

LETTER • OPEN ACCESS

A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds

To cite this article: Jaume Ramon *et al* 2021 *Environ. Res. Lett.* **16** 054010

View the [article online](#) for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds

OPEN ACCESS

RECEIVED
30 July 2020REVISED
29 January 2021ACCEPTED FOR PUBLICATION
9 February 2021PUBLISHED
16 April 2021

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Jaume Ramon¹ , Llorenç Lledó¹ , Pierre-Antoine Bretonnière¹, Margarida Samsó¹
and Francisco J Doblas-Reyes^{1,2}¹ Barcelona Supercomputing Center (BSC), c/ Jordi Girona, 29, Barcelona 08034, Spain² ICREA, Pg. Lluís Companys 23, Barcelona 08010, SpainE-mail: jaume.ramon@bsc.es**Keywords:** statistical downscaling, Euro-Atlantic Teleconnections, seasonal prediction, multi-systemSupplementary material for this article is available [online](#)**Abstract**

This work provides a new methodology based on a statistical downscaling with a perfect prognosis approach to produce seasonal predictions of near-surface wind speeds at the local scale. Hybrid predictions combine a dynamical prediction of the four main Euro-Atlantic Teleconnections (EATC) and a multilinear statistical regression, which is fitted with observations and includes the EATC as predictors. Once generated, the skill of the hybrid predictions is assessed at 17 tall tower locations in Europe targeting the winter season. For comparative purposes, hybrid predictions have also been produced and assessed at a pan-European scale, using the ERA5 100 m wind speed as the observational reference. Overall, results indicate that hybrid predictions outperform the dynamical predictions of near-surface wind speeds, obtained from five prediction systems available through the Climate Data Store of the Copernicus Climate Change Service. The performance of a multi-system ensemble prediction has also been assessed. In all cases, the enhancement is particularly noted in northern Europe. By being more capable of anticipating local wind speed conditions in higher quality, hybrid predictions will boost the application of seasonal predictions outside the field of pure climate research.

1. Introduction

Recent advances in the fields of climate modelling and seasonal prediction have resulted in skilful seasonal predictions of surface variables over the extratropics (Merryfield *et al* 2020). This has, in turn, led to the development of climate services that inform weather-and-climate-vulnerable socio-economic sectors of seasonal anomalies a few months ahead (Buontempo *et al* 2018). The energy sector takes advantage of such valuable information since energy production and demand are strongly linked to climate variability. In particular, the renewable energy industry can profit from seasonal predictions of surface wind speed (Clark *et al* 2017, Torralba *et al* 2017) and wind power generation (Lledó *et al* 2019) to anticipate revenues, balance electricity supply and demand or schedule maintenance activities among others. However, those predictions still suffer from some limitations, mainly due to (1) the limited skill

levels on surface variables available from current seasonal prediction systems and (2) its relatively coarse spatial scales.

Generally, seasonal anomalies of atmospheric variables arise from large-scale forcings that other components of the Earth system exert as boundary conditions, such as anomalies of sea ice extent, sea surface temperature or soil moisture. These boundary-condition forcings can be adequately represented in coarse-scale coupled models—often delivered in grids of tens of square kilometres—leading to some skill in the predictions. However, the absolute values that are experienced near the surface at the local scales can be highly affected by local effects and vary substantially even at short distances. Values of surface temperature or precipitation are affected by the local topography, particularly in complex-terrain regions (e.g. Anders *et al* 2006). Near-surface wind speeds are affected not only by topography but also by surface roughness, buildings and obstacles. For

instance, near-surface wind speed conditions can be very different at the top of a ridge, at a mountain pass or at a valley floor. These differences in magnitude are especially relevant for deriving indicators that are non-linear and therefore sensitive to absolute magnitudes, such as the capacity factor (CF) of wind power (Pickering *et al* 2020).

To transfer climate information from coarser to finer scales, many downscaling techniques have been developed and employed in weather and climate studies to refine model outputs. There are essentially two different downscaling approaches. Firstly, dynamical downscaling couples a Regional Climate Model (RCM) to a Global Circulation Model (GCM) over a limited region within the global domain, using the data from the GCM as boundary conditions. The computational costs of dynamical downscaling are rather high and the additional skill is sometimes negligible (Robertson *et al* 2012), which explains its limited use in seasonal predictions (García-Díez *et al* 2015, Schwitalla *et al* 2020).

Secondly, statistical downscaling relies on the assumption that a relationship exists between the large-scale information provided by a GCM and the fine-scale variable. Once a statistical relationship is built, local values—predictands—are inferred using large-scale information—predictors—. Then, future dynamical predictions of the large-scale variables (i.e. those generated using the physically-based equations of the dynamics of the atmosphere) can be inserted as predictors into the statistical relationship to produce local-scale predictions. Statistical downscaling techniques (see Gutierrez *et al* 2013 for a review) can be in turn subdivided depending on whether the statistical model is fitted using observational data for both predictors and predictands (known as Perfect Prognosis or PP; Klein *et al* 1959) or using data from the GCM itself (often referred to as Model Output Statistics or MOS; Glahn and Lowry 1972).

The statistical downscaling approach is relatively easy to implement with climate prediction systems containing several ensemble members and has already been employed in some studies for downscaling temperature and precipitation forecasts at seasonal timescales (e.g. Pavan and Doblas-Reyes 2013, Manzanas *et al* 2018). However, to the best of the authors' knowledge, no attempt has yet been made to downscale seasonal predictions of wind speed.

The selection of the employed predictors is vital for the success of the statistical downscaling method. Not only do the predictors need to be strongly related to the predictand, but also predictable from the dynamical model. Teleconnection indices that summarise the state of the atmospheric circulation are optimal for this purpose. In this work, four Euro-Atlantic Teleconnection (EATC) indices (namely the North Atlantic Oscillation (NAO), East Atlantic (EA), East Atlantic/Western Russia (EAWR) and Scandinavian Pattern (SCA)) are employed as predictors

to anticipate near-surface wind speed conditions in Europe. Those teleconnection indices are strongly related to wind speed conditions in Europe (Zubiate *et al* 2017) and wind power generation (Yang *et al* 2020), and have been recently shown to be predictable (Lledó *et al* 2020). Since the downscaled predictions combine a dynamical forecast of a circulation variable and a statistical relationship with a second variable of interest, they are referred to as hybrid predictions (see chapter 2 in WMO 2020), to differentiate them from purely statistical seasonal forecasts that employ observed values of potential forcing fields to derive the predictions (Kämäräinen *et al* 2019). Hybrid predictions take advantage of the predictability of the EATC indices from dynamical predictions, especially in winter, and thus help to overcome limitation (1). At the same time, the downscaling allows for transferring such information to a finer grid scale, circumventing limitation (2).

The objective of this work is to generate and assess the quality of a hybrid seasonal prediction of near-surface wind speeds and wind power CF by applying a statistical downscaling with a PP approach to a set of dynamical predictions of EATC indices. Sections 2 and 3 describe the data and methodology employed, respectively. Results are presented in section 4 while conclusions are drawn in section 5.

2. Datasets

The hindcasts from five different operationally-produced seasonal prediction systems have been used in this study: the System2 from Deutscher Wetterdienst (DWD2, Deutscher Wetterdienst 2019), the GloSea5-GC2 from the UK Met Office (GS5GC2, Maclachlan *et al* 2015, Williams *et al* 2015), the System 6 from Météo France (MF6, Dorel *et al* 2017), the SEAS5 (Johnson *et al* 2019) from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Seasonal Prediction System 3 from Centro Euro-Mediterraneo sui Cambiamenti Climatici (SPS3, Sanna *et al* 2017). All five prediction systems have been retrieved from the Climate Data Store data portal in a regular grid of $1^\circ \times 1^\circ$ of spatial resolution and covering the 1993–2016 period. Particular details of the employed seasonal prediction systems, as well as the two observational references, can be found in table 1.

The ERA5 HRES (hereafter ERA5) reanalysis dataset (Hersbach *et al* 2020) produced by the ECMWF has been used as the gridded observational reference. The dataset has been downloaded through the ECMWF retrieval system (MARS) in its native grid (i.e. 0.3° approximately), and at 1-hourly time resolution. Then, the ERA5 data has been horizontally interpolated using a conservative approach to match the spatial resolution of the predictions, allowing for bias adjustment and verification at the grid level.

Table 1. Specific details of the datasets employed.

Dataset	Type of dataset	Available period	Time resolution	Horizontal grid spacing	Ensemble members
DWD2	Seasonal prediction	1993–2016	6 h	1° × 1°	30
GS5GC2	Seasonal prediction	1993–2016	6 h	1° × 1°	28
MF6	Seasonal prediction	1993–2016	6 h	1° × 1°	25
SEAS5	Seasonal prediction	1993–2016	6 h	1° × 1°	25
SPS3	Seasonal prediction	1993–2016	6 h	1° × 1°	40
multi-system	Seasonal prediction	1993–2016	Seasonal	1° × 1°	148
ERA5	Reanalysis	1950–present	Hourly	0.3° × 0.3°	—
TTD	<i>In-situ</i> observations	1984–2017	Sub-hourly	Irregular	—

Table 2. Particular details of the 17 tall towers employed in this study. r represents the Pearson correlation coefficient between the seasonal tall tower winds and the ERA5 100 m winds from closest grid point to each tall tower location.

ID	Name	Longitude (deg east)	Latitude (deg north)	Offshore	Measuring height (m)	Original time span ^a	r
T1	Braschaat	4.52	51.31	No	41	1996–2015	0.82
T2	Cabauw	4.92	51.97	No	80	1986–2017	0.91
T3	Cardington	−0.42	51.10	No	50	2004–2013	0.98
T4	Fino1	6.59	55.01	Yes	100	2004–2017	0.96
T5	Fino2	13.15	55.01	Yes	102	2007–2017	0.90
T6	Fino3	7.16	55.20	Yes	100	2009–2017	0.94
T7	Greater Gabbard MMZ	1.92	51.94	Yes	82	2005–2015	0.97
T8	Hamburg University	10.10	53.52	No	110	2004–2017	0.93
T9	Hegyhatsal	16.65	49.96	No	115	1994–2016	0.21 ^b
T10	Inner Dowsing	0.44	53.13	Yes	43	1999–2008	0.83
T11	Juelich	6.22	50.93	No	100	2011–2017	0.88
T12	Lindenberg	14.12	52.17	No	98	1999–2017	0.98
T13	Lutjewad	6.35	53.40	No	60	2001–2017	0.89
T14	Malin Head	−7.33	55.35	No	22	1988–2017	0.89
T15	Obninsk	36.60	55.11	No	121	2007–2016	0.90
T16	Puijo	27.65	62.91	No	75	2005–2016	0.76
T17	Sodankyla	26.64	67.36	No	24	2000–2015	0.83

^a May contain no-data periods.

^b T9 will not be included in the results.

At the local scale, wind speeds measured *in-situ* at 17 tall tower locations over Europe have been considered (see details in table 2 and their spatial distribution in figure S1 (available online at stacks.iop.org/ERL/16/054010/mmedia)). Those observations have been obtained from the Tall Tower Dataset (TTD, Ramon *et al* 2020), a quality-controlled collection of wind data taken at tall meteorological masts of 20 to more than 200 m height. Since these structures measure winds simultaneously at several heights above ground, we have selected at each of the 17 locations the wind speed series which is closest to the 100-metre height. Modern wind turbines are placed at those heights since the wind flow is notably less affected by surface roughness than at surface level. The 17 time series span from 6 to 30 years within the 1984–2017 period. To unify the timespan of the series, and ensure the representativeness of the comparisons against predictions, the 17 time series have been averaged into hourly values and reconstructed to cover the entire 1981–2017 period. To this end, a Measure-Correlate-Predict approach

with a simple linear regression has been employed (see Carta *et al* 2013 for further details), using as the reference series the hourly 100 m wind series of the ERA5's closest grid point to each tall tower location.

3. Methods

3.1. Hybrid predictions

Hybrid predictions for the boreal winter (December–January–February, DJF) have been produced using the PP methodology as represented in figure 1. Once the dynamical forecasts of the predictors (i.e. the EATC indices) are generated, they are used in a statistical model that accounts for variations in wind speed related to variations in the EATC indices. The statistical model has been previously built solely on observations of wind speed and EATC indices. For the purposes of our work, the PP approach represents an advantage over MOS, because (1) it uses one single statistical relationship that can be applied over various dynamical prediction systems, and (2) the amount of data available for fitting the relationship

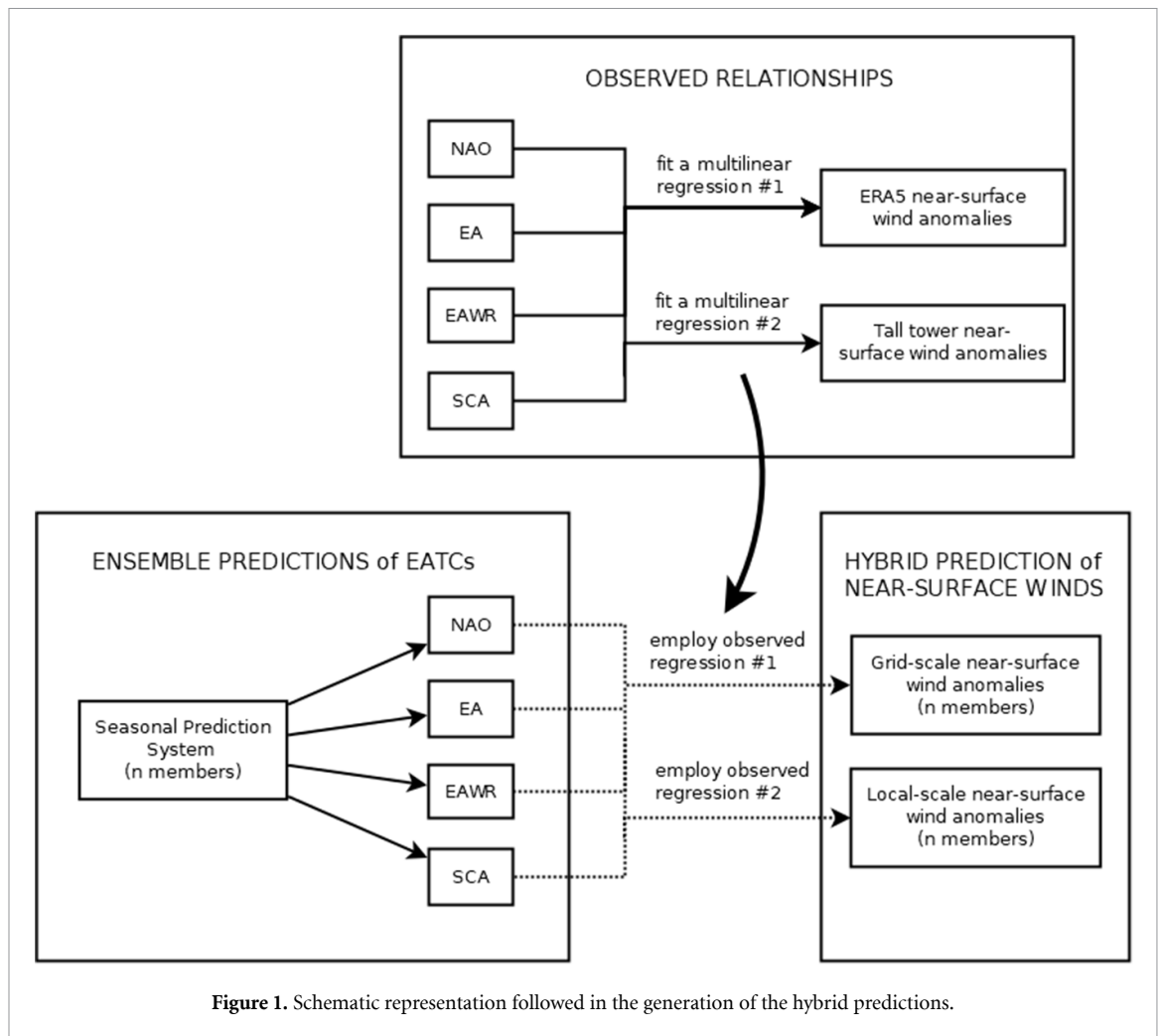


Figure 1. Schematic representation followed in the generation of the hybrid predictions.

is not limited to the length of the hindcast, but to the timespan of the observational series (Marzban *et al* 2006). A more precise description of the different steps of the PP and the generation of the hybrid predictions follows.

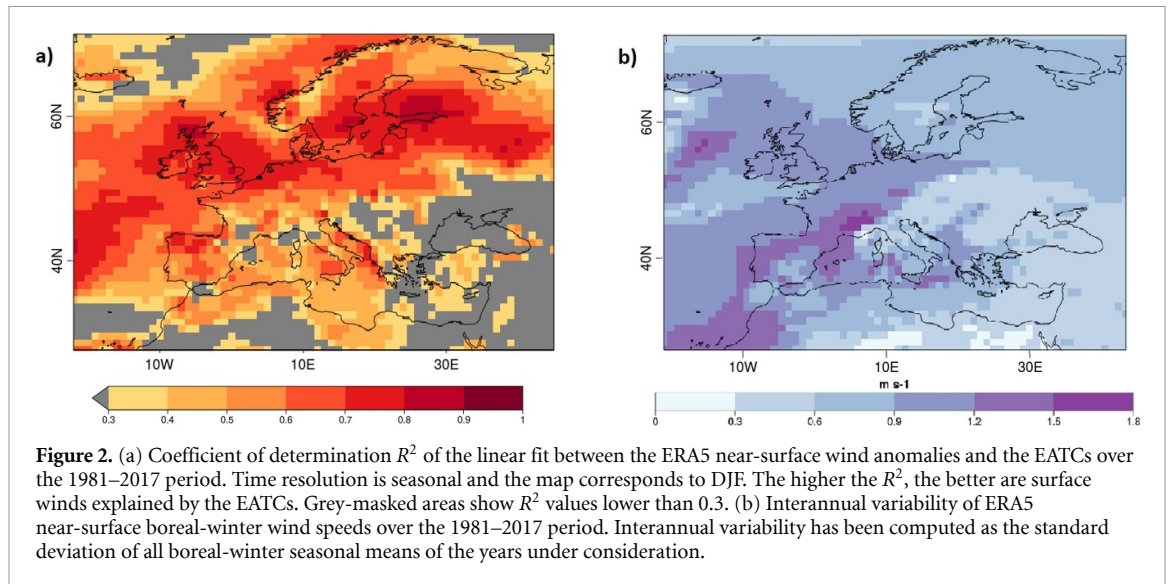
Firstly, four EATC indices are computed as described in Lledó *et al* (2020). The EATC patterns and indices have been derived from the 500 hPa geopotential height field anomalies employing a Rotated Empirical Orthogonal Function (REOF) analysis over the Euro-Atlantic domain [90° W–60° E; 20° N–80° N]. The four teleconnections obtained correspond to the North Atlantic Oscillation (NAO), East Atlantic (EA), East Atlantic/Western Russia (EAWR) and Scandinavian pattern (SCA). This procedure has been followed to obtain both observed—using the ERA5 anomalies—and predicted—employing the anomalies from DWD2, GS5GC2, MF6, SEAS5 and SPS3—EATC indices. The observed EATC patterns are shown in figure 1 in Lledó *et al* (2020).

Then, a statistical model that relates seasonal anomalies of near-surface wind speed and the EATC indices is built from historical observations. A very simple multilinear regression model (equation (1)) has been used here, due to the rather small sample

size available for fitting the model. This method has already been used in Rust *et al* (2015) to model European temperatures from several teleconnections. A model that expresses anomalies of near-surface wind speeds (predictand: w') as a linear combination of the EATC indices (predictors: NAO, EA, EAWR, SCA) is built separately at each grid point or tall tower location (x, y) . The fit adjustment parameters a_n are obtained employing an ordinary least squares method (see their spatial distribution in figure S2). The reference period that is used in all the model fits is 1981–2017. Additionally in the generation of the multilinear models, a leave-one-out cross-validation approach has been considered. The EATC observed indices and its corresponding wind observation of the year under consideration are excluded from the sample used to estimate the fit adjustment parameters. In this way, they can be used later for verification

$$\begin{aligned}
 w'(x, y, t) = & a_0(x, y) + a_1(x, y) * NAO(t) \\
 & + a_2(x, y) * EA(t) + a_3(x, y) * EAWR(t) \\
 & + a_4(x, y) * SCA(t). \quad (1)
 \end{aligned}$$

To avoid overfitting in the statistical model, a selection of the best subset of predictors that retains



the maximum information in the model without necessarily keeping all the predictors is made at each location by using the Akaike Information Criterion and a backward stepwise selection (James *et al* 2013). Albeit using a relatively simple statistical model, the coefficient of determination (R^2) of the fit presented in figure 2(a) shows that the EATCs explain most of the year-to-year variability (also known as interannual variability) in the near-surface winds over extended areas of Europe (figure 2(b)).

In order to obtain an ensemble of hybrid near-surface wind anomaly predictions, the individual-member predictions of EATC indices are inserted into the multi-linear regressions, both at gridded and local scales. The seasonal ensemble predictions of the EATC indices are initialised at the beginning of winter (December) and one, two and three months in advance (i.e. November, October and September, respectively). In this work, lead-zero predictions will refer to those initialised in December, lead-one predictions will be those initialised in November, and so on. Finally, all members from the five prediction system are pooled together to create a new dataset, the multi-system henceforth, with a total of 148 members. Multi-system ensemble predictions can outperform individual-system predictions (Athanasiadis *et al* 2017).

3.2. Wind capacity factor

The wind-based CF index is obtained using the 6 h wind speed data from the predictions, and the 1-hourly winds from the ERA5 and TTD. The conversion between wind speed and power output has been made employing a power curve, which takes into account the specific efficiency characteristics of the wind turbine. Specifically, a power curve for the turbine Type I defined in the IEC-61400-12-1 international standard has been considered (see IEC 2017 and Lledó *et al* 2019 for further information).

Although this turbine type might not be the most suitable for all the investigated locations, it serves the purpose of investigating whether the non-linearities of its power curve affect the quality of the hybrid predictions. Once the conversion is made, CF values are obtained dividing by the nominal power capacity of the turbine. Lastly, seasonal anomalies are calculated.

Hybrid predictions of CF, which might be of particular interest at a turbine or wind farm level within the wind industry, are studied in detail at one tall tower location where local wind effects represent a huge proportion of the seasonal mean wind speed value, and subsequently the seasonal CF value. The relatively low r obtained for Puijo tall tower (T16, table 2) envisages that local wind effects are likely to occur there, and a comparison against a surface station located two kilometres away reveals so (Leskinen *et al* 2009).

3.3. Verification metrics

The quality of the hybrid predictions has been assessed employing both gridded and local-scale observations. Multiple verification scores have been considered to account for different aspects of forecast quality: association, discrimination and reliability (Jolliffe and Stephenson 2012, Mason 2018). In some of those scores, the performance of the hybrid predictions is compared to that of a benchmark prediction. Two different benchmarks have been employed: the climatological forecast (i.e. a 33% of probability for all tercile categories) and the dynamical predictions of near-surface wind speed from the considered systems (table 1). Skill scores using the climatological forecast as a reference are identified with the sub-index c while those using the dynamical prediction use d .

To prepare the dynamical prediction benchmark, seasonal anomalies of surface (10 m) wind speeds for the 1993–2016 period have been obtained at gridded and local scales (in the latter case using a

bilinear interpolation) and then bias-adjusted using a simple bias correction approach (Torralba *et al* 2017). The method adjusts predictions to have an equivalent standard deviation and mean to that of the reference dataset, which has been the ERA5 reanalysis near-surface wind speeds. A leave-one-out cross-validation approach has been again used: the prediction to be adjusted and its corresponding observation are excluded from the sample used to estimate the adjustment parameters (see equations (1)–(4) in Torralba *et al* 2017). The multi-system of the dynamical predictions is also generated by pooling together all the bias-corrected anomalies from the five prediction systems.

The considered scores for the skill assessment are both deterministic and probabilistic, and the R packages `easyVerification` and `SpecsVerification` have been used for their computation:

- The Ensemble Mean Correlation (EMC) quantifies the association (i.e. linear dependency) between observed and predicted wind speeds. The EMC ranges from -1 to 1 , with a value of 1 indicating a perfect association. A Student's t -test at the 95% of confidence level has been applied to emphasise statistically significant areas.
- The Relative Operating Characteristic Skill Score (ROCSS) assesses the discrimination of probabilistic single-category forecasts. Here, predictions are prepared in the form of probability of occurrence of three categories defined by the 33rd and 66th percentiles of the hindcast values. The ROCSS measures the proportion of hits (i.e. correct predictions) versus false alarms (i.e. non-occurrences that were incorrectly predicted) for each of the three categories. The ROCSS ranges from -1 to 1 , with negative values indicating a weaker discrimination capacity than that of the benchmark prediction.
- The Rank Histogram (RH) tests the reliability of the probabilistic predictions, by comparing how the observations rank with respect to the ensemble members of the predictions. Reliable ensemble predictions show a flat RH, which has been statistically assessed with a decomposed Pearson's χ^2 test as in Jolliffe and Primo (2008). When the sample size is small in comparison with the number of ranks available (i.e. the ensemble size), non-flat rank histograms are likely to occur due to randomness, which is not desirable. To prevent this from happening, counts from every ten adjacent bins have been grouped so that the number of ranks has been reduced by a factor of ten.
- The Continuous Ranked Probability Skill Score (CRPSS) measures the quality of the cumulative forecast probability distribution by measuring the distance between the observed and predicted probability distributions. The CRPSS penalises both reliability and resolution—the latter is closely related to discrimination—errors. It ranges from

$-Inf$ to 1 , and positive values indicate an increased skill compared to the benchmark forecast. The Diebold–Mariano test (Diebold and Mariano 1995) has been applied to explore the statistical significance of the differences between the CRPSSs of hybrid and dynamical predictions.

Finally, areas where the hybrid model shows a poor performance based on the R^2 of the statistical fit being smaller than 0.3 —grey areas in figure 2(a)—have been omitted in the verification. Those areas are located around the Black Sea and scattered around the northern Mediterranean, where low values of interannual variability are noted (figure 2(b)). There, winds respond mainly to mesoscale systems rather than large-scale circulation patterns, which may explain the inability of the hybrid model in reproducing the year-to-year variations of near-surface wind speeds. T9 has been omitted in the results as well since the local winds correlate very poorly with the ERA5 winds (table 2), thus not giving robustness to the Measure-Relate-Predict reconstruction.

4. Results

In the following sections, we analyse the skill of the hybrid predictions at the local scale. We complement these results with the verification of grid-scale hybrid forecasts (i.e. adjusted to reanalysis data instead of tower observations) at a pan-European scale [27° N– 72° N; 22° W– 45° E]. This is important because potential users of hybrid predictions may face the limitation of the unavailability of *in-situ* local data needed to generate the predictions. The verification focuses on three key attributes of a probabilistic prediction: association, discrimination and reliability. For the sake of simplicity, results are shown only for the multi-system prediction. Remaining results for the individual systems are available from the authors upon request. We focus on the winter season, when wind speed variability is highest, and so is the importance of its anticipation.

4.1. Do hybrid predictions improve their dynamical counterparts?

The association between the observed and hybrid-predicted near-surface wind anomalies is measured by the EMC and illustrated in figures 3(a)–(d) at both local and grid scales. The EMC is a deterministic metric which is insensitive to forecast errors in the magnitudes and the spread of the ensemble, so only some association with the observations is required for a forecast to be skilful. In this regard, the negative correlation values noted across the Mediterranean basin anticipate a poor performance of the hybrid prediction over there. Conversely, positive and significant correlations above 0.6 have been obtained for lead month zero across northern Europe (figure 3(a)).

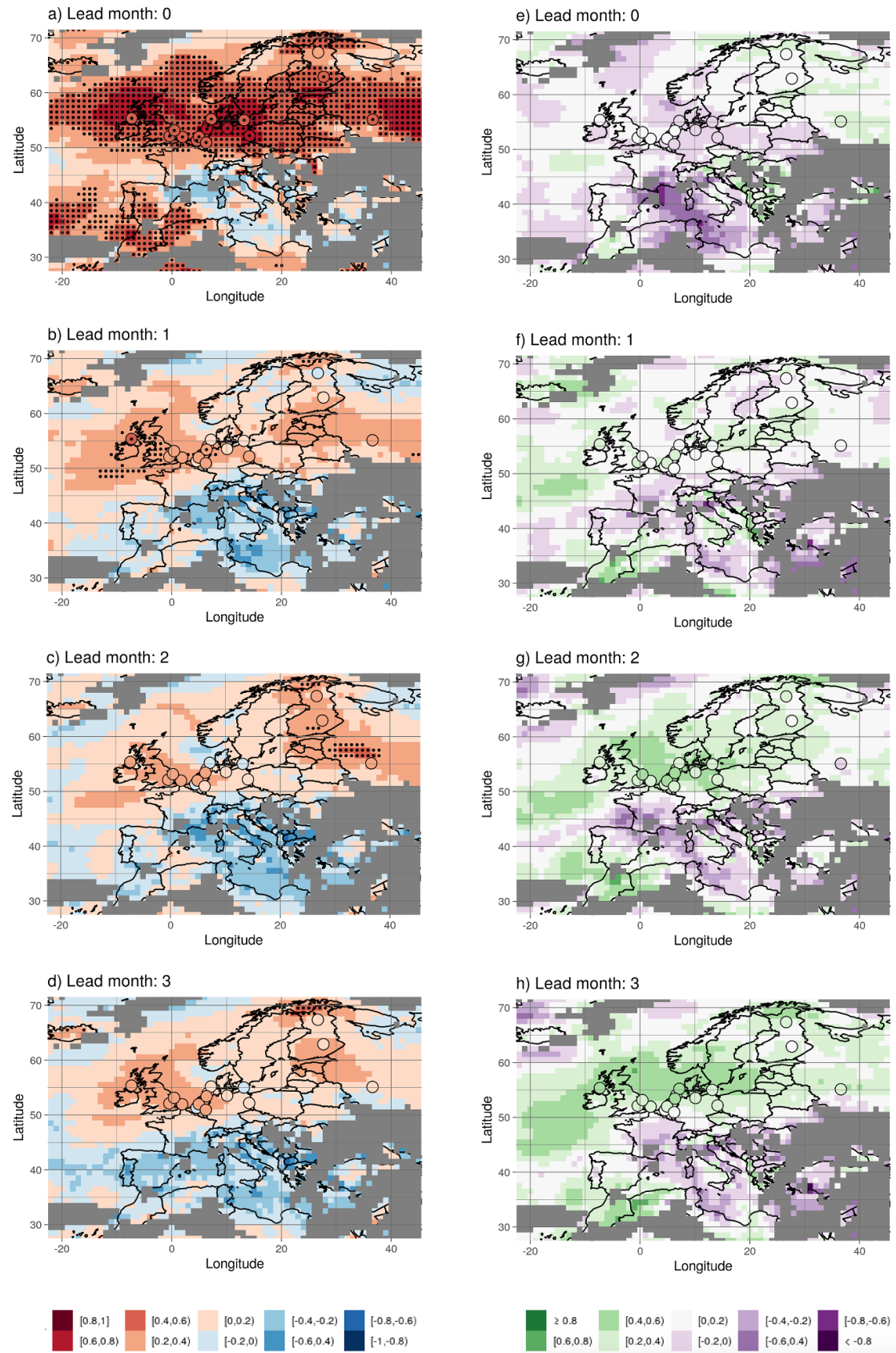
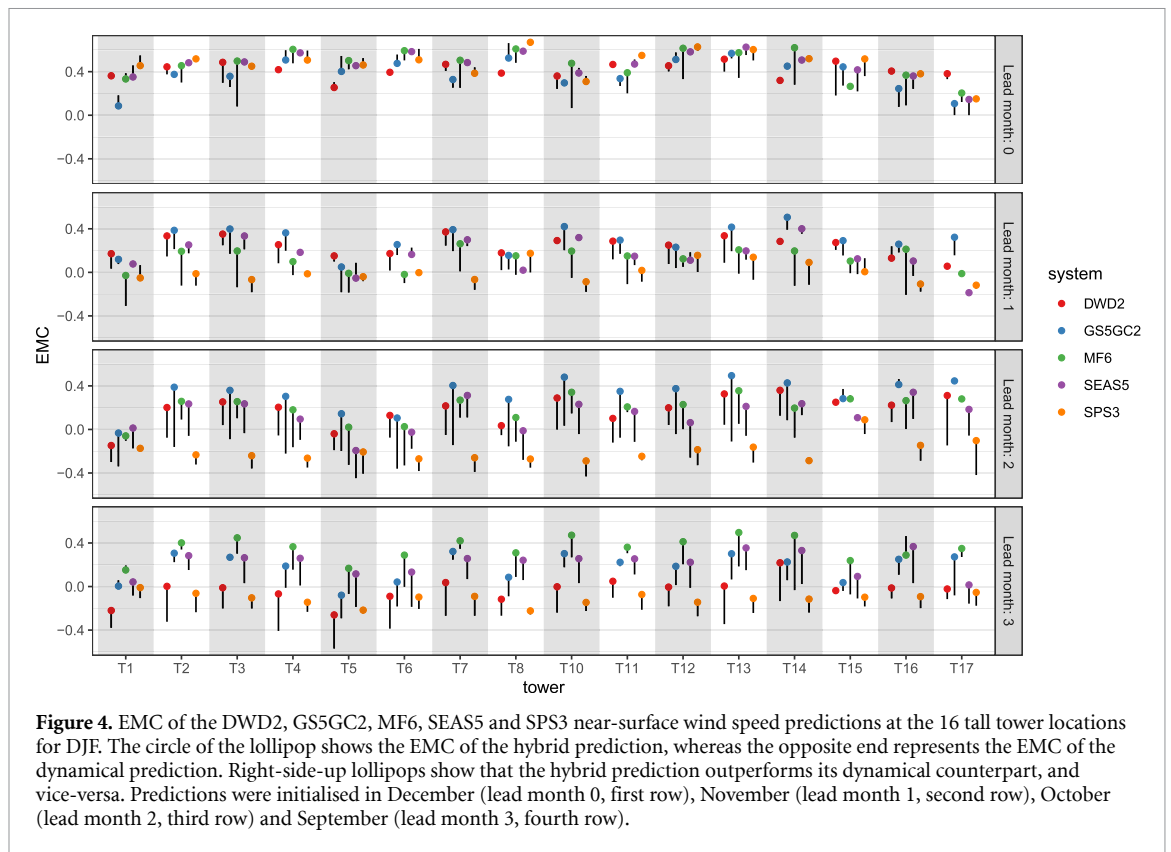


Figure 3. (a)–(d) EMC of the multi-system near-surface wind speed hybrid predictions over Europe for DJF. Black dots highlight areas statistically significant at the 95% of confidence level. (e)–(h) Differences between the EMC of the hybrid predictions and the EMC of the dynamical predictions. Filled points indicate the EMC of the hybrid predictions ((a)–(d)) and the differences in EMC ((e)–(h)) at the 16 tall tower locations. Grey-masked areas indicate regions where the R^2 of the multi-linear regression is below 0.3.



At longer leads, correlations decrease but still depict positive values above 0.4 in the British Isles and the east of the Baltic sea. For the latter region, we observe increased and statistically significant EMC values at lead month two, which are not seen at lead months one and three. This improvement in the hybrid prediction at that particular lead month responds to an increase in the skill values of the EATC predictions. More specifically, the SCA index has the greatest weight in the hybrid model over that region (figure S2), and shows a relative maximum in correlation at lead month two (i.e. 0.42; see table S1). The differences in EMC between the hybrid and dynamical predictions (figures 3(e)–(h)) reveal that the highest gains in skill are seen at the longest leads. While the dynamical forecast offers skill only at leads zero and one (see figure S3), the hybrid prediction shows positive EMCs at all lead times. The increased scores for predictions based on the circulation patterns in the hybrid method appear to match the increase in skill seen in other recent studies (e.g. Scaife *et al* 2014, Baker *et al* 2017).

Results are similar at the local scale. The lollipop plots (figure 4) depict the most noticeable differences between hybrid and dynamical predictions at longer leads, where the improvement of the hybrid prediction is substantial for all systems but the SPS3.

The sensitivity of the predictions to discriminate between observations belonging to different categories has been explored with the ROCSS. The ROCSS_c of the lower-tercile category for the multi-system hybrid

and dynamical predictions is compared in figure 5. While both hybrid and dynamical predictions show similar skill score values at lead zero (panel (a); the density is centred around the $y = x$ line), it is noted that hybrid predictions enhance the discrimination ability at leads one, two and three (panels (b), (c) and (d); most of the density is found above the $y = x$ line). Furthermore, this improvement is not only restricted to a particular region but positive ROCSS_d values are observed all over Europe (not shown).

Analogous results are obtained for the predictions of the upper-tercile category (figure S4). On the other hand, neither hybrid nor dynamical predictions show skill for the central-tercile category (figure S5). The lack of skill in predictions for near-normal is a recurrent issue which has already been addressed in the literature and stems from the definition of the skill scores itself, thus not requiring any physical or dynamical explanation (Van Den Dool and Toth 1991).

To gain more insight into the performance of the hybrid predictions at the local scale, we have selected four tall tower locations to evaluate the reliability of the ensemble predictions by exploring their rank histograms (figure 6). The set of four locations include T2 (Cabauw, The Netherlands), T5 (Fino2, Germany), T15 (Obninsk, Russian Federation) and T16 (Puijo, Finland), which are located in both continental—flat and complex terrain—and offshore platforms across northern Europe.

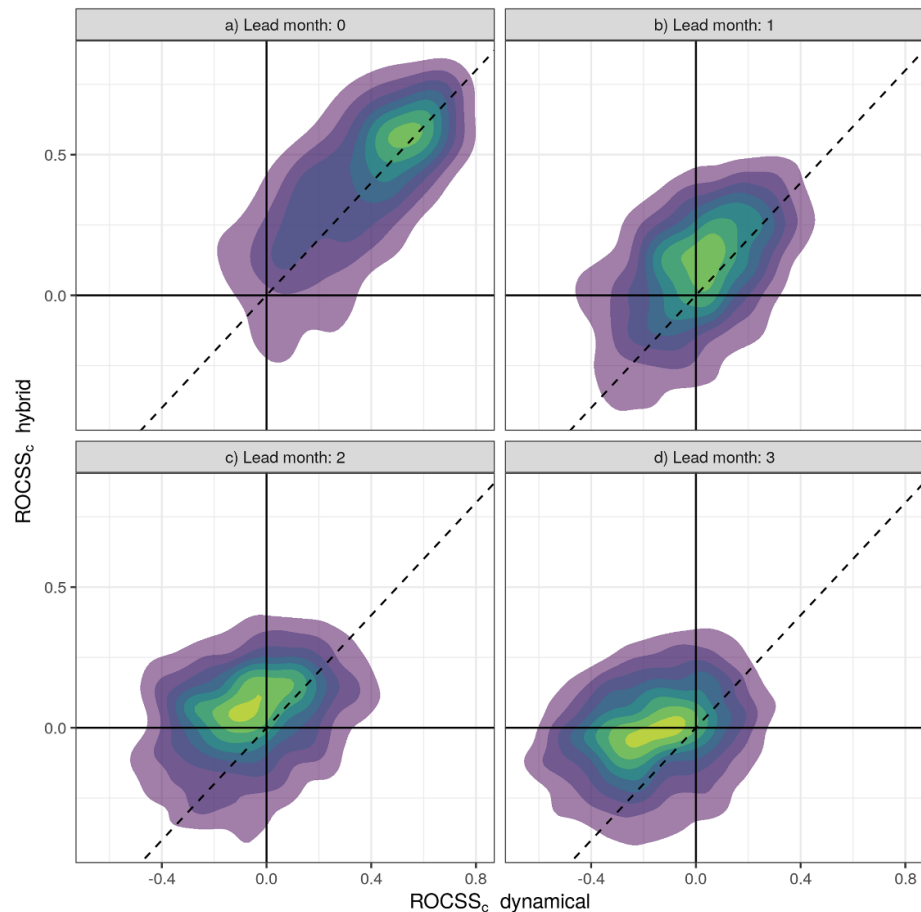


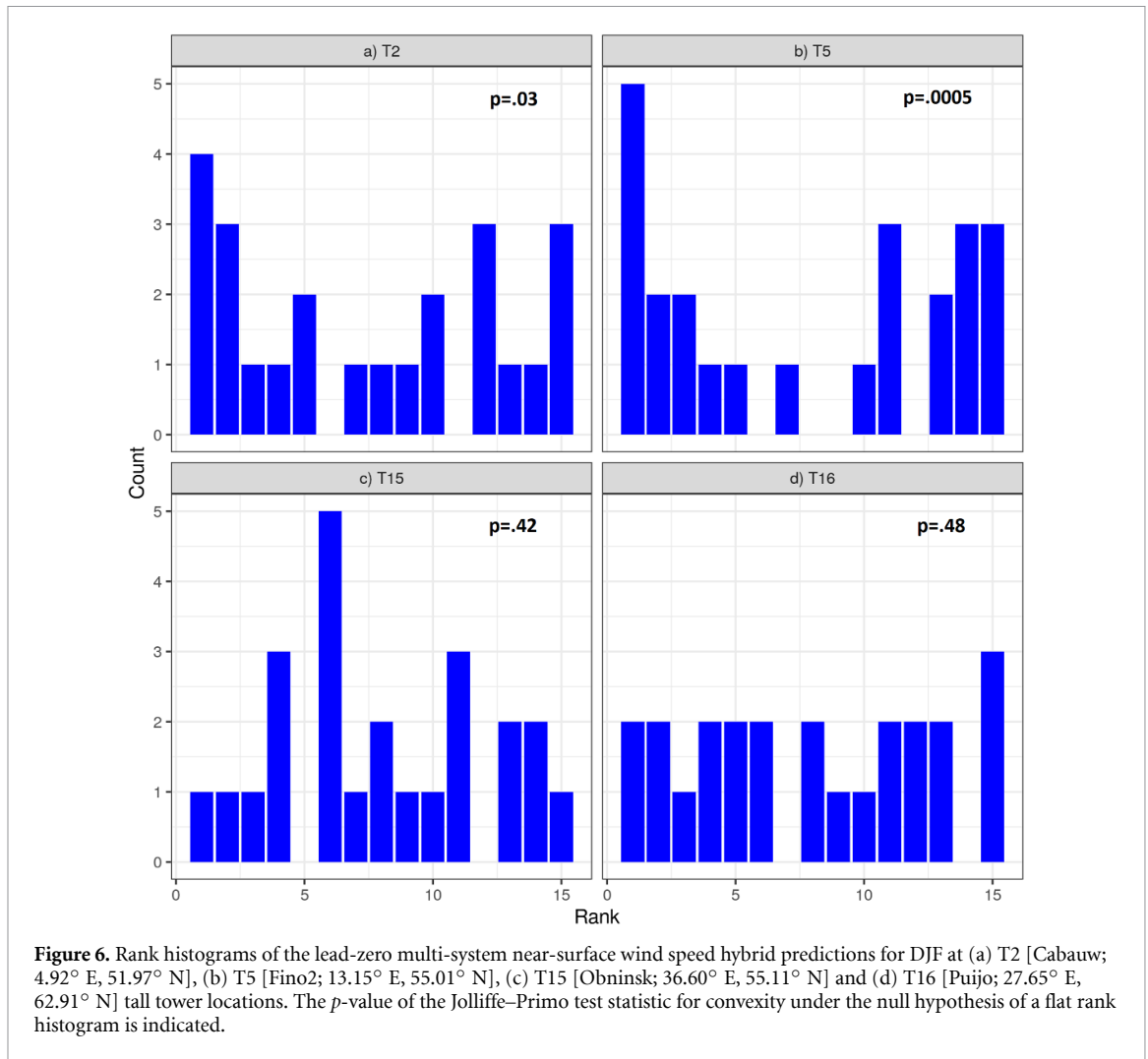
Figure 5. Two-dimensional density plots showing the $ROCSS_c$ of the lower-tertile hybrid and dynamical predictions from the multi-system at all the grid points within the pan-European domain. On a linear scale, greens represent the highest density of points, whereas purples depict the lowest density estimates. The region above the $y = x$ dashed line indicates an improvement of $ROCSS_c$ of the hybrid forecast over its dynamical counterpart. Predictions were initialised in (a) December (lead month 0), (b) November (lead month 1), (c) October (lead month 2) and (d) September (lead month 3). Grid points where the R^2 of the multi-linear regression is below 0.3 have not been included.

Focusing on lead zero, the RHs of the hybrid predictions at T2 and T5 (figures 6(a) and (b), respectively) are both U-shaped, mirroring an overpopulation of the outermost ranks which can occur due to either a lack of ensemble mean signal or a lack of spread around the ensemble mean (Eade *et al* 2014) in the hybrid prediction. The non-flatness of the RH is statistically supported by the p -values of the Jolliffe–Primo statistical test—at the 95% of confidence level. Conversely, the RH at T15 (figure 6(c)) depicts an opposite convexity (i.e. overdispersion), but this outcome is not statistically significant. These results envisage a poor reliability of the multi-system hybrid predictions at these particular locations, which can also be noted for the individual systems (figures S6–S10), especially at T5. The unreliability of the multi-system hybrid predictions is observed in the RHs of 10 out of the 16 tall tower locations, while the other 6 locations show a flatter plot such as that observed at T16 (figure 6(d)). This indicates that the probability distribution of the ensemble at these six locations is in agreement with the observed values. Similar results are obtained for the other leads

(not shown). The performance of the hybrid predictions could be improved further by employing calibration methods (Doblas-Reyes *et al* 2005, Manzanos *et al* 2019) or performing variance corrections to the ensemble mean and members (Eade *et al* 2014).

To complete the skill assessment we compute the CRPSS, a restrictive quality metric of the ensemble distribution that accounts for both discrimination and reliability at the same time. Figure 7 presents the $CRPSS_d$, highlighting areas where the hybrid approach improves (positive values) or degrades (negative values) the dynamical prediction. In general, the results match those discussed for the EMC and ROCSS (figures 3 and 5, respectively) with the highest gains seen for leads two and three. However, the corresponding $CRPSS_c$ values of the hybrid predictions are mostly negative (figure S11). Positive $CRPSS_c$ values are only noted for the lead-zero predictions and, in the case of MF6, the hybrid forecast is the only that offers skill (figure S12).

According to Mason (2004), some scores such as the Ranked Probability Skill Score—and thus the CRPSS—are often too harsh when the climatological



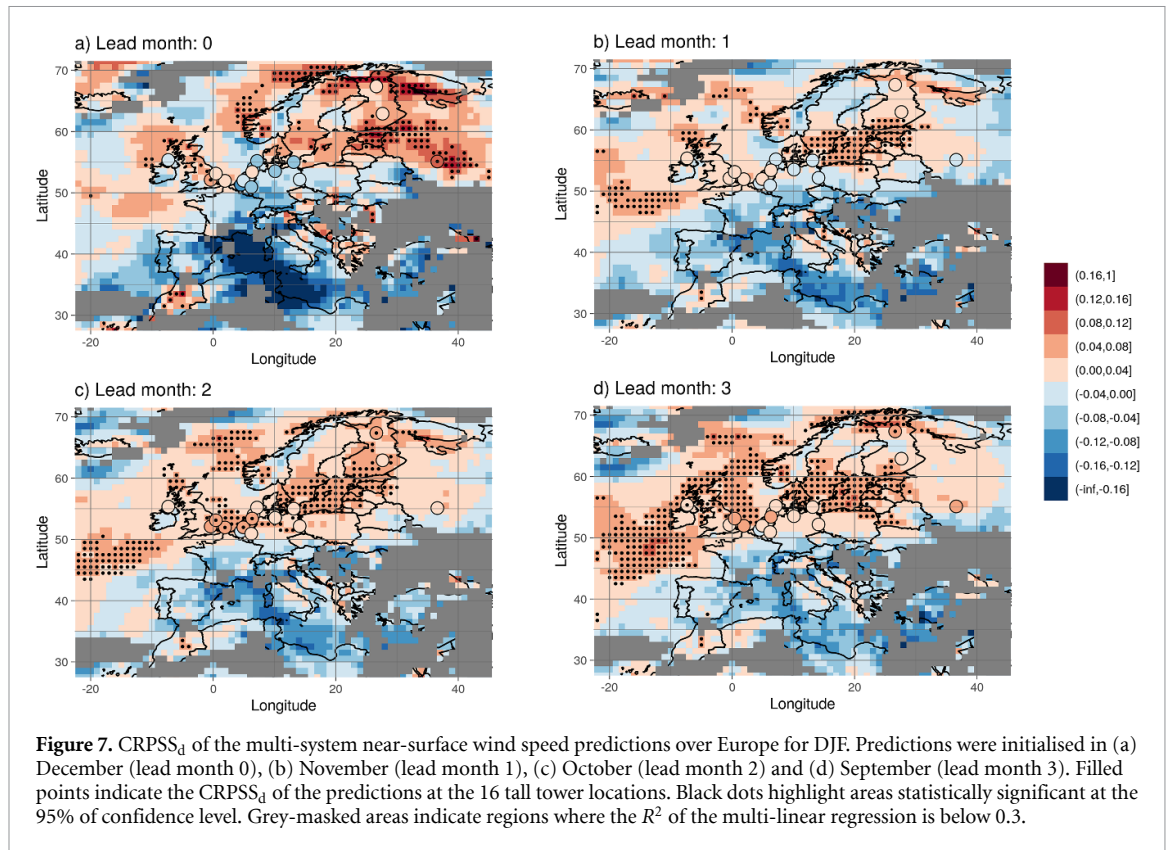
forecast is considered as benchmark. This gives high chances of getting negative values even when predictions provide useful information. Such is the case observed here: although the CRPSS_c is generally negative, we observe gains in association and discrimination and, in some cases, hybrid predictions are reliable. Therefore, one should not rely solely on a single skill score but take into account the whole verification.

4.2. Can hybrid predictions always be trusted at a local scale?

At this point in the results, it has been shown that hybrid predictions improve the dynamical in many aspects, primarily in northern Europe. However, little has been discussed about how hybrid predictions perform at the micro-scale level, especially when local wind effects occur. In the following, we illustrate how hybrid predictions could be applied to predict the absolute values of the wind CF at a location where local wind effects have been reported, and quantify the error made when reanalysis gridded data—which sometimes misrepresent those effects—are used to fit the hybrid model.

The ensemble predictions of CF for Puijo site are presented in figure 8 in the form of Probability Density Functions (PDF). We note that the direct output of the grid-scale hybrid predictions is considerably biased, being the seasonal mean CF systematically underestimated (figure 8(a)). A CRPSS_c value of -4.215 indicates that the prediction is completely useless. A later bias adjustment of this prediction (figure 8(b)) removes the bias and adjusts the variability to that observed at Puijo—though the skill score of the prediction is still negative (-0.046), indicating a similar performance to that of a climatological forecast. Finally, the hybrid prediction fitted with *in-situ* data also adjusts well to the observed CFs, and the CRPSS increases a bit more, up to a positive value of 0.0007 , indicating that the use of local observations with the hybrid method provides the most accurate prediction of seasonal CF values.

The important bias in the grid-scale predictions in figure 8(a) responds to the fact that gridded data are a representation of the average value within a grid cell of hundreds of square kilometres. Therefore, values of variables with high spatial variability such as wind speed in complex terrain regions may differ substantially from the actual values observed at



different locations within the grid cell. This misrepresentation of local values is said to produce representativeness errors. In the case of wind, local effects such as katabatic winds over complex terrain regions may account for a large proportion of the mean wind speed value, thus enlarging the representativeness error of the wind speeds in the reanalysis. These errors are propagated to the CF values, and eventually to the hybrid predictions. Hence, reanalysis gridded data are sometimes not suitable to generate hybrid predictions because these datasets are unable to represent local wind effects occurring at much finer scales, such as those observed at Puijo. A later bias-correction may enhance the grid-scale hybrid predictions, but this post-processing can only be carried out where *in-situ* measurements are available.

5. Summary and conclusions

This research proposes and applies a methodology to overcome two main restraints of seasonal predictions that jeopardises every decision based upon them. The first impediment is the limited skill levels observed in the prediction of surface variables such as wind speed, while the second is the lack of adaptation to the local scale due to the relatively coarse scales in which forecasts are delivered.

Results show that hybrid predictions of near-surface wind speed based on a PP statistical downscaling technique help reduce the effects of both issues simultaneously. Using the indices of the four

main EATCs as predictors, the hybrid predictions proposed here have been shown to improve the skill of the same predictions obtained from a dynamical approach. Besides, the statistical downscaling has enabled to transfer the coarse-scale predictions to a station-scale level, and the comparison with station-based observations has revealed certain level of agreement even when local wind effects play an important role. In particular:

- Hybrid predictions enhance the skill scores of the dynamical predictions at both local and pan-European scales.
- In general, hybrid predictions are able to provide skill at leads two and three, while dynamical forecasts cannot.
- The highest gains in quality are observed in the association with the observations and the discrimination against the different observed outcomes.
- Although hybrid predictions can also be built using reanalyses, it is advisable not to use gridded data to build the statistical model over areas where local effects are considerable.
- EATC predictions—and thus hybrid predictions—provide no added value in the Mediterranean basin.

Hybrid forecasts foster the information available in the EATC predictions to anticipate near-surface wind speed or CF anomalies. The derived predictions are consistent with the main features of the atmospheric circulation, which are summarised in

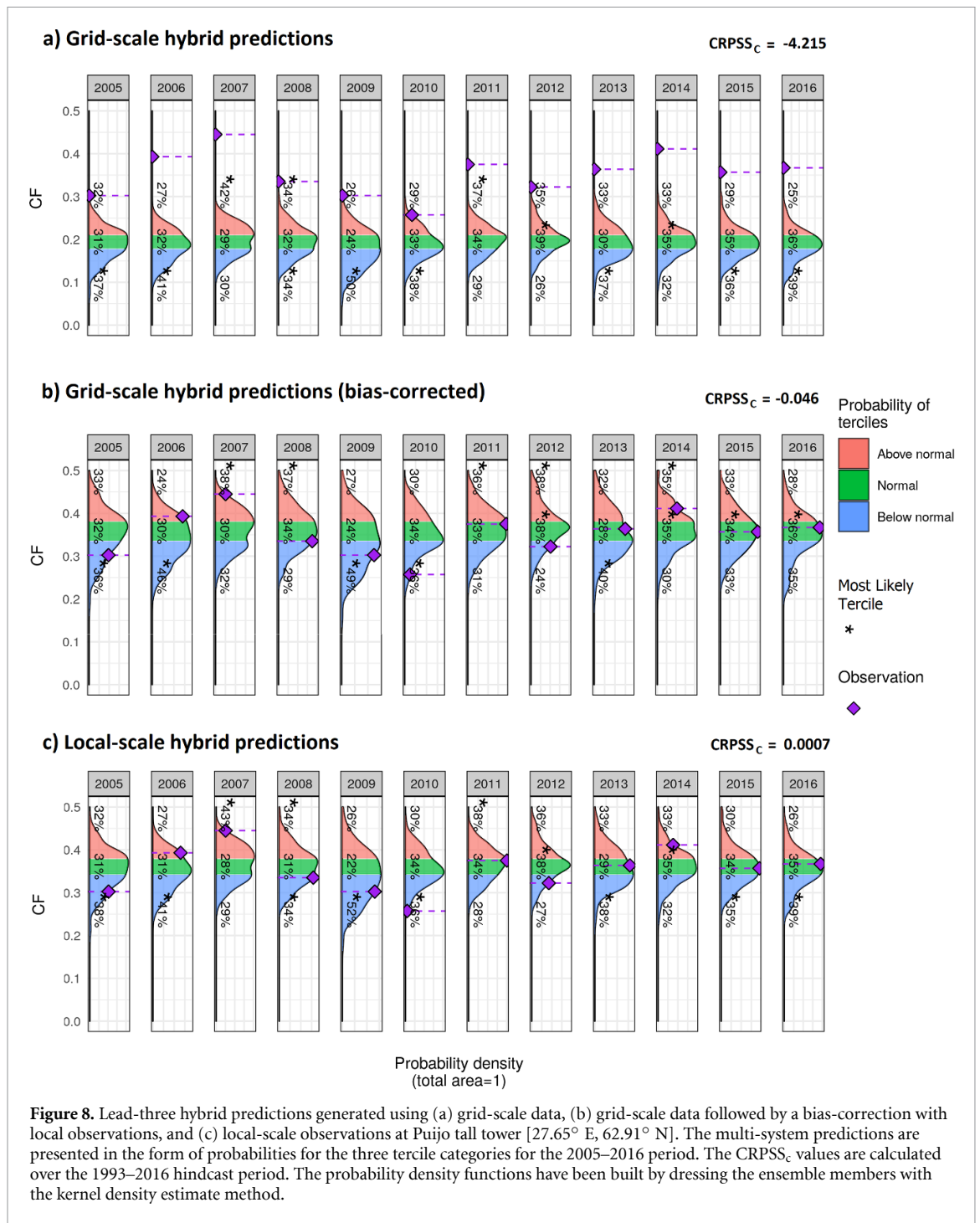


Figure 8. Lead-three hybrid predictions generated using (a) grid-scale data, (b) grid-scale data followed by a bias-correction with local observations, and (c) local-scale observations at Puijo tall tower [27.65° E, 62.91° N]. The multi-system predictions are presented in the form of probabilities for the three tercile categories for the 2005–2016 period. The CRPSS_c values are calculated over the 1993–2016 hindcast period. The probability density functions have been built by dressing the ensemble members with the kernel density estimate method.

the status of the EATCs. This provides interpretability of the results, which enables users to make more informed decisions. For example, one can link higher winds across the UK and the North Sea to a positive NAO phase.

The wind power industry is one of the potential users that can profit most from hybrid predictions. Wind and CF forecasts have been proven to offer useful results at a wind farm scale, provided that site observations from a met mast are available. Moreover, the skilfulness is not restricted to the shortest leads—as it is often the case of the dynamical forecasts—but hybrid predictions issued two or three months in

advance can already anticipate understanding of the conditions for the coming season.

The PP is a simple and effective approach but also suffers from some limitations. For instance, the proposed hybrid model does not account for the biases in the EATC predictions. Future work may look into existing post-processing methods like calibration techniques to bias-correct the model output. Besides, the optimal number of EATCs employed to explain the wind variability can be tuned for each region as in Bastien (2018) (chapter 3), who found varying results over France. Investigating whether these improvements lead to a marginal or substantial increase in

skill would be valuable for any potential user of the hybrid predictions.

Data availability statement

Most of the data used in this article can be accessed from publicly available sources: [Climate Data Store](#) and [Tall Tower Dataset](#).

Acknowledgments

Authors acknowledge the funding support from project INDECIS co-funded by the H2020 ERA-net ERA4CS (GA 690462) and the MICINN Grant No. BES-2017-082216 ('Ayudas para contratos predoctorales'). The Copernicus Climate Change Service (C3S) is also acknowledged for providing seasonal predictions from several European meteorological centres, and the ERA5 reanalysis. We also thank the developers of the R packages SpecsVerification, easyVerification and CSTools. Finally, we want to express our gratitude to the 17 tall tower data providers, namely, the CESAR observatory, the Crown State, Met Éireann, the BMWi (Bundesministerium fuer Wirtschaft und Energie, Federal Ministry for Economic Affairs and Energy), the PTJ (Projekttraeger Juelich, project executing organization); and Principal Investigators Dr Fred Bosveld, Dr Ingo Lange, Dr Laszlo Haszpra, Dr Jan Schween and Dr Frank Beyrich. This article benefited from useful comments and feedback from three anonymous reviewers.

ORCID iDs

Jaume Ramon  <https://orcid.org/0000-0003-2818-5206>

Llorenç Lledó  <https://orcid.org/0000-0002-8628-6876>

References

- Anders A M, Roe G H, Hallet B, Montgomery D R, Finnegan N J and Putkonen J 2006 Spatial patterns of precipitation and topography in the Himalaya *Geol. Soc. Am.* **398** 39–53
- Athanasiadis P J, Bellucci A, Scaife A A, Hermanson L, Matera S, Sanna A, Borrelli A, MacLachlan C and Gualdi S 2017 A multisystem view of wintertime NAO seasonal predictions *J. Clim.* **30** 1461–75
- Baker L H, Shaffrey L C and Scaife A A 2017 Improved seasonal prediction of UK regional precipitation using atmospheric circulation *Int. J. Climatol.* **38** e437–53
- Bastien A 2018 Seasonal forecasting of wind energy resource and production in France, and associated risk PhD Thesis
- Buontempo C et al 2018 What have we learnt from EUPORIAS climate service prototypes? *Clim. Serv.* **9** 21–32
- Carta Je A, Velázquez S and Cabrera P 2013 A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site *Renew. Sustain. Energy Rev.* **27** 362–400
- Clark R T, Bett P E, Thornton H E and Scaife A A 2017 Skilful seasonal predictions for the European energy industry *Environ. Res. Lett.* **12** 119602
- Deutscher Wetterdienst 2019 Seasonal forecasting with the german climate forecast system (available at: https://www.dwd.de/EN/ourservices/seasonals_forecasts/project_description.html?nn=641552&lsId=619784)
- Diebold F X and Mariano R 1995 Comparing predictive accuracy *J. Bus. Econ. Stat.* **13** 253–63
- Doblas-Reyes F J, Hagedorn R and Palmer T N 2005 The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination *Tellus A* **57** 234–52
- Dorel L, Ardilouze C, Déqué M, Batté L and Guérémy J F 2017 Documentation of the METEO-FRANCE pre-operational seasonal forecasting system METEO-FRANCE *Technical Report* (available at: http://seasonal.meteo.fr/sites/data/Documentation/doc_modele/Model_MF-S6_C3S_technical_en.pdf)
- Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L and Robinson N 2014 Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.* **41** 5620–8
- García-Díez M, Fernández J, San-Martín D, Herrera S and Gutiérrez J M 2015 Assessing and improving the local added value of WRF for wind downscaling *J. Appl. Meteorol. Climatol.* **54** 1556–68
- Glahn H R and Lowry D A 1972 The use of model output statistics (MOS) in objective weather forecasting *J. Appl. Meteorol.* **11** 1203–11
- Gutierrez J M, Bedia J, Benestad R and Pagé C 2013 Local predictions based on statistical and dynamical downscaling *Technical Report* 308378 (SPECS)
- Hersbach H et al 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1–51
- IEC 2017 International Standard—IEC61400-12-1 *Technical Report* (International Electrotechnical Commission) (available at: <https://webstore.iec.ch/publication/26603>)
- James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning (Springer Texts in Statistics vol 103)* (New York: Springer)
- Johnson S J et al 2019 SEAS5: the new ECMWF seasonal forecast system *Geosci. Model Dev.* **12** 1087–117
- Jolliffe I T and Primo C 2008 Evaluating rank histograms using decompositions of the Chi-square test statistic *Mon. Weather Rev.* **136** 2133–9
- Jolliffe I T and Stephenson D B 2012 *Forecast Verification* (Oxford: Wiley)
- Kämäräinen M, Uotila P, Karpechko A Y U, Hyvärinen O, Lehtonen I and Räisänen J 2019 Statistical learning methods as a basis for skillful seasonal temperature forecasts in Europe *J. Clim.* **32** 5363–79
- Klein W H, Lewis B M and Enger I 1959 Objective prediction of five-day mean temperatures during winter *J. Meteorol.* **16** 672–82
- Leskinen A, Portin H, Komppula M, Miettinen P, Arola A, Lihavainen H, Hatakka J, Laaksonen A and Lehtinen K E J 2009 Overview of the research activities and results at Puijo semi-urban measurement station *Boreal Environ. Res.* **14** 576–90 (<https://helda.helsinki.fi/handle/10138/233514>)
- Lledó L, Cionni I, Torralba V, Bretonnière P-A and Samsó M 2020 Seasonal prediction of Euro-Atlantic teleconnections from multiple systems *Environ. Res. Lett.* **15** 074009
- Lledó L, Torralba V, Soret A, Ramon J and Doblas-Reyes F J 2019 Seasonal forecasts of wind power generation *Renew. Energy* **143** 91–100
- MacLachlan C et al 2015 Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system *Q. J. R. Meteorol. Soc.* **141** 1072–84
- Manzanas R, Gutiérrez J M, Bhend J, Hemri S, Doblas-Reyes F J, Torralba V, Penabaz E and Brookshaw A 2019 Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset *Clim. Dyn.* **53** 1287–305
- Manzanas R, Lucero A, Weisheimer A and Gutiérrez J M 2018 Can bias correction and statistical downscaling methods improve

- the skill of seasonal precipitation forecasts? *Clim. Dyn.* **50** 1161–76
- Marzban C, Sandgathe S and Kalnay E 2006 MOS, perfect prog and reanalysis *Mon. Weather Rev.* **134** 657–63
- Mason S J 2004 On using ‘climatology’ as a reference strategy in the Brier and the ranked probability skill scores *Mon. Weather Rev.* **132** 1891–5
- Mason S J 2018 Guidance on verification of operational seasonal climate forecasts *Technical Report* (WMO) (available at: [www.seevccc.rs/SEECOF/SEECOF-10/SEECOF-LRF-TRAINING/November 13th 2013/CCI verification recommendations.pdf](http://www.seevccc.rs/SEECOF/SEECOF-10/SEECOF-LRF-TRAINING/November%2013th%202013/CCI%20verification%20recommendations.pdf))
- Merryfield W J et al 2020 Current and emerging developments in subseasonal to decadal prediction *Bull. Am. Meteorol. Soc.* **101** 1–90
- Pavan V and Doblas-Reyes F J 2013 Calibrated multi-model ensemble summer temperature predictions over Italy *Clim. Dyn.* **41** 2115–32
- Pickering B, Grams C M and Pfenninger S 2020 Sub-national variability of wind power generation in complex terrain and its correlation with large-scale meteorology *Environ. Res. Lett.* **15** 044025
- Ramon J, Lledó L, Pérez-Zañón N, Soret A and Doblas-Reyes F J 2020 The tall tower dataset: a unique initiative to boost wind energy research *Earth Syst. Sci. Data* **12** 429–39
- Robertson A W, Qian J H, Tippett M K, Moron V and Lucero A 2012 Downscaling of seasonal rainfall over the philippines: dynamical versus statistical approaches *Mon. Weather Rev.* **140** 1204–18
- Rust H W, Richling A, Bissolli P and Ulbrich U 2015 Linking teleconnection patterns to European temperature—a multiple linear regression model *Meteorol. Z.* **24** 411–23
- Sanna A, Borrelli A, Athanasiadis P J, Matera S, Storto A, Navarra A, Tibaldi S and Gualdi S 2017 RP0285—CMCC-SPS3: the CMCC seasonal prediction system 3 *Technical Report* (Centro Euro-Mediterraneo sui Cambiamenti Climatici) (available at: www.cmcc.it/publications/rp0285-cmcc-sps3-the-cmcc-seasonal-prediction-system-3)
- Scaife A A et al 2014 Skillful long-range prediction of European and North American winters *Geophys. Res. Lett.* **41** 2514–19
- Schwitalla T, Warrach-Sagi K, Wulfmeyer V and Resch M 2020 Near-global-scale high-resolution seasonal simulations with WRF-Noah-MP v.3.8.1 *Geosci. Model Dev.* **13** 1959–74
- Torralla V, Doblas-Reyes F J, MacLeod D, Christel I and Davis M 2017 Seasonal climate prediction: a new source of information for the management of wind energy resources *J. Appl. Meteorol. Climatol.* **56** 1231–47
- Van Den Dool H M and Toth Z 1991 Why do forecasts for ‘near normal’ often fail? *Weather Forecast.* **6** 76–85
- Williams K D et al 2015 The met office global coupled model 2.0 (GC2) configuration *Geosci. Model Dev.* **8** 1509–24
- WMO 2020 *Guidance on Operational Practices for Objective Seasonal Forecasting* 1246 (available at: https://library.wmo.int/doc_num.php?explnum_id=10314)
- Yang W, Foster K, Llorenç Lledó, Torralba V, Cortesi N, Schaller N, Cionni I, De Felice M, Brayshaw D and Bloomfield H 2020 Modes of variability in Europe and their impact on the energy indicators S2S4E project (available at: https://s2s4e.eu/sites/default/files/2020-06/s2s4e_d32.pdf)
- Zubieta L, McDermott F, Sweeney C and O’Malley M 2017 Spatial variability in winter NAO—wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns *Q. J. R. Meteorol. Soc.* **143** 552–62