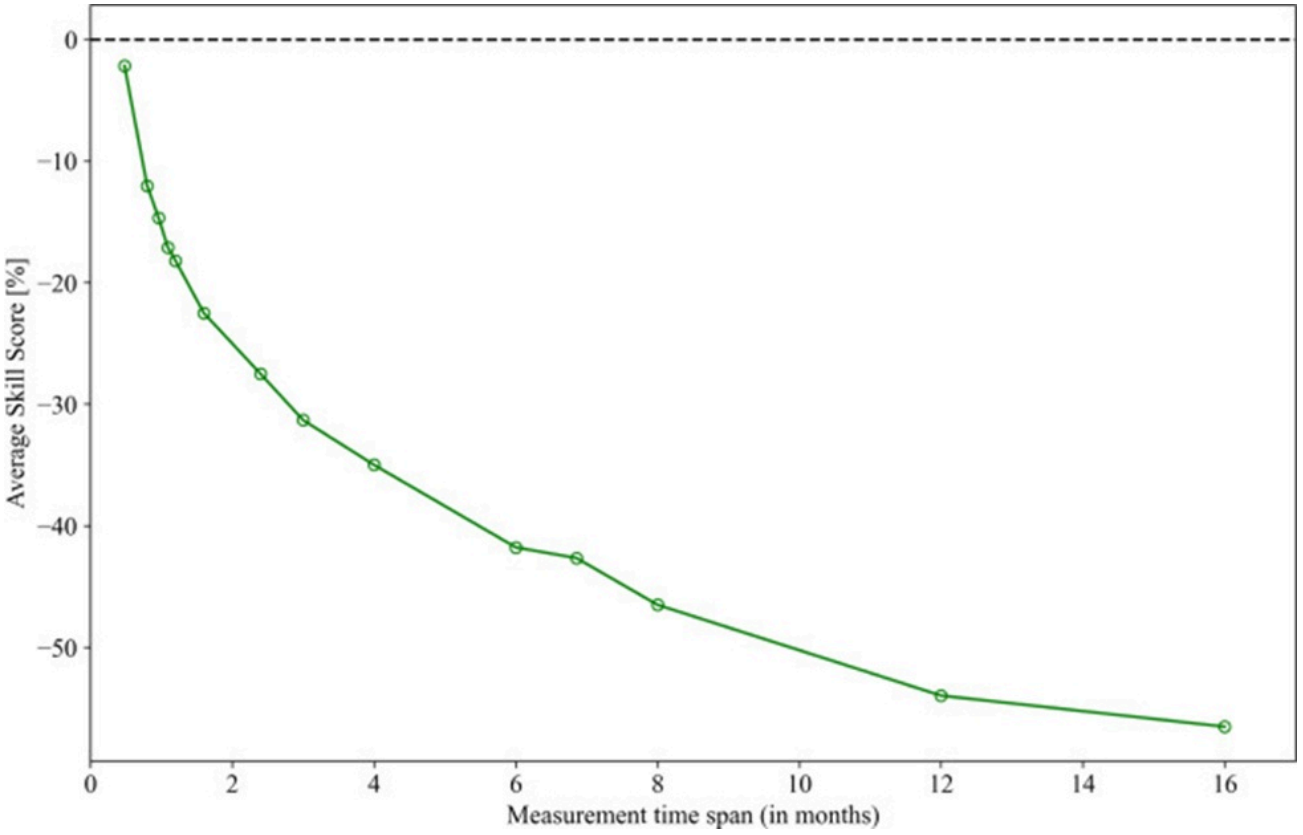


Figure 3. Skill score of Δh, averaged over all SYNOP stations, using climatology as reference. The abscissa reports different durations of the hypothetical measurement campaign (in months) where wind speed measurements are supposed to be available.

Calibration Strategies: Quantile Random Forest (QRF), Random Forest (RF), And Ensemble Model Output Statistics (EMOS)

While the findings in the preceding section may initially seem unexpected, a global modeling approach inherently relies on a coarse spatial resolution and therefore struggles to capture localized phenomena. Such microscale effects are undeniably significant across the Italian landscape, which is heavily influenced by the complex interaction of orography and sea-induced flows. Our objective here is to demonstrate that these localized factors can be successfully integrated into the reanalysis through appropriate Machine-Learning (ML) techniques, thereby creating tangible added value. We have selected three established methods: the Quantile Random Forest (QRF) [19, 20]—a supervised ML technique that is relatively novel in wind energy applications—the Random Forest (RF) [19], and the widely adopted statistical calibration approach known as Ensemble Model Output Statistics (EMOS) [24-27]. In this context, RF and EMOS serve as baseline benchmarks due to their accessibility, simplicity, and widespread adoption.

Results on the Calibrated Reanalysis

Following the conceptual framework, assume we isolate a single year (out of the 20) containing wind speed measurements and use these metrics to train both the QRF and the RF models. Just as we generated 20 distinct climatological predictions, we can similarly produce 20 different QRF calibrations depending on the specific year chosen for training. Let us designate these 20 predictions for the total accumulated wind energy over the two-decade span as

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

E^i_{QRF}, i = 1, 20. The value of E^i_{QRF} for each respective training year, denoted by index i, is derived from the predicted QRF wind speed ensemble by first converting each wind speed member into a wind power member and subsequently into a wind energy member. Once the full distribution of wind energy members is established for a given year, it is straightforward to calculate the distribution's mean and sum it to find the total accumulated energy across the 20-year timeframe. The resulting overall energy derived using the ith year for training is precisely E^i_{QRF} as previously defined. With the predictions E^i_{QRF} in hand, the error relative to the true value can be established in analogy to **Eq. 1:**

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

The enhanced performance achieved by the QRF-calibrated reanalysis is illustrated in **Figure 4**, which plots the skill score $((\Delta_c - \Delta_{QRF})/\Delta_c) \times 100$ (filled squares, red line) across all locations, using the climatology-derived error as the benchmark. For 23 of the 28 sites, this index yields a positive value, signaling a clear performance advantage over the climatological forecast. For comparative purposes, the blue line (plus symbols) illustrates the identical skill score metric computed for the EMOS-based forecast. As anticipated, QRF visibly surpasses the standard EMOS method.

It is now compelling to further examine how the duration of the measurement interval (one year in the prior analysis) influences the ultimate predictive skill of the calibrated reanalysis. This interval functions not only as the training period for the EMOS, RF, and QRF models but also as the foundation to build the climatology-based prediction. The outcome of this is not a priori obvious. Indeed, if expanding the training dataset corresponds to a more robust calibration, an extended measurement timeframe is simultaneously expected to yield larger skills for the baseline climatological prediction.

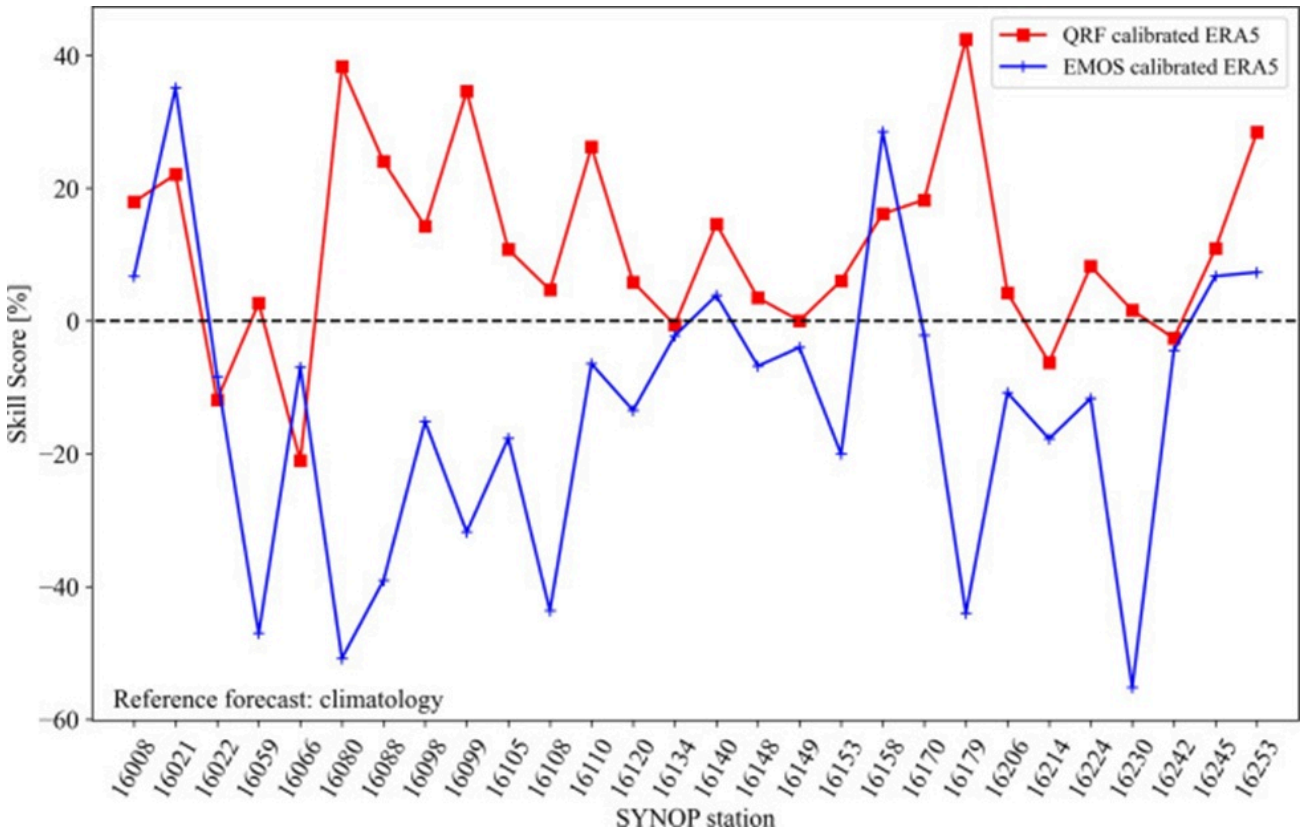


Figure 4. Skill score of ΔQRF using Δc as a reference (red line, filled squares) for all considered SYNOP stations (along the abscissa). The blue line (crosses) is analogous. Still, it refers to the EMOS-based calibration (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

As depicted in **Figure 5**, an optimal training duration emerges at approximately 7 months, where the skill score comparing the QRF error against the climatology-based prediction error $((\Delta_c - \Delta_{QRF})/\Delta_c) \times 100$, averaged across all evaluated locations, achieves its maximum (red line, filled squares). With extended training periods, the climatological approach becomes more significant, though it continues to underperform compared to the calibrated reanalysis. Conversely, the EMOS methodology exhibits no such peak; its skill score trajectory remains consistently negative and decreases monotonically as the training dataset expands. Consequently, the EMOS approach yields no supplementary benefit across any of the evaluated training durations.

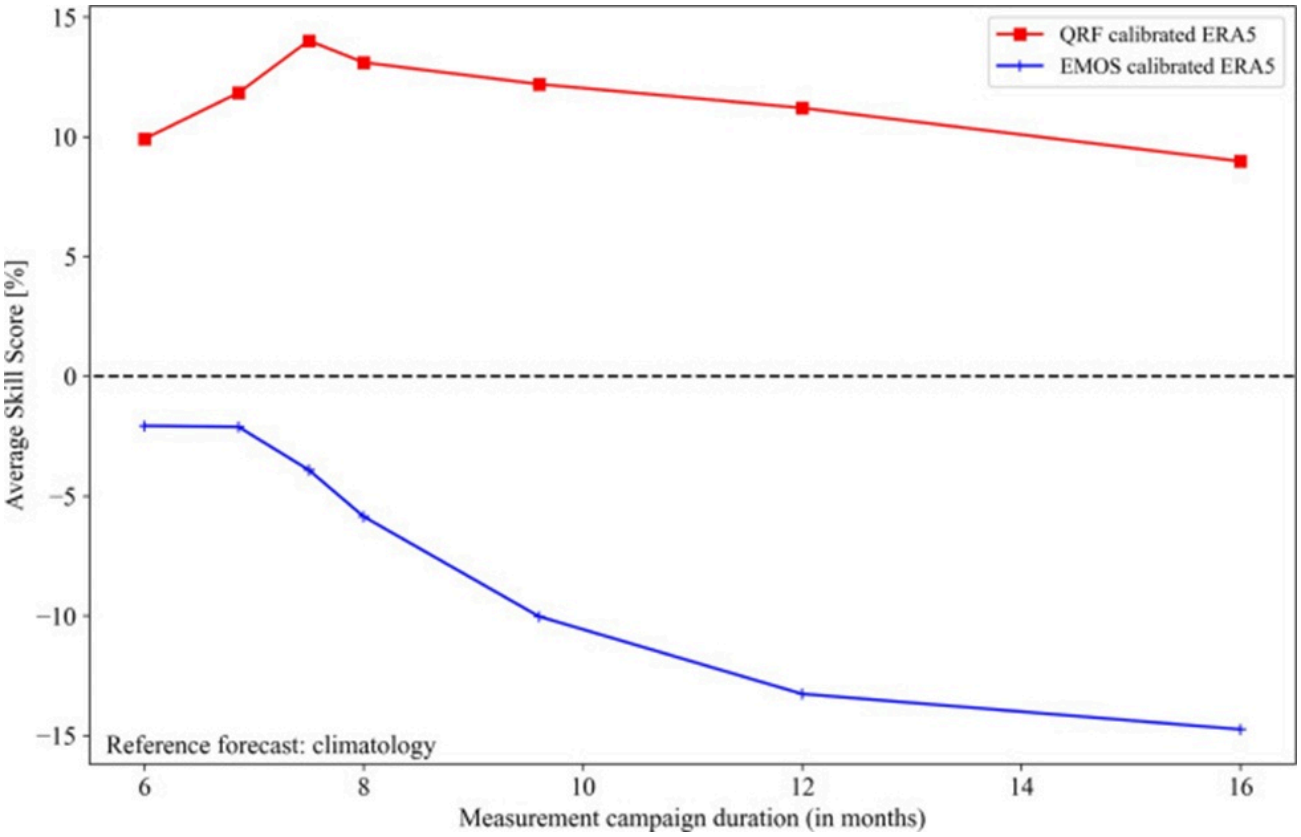


Figure 5. Skill score of ΔQRF, averaged over all SYNOP stations, using climatology as reference (red line, filled square). The abscissa reports different durations of the hypothetical measurement campaign (in months) during which wind speed measurements are assumed to be available. The blue line (crosses) 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).

At this stage, we demonstrate the performance advantage of the QRF framework over the standard RF approach for forecasting cumulative wind energy over the 20 years. Similar to **Figure 5**, **Figure 6** displays the skill score of ΔQRF, averaged across all SYNOP sites, with the RF-derived estimates acting as the baseline (main frame). The horizontal axis denotes varying durations of the observational campaign during which wind speed data are assumed to be accessible.

The specific behavior of the skill score across individual SYNOP locations is detailed in the inset of **Figure 6**. The observed skill scores are substantially higher, reaching up to 40%, which underscores the persistent operational benefit provided by the QRF method compared to the more conventional and broadly utilized RF technique.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

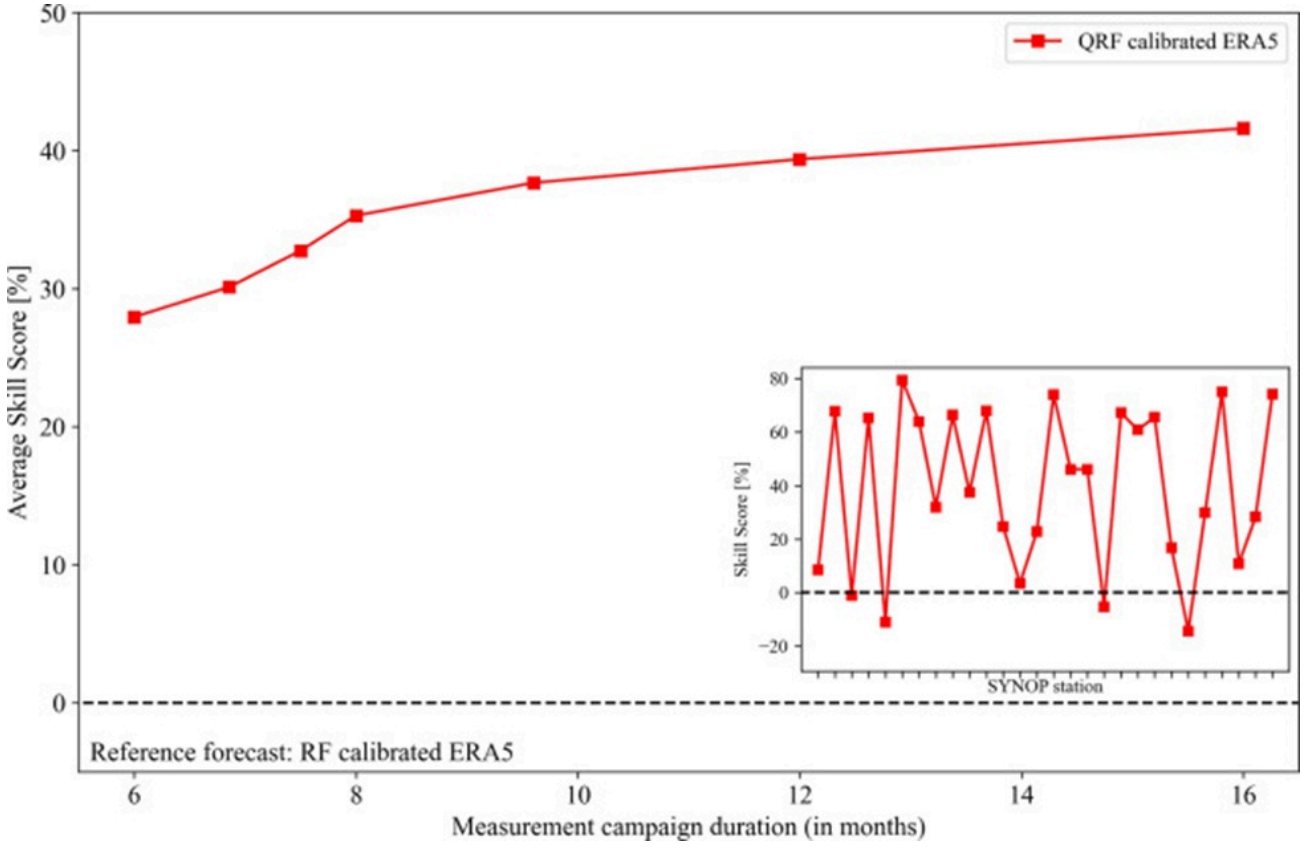


Figure 6. Main frame: skill score of ΔQRF averaged over all SYNOP stations, using the RF-based predictions as reference. The abscissa reports different durations of the hypothetical measurement campaign (in months) during which wind speed measurements are assumed to be available. Inset: for a training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

Following the demonstration that QRF-calibrated reanalysis exceeds the accuracy of both climatological and RF-derived estimates for 20-year cumulative wind energy, we evaluate whether QRF calibration similarly enhances the raw reanalysis with respect to the interannual variability of wind energy over the n = 20 years. To investigate this, we maintain a fixed 1-year assessment window for evaluating wind energy fluctuations while varying the training-phase duration, which represents the period of available observational data. Building on the definition of E as the aggregate wind energy over the 2-decade period, we denote the wind energy for each respective year as Ei, i = 1, ⋯, 20. Utilizing Ei, the year-to-year wind energy variability can be quantified via the standard deviation;

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

The objective is to evaluate the equivalent standard deviations derived from the calibrated ERA5 data (utilizing QRF and EMOS) against the ground truth. Since a 1-year climatological baseline inherently lacks inter-annual fluctuation, it is excluded from this analysis. Regarding the uncalibrated reanalysis, σh is formulated correspondingly to (3) by substituting (i) the raw reanalysis estimate for the ith year in place of Ei, and (ii) the mean annual wind energy derived from the reanalysis in place of $\langle Ei \rangle$. For the calibrated dataset, M potential standard deviations emerge, contingent upon the placement of the chosen timeframe (spanning 20 years/M) that acts as the training phase. Specifically, we establish:

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\sigma_{QRF}^{(j)} = \sqrt{\frac{\sum_{i=1}^{N}\left(E_{QRF\,i}^{(j)} - \langle E_{QRF\,i}^{(j)}\rangle\right)^2}{N}} \qquad \langle E_{QRF\,i}^{(j)}\rangle = \frac{\sum_{i=1}^{N} E_{QRF\,i}^{(j)}}{N} \quad (4)$$

Here, $E_{QRF\,i}^{(j)}$ represents the QRF-derived wind energy forecast for the ith year (out of the 20 years) associated with the jth training subset (out of the M possible distinct training subsets). Similarly, we introduce $\sigma_{EMOS}^{(j)}$, deviating from (4) only in that the forecasts are generated via the EMOS calibration framework. Once σ_h, $\sigma_{QRF}^{(j)}$, $\sigma_{EMOS}^{(j)}$, $j = 1, \cdots, M$ are established, the discrepancies relative to the true value σtr can be measured. This comparison is straightforwardly executed through a normalized error metric, such as:

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

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

$$\Delta_{\sigma_h} = \frac{\left|\sigma_h - \sigma_{tr}\right|}{\sigma_{tr}} \quad (7)$$

The dynamics of $\Delta_{\sigma_{QRF}}$, $\Delta_{\sigma_{EMOS}}$, and Δ_{σ_h} across all evaluated SYNOP locations (abscissa) are depicted in **Figure 7a.**

This visual representation distinctly highlights the dominance of the QRF calibration (red line, filled squares) in modeling inter-annual wind energy fluctuations (effective for 22 of the 28 sites), outperforming both the uncalibrated reanalysis (red line, open circles) and the EMOS technique (blue line, crosses). Panel (b) illustrates the inter-annual variability skill score derived from (i) the QRF model ((Δ_{σ_h} -$\Delta_{\sigma_{QRF}}$)/Δ_{σ_h}) compared against the baseline raw reanalysis (red line, filled squares), alongside (ii) the EMOS framework ((Δ_{σ_h}-$\Delta_{\sigma_{EMOS}}$)/ Δ_{σ_h}) also assessed against the uncalibrated reanalysis (blue line, crosses). The operational superiority of the QRF calibration compared to EMOS is striking. Specifically, the QRF approach yields a positive skill score for 22 out of 28 locations (indicating that calibration enhances the modeling of year-to-year variability compared to raw data). In contrast, the EMOS method improves only 11 of 28 stations. We finalize this assessment by exploring how the skill scores from **Figure 7**, averaged across all SYNOP sites, correlate with the size of the training dataset. As evidenced in **Figure 8**, the performance of both calibration methods remains relatively stable, reflecting their reliability in estimating inter-annual wind energy dynamics across the tested training durations.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

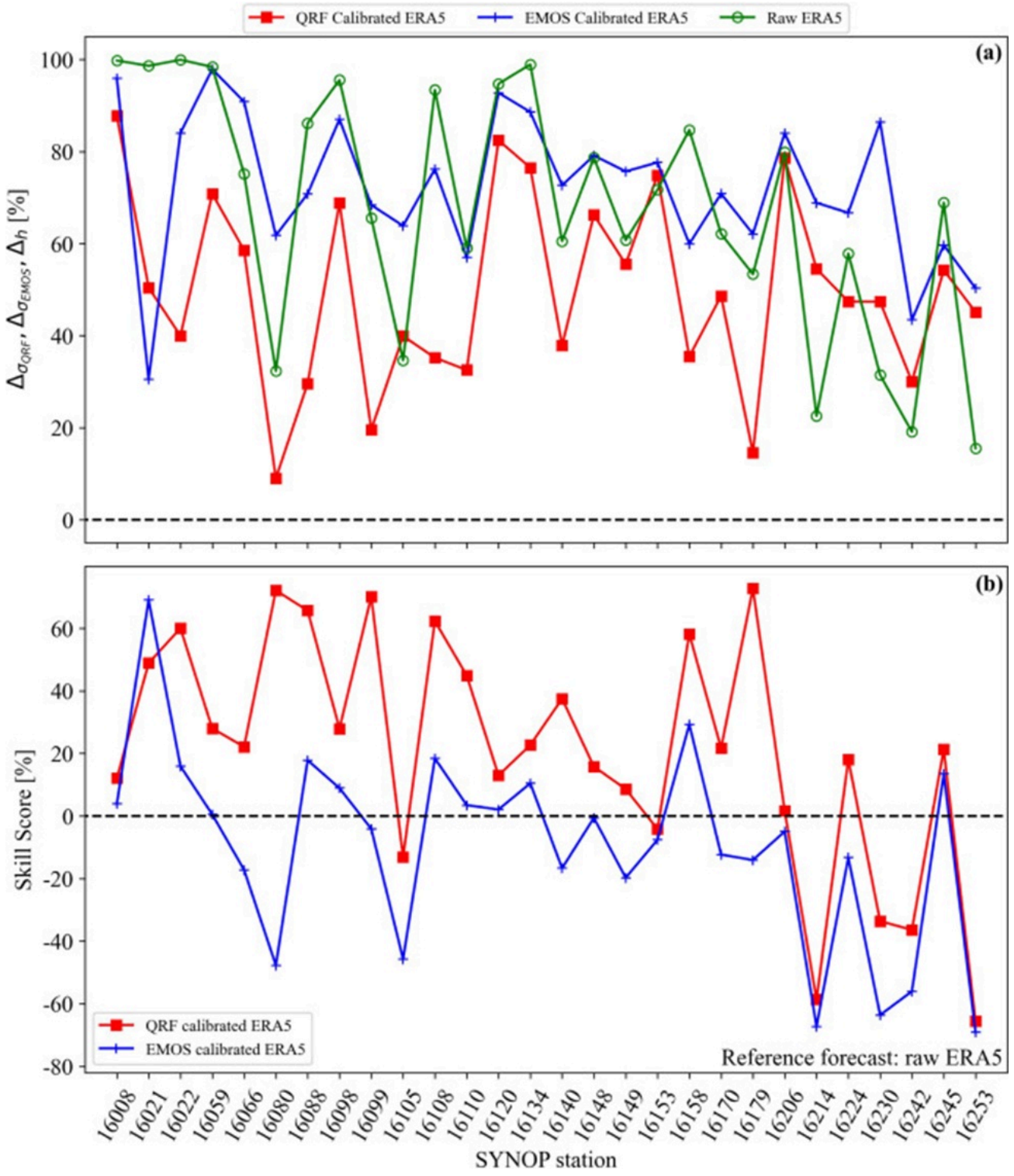


Figure 7. Panel (a): Δ_{σ_h}, $\Delta_{\sigma_{QRF}}$, and $\Delta_{\sigma_{EMOS}}$ are shown for all analyzed SYNOP stations. Panel (b): Skill score of $\Delta_{\sigma_{QRF}}$ using Δ_{σ_h} as a reference (red line, filled squares) for all considered SYNOP stations (along the abscissa); the blue line (crosses) is analogous, 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).

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

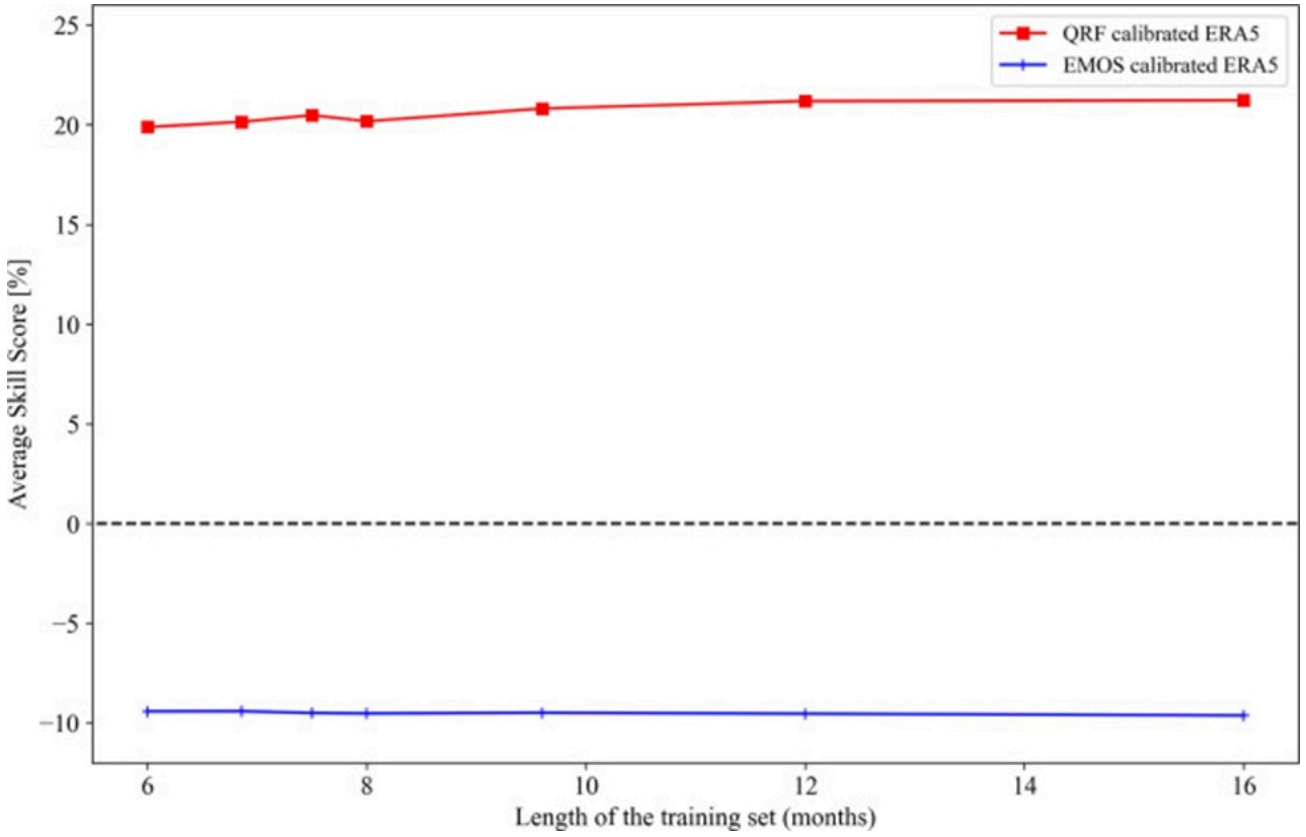


Figure 8. Skill scores of **Figure 7**, averaged over all SYNOP stations, as a function of the length of the training set.

Does High-Resolution Downscaling Always Mean High Accuracy?

Here, we investigate whether high-resolution downscaling can match the performance of our optimal ML-driven calibration. This inquiry stems from the operational demands of wind energy evaluation, wherein high-resolution hindcasts are generally generated through dynamical downscaling processes. In such setups, broader models like ERA5 supply the requisite boundary and initial conditions for localized, high-resolution modeling. To address this, we executed a downscaling procedure based on ERA5 data over the previously analyzed timeframe. For this task, we deployed the Weather Research and Forecasting (WRF) framework [18]. The structural setup mirrors that of Ferrari *et al*. [28], except for the spatial resolution, which is refined to 3 km for the primary domain shown in **Figure 1** and 10 km for the broader European territory. The findings are illustrated in **Figure 9**. As in **Figure 4**, this chart plots the skill score of ΔQRF for the calibrated ERA5 dataset, this time employing WRF Δ as the baseline metric—evaluating both the uncalibrated (dashed line) and calibrated (continuous line) iterations. The WRF calibrations utilized the identical QRF methodology applied to ERA5, incorporating the same input features. The data provides a definitive conclusion: the QRF-calibrated ERA5 model outperforms the uncalibrated WRF output, despite the latter's fine 3 km resolution. Furthermore, any performance discrepancies between the calibrated WRF forecasts and the calibrated ERA5 estimates are minimal and virtually inconsequential in practical scenarios.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

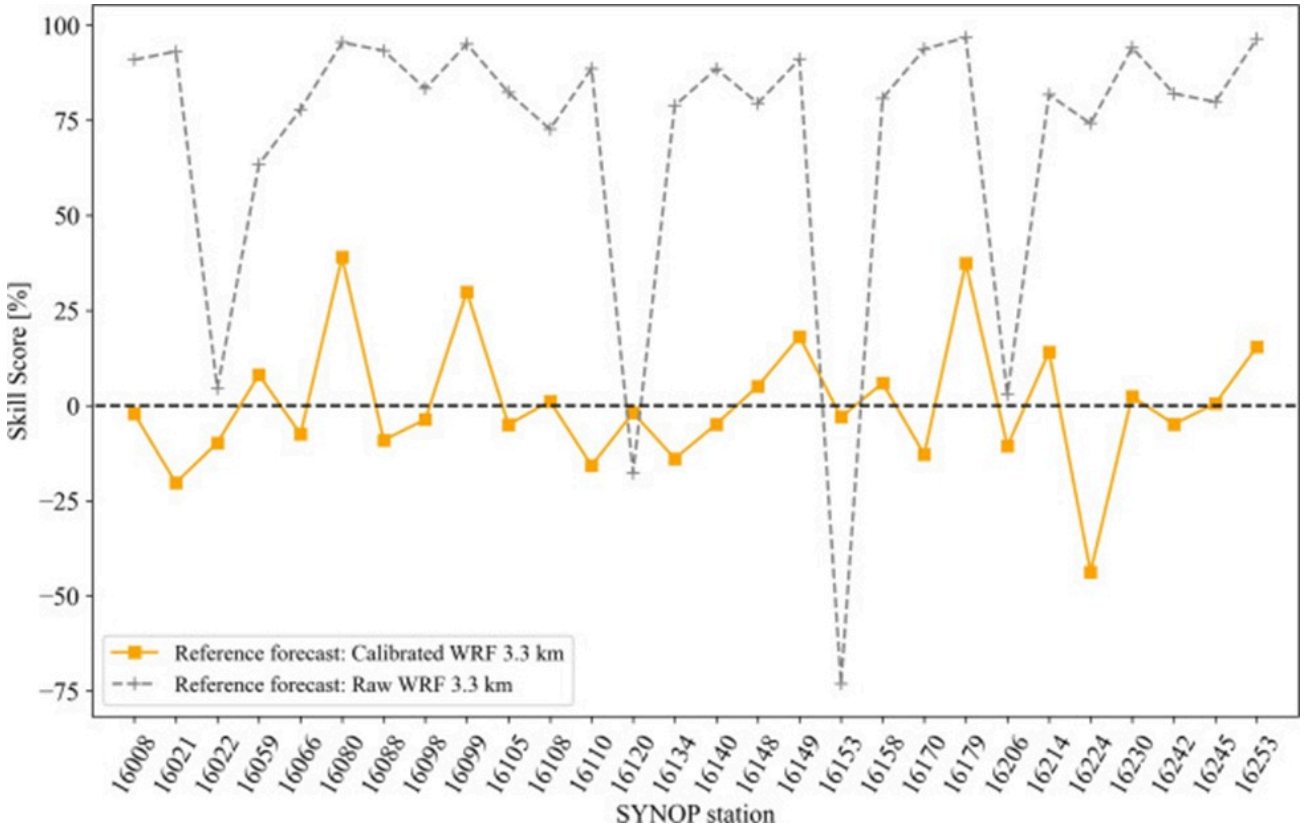


Figure 9. Skill score of ΔQRF using Δ from the raw WRF simulations as a reference (gray dashed line, crosses) and from the calibrated WRF simulations (orange line, filled squares). The skill scores have been calculated for all considered SYNOP stations (along the abscissa) (For interpretation of the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

Conclusion

This study investigated the potential benefits of utilizing the ERA5 dataset compared to a 3 km high-resolution WRF downscaling for approximating cumulative wind energy over 20 years. Observational data from 28 Italian SYNOP stations spanning 2001 to 2020 served as the baseline truth to assess various forecasting frameworks. These included two machine learning approaches—Random Forest (RF) and Quantile Random Forest (QRF)—alongside an established, streamlined parametric method called EMOS, and a baseline climatology forecast. The climatological method estimates the total 20-year wind energy yield merely by multiplying a single year's observational data by 20. The findings indicate that the rudimentary climatological projection significantly exceeds the accuracy of the uncalibrated ERA5 forecast. This underperformance of the raw ERA5 data is likely attributable to Italy's intricate topography, which generates complex, sub-synoptic dynamical patterns that broad-scale spatial models fail to resolve accurately.

Driven by these findings, we deployed a machine learning model designed to yield the complete probability density function (PDF) of wind velocities at specific stations, rather than merely calculating average wind speeds. With this PDF established, it becomes straightforward to derive the corresponding wind power PDF, facilitating the calculation of average wind power and aggregate wind energy. For this ensemble calibration, we utilized the Quantile Random Forest technique. Our evaluation demonstrates that the QRF calibration unlocks the full potential of the ERA5 dataset, achieving substantial performance gains over the baseline climatology, the standard RF model, and the uncalibrated reanalysis. Notably, the enhancements provided by QRF extend beyond calculating the total 20-year wind energy yield; they also significantly improve the estimation of year-to-year wind energy fluctuations across single-year intervals.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Furthermore, the QRF-generated forecasts were evaluated against the more streamlined EMOS and RF alternatives, revealing that the QRF methodology delivers a marked operational advantage over both comparative techniques.

Ultimately, we benchmarked our optimal ERA5 calibration against the 3-km dynamic WRF downscaling, testing both the uncalibrated WRF output and a version subjected to the identical ERA5 calibration process. The outcomes reveal that a calibrated ERA5 dataset yields superior 20-year cumulative wind energy estimates compared to the raw, high-resolution WRF downscaling. The performance gap essentially disappears only when the raw WRF model is calibrated using the exact procedure applied to ERA5. Considering the substantial computational costs and extended processing times required to generate high-resolution hindcasts, employing a rigorously calibrated, lower-resolution hindcast emerges as the more practical and efficient alternative.

We conclude by outlining the relevance of these findings to the commercial wind energy sector, acknowledging that actual wind turbines function at elevations well above 10 m a.g.l. One potential critique is that surface terrain and local land-use heavily bias SYNOP station readings. However, even if localized anomalies skew these measurements, this characteristic is arguably beneficial for wind energy contexts. Operational wind farms are environments heavily disrupted by wake phenomena. The capacity to model such complex, non-local disturbances using targeted AI algorithms reinforces the direct industrial applicability of our findings. A second foreseeable critique relates to the 10-minute sampling interval of turbine nacelle anemometers, which is significantly finer than the resolution of SYNOP records. This discrepancy is ultimately negligible, as the AI framework deployed in this study functions independently of the underlying wind speed sampling frequency. Finally, we emphasize that these methodologies hold promise for broader energy and environmental sectors beyond wind power. Developing precise calibrations for high-resolution downscaled data to maximize their utility over broader hindcasts—particularly regarding 10-m wind speeds—is crucial for marine and meteorological applications. In these fields, 10-m wind parameters drive the wave simulation models used to calculate marine wave energy potential [29-33], investigate climate change impacts [34], and analyze complex atmospheric-oceanic coupling dynamics [35].

Acknowledgements

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.

Conflict of interest

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

Financial support

None

Ethics statement

None

Received: 08 Aug 2023 Revised: 13 Oct 2023 Accepted: 30 Oct 2023
Published online: 15 June 2025

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

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, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, et al. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc. 1996;77(3):431-7.

3. Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M, et al. The ERA-40 re-analysis. Q J R Meteorol Soc. 2005;131(612):2961-3012.

4. Onogi K, Koide H, Sakamoto M, Kobayashi S, Tsutsui J, Hirai M, et al. JRA-25: Japanese 25-year re-analysis project – progress and status. Q J R Meteorol Soc. 2005;131(612):3259-68.

5. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc. 2011;137(656):553-97.

6. Kaspar F, Niermann D, Borsche M, Fiedler S, Keller J, Potthast R, et al. 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.

7. Gregow H, Jylhä K, Mäkelä H, Aalto J, Manninen T, Karlsson P, 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(4):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; 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 GJ. 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):1118.

14. Kalverla PC, Duncan JB Jr, Steeneveld GJ, 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, et al. 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.

18. Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, et al. A description of the Advanced Research WRF version 2. Boulder (CO): National Center for Atmospheric Research, Mesoscale and Microscale Meteorology Division; 2005. Technical report.

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):983-99.

21. World Meteorological Organization. Guide to instruments and methods of observation. Geneva: WMO; 2018.

22. Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020;146(730):1999-2049.

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.

24. Gneiting T, Raftery AE, Westveld AH III, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon 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 Am Stat Assoc. 2006;101(475):968-79.

26. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Ser B Stat Methodol. 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.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

35. Rizza U, Canepa E, Miglietta MM, Passerini G, Morichetti M, Mancinelli E, et al. Evaluation of drag coefficients under medicane conditions: coupling waves, sea spray and surface friction. Atmos Res. 2021;247:105207.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press