



HAL
open science

How Skillful Are the European Subseasonal Predictions of Wind Speed and Surface Temperature?

Naveen Goutham, Riwal Plougonven, Hiba Omrani, Sylvie Parey, Peter Tankov, Alexis Tantet, Peter Hitchcock, Philippe Drobinski

► **To cite this version:**

Naveen Goutham, Riwal Plougonven, Hiba Omrani, Sylvie Parey, Peter Tankov, et al.. How Skillful Are the European Subseasonal Predictions of Wind Speed and Surface Temperature?. *Monthly Weather Review*, 2022, 150, pp.1621-1637. 10.1175/MWR-D-21-0207.1 . insu-03778105

HAL Id: insu-03778105

<https://insu.hal.science/insu-03778105v1>

Submitted on 10 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How Skillful Are the European Subseasonal Predictions of Wind Speed and Surface Temperature?

NAVEEN GOUTHAM,^{a,b} RIWAL PLOUGONVEN,^b HIBA OMRANI,^a SYLVIE PAREY,^a PETER TANKOV,^c ALEXIS TANTET,^b PETER HITCHCOCK,^d AND PHILIPPE DROBINSKI^b

^a EDF Laboratory Paris-Saclay, Palaiseau, France

^b Laboratoire de Météorologie Dynamique-IPSL, Ecole Polytechnique, Institut Polytechnique de Paris, ENS, PSL Research University, Sorbonne Université, CNRS, France

^c CREST/ENSAE, Institut Polytechnique de Paris, Palaiseau, France

^d Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York

(Manuscript received 4 August 2021, in final form 3 March 2022)

ABSTRACT: Subseasonal forecasts of 100-m wind speed and surface temperature, if skillful, can be beneficial to the energy sector as they can be used to plan asset availability and maintenance, assess risks of extreme events, and optimally trade power on the markets. In this study, we evaluate the skill of the European Centre for Medium-Range Weather Forecasts' subseasonal predictions of 100-m wind speed and 2-m temperature. To the authors' knowledge, this assessment is the first for the 100-m wind speed, which is an essential variable of practical importance to the energy sector. The assessment is carried out on both forecasts and reforecasts over European domain gridpoint wise and also by considering several spatially averaged domains, using several metrics to assess different attributes of forecast quality. We propose a novel way of synthesizing the continuous ranked probability skill score. The results show that the skill of the forecasts and reforecasts depends on the choice of the climate variable, the period of the year, and the geographical domain. Indeed, the predictions of temperature are better than those of wind speed, with enhanced skill found for both variables in the winter relative to other seasons. The results also indicate significant differences between the skill of forecasts and reforecasts, arising mainly due to the differing ensemble sizes. Overall, depending on the choice of the geographical domain and the forecast attribute, the results show skillful predictions beyond 2 weeks, and in certain cases, up to 6 weeks for both variables, thereby encouraging their implementation in operational decision-making.

KEYWORDS: Europe; Subseasonal variability; Forecast verification/skill; Temperature


1. Introduction

Subseasonal to seasonal (S2S) predictions (Vitart et al. 2017; Robertson and Vitart 2018), which refer to predictions beyond 2 weeks and up to a season, are influenced by both atmospheric initial conditions and boundary forcings (Hoskins 2012). Issuing skillful predictions on S2S time scale used to be considered difficult as it was thought that this time scale was both too long for the memory in the initial conditions to persist and too short for the changes in the boundary conditions to have a significant impact (Molteni et al. 1986; Robertson and Vitart 2018). However, recent studies (Hoskins 2012; Robertson and Vitart 2018) have shown otherwise by identifying the key sources of predictability on S2S time scale, which are the Madden-Julian oscillation (MJO) (e.g., Jones et al. 2004a,b), soil moisture (e.g., Koster et al. 2011; van den Hurk et al. 2012), snow cover (e.g., Sobolowski et al. 2010; Lin and Wu 2011), stratosphere-troposphere interaction (e.g., Baldwin et al. 2003), and ocean conditions (e.g., Woolnough et al. 2007; Fu et al. 2007). Although the predictability of small-scale phenomena and intraday variations on S2S time scales

remains poor (Robertson and Vitart 2018), predictability may persist for large scale phenomena. It is thus critical to aggregate/average values on relevant spatiotemporal scales in order to extract the predictable component of the signal by filtering out motions that behave like noise (Zhu et al. 2014).

In practice, both weather and seasonal predictions are carried out from imperfect initial conditions using imperfect numerical models (Robertson and Vitart 2018). S2S predictions fall beyond the theoretical limit of deterministic predictability (i.e., 10 days) (Lorenz 1965; Jifan 1989; Zhang et al. 2019), and hence these forecasts are produced using ensembles of numerical integrations: a future state of the atmosphere is then a range of possibilities. This transition from deterministic to probabilistic approach has been a major breakthrough in extending the skill horizon of S2S forecasts (Palmer 2012).

A continuously growing share of renewable power systems in the energy mix (International Energy Agency 2020), and changing frequency and intensity of extreme events in the form of storms, heat waves, and cold spells (Seneviratne et al. 2012) make the energy sector one of the most prominent potential end-users of S2S forecasts (White et al. 2017). The energy industry can greatly benefit from skillful S2S forecasts of geophysical variables as they can be used to plan asset availability and maintenance, assess and allocate risks of extreme events on production and consumption several weeks in advance in the framework of the "Ready-Set-Go!" approach

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Naveen Goutham, naveen.goutham@edf.fr or naveen.goutham@polytechnique.edu

DOI: 10.1175/MWR-D-21-0207.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

(White et al. 2017), improve grid efficiency, and optimally trade power on the markets. In the recent years, several studies have been conducted to assess the skill of S2S forecasts: Lynch et al. (2014) evaluated the skill of the European Centre for Medium-Range Weather Forecasts' (ECMWF) extended-range forecasts of 10-m wind speed between 2008 and 2013 for the winter months at the weekly time scale over Europe, and they found statistically significant skill beyond 14 days. Lledó and Doblas-Reyes (2020) assessed the impacts of strong MJO events on 10-m wind speed over Europe and developed a hybrid statistical–dynamical model to better predict 10-m wind speed conditioned on the MJO status. Büeler et al. (2020) studied windows of opportunity to have enhanced skill for ECMWF's S2S predictions of country and month-ahead-averaged quantities of 10-m wind speed, 2-m temperature, and precipitation following anomalous stratospheric polar vortex (SPV) events, and they found enhanced/reduced skill over certain regions in Europe following strong SPV events. Monhart et al. (2018) assessed the skill of the ECMWF's subseasonal forecasts of surface temperature and precipitation against several ground based-station data across Europe and found higher skill for temperature forecasts as compared with precipitation forecasts. They also demonstrated that the skill of temperature across Europe shows a seasonal pattern with higher skill observed in winter relative to other seasons, and a spatial pattern with improved skill observed in northern Europe. Vigaud et al. (2019) evaluated the skill of the surface temperature predictions from several forecasting systems over North America, and found skillful predictions beyond 2 weeks. Diro and Lin (2020) assessed the skill of the S2S forecasts of snow water equivalent and surface temperature from several models within the subseasonal experiment project. They also built a link between the two variables concluding that the weak snow–temperature coupling strength in the models is one of the contributing factors for lower skill of temperature forecasts. Dorrington et al. (2020) quantified the skill of S2S forecasts of surface temperature averaged across France from an end-user perspective, and emphasized basing the assessment of forecasts keeping potential end-user applications in mind.

For the energy sector, the 100-m wind speed forecasts are crucial to estimate the energy extracted from the wind (Jourdiér 2015). Nevertheless, as per our knowledge, there is no published peer-reviewed work on the assessment of S2S 100-m wind speed forecast skill. Because the 100-m wind speed is both closer to the turbine hub height and better represented in the ECMWF model relative to the 10-m wind speed (Alonzo et al. 2018), and since the vertical extrapolation of wind speed from 10 m to the turbine hub height could lead to significant errors (Jourdiér 2015), it is important to assess the skill of S2S forecasts of 100-m wind speed and understand their predictability limits. A vast majority of the published peer-reviewed research to date on the assessment of S2S surface temperature forecast skill are either limited to some ground based stations (e.g., Monhart et al. 2018) or a specific geographic domain (e.g., Vigaud et al. 2019) or restricted by a single metric (e.g., Diro and Lin 2020). In addition, the fast pace of change and improvement of S2S prediction systems is such that it is necessary to regularly revisit and update the assessment of their skill (Vitart 2014).

Although we agree with Dorrington et al. (2020) on the need and value of assessing forecasts based on end-user applications, it is also useful and complementary to assess the skill of the S2S forecasts of purely meteorological variables: this provides a baseline measure of the general skill of the forecasts, indicative independently of specific applications. This serves as a reference for further attempts to improve forecasts.

This study examines the skill of the S2S forecasts and reforecasts [note that for brevity “(re)forecasts” will be used hereinafter to indicate “forecasts and reforecasts” when referring to both at once] of 100-m wind speed and 2-m temperature at the weekly time scale in the recent versions of the ECMWF's S2S prediction system to understand the differences of skill that may arise due to differing ensemble sizes between the forecasts and the reforecasts. The assessment is carried out systematically across the European domain gridpoint wise and also by considering several spatially averaged country-sized domains to identify geographical regions with enhanced/reduced skill, using several metrics (Coelho et al. 2018) for providing a comprehensive overview of the forecast quality. The seasonal cycle of skill is also investigated in the study. The article is organized as follows: section 2 outlines the data used, section 3 describes in detail the method employed to evaluate the skill of (re)forecasts, section 4 explains the results obtained, and section 5 discusses the key findings and provides concluding remarks.

2. Data

a. Forecasts and hindcasts

The operational S2S predictions (Vitart et al. 2017) from the ECMWF model (Vitart et al. 2019) are produced by extending the medium range forecasts (i.e., up to 2 weeks) to 46 days 2 times per week (at 0000 UTC on Mondays and Thursdays). These are ensemble predictions resulting from coupled ocean–atmosphere integrations. The ensemble is composed of 51 members (50 perturbed + control) obtained using singular vectors (Leutbecher 2005). Model uncertainty is represented through the stochastically perturbed parameterization tendencies scheme (Buizza et al. 1999; Palmer et al. 2009). These predictions are originally produced at a horizontal resolution of Tco639L91 (~18 km) up to a lead time of 15 days, and Tco319L91 (~36 km) thereafter (Robertson and Vitart 2018).

As a result of the imperfect representation of the physical processes and the inherent atmospheric unpredictability (Zhang et al. 2019; Žagar and Szunyogh 2020) in the prediction models, these models tend to drift significantly from the reality after approximately 1 week (or at most 10 days) of integrations. This drift needs to be corrected to obtain the maximum value out of forecasts. To do this, ECMWF produces a set of 20 hindcasts (or reforecasts) with 11 (10 perturbed + control) ensemble members each. These hindcasts are initialized using ERA5 reanalysis, and are issued for each of the past 20 years starting from the same date as the operational forecast. To illustrate, if the operational ensemble forecast with 51 members is initiated on 14 February 2019, the hindcast

set consists of 11 ensemble members starting on 14 February 1999, 14 February 2000, ..., 14 February 2018 (Fig. 1). This hindcast set with 220 integrations (20 years \times 11 members) allows us to evaluate the model climatology, and is used to calibrate the operational version.

For the purpose of this study, only the perturbed members (refer data statement to learn about the missing control) of forecasts and their corresponding hindcasts of temperature at 2 m, zonal and meridional components of 100-m wind are retrieved between December 2015 and November 2019 at a temporal resolution of 6 h (i.e., instantaneous values at 0000, 0600, 1200, and 1800 UTC) and a spatial resolution of 0.9° over Europe ($34^\circ\text{--}74^\circ\text{N}$, $13^\circ\text{W--}40^\circ\text{E}$). The forecasts and hindcasts data are retrieved from the ECMWF's Meteorological Archival and Retrieval System (MARS). The 100-m wind speed is computed as the square root of the sum of the squares of the zonal and meridional components. The forecast model underwent several cycles of improvement during the same period, and hence, the dataset consists of forecasts and hindcasts from the versions CY41R1 (1 December 2015–7 March 2016), CY41R2 (8 March 2016–21 November 2016), CY43R1 (22 November 2016–10 July 2017), CY43R3 (11 July 2017–5 June 2018), CY45R1 (6 June 2018–10 June 2019), and CY46R1 (11 June 2019–30 November 2019). Only two of them included important changes: CY41R2 benefited from an increased atmospheric resolution, whereas CY43R1 included an increase in the oceanic resolution and the addition of dynamic sea ice. Nevertheless, the differences in statistics between the different versions of the model are marginal (see appendix A).

b. Reference

Ideally, it is preferred to verify the quality of forecasts against observations. In the absence of a serially complete, spatially coherent observed dataset, it is a common practice to verify the forecasts against reanalysis (Kalnay 2003). ECMWF produces its own fifth-generation high-resolution (1 h; 31 km) reanalysis named ERA5 using 4D-Var data assimilation and the CY41R2 cycle of the Integrated Forecast System (Hersbach et al. 2020). In this work, ERA5 data of temperature at 2 m and zonal and meridional components of wind at 100 m are retrieved from January 1979 to 2020 over Europe ($34^\circ\text{--}74^\circ\text{N}$, $13^\circ\text{W--}40^\circ\text{E}$) at a spatial resolution of 0.25° and a temporal resolution of 6 h (i.e., instantaneous values at 0000, 0600, 1200, and 1800 UTC) from the Copernicus Climate Change Services' Climate Data Store (Raoult et al. 2017). The data are then regridded to 0.9° using bilinear interpolation (Cionni et al. 2018, 14–21) to match the resolution of the forecasts/hindcasts. The choice of the method is motivated by the fact that ECMWF uses bilinear interpolation as the default method for the interpolation of continuous variables of ERA5 data (<https://tinyurl.com/v3d47mw8>). The 100-m wind speed is computed from the wind components as previously described. Despite the biases, ERA5 reanalysis represents the surface wind speed (e.g., Ramon et al. 2019; Jourdier 2020; Brune et al. 2021) and surface temperature (e.g., Simmons et al. 2021) well, with small errors that are acceptable for

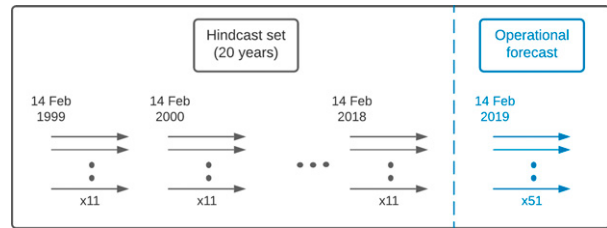


FIG. 1. Illustration of the hindcast set for the operational forecast issued on 14 Feb 2019. The ensemble size is indicated below the arrow.

verification purposes. Consequently, the ERA5 reanalysis data of 2-m temperature and 100-m wind speed act as reference/truth in this study against which the forecasts and hindcasts are verified.

3. Method

The forecasts data under consideration span and represent only 4 years of climatic variability, i.e., from December 2015 to November 2019. Although the forecasts, initialized using operational analysis with ensemble size 5 times as large as the hindcasts, are expected to better represent uncertainty in initial conditions and predictions, conclusions obtained from the verification of forecasts alone may be misleading because the climate during this period of time may have been more (or less) favorable for skillful predictions (Jung et al. 2011). In contrast, hindcasts, spanning 23 years from December 1995 to November 2018, represent climate variability that is 6 times as long as that of the forecasts, and can be used to perform a robust model skill assessment. However, reforecasts are likely less reliable because the ensemble size is smaller by a factor of 5 relative to forecasts and because of the way they are initialized. Hence, both forecasts and reforecasts are assessed in this study so as to understand the skill differences.

In the absence of reliable forecasts, for end-users or for different applications, a common practice is to use observed climatology, a long term average estimated from available historical observed data for the area and time period of the year concerned, as the expected weather. Therefore, it is often encouraged to not just assess the quality of forecasts, but also their relative value with respect to observed climatology. In this work, the observed climatology for each of the evaluated forecasts in any given time of the year is constructed from ERA5 reanalysis by taking the values of each of the past 35 years for the same time period of the year under consideration. To illustrate, for the forecasts issued on 14 February 2019, the observed climatology consists of weekly averaged ERA5 data starting on 14 February 1984, 14 February 1985, ..., 14 February 2018. This also implies that the forecasts issued in 2015 and 2018 have different observed climatology (i.e., rolling climatology) in order to take into account the climatic trend. However, because of the limited availability of ERA5 (i.e., from January 1979 onward at the time of commencement of this study), each of the reforecasts within a given hindcast set has the

same observed climatology as the corresponding forecast (e.g., all the hindcasts demonstrated in Fig. 1 have the same observed climatology as that of the forecast of 14 February 2019). The choice of observed climatology for the reforecasts may have a consequence on the skill of the reforecasts in a sense that the climatology may be favored over reforecasts while computing the skill score in the case of an extreme event because the information about the event is already present in the observed climatology. Nevertheless, the likelihood of witnessing such events is very low. Anyhow, this problem could be averted by using the back-extended ERA5 (i.e., from 1950 onward) and constructing climatology for the reforecasts in the same way as that of the forecasts.

a. Bias adjustment of (re)forecasts

Calibration is a joint property of forecasts and observations. A probabilistic forecast is perfectly calibrated, if, when averaged over several forecasts, the forecast probabilities match the observed frequencies. In addition to the chaotic nature of the atmosphere, the lack of reliable and calibrated forecasts may arise due to either one or a combination of initialization errors, model errors, model parameterizations, truncation errors, and missing physical processes. The lack of forecast calibration could be observed in the form of forecast mean bias, dispersion error, lack of association with reality, lack of reliability, or imperfect representations of trend and variability, to quote a few. If the forecast is well calibrated, the average error in the ensemble mean should be indicative of the ensemble spread, and the variance of the forecast model climatology should be equivalent to that of the climatological truth (Wilks 2019). There exist several methods with varying levels of sophistication to calibrate forecasts (Manzanas et al. 2019). Manzanas et al. (2019) have shown that simple bias adjustment methods such as mean and variance adjustment (MVA) can perform as well as the sophisticated calibration techniques such as nonhomogeneous Gaussian regression in correcting model biases. Their study also highlighted that the additional value gained by using sophisticated calibration techniques over simple bias adjustment methods are only marginal and are limited to certain geographical regions (e.g., tropics) and/or seasons. In this study, the bias adjustment of the (re)forecasts is carried out using the MVA method as described in Leung et al. (1999), Torralba et al. (2017), and Manzanas et al. (2019). The bias adjusted ensemble member j of any forecast at any given lead time is given by

$$x_j^* = (x_j - \bar{x}_e) \frac{\sigma_{\text{ref}}}{\sigma_e} + \bar{o}_{\text{ref}}, \quad (1)$$

where x_j is the member whose bias needs to be adjusted; \bar{x}_e and σ_e are the mean and the standard deviation, respectively, of all the members of all the hindcasts corresponding to the forecast; and \bar{o}_{ref} and σ_{ref} are the mean and the standard deviation, respectively, of the truth (or observations) corresponding to the hindcasts. For the bias adjustment of any given reforecast within a hindcast set, the remaining 19 years of hindcasts are used to adjust the mean and the

spread through a leave-one-out approach to prevent overfitting.

b. Measures of predictive skill

The predictive skill of a point forecast could be evaluated by measuring the correspondence between the forecast and the observation through simple scores such as the mean absolute error, the mean squared error, or the root mean squared error (Jolliffe and Stephenson 2003). To assess the skill of the probabilistic forecasts, several scores have been proposed in the literature, each one assessing a specific attribute of forecast quality (Wilks 2019; Coelho et al. 2019). One of the important and most widely used scores to evaluate the skill of the full predictive distribution of the probabilistic forecasts of continuous predictands is the continuous ranked probability score (CRPS) (Matheson and Winkler 1976; Unger 1985; Hersbach 2000). The CRPS is the area under the curve that is formed by computing the squared difference between the cumulative distribution functions (CDFs) of the forecast and the observation. When the observation is a single number, as is often the case, its CDF is a Heaviside step function centered on that value. This implies that if the forecast is deterministic, CRPS simplifies to the absolute error between the forecast and the observation:

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_e(y)]^2 dy, \quad (2)$$

where $F(y)$ is the empirical CDF of forecasts/(re)forecasts computed by taking weekly mean of each of the ensemble members and

$$F_e(y) = \begin{cases} 0, & \text{if } y < e \\ 1, & \text{if } y \geq e \end{cases} \quad (3)$$

is the CDF of observation for the observed weekly mean e , denoted as a step function that jumps from 0 to 1 at the point where the forecast is equal to the observation.

The CRPS is negatively oriented (i.e., smallest values indicate more accurate forecasts), and it rewards those forecasts whose probabilities are concentrated around the observation. As the lead time increases, the ability to predict finer-scale features in time and space quickly diminishes. Consequently, in this work, the CRPS is computed for weekly averaged quantities by considering (re)forecasts and observations depending on the lead time and start date [Eq. (2)]. Spatial averaging, wherever applicable, is performed by taking the mean of cosine-latitude weighted gridpoint values in order to obtain a single scalar time series over the domain, for which all the metrics are computed (Dorrington et al. 2020). The CRPS has the same units as the physical quantity being assessed. The CRPS can also be calculated for the observed climatology: the CDF is then obtained from the weekly means of the targeted time period of the year in each of the years covered.

The relative value of the forecasts/(re)forecasts with respect to climatology is measured using the continuous ranked probability skill score (CRPSS) as described in Eq. (4). It can be

observed from Eq. (4) that skillful (re)forecasts should have a CRPSS > 0. The CRPSS is bounded above by 1, but negative values are unbounded;

$$CRPSS = 1 - \frac{CRPS_{(re)forecasts}}{CRPS_{climatology}}. \quad (4)$$

The standard practice in forecast verification is to compute the average CRPS of several reforecasts at each of the lead times considered and compare it with the average CRPS of climatology to obtain an average CRPSS. The averaging is done in order to assess the reliability component of forecasts (Hersbach 2000), which can only be assessed with multiple forecast instances. However, as we show in section 4a, averaging ratios (i.e., CRPSS) overemphasizes negative instances, and can therefore be misleading in evaluating forecast skill. Besides, within the framework of the S2S forecasts, Coelho et al. (2019) recommended to use novel verification metrics that are meaningful to the end-users. Keeping in mind the S2S end-users in the energy sector, we propose “proportion of skillful (re)forecasts” as a novel way of synthesizing the CRPSS to measure skillfulness of the (re)forecasts. As the name suggests, the proportion of skillful (re)forecasts measures the proportion of (re)forecasts that have CRPS lower than that of climatology. To compute the proportion of skillful (re)forecasts [Eq. (5)], we first compute CRPSS of each of the (re)forecasts separately at each of the specified lead times, and then compute at each of the specified lead times the ratio of the number of (re)forecasts with CRPSS greater than zero to the total number of (re)forecasts. Accordingly, skillful (re)forecasts should have values > 50%. In general, the proportion of skillful (re)forecasts is consistent with the median of the CRPSS of (re)forecasts. The proportion of skillful (re)forecasts is flexible in the sense that the threshold of “CRPSS greater than zero” can be adjusted (i.e., increased), in particular in situations where it would be useful to have CRPSS above a given threshold, not just above zero. The 95% confidence intervals for the proportion of skillful (re)forecasts in this study are computed using the standard parametric approach by assuming a normal distribution for the underlying data (Machin et al. 2013):

$$\begin{aligned} &\text{proportion of skillful (re)forecasts} \\ &= \frac{\text{no. of (re)forecasts with CRPSS} > 0}{\text{total no. of (re)forecasts}} \times 100. \end{aligned} \quad (5)$$

Assessing the skill of probabilistic forecasts generally involves assessing several forecast attributes (Coelho et al. 2019). Another commonly used score is the anomaly correlation coefficient (ACC), which measures the linear association between the ensemble mean and the observations. ACC is a deterministic score that is computed as the usual Pearson’s correlation between forecast ensemble mean and observed anomaly pairs of several independent forecasts (Namias 1952). The ACC for weekly averaged quantities of n independent (re)forecasts at any given place or across any given domain is given by

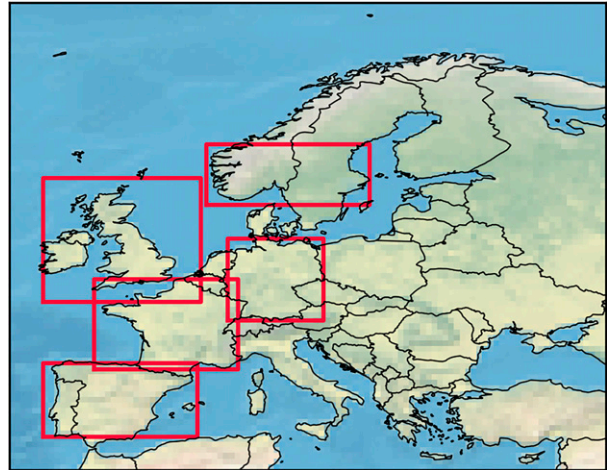


FIG. 2. Illustration of the five countrywide domains and the European domain considered in this study.

$$ACC = \frac{\text{covariance}(y', o')}{\sigma_y \sigma_o}, \quad (6)$$

where y' is the (re)forecast anomaly computed by removing the weekly mean climatology from the bias adjusted (re)forecast weekly ensemble mean, o' is the observed anomaly computed by removing the weekly mean climatology from the observed weekly mean, and σ is the standard deviation. The 95% confidence intervals for the ACC, wherever applicable, are computed through a nonparametric bootstrap approach carried out 1000 times. In general, for positioning of the synoptic scale features, the skillful forecasts should have ACC > 60%, below which the value in the forecasts becomes only marginally useful (Robertson and Vitart 2018). Other important attributes of forecast quality such as reliability (i.e., a measure of calibration of the issued forecast probabilities), resolution (i.e., a measure of how the frequency of occurrence of the event varies as the issued forecast probability changes), and sharpness (i.e., a measure of the ability of the forecasts to produce concentrated predictive distributions that are different from the climatological probabilities) are also assessed for the weekly mean tercile, quartile, and decile forecasts through the use of reliability diagrams (Sanders 1963; Jolliffe and Stephenson 2003; Wilks 2019), which are discussed alongside the results in section 4b.

4. Results

The first part of this section presents the general skill assessment of 2-m temperature and 100-m wind speed reforecasts averaged across the European domain (Fig. 2). Subsequently, the countrywide assessments of reforecast skill are also carried out. Moreover, the reforecast skill assessment at the gridpoint scale is investigated to explore geographical variations of skill. In the second part of this section, we compare the skill of reforecasts with the skill of forecasts to understand the skill differences considering different ensemble sizes.

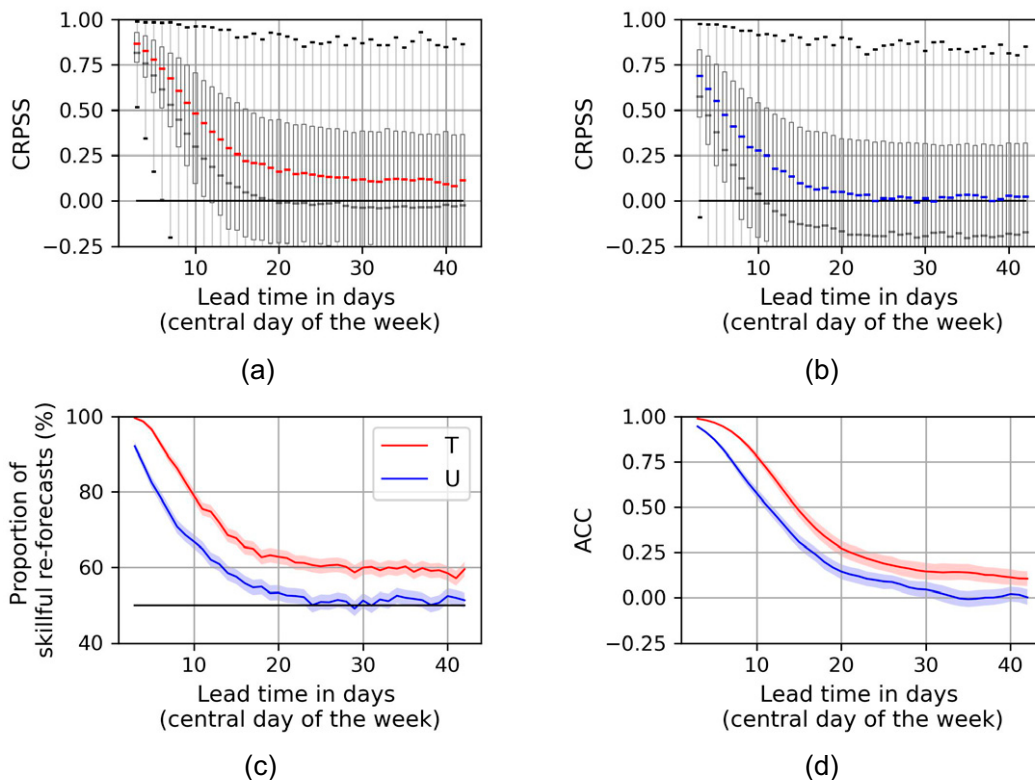


FIG. 3. Reforecast quality assessment averaged across the European domain (34° – 74° N, 13° W– 40° E), showing the temporal evolution of CRPSS for (a) 2-m temperature T and (b) 100-m wind speed U), demonstrated as standard boxplots with colored bars indicating the median, gray bars indicating the mean, the gray box indicating the first and the third quartiles, whiskers indicating the end points, and outliers hidden. Values above 0 indicate skillfulness. Lead time is indicated as central day of the week (as an illustration, day 10 corresponds to the week between days 7 and 13). (c) The temporal evolution of the proportion of skillful reforecasts for the same variables. Values above 50% (black horizontal line) indicate skillfulness. (d) The temporal evolution of ACC for the same variables. Shaded regions in (c) and (d) correspond to the 95% confidence intervals.

Last, we assess reliability, resolution, and sharpness of the forecasts through the aid of reliability diagrams.

a. Reforecast skill assessments

1) GENERAL ASSESSMENT OVER EUROPE

Figure 3 compares the temporal evolution of 2-m temperature and 100-m wind speed reforecast skill for weekly mean values averaged across the European domain (Fig. 2). Figures 3a and 3b show that the mean values (i.e., gray bars) of CRPSS drop below zero after 18 and 10 days, respectively, for temperature and wind speed. On the other hand, the median (i.e., colored bars) largely stays above zero throughout all the leads, although more significantly for temperature. By definition, the skill score is bounded by 1 above, but its negative values are unbounded such that the average can be sensitive to rare, strong negative values. Hence, instead of calculating the average of the CRPSS, we compute the proportion of skillful (re)forecasts as a measure of (re)forecast skill. Figure 3c shows the temporal evolution of the proportion of skillful reforecasts with increasing lead time.

It is conspicuous that the model performs better in predicting 2-m temperature than 100-m wind speed at all lead times. While temperature reforecasts are skillful at all lead times, wind speed reforecasts are skillful until approximately day 24. Figure 3d displaying the time evolution of ACC for temperature and wind speed confirms that temperature reforecasts are more skillful than wind speed reforecasts. The ACC for temperature falls below 0.6 around day 13, and below 0.25 around day 22, whereas the ACC for wind speed falls below 0.6 around day 8, and below 0.25 around day 17. Seasonal variations of skill are discussed in the following section.

2) SEASONAL VARIATIONS OF SKILL

Seasonal variations of temperature and wind speed reforecast skill averaged across the European domain are shown in Fig. 4. It can be noticed that the proportion of skillful reforecasts is larger in the Northern Hemisphere winter [December–February (DJF)] and summer [June–August (JJA)] than in the transition seasons for both temperature (Fig. 4a) and wind speed (Fig. 4b). The ACC for temperature (Fig. 4c) is larger in

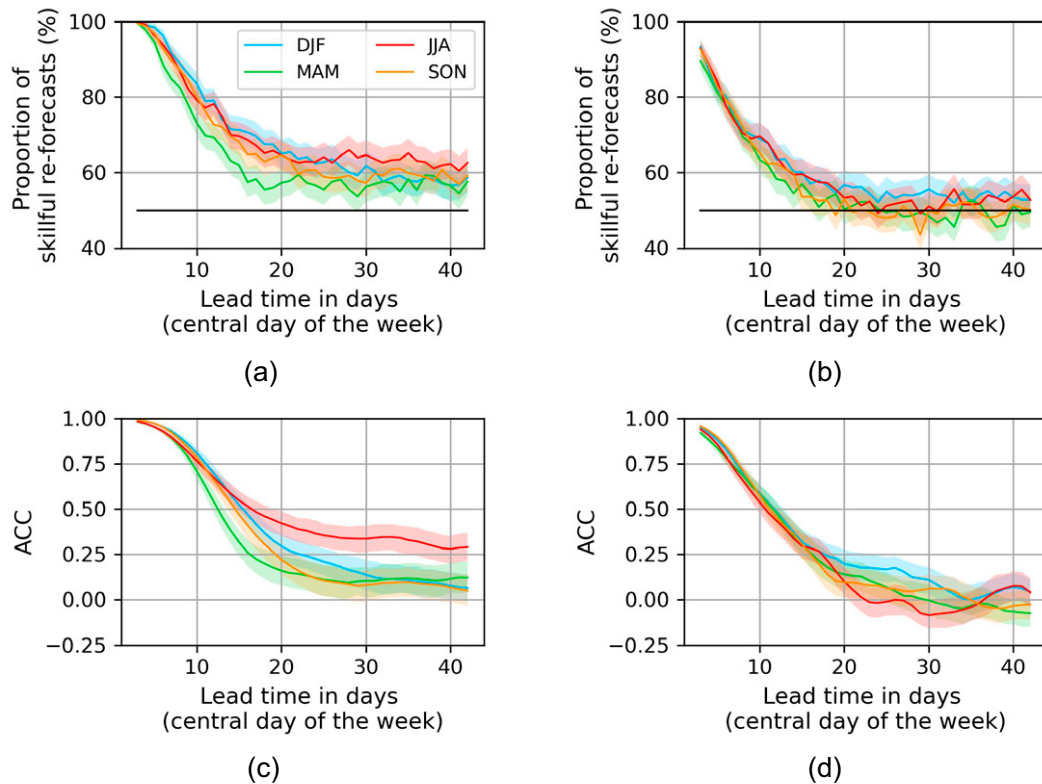


FIG. 4. Seasonal variations of reforecast quality assessment averaged across the European domain (34°–74°N, 13°W–40°E), showing the temporal evolution of the proportion of skillful reforecasts for (a) 2-m temperature and (b) 100-m wind speed. Values above 50% (black horizontal line) indicate skillfulness. Also shown are the temporal evolution of ACC for (c) 2-m temperature and (d) 100-m wind speed. Shaded regions correspond to the 95% confidence intervals.

the summer relative to other seasons after a lead time of around 17 days. In contrast, the ACC for wind speed is larger in the winter relative to other seasons between 18 and 31 days (Fig. 4d). The improved skill in winter and/or summer may arise from stronger boundary conditions (e.g., sea surface temperature gradients), reinforced coupling (e.g., troposphere–stratosphere), enhanced memory of the initial conditions (e.g., soil moisture) among others (Robertson and Vitart 2018). However, addressing the reasons for enhanced skill in certain seasons is beyond the scope of this study. The seasons of winter and summer, in addition to demonstrating enhanced skill, also have a significant impact on the energy sector in the form of increased demand driven by heating and cooling, respectively. Therefore, only the results corresponding to winter and summer will be shown in the following sections.

3) COUNTRYWISE SKILL ASSESSMENT

The assessment of reforecast skill averaged across the European domain may have limited application. In contrast, countrywise average skill of wind speed and temperature reforecasts may be closer to the scale on which end-users may need forecasts for decision-making (e.g., transmission system operators). This section investigates variations in skill over

domains typically a thousand kilometers across, e.g., over a country like the United Kingdom. Table 1 and Fig. 2 present the domains considered in this study. The choice of the domains is not just motivated by their geography (inland versus coastal, location relative to the climatological storm tracks) offering different sampling conditions, but also by their considerable share of wind power in the energy mix (International Energy Agency 2020). Figures 5 and 6 illustrate the differences in the temporal evolution of the proportion of skillful reforecasts for a selection of domains. For temperature (Fig. 5), in both DJF and JJA, predictions over Germany (e.g., proportion > 60% up to about 27 days in DJF and about 15 days in JJA) are more skillful than predictions over France (proportion > 60% up to about 20 days in DJF and about

TABLE 1. Description of the domains.

Domain region	Lower-left bound	Upper-right bound
France	43.0°N, 5.5°W	51.0°N, 7.3°E
Germany	47.3°N, 6.4°E	54.6°N, 14.9°E
Southern Scandinavia	57.6°N, 4.5°E	63.0°N, 19.0°E
Spain and Portugal	37.0°N, 10.0°W	43.5°N, 3.7°E
United Kingdom	49.0°N, 10.0°W	60.0°N, 4.0°E

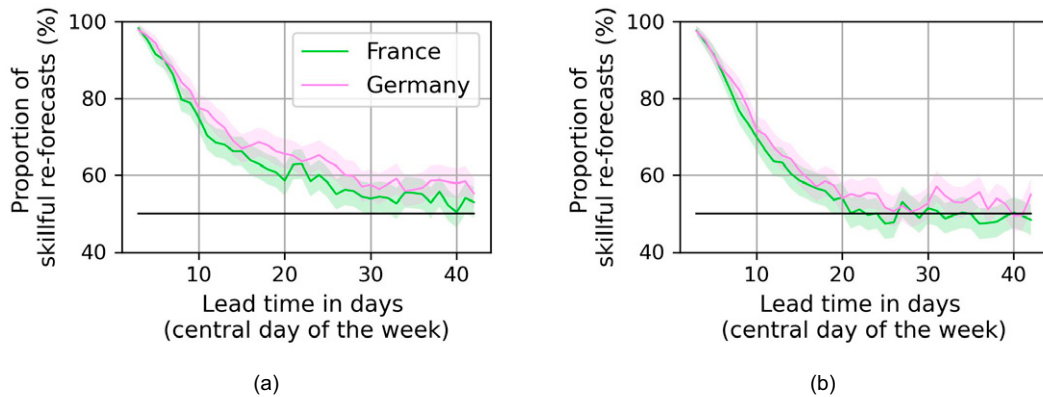


FIG. 5. Comparison of the temporal evolution of proportion of skillful 2-m temperature reforecasts between France and Germany for (a) DJF and (b) JJA. Shaded region correspond to the 95% confidence intervals. Values above 50% (black horizontal line) indicate skillfulness.

14 days in JJA). The reforecasts are more skillful in the winter as compared with the summer across both domains. The ACC for temperature also shows similar behavior to that of the proportion of skillful reforecasts (not shown). The skill of winter temperature reforecasts for Spain and Portugal and the United Kingdom (not shown) domains are similar to that for France, and the skill across southern Scandinavia (not shown) is at least as good as, if not slightly better than Germany. Overall, in DJF, the skill across the most skillful country-sized domains (e.g., Germany) is almost as good as the skill across the European domain (Fig. 4a). Whereas in JJA, the skill across the European domain is better than the skill of the most skillful country-sized domains.

The wind speed reforecasts (Fig. 6) across the United Kingdom (e.g., proportion $> 60\%$ up to about 17 days) are more skillful than that of France (proportion $> 60\%$ up to about 11 days) in DJF. However, in JJA, France demonstrates marginally larger skill than the United Kingdom after about 10 days. The ACC again displays a similar pattern to that of the proportion of skillful reforecasts (not shown). In winter,

the skill across Germany and Spain and Portugal (not shown) are comparable to that of France, and the skill across southern Scandinavia (not shown) is at least as good as, if not marginally better than the United Kingdom. In DJF the skill across the United Kingdom is better than the skill across the European domain (Fig. 4b), whereas in JJA the opposite is true.

4) GRIDPOINT SKILL ASSESSMENT

Although countrywide domains are useful in predicting national averages of the variables, gridpoint assessment of skill are more appropriate to explore the geographical variations of skill. The spatial resolution of the data used in this study is about 90 km (i.e., 0.9°). This resolution is coarse enough for the S2S prediction models to still hold prediction skill on the subseasonal time scales (Buizza and Leutbecher 2015), and fine enough to be useful for a range of applications.

Figure 7 illustrates the maps of temporal evolution of the proportion of skillful reforecasts and ACC for temperature across the European domain. The maps for winter (top row of Fig. 7a) show the presence of a zonal (i.e., east–west) pattern

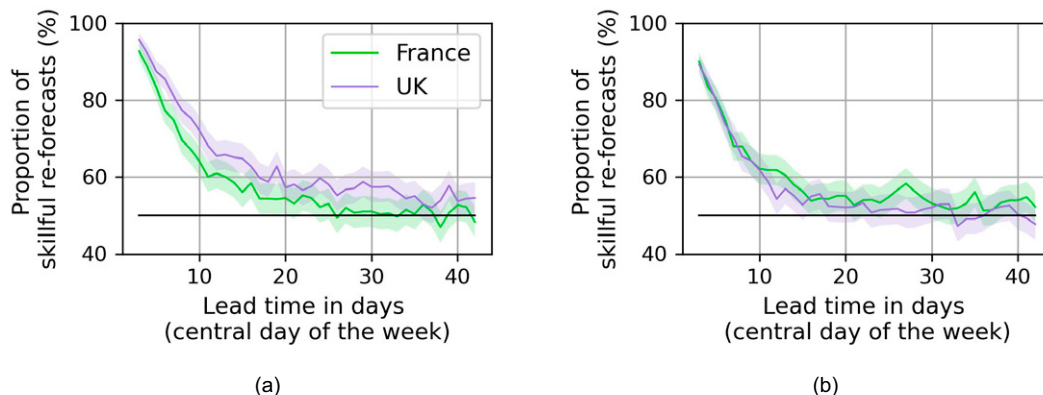


FIG. 6. Comparison of the temporal evolution of proportion of skillful 100-m wind speed reforecasts between France and the United Kingdom for (a) DJF and (b) JJA. Shaded regions correspond to the 95% confidence intervals. Values above 50% (black horizontal line) indicate skillfulness.

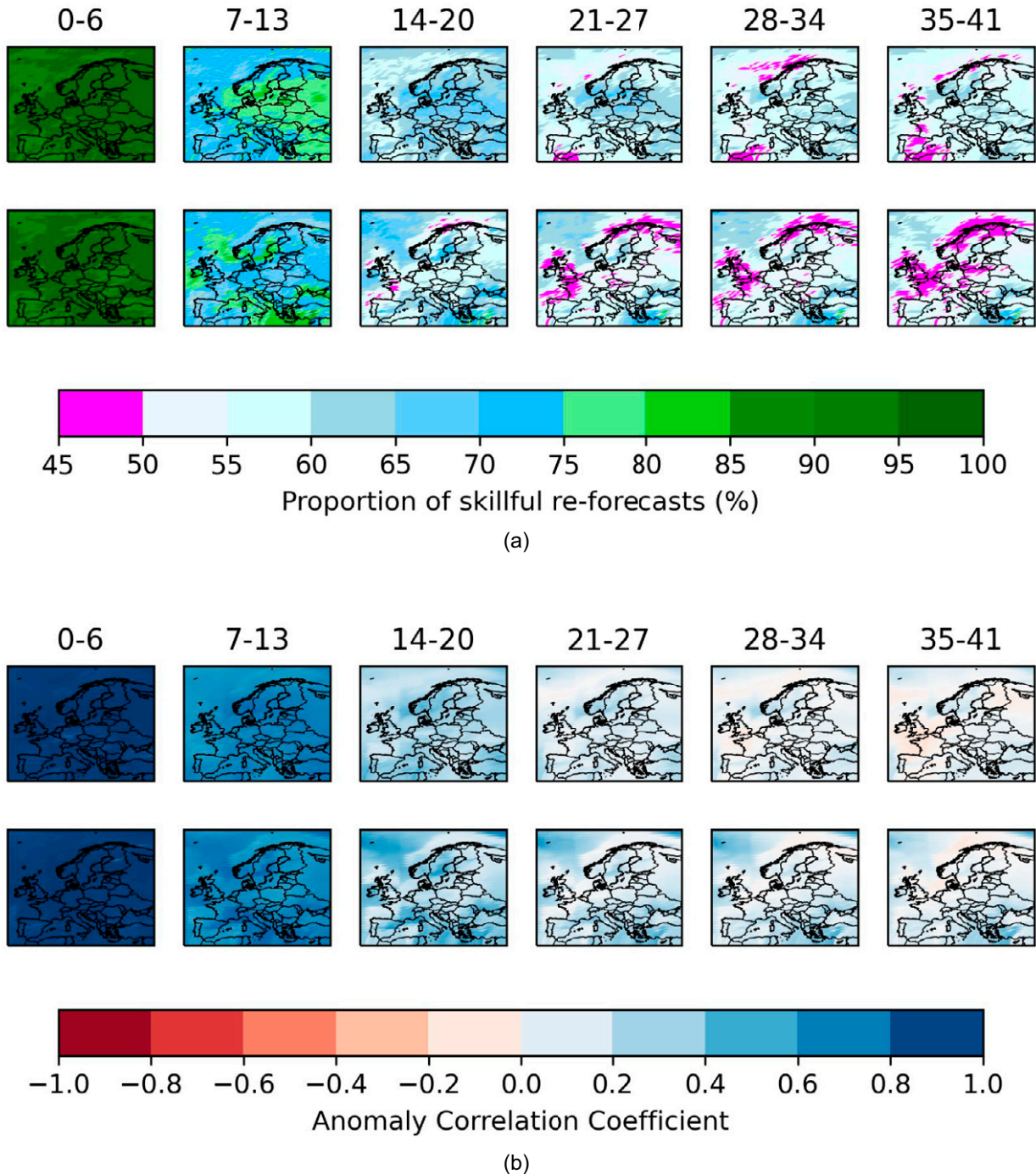


FIG. 7. Maps of (a) proportion of skillful reforecasts and (b) ACC for 2-m temperature over Europe. In (a) and (b), the top row is for DJF and the bottom row is for JJA. Columns from left to right show lead times centered on days 3, 10, 17, 24, 31, and 38. Values above 50% in (a) indicate skillfulness.

between second and fourth weeks indicating that temperature predictions are generally more skillful over central/eastern Europe than western Europe. In central and eastern Europe, the reforecasts are skillful even at a lead time of 6 weeks in

winter, encouraging their use in the decision-making value chain across sectors. The reforecasts are less skillful in summer in general with proportion of skillful reforecasts converging toward climatology beyond 3 weeks. The ACC (Fig. 7b)

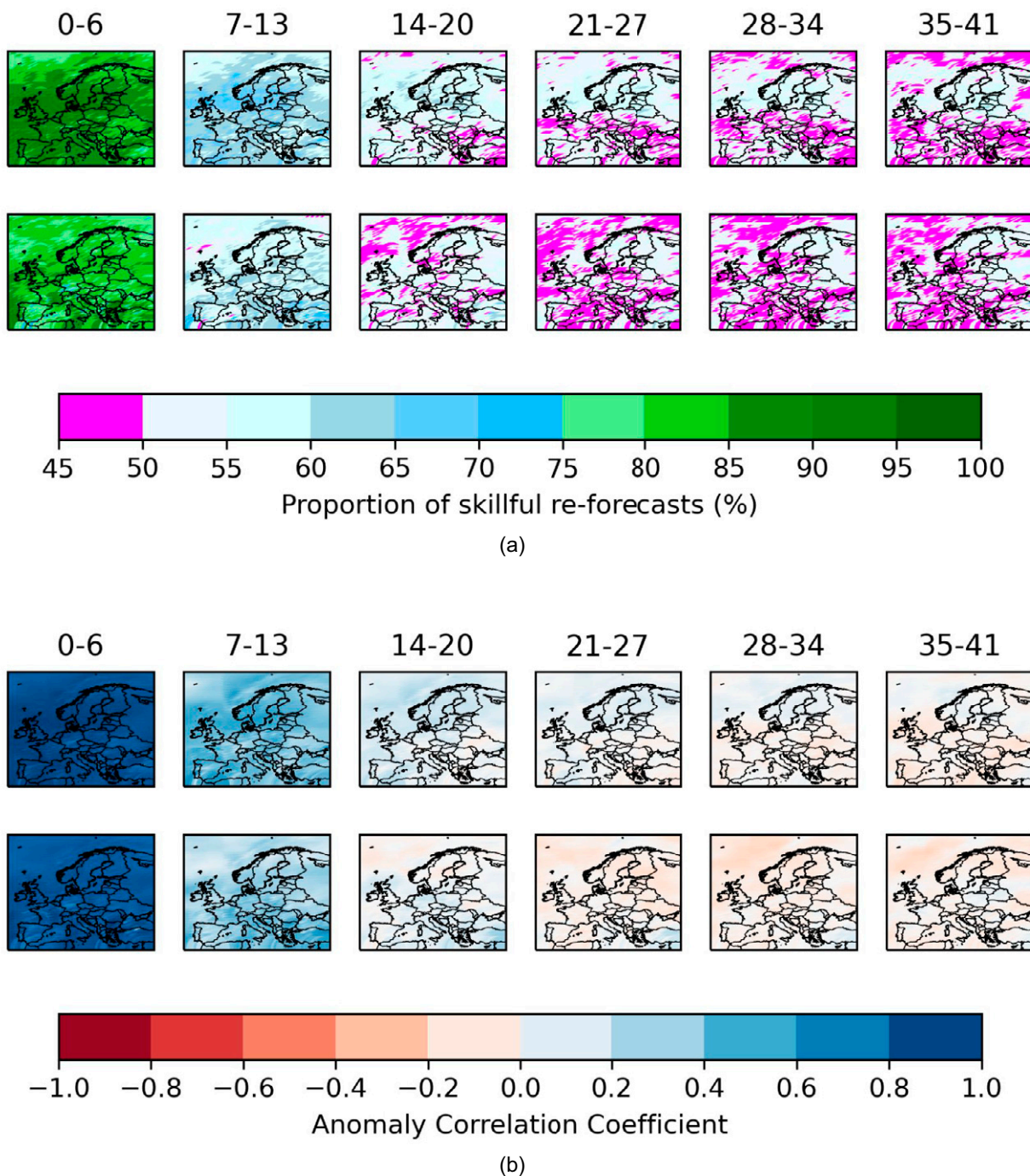


FIG. 8. As in Fig. 7, but for 100-m wind speed.

drops below 60% beyond 2 weeks showing no noticeable differences between seasons.

The maps of the proportion of skillful reforecasts and ACC for wind speed are shown in Fig. 8. Overall, the reforecasts are more skillful in winter. In addition, there exists a meridional (i.e., north-south) pattern of skill in winter (top rows in Figs. 8a,b) between second and fifth weeks indicating that

wind speed predictions are generally more skillful over northern than southern Europe. Across Scandinavia, the proportion of skillful reforecasts still exceeds 50% after 5 weeks. However, in summer, the proportion of skillful reforecasts for wind speed drops below 50% over a large part of the domain beyond 3 weeks. The ACC drops to significantly lower levels than that of temperature beyond a week in both winter and

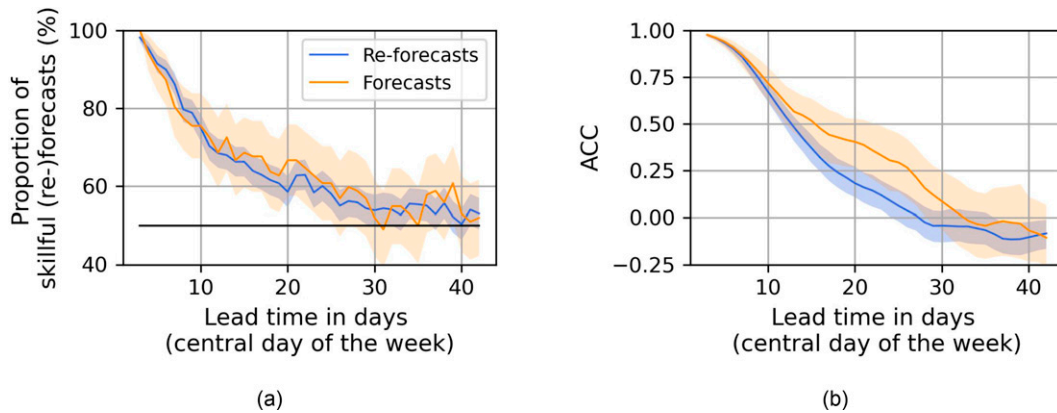


FIG. 9. Comparison of the temporal evolution of 2-m temperature forecast and reforecast skills averaged across France for DJF, showing (a) proportion of skillful reforecasts (blue) and forecasts (orange) and (b) ACC. Values above 50% (black horizontal line) indicate skillfulness. Shaded regions correspond to the 95% confidence intervals.

summer. The reasons for enhanced skill witnessed across certain regions in Figs. 7 and 8 may be related to the differences in the regional climate (i.e., maritime vs continental), low-frequency oscillations (Ardilouze et al. 2021), and to the role of circulation features, in particular the storm tracks. However, investigating the reasons for the existence of spatial pattern of skill is beyond the scope of this work.

b. Forecast skill assessments

The operational forecasts with 50 ensemble members each are expected to better represent uncertainty in the initial conditions and model parameterizations as compared with the reforecasts with only 10 members (Robertson and Vitart 2018). Figure 9 compares the skill between the winter forecasts and reforecasts of temperature for weekly means averaged across France (Table 1 and Fig. 2). Overall, the forecasts are more skillful than the reforecasts. The proportion of skillful forecasts (e.g., values > 60% up to about 25 days) is essentially greater than the proportion of skillful reforecasts (values > 60% up to about 19 days). However, the confidence intervals for the forecasts are 2 times as wide as those of the reforecasts due to a smaller sample size. The ACC of

forecasts (values > 0.5 up to about 16 days, and > 0.25 up to about 26 days) has a longer skill horizon when compared with that of the reforecasts (values > 0.5 up to about 13 days, and > 0.25 up to about 17 days). A similar pattern can also be observed with respect to other seasons and domains (not shown). The differences between the skill of the forecasts and the reforecasts of wind speed in winter are shown in Fig. 10 for the same domain. The behavior is overall comparable to that of the temperature. The significant differences between the skill of the forecasts and the reforecasts are mainly due to the differing ensemble sizes between the two (see appendix B). Even though the reforecasts (23 yr) represent a longer climatic variability than the forecasts (4 yr) and hence a better estimation of the overall skill of the model, given that they have an ensemble size that is smaller by a factor of 5 relative to the forecasts, the skill of the reforecasts should only be considered as a lower bound for the skill of the operational forecasts.

The CRPS and its respective skill score give one measure of the agreement between the forecasts and the observations. However, a thorough appreciation of the quality of forecasts requires the use of the full joint distribution of forecasts and

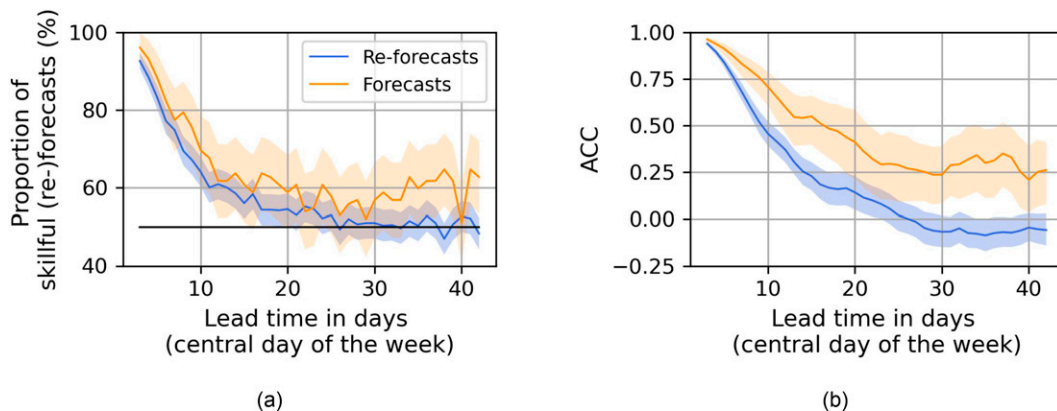


FIG. 10. As in Fig. 9, but for 100-m wind speed.

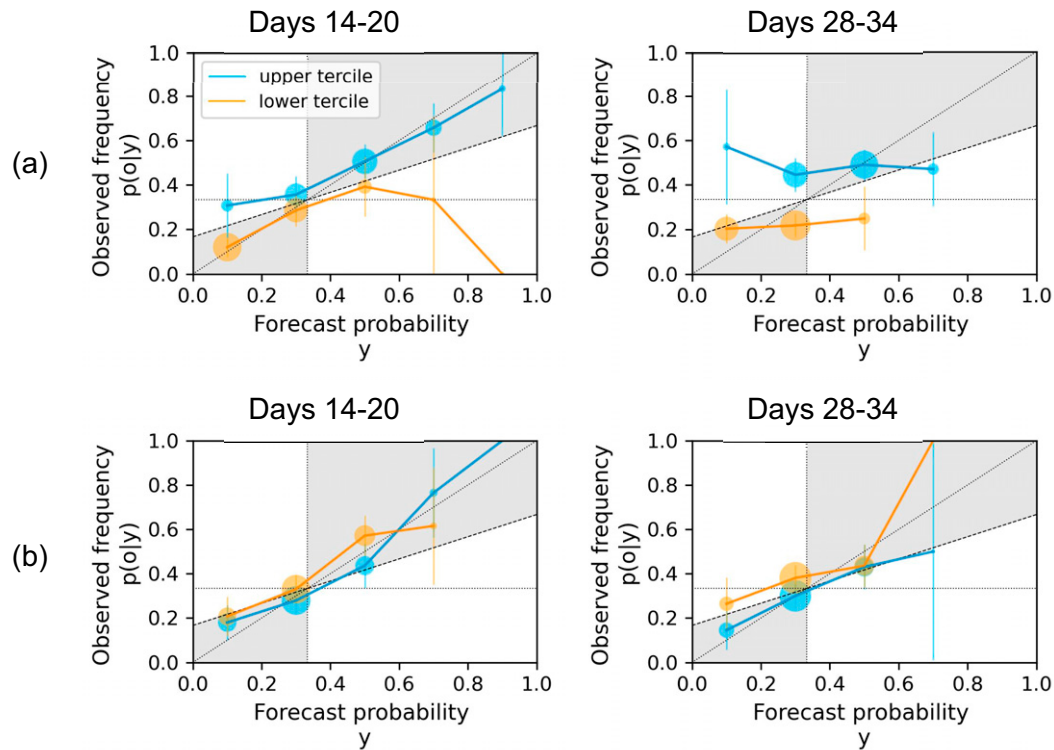


FIG. 11. Reliability diagrams for upper and lower terciles of weekly mean forecasts for the weeks centered on (left) day 17 (i.e., days 14–20) and (right) day 31 (i.e., days 28–34) averaged across France for (a) 2-m temperature and (b) 100-m wind speed. The forecasts are stratified into five bins of equal width. The size of the points is proportional to the number of forecasts in the respective bins. The vertical bars refer to the 95% confidence intervals computed through the standard parametric approach. The vertical and horizontal dotted lines indicate the climatological tercile probabilities (theoretically, the value is $1/3$) in the forecasts and observations, respectively. Perfectly reliable forecasts fall on the dotted diagonal line connecting the points (0, 0) and (1, 1). The points located within the gray area contribute positively to the skill.

observations. The reliability diagram (Sanders 1963; Jolliffe and Stephenson 2003; Wilks 2019) is a graphical tool to comprehend the full joint distribution of forecasts and observations for probabilistic forecasts of a dichotomous predictand (i.e., predictand with a binary outcome). Perfectly reliable (i.e., calibrated) forecasts have observed frequencies essentially equal to forecast probabilities. Since the forecasts considered in this study have a larger ensemble size and hence a better representation of uncertainty relative to the reforecasts, the reliability diagrams are produced only for the forecasts. Figure 11 demonstrates the reliability diagrams for upper and lower terciles of weekly mean temperature and wind speed forecasts for the week centered on day 17 (i.e., days 14–20) and day 31 (i.e., days 28–34) for the France domain described in Table 1. In the figure, the lines connecting the points show no persistent offset from the 1:1 diagonal line (45°) illustrating the absence of unconditional biases. In a reliability diagram, the smaller the vertical distance between the points and the diagonal line, and the larger the vertical distance between the points and the climatological line (dotted horizontal line in the figure), the higher are the forecast reliability and resolution, respectively. Conversely, the larger

the vertical distance between the points and the diagonal line, and the smaller the distance between the points and the horizontal climatological line, the lower are the reliability and resolution, respectively. The dashed line located midway between the perfect reliability line and the horizontal climatological line represents the *no skill line*. Accordingly, the points located in the gray area contribute positively to the skill of the forecasts. For the third week of temperature forecasts (Fig. 11a), the upper-tercile forecasts are more reliable than the lower counterparts. In contrast, the reliability of the upper- and the lower-tercile wind speed forecasts are virtually comparable for both weeks (Fig. 11b). The upper tercile of temperature forecasts for the third week exhibit underforecasting biases associated with low probabilities, and marginal overforecasting biases associated with high probabilities. Furthermore, the lower tercile of temperature forecasts generally show significant overforecasting associated with high probabilities, indicating poor resolution and overconfidence. For temperature in the third week, the maximum number of forecasts is located in the bins beside the climatological probability (dotted vertical line) bin (i.e., 0.2–0.4), indicating reasonable sharpness of these forecasts. In contrast, for

wind speed for both weeks, the maximum number of forecasts are concentrated in the climatological bin suggesting that the forecasts have low sharpness. For the fifth week of temperature forecasts, both the upper- and the lower-tercile forecasts are less reliable, and have a poorer resolution and sharpness as compared with the third week. While for wind speed forecasts in the fifth week, the reliability, resolution, and sharpness are comparable to that of the third week. The number of events falling within the bins are typically concentrated near low forecast probabilities. Therefore, the confidence intervals are generally narrower for lower forecast probabilities as compared with higher forecast probabilities for both upper and lower terciles. The fact that the upper-tercile temperature forecasts for the third week being more reliable and having a higher resolution than the lower-tercile forecasts, and the comparable reliability and resolution of the upper- and lower-tercile wind speed forecasts holds true for the other domains (not shown). The reliability diagrams for the upper and lower quartiles and deciles of temperature (see [appendix C](#)) and wind speed (not shown) forecasts are less reliable, and have significantly lower resolution, especially for larger forecast probabilities, relative to that of the terciles. Overall, the forecasts of temperature and wind speed carry valuable information in predicting terciles even beyond 2 weeks, encouraging their implementation in operational decision-making on this time horizon.

5. Conclusions

In this study, the skill of the subseasonal forecasts and reforecasts of 2-m temperature and 100-m wind speed was evaluated against ERA5 reanalysis across the European domain. The bias adjustment of the (re)forecasts was carried out using mean and variance adjustment method. To account for the different aspects of (re)forecast quality (i.e., accuracy, association, reliability, resolution, and sharpness), several metrics were applied, providing evidence that

- 1) the model generally performs better in predicting 2-m temperature than 100-m wind speed,
- 2) the skill over Europe displays a seasonal pattern with winter showing more skillful forecasts, which is followed by summer for temperature and summer/fall for wind speed,
- 3) the skill also displays a spatial pattern for temperature having more skill for eastern than for western Europe and for wind speed having more skill in northern than southern Europe,
- 4) the skill of the reforecasts should only be considered as a lower bound, and the forecasts due to their larger ensemble size represent uncertainty better and hence perform better, and
- 5) depending on the geographical domain, climate variable, and forecast attribute of choice, the weekly mean forecasts can be skillful even up to 6 weeks, encouraging their implementation in the decision-making value chain.

This study evaluated the skill of the (re)forecasts of a model originating from a single weather forecasting center (i.e.,

ECMWF). This choice was motivated by the fact that the skill of the forecasts of temperature (at 2 m and at 850 hPa) of the ECMWF model compares to or even outperforms the skill of a multimodel combination ([Hagedorn et al. 2012](#)). Nevertheless, investigation of the skill of a multimodel ensemble is an important next step. The reference data (i.e., ERA5 reanalysis) used in this study also originates from the ECMWF, produced using one of the same models (CY41R2) that is used to produce the (re)forecasts. However, verifying (re)forecasts against reanalysis produced from the same model may contribute to enhancing the skill of (re)forecasts. Hence, it is important to assess the skill of the (re)forecasts against observations or other global/regional reanalysis datasets produced using a different model. To aid in the further development of the prediction model, it is essential to understand the potential sources of predictability and the origin of model biases. The authors did not assess the skill of the (re)forecasts of other variables that are critical for the renewable energy sector such as the solar radiation and the precipitation. The authors propose to undertake these explorations in a future study.

Acknowledgments. Naveen Goutham thanks Association Nationale de la Recherche et de la Technologie (ANRT) for the Convention industrielle de formation par la recherche (CIFRE) fellowship of his Ph.D. This work contributes to the Energy4Climate Interdisciplinary Centre (E4C) of Institut Polytechnique de Paris and Ecole des Ponts ParisTech, supported by third Programme d'Investissements d'Avenir (ANR-18-EUR-0006-02). The authors acknowledge the data center Ensemble de services pour la recherche à l'Institut Pierre-Simon Laplace (ESPRI) for their help in storing and accessing the data. The authors thank the two anonymous reviewers who contributed significantly in improving the quality of the paper through their critical assessment.

Data availability statement. The archived ECMWF extended-range forecasts and hindcasts are published under a Creative Commons Attribution 4.0 International (CC BY 4.0). However, a data access fee may be applicable. Further information for the reader is available online (<https://www.ecmwf.int/>). The ECMWF ERA5 reanalysis data are publicly available and can be accessed through the Copernicus Climate Change Services' Climate Data Store upon registration. The control of the ensemble should be treated as another indistinguishable ensemble member. However, because of the unavailability of the control member in the internal database of ESPRI as a result of an unintentional human error, we had to use only the perturbed members.

APPENDIX A

Comparison of Skill between Different Cycles of the ECMWF Model

Figure A1 shows the temporal evolution of CRPSS of reforecasts of 2-m temperature averaged across Germany ([Table 1](#)) between different cycles of the ECMWF model for four seasons

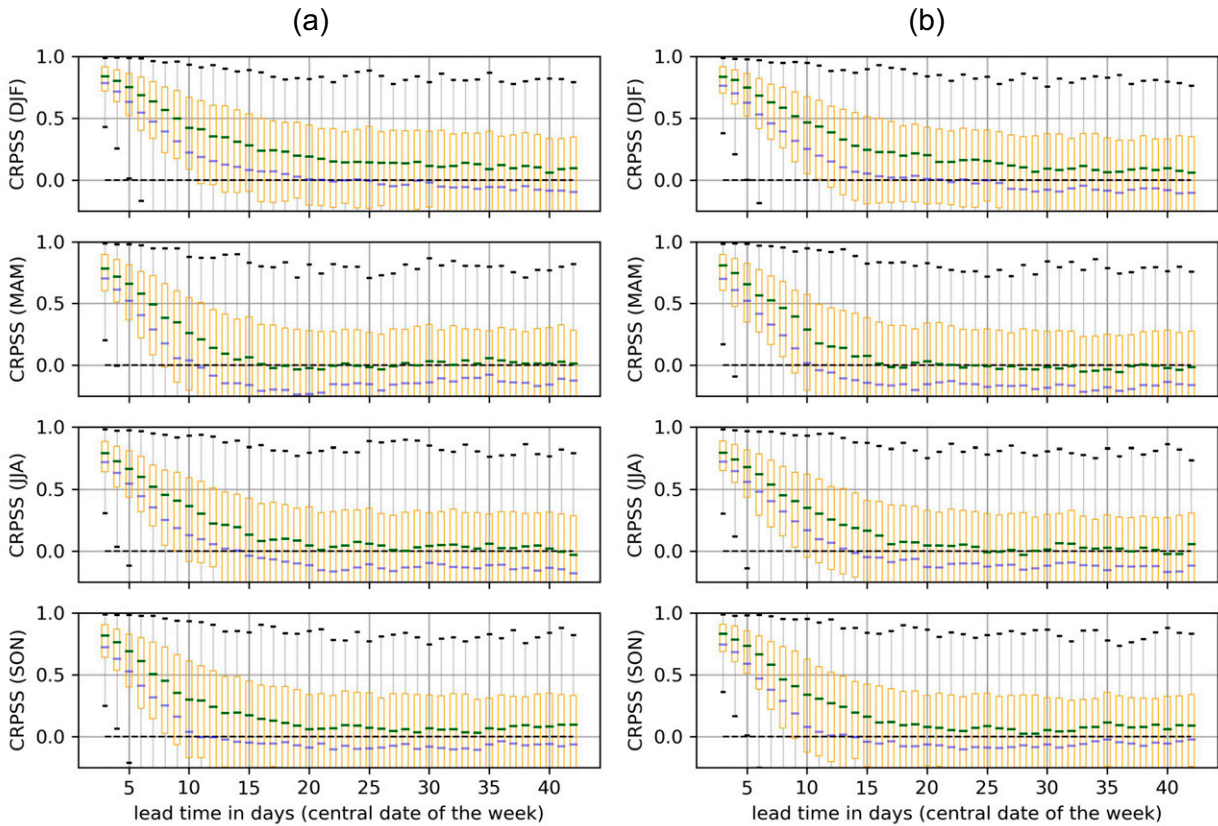


FIG. A1. Comparison of CRPSS of reforecasts of 2-m temperature averaged across Germany between different cycles of the ECMWF model for four seasons (DJF, MAM, JJA, and SON), demonstrated as standard boxplots with green bars indicating the median, blue bars indicating the mean, the orange box indicating the first and the third quartiles, whiskers indicating the end points, and outliers hidden. Values above 0 indicate skillfulness. Lead time is indicated as central day of the week (as an illustration, day 10 corresponds to the week between days 7 and 13). The reforecasts correspond to the forecasts (a) between 1 Dec 2015 and 30 Nov 2016 and (b) between 1 Dec 2018 and 30 Nov 2019.

[DJF, March–May (MAM), JJA, and September–November (SON)]. The data used in Fig. A1a consist of reforecasts mostly from the cycles CY41R1 and CY41R2, with only a few reforecasts from the cycle CY43R1. Whereas, Fig. A1b is

produced from the reforecasts data involving cycles CY45R1 and CY46R1. Although the data consists of a combination of several cycles, we can compare two versions (CY41R1 and CY46R1) by isolating DJF (top row in Fig. A1). It is very

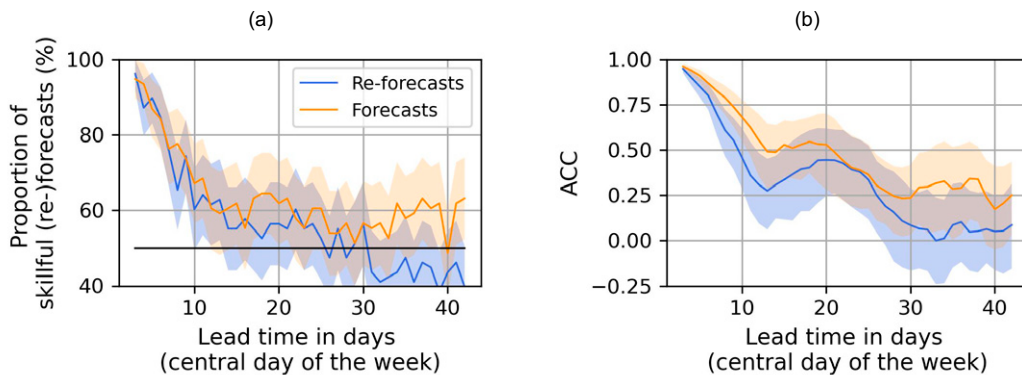


FIG. B1. Comparison of the temporal evolution of 100-m wind speed forecast and reforecast skills averaged across France for DJF between December 2015 and February 2018 (three winters), showing (a) proportion of skillful reforecasts (blue) and forecasts (orange) and (b) ACC. Shaded regions correspond to the 95% confidence intervals. In (a), values above 50% (black horizontal line) indicate skillfulness.

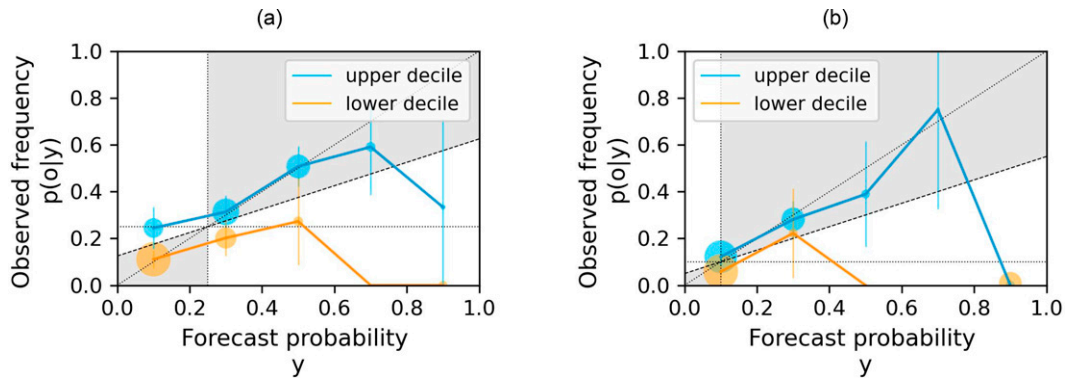


FIG. C1. Reliability diagram for upper and lower (a) quartiles and (b) deciles of weekly mean 2-m temperature forecasts for the week centered on day 17 (i.e., days 14–20) averaged across France. The forecasts are stratified into five bins of equal width. The size of the points is proportional to the number of forecasts in the respective bins. The vertical bars refer to the 95% confidence intervals computed through the standard parametric approach. The vertical and horizontal dotted lines indicate the climatological quartile/decile probabilities (theoretically, the value is $1/4$ for a quartile and $1/10$ for a decile) in the forecasts and observations, respectively. Perfectly reliable forecasts fall on the dotted diagonal line connecting the points (0, 0) and (1, 1). The points located within the gray area contribute positively to the skill.

difficult to say that the version CY46R1 is better than CY41R1, or vice versa. Similar observations were made with respect to other domains for both 2-m temperature and 100-m wind speed.

APPENDIX B

Comparison of Skill between Forecasts and Reforecasts

Figures 9 and 10 indicated that the forecasts are more skillful than the reforecasts. The improved skill of the forecasts may be a result of one or a combination of the way forecasts and reforecasts are initialized (forecasts are initialized using operational analysis, whereas reforecasts are initialized using ERA5 reanalysis), the difference in ensemble size (50 for forecasts and 10 for reforecasts), or the period of the sample considered (December 2015–November 2019 for forecasts and December 1995–November 2018 for reforecasts). Through this section, we try to understand the reasons for improved skill by isolating one or several factors. Figure B1 compares the temporal evolution of skill between 100-m wind speed forecasts and reforecasts similar to Figs. 9 and 10, but for the same period (i.e., DJF 2015/16, 2016/17, and 2017/18). Overall, the behavior is comparable to that of Fig. 10, with forecasts being more skillful than the reforecasts. In addition, the behavior of the ACC (Fig. B1b) of the reforecasts is similar to that of the forecasts but with lower values, indicating the importance of the role of ensemble size and the way the (re)forecasts are initialized on the skill of the (re)forecasts. In this study, since ERA5 reanalysis is used as reference against which the (re)forecasts are verified, the (re)forecast skill may not necessarily be dependent on the way (re)forecasts are initialized, thereby leaving greater weight on the ensemble size.

APPENDIX C

Reliability Diagrams for Quartiles and Deciles of Weekly Mean Temperature Forecasts

Figure C1 shows that the reliability diagrams for the upper and lower quartiles and deciles of temperature forecasts are less reliable and have significantly lower resolution, especially for larger forecast probabilities, relative to that of the terciles.

REFERENCES

- Alonzo, B., R. Plougonven, M. Mougeot, A. Fischer, A. Dupré, and P. Drobinski, 2018: From numerical weather prediction outputs to accurate local surface wind speed: Statistical modeling and forecasts. *Renewable Energy: Forecasting and Risk Management*, Springer Proceedings in Mathematics and Statistics, P. Drobinski et al., Eds., Springer International Publishing, 23–44, https://doi.org/10.1007/978-3-319-99052-1_2.
- Ardilouze, C., D. Specq, L. Batté, and C. Cassou, 2021: Flow dependence of wintertime subseasonal prediction skill over Europe. *Wea. Climate Dyn.*, **2**, 1033–1049, <https://doi.org/10.5194/wcd-2-1033-2021>.
- Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301**, 636–640, <https://doi.org/10.1126/science.1087143>.
- Brune, S., J. D. Keller, and S. Wahl, 2021: Evaluation of wind speed estimates in reanalyses for wind energy applications. *Adv. Sci. Res.*, **18**, 115–126, <https://doi.org/10.5194/asr-18-115-2021>.
- Büeler, D., R. Beerli, H. Wernli, and C. M. Grams, 2020: Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quart. J. Roy. Meteor. Soc.*, **146**, 3675–3694, <https://doi.org/10.1002/qj.3866>.

- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, <https://doi.org/10.1002/qj.2619>.
- , M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- Cionni, I., J. Ramon, L. Lledó, H. Loukos, and T. Noël, 2018: Validation of observational dataset and recommendations to the energy users. Research and Innovation action H2020-SC5-2017 S2S4E Climate Services for Clean Energy Tech. Rep. 3.1, 106 pp., https://s2s4e.eu/sites/default/files/2020-06/s2s4e_d31.pdf.
- Coelho, C. A., M. A. Firpo, and F. M. de Andrade, 2018: A verification framework for South American sub-seasonal precipitation predictions. *Meteor. Z.*, **27**, 503–520, <https://doi.org/10.1127/metz/2018/0898>.
- , B. Brown, L. Wilson, M. Mittermaier, and B. Casati, 2019: Forecast verification for S2S timescales. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. Robertson and F. Vitart, Eds., Elsevier, 337–361.
- Diro, G. T., and H. Lin, 2020: Subseasonal forecast skill of snow water equivalent and its link with temperature in selected SubX models. *Wea. Forecasting*, **35**, 273–284, <https://doi.org/10.1175/WAF-D-19-0074.1>.
- Dorrington, J., I. Finney, T. Palmer, and A. Weisheimer, 2020: Beyond skill scores: Exploring sub-seasonal forecast value through a case-study of French month-ahead energy prediction. *Quart. J. Roy. Meteor. Soc.*, **146**, 3623–3637, <https://doi.org/10.1002/qj.3863>.
- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of atmosphere–ocean coupling on the predictability of monsoon intraseasonal oscillations. *J. Atmos. Sci.*, **64**, 157–174, <https://doi.org/10.1175/JAS3830.1>.
- Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2012: Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 1814–1827, <https://doi.org/10.1002/qj.1895>.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- , and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hoskins, B., 2012: Predictability beyond the deterministic limit. World Meteorological Organization, <https://public.wmo.int/en/bulletin/predictability-beyond-deterministic-limit>.
- International Energy Agency, 2020: World energy outlook 2020. IEA Rep., 464 pp., <https://iea.blob.core.windows.net/assets/a72d8abf-de08-4385-8711-b8a062d6124a/WEO2020.pdf>.
- Jifan, C., 1989: Predictability of the atmosphere. *Adv. Atmos. Sci.*, **6**, 335–346, <https://doi.org/10.1007/BF02661539>.
- Jolliffe, I., and D. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 254 pp.
- Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004a: Global occurrences of extreme precipitation and the Madden–Julian Oscillation: Observations and predictability. *J. Climate*, **17**, 4575–4589, <https://doi.org/10.1175/3238.1>.
- , —, —, and —, 2004b: The Madden–Julian Oscillation and its impact on Northern Hemisphere weather predictability. *Mon. Wea. Rev.*, **132**, 1462–1471, [https://doi.org/10.1175/1520-0493\(2004\)132<1462:TMOAII>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1462:TMOAII>2.0.CO;2).
- Jourdir, B., 2015: Wind resource in metropolitan France: Assessment methods, variability and trends. Ph.D. thesis, Ecole Polytechnique, 229 pp., http://inis.iaea.org/search/search.aspx?orig_q=RN:48072202.
- , 2020: Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France. *Adv. Sci. Res.*, **17**, 63–77, <https://doi.org/10.5194/asr-17-63-2020>.
- Jung, T., F. Vitart, L. Ferranti, and J.-J. Morcrette, 2011: Origin and predictability of the extreme negative NAO winter of 2009/10. *Geophys. Res. Lett.*, **10**, L16815, <https://doi.org/10.1029/2011GL046786>.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- Koster, R. D., and Coauthors, 2011: The second phase of the global land–atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill. *J. Hydrometeorol.*, **12**, 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- Leung, L. R., A. F. Hamlet, D. P. Lettenmaier, and A. Kumar, 1999: Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River Basin. *Bull. Amer. Meteor. Soc.*, **80**, 2313–2330, [https://doi.org/10.1175/1520-0477\(1999\)080<2313:SOTEHS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2313:SOTEHS>2.0.CO;2).
- Leutbecher, M., 2005: On ensemble prediction using singular vectors started from forecasts. ECMWF Tech. Memo. 462, 11 pp., <https://doi.org/10.21957/xuyeqtxv>.
- Lin, H., and Z. Wu, 2011: Contribution of the autumn Tibetan Plateau snow cover to seasonal prediction of North American winter temperature. *J. Climate*, **24**, 2801–2813, <https://doi.org/10.1175/2010JCLI3889.1>.
- Lledó, L., and F. J. Doblas-Reyes, 2020: Predicting daily mean wind speed in Europe weeks ahead from MJO status. *Mon. Wea. Rev.*, **148**, 3413–3426, <https://doi.org/10.1175/MWR-D-19-0328.1>.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333, <https://doi.org/10.1111/j.2153-3490.1965.tb01424.x>.
- Lynch, K. J., D. J. Brayshaw, and A. Charlton-Perez, 2014: Verification of European subseasonal wind speed forecasts. *Mon. Wea. Rev.*, **142**, 2978–2990, <https://doi.org/10.1175/MWR-D-13-00341.1>.
- Machin, D., T. Bryant, D. Altman, and M. Gardner, 2013: *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. John Wiley and Sons, 254 pp.
- Manzanas, R., J. M. Gutiérrez, J. Bhend, S. Hemri, F. J. Doblas-Reyes, V. Torralba, E. Penabad, and A. Brookshaw, 2019: Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the C3S dataset. *Climate Dyn.*, **53**, 1287–1305, <https://doi.org/10.1007/s00382-019-04640-4>.
- Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. *Manage. Sci.*, **22**, 1087–1096, <https://doi.org/10.1287/mnsc.22.10.1087>.
- Molteni, F., U. Cubasch, and S. Tibaldi, 1986: 30- and 60-day forecast experiments with the ECMWF spectral models. *Workshop on Predictability in the Medium and Extended Range*, Shinfield Park, Reading, ECMWF, 51–107, <https://www.ecmwf.int/node/11210>.

- Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of subseasonal forecasts in Europe: Effect of bias correction and downscaling using surface observations. *J. Geophys. Res. Atmos.*, **123**, 7999–8016, <https://doi.org/10.1029/2017JD027923>.
- Namias, J., 1952: The annual course of month-to-month persistence in climatic anomalies. *Bull. Amer. Meteor. Soc.*, **33**, 279–285, <https://doi.org/10.1175/1520-0477-33.7.279>.
- Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138**, 841–861, <https://doi.org/10.1002/qj.1923>.
- , R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp., <https://doi.org/10.21957/ps8gbwbdv>.
- Ramon, J., L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2019: What global reanalysis best represents near-surface winds? *Quart. J. Roy. Meteor. Soc.*, **145**, 3236–3251, <https://doi.org/10.1002/qj.3616>.
- Raoult, B., C. Bergeron, A. L. Alós, J.-N. Thépaut, and D. Dee, 2017: Climate service develops user-friendly data store. *ECMWF Newsletter*, No. 151, ECMWF, Reading, United Kingdom, <https://www.ecmwf.int/en/newsletter/151/meteorology/climate-service-develops-user-friendly-data-store>.
- Robertson, A., and F. Vitart, 2018: *Sub-seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*. Elsevier, 512 pp.
- Sanders, F., 1963: On subjective probability forecasting. *J. Appl. Meteor.*, **2**, 191–201, [https://doi.org/10.1175/1520-0450\(1963\)002<0191:OSPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1963)002<0191:OSPF>2.0.CO;2).
- Seneviratne, S., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230, https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3_FINAL-1.pdf.
- Simmons, A., and Coauthors, 2021: Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets. ECMWF Tech. Memo. 881, 99 pp., <https://doi.org/10.21957/ly5vbtbfd>.
- Sobolowski, S., G. Gong, and M. Ting, 2010: Modeled climate state and dynamic responses to anomalous North American snow cover. *J. Climate*, **23**, 785–799, <https://doi.org/10.1175/2009JCLI3219.1>.
- Torralba, V., F. J. Doblas-Reyes, D. MacLeod, I. Christel, and M. Davis, 2017: Seasonal climate prediction: A new source of information for the management of wind energy resources. *J. Appl. Meteor. Climatol.*, **56**, 1231–1247, <https://doi.org/10.1175/JAMC-D-16-0204.1>.
- Unger, D. A., 1985: A method to estimate the continuous ranked probability score. *Conf. Prob. Stat. Atmos. Sci.*, **9**, 206–213, https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=200902092398162270.
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dyn.*, **38**, 349–362, <https://doi.org/10.1007/s00382-010-0956-2>.
- Vigaud, N., M. K. Tippett, J. Yuan, A. W. Robertson, and N. Acharya, 2019: Probabilistic skill of subseasonal surface temperature forecasts over North America. *Wea. Forecasting*, **34**, 1789–1806, <https://doi.org/10.1175/WAF-D-19-0117.1>.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, <https://doi.org/10.1002/qj.2256>.
- , and Coauthors, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- , and Coauthors, 2019: Extended-range prediction. ECMWF Tech. Memo. 854, 60 pp., <https://doi.org/10.21957/pdivp3t9m>.
- White, C. J., and Coauthors, 2017: Potential applications of sub-seasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Sciences*. 4th ed. Elsevier, 840 pp.
- Woolnough, S. J., F. Vitart, and M. A. Balmaseda, 2007: The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Quart. J. Roy. Meteor. Soc.*, **133**, 117–128, <https://doi.org/10.1002/qj.4>.
- Žagar, N., and I. Szunyogh, 2020: Comments on “What is the predictability limit of midlatitude weather?”. *J. Atmos. Sci.*, **77**, 781–785, <https://doi.org/10.1175/JAS-D-19-0166.1>.
- Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76**, 1077–1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.
- Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Wea. Rev.*, **142**, 1556–1569, <https://doi.org/10.1175/MWR-D-13-00222.1>.