

Evaluating the Skill of Seasonal Wind Energy Forecasts in Germany

Abhinav Tyagi, Lukas Pauscher, Alina Happ,
Muhammad Arsalan, Axel Braun,
Martin Braun, Malte Siefert
Fraunhofer IEE
Kassel, Germany

Abhinav Tyagi, Lukas Pauscher, Martin Braun
Sustainable Electrical Energy Systems
University of Kassel
Kassel, Germany

Jan Wandel, Andreas Paxian
Deutscher Wetterdienst
Offenbach, Germany

Lukas Pauscher
Acoustics & Vibration Research Group
Vrije Universiteit Brussel
Elsene, Belgium

Abstract—Seasonal renewable wind energy forecasts predict energy production, typically up to six months, with possible applications in supporting energy trading and optimizing operations and reducing grid risks. This study specifically develops the first seasonal energy forecasts tailored for Germany by applying a wind power model to seasonal wind speed forecasts from the German Climate Forecast System Version 2.1 (GCFS2.1). The hindcasts for years 1990-2000, with reanalysis data from ERA5 used as a baseline, are used for skill evaluation. The verification with three complementary metrics, probabilistic and deterministic, reveal that, despite the overall low skill and limited statistical significance, there might exist a few ‘opportunity windows’ where the seasonal forecasts could be used as an additional input in a decision-making process. Concretely, mid-summer (June-August) and early winter (November-January) seasons in south Germany show a slight better performance, which may be connected to the predictability of large-scale circulations in these regions and seasons. The analysis also highlights the need for further calibration techniques that might enhance the forecast skills.

Index Terms—Power curve, Ensemble methods, Seasonal energy forecasting, Wind energy conversion, Wind energy forecasting

I. INTRODUCTION

Renewable energy is becoming increasingly significant in the global energy supply. In the past year, wind energy constituted roughly 30% of the total electricity production of Germany, reaching roughly 140 TWh, ahead of all other sources of energy, conventional or renewable [1].

Renewable energy production’s dependency on weather introduces uncertainties in grid operations and planning, necessitating additional safeguards and incurring extra costs. Short-term energy forecasts (up to a few days in advance) have aided in addressing this uncertainty. Although weather forecasts have drastically improved in the last few decades [2], the chaotic nature of atmospheric flows [3] makes weather forecasts beyond two weeks difficult [4]. Seasonal forecasts are dynamic and derive predictability beyond two weeks due to the boundary conditions derived from slowly changing climate

variables like soil moisture, snow cover, sea-ice, and ocean temperatures [5].

Thus, skilful seasonal energy forecasts could complement the short-term forecasts and help mitigate uncertainties by allowing energy traders to address trading risks earlier and enabling grid operators to anticipate operational risks and take proactive measures. These forecasts also support strategic resource planning and facilitate international energy trade by predicting market dynamics. Case studies, such as the 2017 European cold wave and the 2015 North American wind drought [6], highlight the potential of these forecasts in guiding decision-making to ensure stable energy supplies.

An increasing number of operational weather services are offering seasonal forecasts [7], in part motivated by the success of precipitation and temperature seasonal forecasts (for example, [8], [9]). A recent example is the Decision Support Tool¹, developed in Sub-seasonal to Seasonal Forecasts for Energy (S2S4E) project, to highlight the trend towards developing high-resolution regional forecasts. Numerous seasonal forecasts datasets are also being developed, most notably from ECMWF but also from Météo-France’s (METFR) System [10], Met Office Global Seasonal forecasting system [11], [12] and German Climate Forecast System (GCFS) service by Deutscher WetterDienst [13]

Various skill analyses focusing on temperature and precipitation have been published in the last decade. However, the existing literature for wind energy related weather variables and even more so for wind energy forecasts is scarce and portrays an unclear picture of their skill. Skill in wind forecasts is typically shown in a few selected regions and/or seasons. Lledo *et al.* [14] performed global capacity factor forecasts and portrayed positive skills for DJF (December-January-February) season in few regions, mostly limited to countries in Northern Europe. Surface wind speed skill analysis by Lee *et al.* [15] using a multimodel approach showed positive skills for the regions extending North America, northern South

America, most of the maritime continent, eastern Africa and northern portion of China. Similarly, Yang *et al.* [16]’s wind power skill analysis over the contiguous United States region demonstrated that the ensemble models initialized in March could describe up to 65% of the variance in the wind power observed during spring (MAM) over the state of Texas. However, all of these studies focus on specific events or regions and limit the analysis to specific regions and seasons. Systematic and comprehensive evaluations across different seasons are lacking.

This paper presents a systematic skill analysis of energy simulations in Germany, utilizing the latest seasonal forecasting model from DWD [13] for the first time (see description of used Datasets in section II-B). The power hindcasts over Germany are aggregated over different seasons and regions in Germany to return the total energy and investigate the temporal and spatial dependence of skill. Three metrics—Anomaly Cross Correlation, Ranked Probability Skill Score, and a new *Mismatch* metric—are used to assess forecast accuracy and significance. These metrics are complementary as they investigate both deterministic and probabilistic skills of the hindcasts and are described in detail in section III.

A key aspect of this analysis is that it aims to be systematic and comprehensive. As previously discussed, the skill of seasonal forecasts is often only communicated for specific seasons and specific regions in the literature. This makes the interpretation of the overall skill of the forecasts difficult. This is avoided in this paper by presenting averaged skills with significance tests w.r.t. target season and lead months for different regions of Germany.

II. DATA

A. Forecasts

Seasonal forecasts of the wind speed are taken from the German Climate Forecasting System (GCFS) Version 2.1 [13]. The forecasts are generated using the Max Planck Institute for Meteorology earth System Model High Resolution (MPI-ESM-HR) [17]) in combination with the statistical downscaling procedure EPISODES [18], reaching a spatial resolution of five kilometres. They are delivered as an ensemble (30 members) of surface wind speeds (at 10 m) averaged over a day for the years between 1990 and 2020. Each ensemble depends on slightly different initial conditions and thus describes a possible scenario for the future development [19]. For each year, the hindcasts are initialised every month and predictions are made six months into the future.

B. Baseline

The ERA5 reanalysis [20] dataset developed by ECWMF and delivered within the Copernicus Climate Change Service (C3S), depicts the global atmosphere, land surface, and ocean waves from 1950 to the present on a grid with a spatial resolution of 31 km and a temporal resolution of one hour. Reanalysis data sets are created by combining measured past observations with weather models. ERA5, based on the Integrated Forecasting System (IFS), provides coherent spatial and

temporal maps of essential climate variables [20], making it ideal for validating seasonal hindcasts. It serves as the baseline (ground truth) for this paper’s experiments to assess uncertainties and error measures. To make the dataset compatible with the hindcasts, linear interpolation is performed spatially, and the wind speeds are averaged over a day.

III. METHODS

A. Conversion of Weather Data into Energy

The wind power output of a turbine relies on weather variables such as wind speed, direction, and air density, with wind speed being the most influential. Due to large uncertainties in seasonal wind speeds, other variables’ effects are often negligible [21]. Conversion of wind speeds into energy in this paper is performed using a power curve, which models turbine response to varying wind speeds. To roughly represent the average wind energy production in Germany [1], Enercon E-70/2300 wind turbine with 2300 kW nominal power and a hub height of 98 m was chosen [22]. The wind turbine is a widely sold wind turbine and should simply serve as a general reference in this paper.

Wind speeds from forecast models are provided at 10 m above ground and extrapolated to hub height using a power law [14], [23] with a shear coefficient of $\alpha = 0.1$ for agricultural terrain [24]. While a more diverse terrain and altitude-specific treatment of power curves (or the type of wind turbine) can be used to reflect the true wind energy production in Germany, this is not expected to affect the skills of the hindcasts, since the same model is used to convert wind speed from both, hindcasts and reanalysis, to power.

A power curve reflects the instantaneous wind turbine response to an input wind speed. However, the wind speeds predicted using GCFS weather models are averages over 24 hours. Owing to the non-linearity of the power curve, average wind speeds do not translate to true wind power averages over a given time window [14]. Macleod *et al.* [21] found that using sub-daily averages did not significantly improve the deviation of the true energies from the simulated energy.

B. Skill Verification

1) *Preprocessing the Energy time series:* Seasonal forecasts are not expected to have skill below sub-monthly temporal resolutions because the forecastability is derived from large-scale circulations that change over a monthly timescale [25]. Therefore, an average over a three-month window is chosen as also recommended by the World Meteorological Organization (WMO) [26].

Additionally, unlike existing literature that typically assesses skills at the finest resolution, this analysis aggregates skills across five regions in Germany: for northwest (NW), northeast (NE), southwest (SW), southeast (SE) and the whole of Germany. This approach aims to provide an average overview of skills in the country. Although, while higher resolution can enhance the depiction of local events and minimize systematic biases [27], [28], it does not generally result in improved

overall forecast skill because of the high spatial correlation across large areas [29].

To enable an easy interpretation of overall skill, skills are calculated for each predicted target season and lead time. Skill at each lead time is calculated by combining the available twelve seasons, i.e. using 372 unique windows (31 years \times 12 seasons). Similarly, skill for each target season is calculated by combining the four lead times, i.e. using 124 unique windows (31 years \times 4 lead times).

The lead-time dependent linear biases that might have been introduced in the forecast model due to model drift from the climatology are accounted for by evaluating the skill scores on the forecasted anomalies and analysis anomalies rather than the absolute values themselves [13], [30]. Anomalies are calculated individually for hindcasts and reanalysis by subtracting the lead-time dependent (for hindcasts) climatology from the absolute values.

2) *Verification Metrics*: Three complementary scores have been used to evaluate the skill of energy hindcasts. The first one is the Anomaly Cross Correlation (*ACC*) which is a deterministic score that evaluates the Pearson correlation coefficient between the mean forecasted anomalies of the ensemble mean and the reanalysis anomalies and provides a measure of much of the variation of the reanalysis can be described by the model.

While *ACC* evaluates the skill of the ensemble mean prediction of the ensemble hindcasts to predict anomalies, the Ranked Probability Skill Score (*RPSS*) quantifies the difference between the forecasted and reanalysis probabilities. This is done using three distinct events: ‘below normal’, ‘normal’ and ‘above normal’ energy output (see the implementation in [13]). Since the terciles are defined as the analysis anomalies, the probabilities for each event j , given by the climatology model, is $\bar{O}^j = 33.\bar{3}\%$. The skill of the climatology model ($RPSS^O$) can be compared with that of the forecast model ($RPSS^P$) using a skill score:

$$RPSS = 1 - \frac{RPSS^P}{RPSS^O}, \quad (1)$$

where a positive *RPSS* indicates a better performance than climatology. A more detailed explanation of the *RPSS* can be found in Mueller *et al.* [31].

The third and final metric is derived from the spread-skill score ($SSR = \epsilon/\sigma$, as implemented in [32], where σ is the ensemble spread and ϵ is the mean displacement of the ensemble from the reanalysis. For easier interpretation, a new quantity, called *Mismatch*, $M = 1 - \frac{1}{SSR}$ has been introduced in this paper. *Mismatch* becomes exactly zero if the mean ensemble spread is equal to the mean ensemble error, implying that the analysis ‘matches’ the hindcasts. Additionally, a negative *Mismatch* implies an under-variant forecast ensemble and a positive *Mismatch* implies an over-variant forecast ensemble.

3) *Significance Estimation*: The significance estimation utilized a two-tailed hypothesis test, with the null hypothesis stating that the evaluated skill score equals zero. The Jackknife sampling test was employed for all three metrics, as

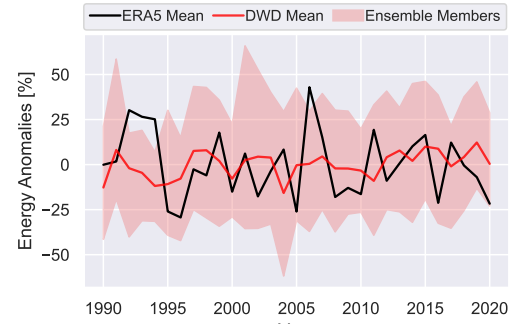


Fig. 1. Ensemble hindcasts, initialized in November (Ensemble members and DWD mean) and Reanalysis (ERA5 mean) plotted as the deviation from climatology in percentage (percentage anomalies) aggregated for season November-January and all of Germany for years 1990-2020.

it does not require a normal sample distribution [33]. This method is less computationally intensive than bootstrapping and is more effective for non-stationary time series, such as weather data [34]. The Jackknife method generated $T = 31$ skill score iterations, corresponding to 31 years of forecast-reanalysis pairs from 1990 to 2020, by omitting the current forecast observation pair in each iteration. A value is positively/negatively statistically significant if 95% of the iterations are positive/negative. However, sampling tests also have drawbacks and may yield inaccurate estimates from small samples and may be affected by the lack of identical and independent distribution. While we could not find a significance test, that would be ideal given the nature of data here, an additional parametric t-test for correlation was conducted (as described in [35]) to ensure robust skill estimates.

IV. RESULTS

The forecast of the ensemble members and reanalysis of the energy production for Germany from November-January (1990-2020) are shown in figure 1. The y-axis displays energy anomalies as a percentage change from climatology: the reanalysis (black curve) and the ensemble mean (red curve), with the shaded area indicating min-max ensemble spread. The reanalysis generally falls within this ensemble spread, albeit with no clear correlation with the ensemble mean, particularly in 1992-93 and 2006, where simulated energy was higher than the climatological mean, while the ensemble predicted close to the climatological mean. Figure 1 represents a single forecast scenario and does not reflect the overall prediction model properties. Therefore, the following will entail a discussion of skill w.r.t different predicted seasons (target season) and lead times for different regions in Germany.

The correlation values w.r.t target season lie in range $ACC \in (-0.2, 0.25)$ (figure 2, top panel), indicating that the forecast model explains roughly 6% of the total variance at its best. However, the two hypothesis tests (as described in III-B3) deliver very contrasting results. The points that are not statistically significant using the jackknife approach have been marked with circles, and the critical threshold above which the correlations can be considered significant according to the

t-test is plotted as the dashed horizontal-line in the figure. The jackknife test is rather liberal and all correlation values, apart from SE Germany in August-October and NW Germany in March-July, are flagged as either significantly positive or significantly negative. In contrast, only the correlation values for southern Germany in April-September and November-January exceed the critical correlation threshold ($r = 0.17$) and can be considered statistically significant according to the parametric t-test.

The skill variation is much more significant w.r.t target seasons when compared with lead months, as will be discussed in detail later. The skill plots w.r.t lead months can be seen in the appendix. The correlation values w.r.t lead months lie in range $ACC \in (-0.05, 0.2)$ (figure 4, top panel) and show contrasting results, similar to the target months plot. The correlations do not have any statistically insignificant points according to the jackknife test. In contrast, only the correlations for lead month zero (model initialized at the beginning of a given season) are above the critical correlation threshold ($r = 0.08$) as derived from the t-test.

The two thresholds derived from the t-test differ in magnitude simply because of the size of the test sample in each case (see section III-B1). It is clear that the statistical significance of the skill scores is starkly different within the two tests performed here and should be interpreted with caution.

RPSS scores (figure 2, middle panel) consistently fall below zero, indicating worse performance than climatology. Furthermore, *Mismatch* (figure 2, bottom panel) highlights the need for further calibration, showing mostly negative values (implying an under-variant ensemble) across all seasons and lead months. Complementary to the linear bias correction performed here, non-linear calibration techniques like quantile mapping [36] and machine learning methods (such as Random Forest [37] and Support Vector Machine [38], that have proven to be especially effective) could be employed in the future to improve the seasonal forecasts' skill.

It is clear that the low absolute skill values and the non-stationarity of the time-series make the significance testing difficult. While the overall skills of seasonal forecasts are low, a few 'opportunity windows' can still be identified where the forecasts are relatively better when compared with other regions/seasons.

1) *Higher Skills in Summer in Southern Germany*: The consistently negative *Mismatch* plotted in the bottom panels of figure 2 indicates that the ensemble is generally under-variant. Regionally, Northern Germany (blue and green curves) exhibits a lower *Mismatch* than southern Germany, particularly pronounced during summer (JAS), where NW Germany (blue curve) shows the highest under-variance in the ensemble. Exceptionally, SE Germany (yellow curve) shows a slight positive mismatch for most seasons and lead months, indicating an over-variant ensemble.

These regional trends can be compared with the seasonal standard deviation in reanalysis-based wind energies, plotted as a percentage of the climatological mean in figure 3). The regions with high *Mismatch* correspond to a higher standard de-

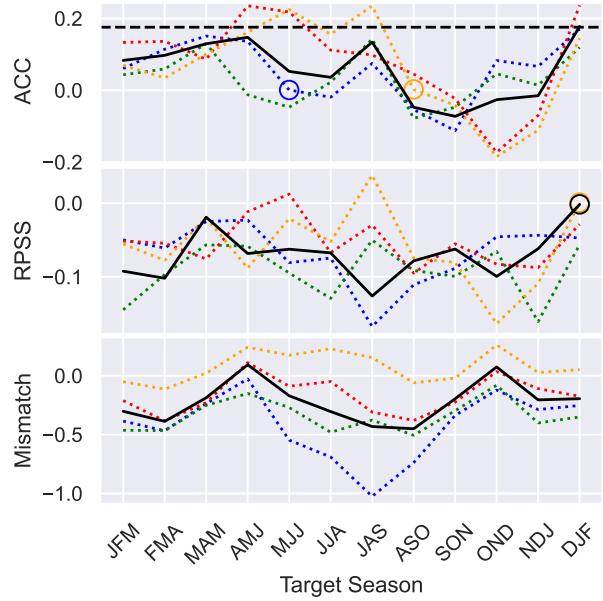


Fig. 2. Skill Metrics plotted against forecasted season. Different curves represent different skill analysis done over different regions (see colour scheme in the legend of figure 3). The circles highlight the skill scores that are flagged statistically insignificant using the Jackknife method. The dashed horizontal line in the correlation plot (top panel) is the critical threshold above which the values are significant according to the parametric t-test.

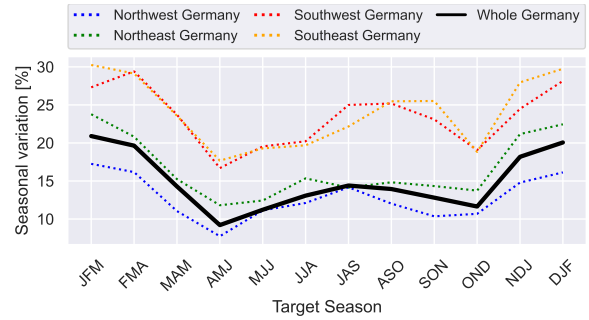


Fig. 3. Inter-annual (1980-2020) seasonal standard deviation (variation) in the reanalysis-based simulated energy production, plotted as percentage of the mean for all seasons of the year.

viation in simulated wind energy production. More concretely, Southern Germany shows greater seasonal variation (20-30%) and a *Mismatch* closer to zero (lower under-variance), while northern Germany displays lower variation (10-20%) and a more negative *Mismatch* (higher under-variance). This suggests the ensemble model might be sensitive to regional seasonal variations.

Although the ensemble is under-variant overall for all seasons and lead times, *Mismatch* shows an increasing tendency during spring (March-May) and autumn (September-November). This tendency towards higher-variance forecast models aligns with the seasons of decreased seasonal standard deviation in the reanalysis dataset (figure 3), suggesting the forecast model might be struggling to adjust its variance appropriately for low-seasonal-variation seasons.

Not only is northern Germany generally more under-variant,

but also shows lower correlations (figure 2, top panel) and more negative *RPSS* scores compared to their southern counterparts, the only exception being the season around October-December.

Furthermore, both *RPSS* and *ACC* show their best performance ($ACC \sim 0.2$; $RPSS > 0$) for SE Germany around summer (March to August). At the same time, the seasonal standard deviation in summer is also at its lowest (10-20%) compared to winter (up to 30%). Both correlations and *RPSS* trend upward toward peak winter months (December-January), where seasonal variation is highest.

The regional and seasonal dependence of the skills might be connected with the predictability patterns of the large-scale circulations. This has been shown for sub-seasonal time scales on a country level in the work of Bloomfield *et al.* [39]. A similar in-depth investigation for seasonal forecasts and at a higher resolution (within Germany) will be done in subsequent analysis.

2) *Strong dependence of forecast skill on target season:* While the low absolute correlation values make statistical significance challenging, trends indicate that correlation varies more across target seasons than lead times. The range for target months ($ACC \in (-0.2, 0.25)$, figure 2, top panel) is almost twice that of lead months ($ACC \in (-0.05, 0.2)$, figure 4, top panel). Additionally, correlation shows strong fluctuations across seasons but remains roughly constant after the first lead month. However, the sample size used for the calculation of skills for each season is much smaller (124 windows) than the sample for lead time (372 windows), as described in section III-B1, which might also be a reason for stronger fluctuations w.r.t the seasons.

Similar trends are observed for both *RPSS* and *Mismatch*, as the variation regarding target seasons is significantly stronger than that for lead months (figures 2 and 4, middle and bottom panels). For both the cases, the variation range for target seasons is nearly double that for lead months. Unlike *ACC*, where correlation decreases after the first lead month and stabilizes, *RPSS* and *Mismatch* show little variation across lead months. This suggests that the skill of the seasonal hindcasts might be influenced more by the forecasted season than by the lead time from initialization.

This phenomenon has been noted in previous literature, such as in Das *et al.* [37]. The likely reason is that forecasting errors saturate quickly within the first few days, indicating a stronger dependence on the forecasted season than on lead time [40].

V. CONCLUSION AND FUTURE PROSPECTS

This paper presents the first systematic skill analysis of seasonal energy hindcasts in Germany. We evaluate simulated energy hindcasts based on a physical wind power model. The seasonal hindcasts and baseline dataset (1980-2020) for surface wind speed at 10 m above ground are sourced from the German Climate Forecast System Version 2.1 (GCFS) [13] and the ERA5 reanalysis [20] respectively. Future work will incorporate 100 m wind speed hindcasts, already planned for

GCFS in upcoming releases, that will help mitigate systematic biases in energy estimates.

Skill evaluation employs three metrics: Anomaly Cross Correlation (*ACC*), Ranked Probability Skill Score (*RPSS*), and *Mismatch*. *Mismatch* assesses variance in ensemble hindcasts, *ACC* measures correlation between forecasted and reanalysis anomalies, and *RPSS* evaluates predicted probabilities of extreme trends. We analyse skills temporally across target seasons and lead months, and spatially by dividing Germany into four quadrants. Overall, hindcasts demonstrate low skill across all metrics, with *ACC* values between -0.2 and 0.25, indicating roughly 6% explained variance in the best-case scenario. *RPSS* values are often negative, performing worse than climatology.

Due to a small test sample comprising non-stationary weather time series, determining the statistical significance of the estimated skills was difficult. Nevertheless, some trends have been identified in the data which might hint at regions or seasons with a relatively higher skill. These ‘opportunity windows’ albeit with low skills, might still be an addition and a useful input for decision-makers. Regionally, northern Germany showed worse skills compared to southern Germany, and seasonally, mid-summer and early winter had better skills than the other seasons. Notably, skill seems to be more affected by the target season than by the lead time from initialization.

While the observed skills suggest that seasonal forecast cannot be integrated into energy applications as easily as forecasts on shorter time scales, future research and development offers two immediate exploration avenues to improve the forecasting skills:

- 1) The *Mismatch* analysis suggests that ensemble hindcasts are consistently under-variant compared to the reanalysis dataset. Non-linear calibration methods, such as Quantile Regression and machine learning techniques, could enhance skill, though their effectiveness is uncertain due to the limited training points available for seasonal forecasts.
- 2) Incorporating the inherent predictability in large-scale atmospheric circulations by subsampling the ensemble forecasts based on each member’s correlation with the teleconnections has shown promising results in the literature [41].

Conclusively, while seasonal forecasting is an emerging research field, and still has vast potential for improvement, they can be consulted as an additional input in the decision-making process for certain seasons and regions in Germany.

VI. ACKNOWLEDGMENT

The work for this paper was co-funded by the European Union (via the WinDTwin project, Grant Agreement number: 101147377) and the German Federal Ministry for Economic Affairs and Energy (via the LongCast project, funding code: 03EI1085A). Additionally, the authors acknowledge the use of OpenAI’s Chatbots for grammar and style suggestions in the preparation of this manuscript.

REFERENCES

- [1] B. W. e.V., “Statistics GermanyBWE e.V.” <https://www.wind-energie.de/english/statistics/statistics-germany/>.
- [2] P. Bauer, A. Thorpe, and G. Brunet, “The quiet revolution of numerical weather prediction,” *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.
- [3] E. N. Lorenz, “Irregularity: A fundamental property of the atmosphere,” *Tellus A: Dynamic Meteorology and Oceanography*, vol. 36, no. 2, p. 98, Jan. 1984.
- [4] T. A. Gleeson, “On theoretical limits of predictability,” *Journal of Applied Meteorology*, vol. 6, pp. 213–215, 1967. [Online]. Available: <https://api.semanticscholar.org/CorpusID:120814703>
- [5] F. J. Doblas-Reyes, J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, “Seasonal climate predictability and forecasting: Status and prospects,” *WIREs Climate Change*, vol. 4, no. 4, pp. 245–268, 2013.
- [6] A. Soret, A. Soret, V. Torralba, V. Torralba, N. Cortesi, N. Cortesi, N. Cortesi, I. Christel, I. Christel, L. Palma, L. Palma, A. Manrique-Suñén, A. Manrique-Suñén, L. Lledó, L. Lledó, N. González-Reviriego, N. González-Reviriego, F. J. Doblas-Reyes, F. J. Doblas-Reyes, and F. J. Doblas-Reyes, “Sub-seasonal to seasonal climate predictions for wind energy forecasting,” *Journal of Physics: Conference Series*, 2019.
- [7] C. D. Hewitt, E. Allis, S. J. Mason, M. Muth, R. Pulwarty, J. Shumake-Guillemot, A. Bucher, M. Brunet, A. M. Fischer, A. M. Hama, R. K. Kolli, F. Lucio, O. Ndiaye, and B. Tapia, “Making Society Climate Resilient: International Progress under the Global Framework for Climate Services,” *Bulletin of the American Meteorological Society*, vol. 101, no. 2, pp. E237–E252, Feb. 2020.
- [8] T. Roy, X. He, P. Lin, H. E. Beck, C. Castro, and E. F. Wood, “Global Evaluation of Seasonal Precipitation and Temperature Forecasts from NMME,” *Journal of Hydrometeorology*, vol. 21, no. 11, pp. 2473–2486, Nov. 2020.
- [9] Y. Lang, A. Ye, W. Gong, C. Miao, Z. Di, J. Xu, Y. Liu, L. Luo, and Q. Duan, “Evaluating Skill of Seasonal Precipitation and Temperature Predictions of NCEP CFSv2 Forecasts over 17 Hydroclimatic Regions in China,” *Journal of Hydrometeorology*, vol. 15, no. 4, pp. 1546–1559, Aug. 2014.
- [10] “Page non trouvée | seasonal.meteo.fr,” <https://seasonal.meteo.fr/content/SF-data>.
- [11] “GloSea5: Met Office seasonal prediction system,” <https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/climate-models/glosea5>.
- [12] K. D. Williams, C. M. Harris, A. Bodas-Salcedo, J. Camp, R. E. Comer, D. Copesey, D. Fereday, T. Graham, R. Hill, T. Hinton, P. Hyder, S. Ineson, G. Masato, S. F. Milton, M. J. Roberts, D. P. Rowell, C. Sanchez, A. Shelly, B. Sinha, D. N. Walters, A. West, T. Woollings, and P. K. Xavier, “The Met Office Global Coupled model 2.0 (GC2) configuration,” Jan. 2015.
- [13] A. Paxian, B. Mannig, M. Tivig, K. Reinhardt, K. Isensee, A. Pasternack, A. Hoff, K. Pankatz, S. Buchholz, S. Wehring, P. Lorenz, K. Fröhlich, F. Kreienkamp, and B. Fröh. “The DWD climate predictions website: Towards a seamless outlook based on subseasonal, seasonal and decadal predictions,” *Climate Services*, vol. 30, p. 100379, Apr. 2023.
- [14] L. Lledó, V. Torralba, A. Soret, J. Ramon, and F. J. Doblas-Reyes, “Seasonal forecasts of wind power generation,” *Renewable Energy*, vol. 143, pp. 91–100, Jan. 2019.
- [15] D. Y. Lee, D. Y. Lee, D. Y. Lee, F. J. Doblas-Reyes, F. J. Doblas-Reyes, F. J. Doblas-Reyes, V. Torralba, V. Torralba, N. González-Reviriego, and N. González-Reviriego, “Multi-model seasonal forecasts for the wind energy sector,” *Climate Dynamics*, 2019.
- [16] X. Yang, T. L. Delworth, L. Jia, N. C. Johnson, F. Lu, and C. McHugh, “Skillful seasonal prediction of wind energy resources in the contiguous United States,” *Communications Earth & Environment*, 2024.
- [17] T. Mauritsen, J. Bader, T. Becker, J. Behrens, M. Bittner, R. Brokopf, V. Brovkin, M. Claussen, T. Crueger, M. Esch, I. Fast, S. Fiedler, D. Fläschner, V. Gayler, M. Giorgetta, D. Goll, H. Haak, S. Hagemann, C. Hedemann, C. Hohenegger, T. Ilyina, T. Jahns, D. de la Cuesta Otero, J. Jungclaus, T. Kleinen, S. Kloster, D. Kracher, S. Kinne, D. Kleberg, G. Lasslop, L. Kornblüeh, J. Marotzke, D. Matei, K. Meraner, U. Mikolajewicz, K. Modali, B. Möbis, W. Müller, J. Nabel, C. Nam, D. Notz, S.-S. Nyawira, H. Paulsen, K. Peters, R. Pincus, H. Pohlmann, J. Pongratz, M. Popp, T. Raddatz, S. Rast, R. Redler, C. Reick, T. Rohrschneider, V. Schemann, H. Schmidt, R. Schnur, U. Schulzweida, K. Six, L. Stein, I. Stemmler, B. Stevens, J.-S. von Storch, F. Tian, A. Voigt, P. de Vrese, K.-H. Wieners, S. Wilkenskjeld, A. Winkler, and E. Roeckner, “Developments in the mpi-m earth system model version 1:2 (mpi-esm1:2) and its response to increasing co2,” *Journal of Advances in Modeling Earth Systems*, vol. 11, pp. 998–1038, 2018.
- [18] F. Kreienkamp, A. Paxian, B. Fröh, P. Lorenz, and C. Matulla, “Evaluation of the empirical-statistical downscaling method episodes,” *Clim. Dyn.*, vol. 52, pp. 991–1026, 2018.
- [19] C. Deser, F. Lehner, K. B. Rodgers, T. R. Ault, T. L. Delworth, P. N. DiNezio, A. M. Fiore, C. Frankignoul, J. C. Fyfe, D. E. Horton, J. E. Kay, J. E. Kay, R. Knutti, N. S. Lovenduski, N. S. Lovenduski, J. Marotzke, K. A. McKinnon, S. Minobe, J. T. Randerson, J. A. Screen, I. R. Simpson, and M. Ting, “Insights from earth system model initial-condition large ensembles and future prospects,” *Nature Climate Change*, vol. 10, pp. 277–286, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:214704538>
- [20] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, A. M. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellán, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hölm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut, “The era5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [21] D. MacLeod, V. Torralba, M. Davis, and F. Doblas-Reyes, “Transforming climate model output to forecasts of wind power production: how much resolution is enough?” *Meteorological Applications*, vol. 25, no. 1, pp. 1–10, 2018. [Online]. Available: <https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.1660>
- [22] “OEP - Topics,” <https://openenergyplatform.org/dataedit/schemas>.
- [23] *Estimating the Resource at hub Height*. John Wiley & Sons, Ltd, 2012, ch. 11, pp. 143–154. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118249864.ch11>
- [24] “Windenergie-Daten der Schweiz,” <https://wind-data.ch/tools/profile.php?h=10&v=5&z0=0.1&abfrage=Update>.
- [25] J. Shukla, “Predictability in the midst of chaos: A scientific basis for climate forecasting,” *Science*, vol. 282 5389, pp. 728–31, 1998. [Online]. Available: <https://api.semanticscholar.org/CorpusID:39914008>
- [26] G. Kim, J.-B. Ahn, V. N. Kryjov, S.-J. Sohn, W.-T. Yun, R. Graham, R. K. Kolli, A. Kumar, and J.-P. Ceron, “Global and regional skill of the seasonal predictions by wmo lead centre for long-range forecast multi-model ensemble,” *International Journal of Climatology*, vol. 36, no. 4, 2016.
- [27] C. Prodhomme, L. Batté, F. Massonnet, P. Davini, O. Bellprat, V. Guemas, and F. J. Doblas-Reyes, “Benefits of increasing the model resolution for the seasonal forecast quality in ec-earth,” *Journal of Climate*, vol. 29, pp. 9141–9162, 2016. [Online]. Available: <https://api.semanticscholar.org/CorpusID:89606779>
- [28] S. Morelli, “The impact of the spatial resolution of wind data on multi-decadal wind power forecasts in germany.” [Online]. Available: <http://arxiv.org/pdf/2410.14681v1>
- [29] J. Roads, S. Chen, and M. Kanamitsu, “U.s. regional climate simulations and seasonal forecasts,” *Journal of Geophysical Research*, vol. 108, p. 8606, 2003. [Online]. Available: <https://api.semanticscholar.org/CorpusID:129320216>
- [30] L. Goddard, A. Kumar, A. Solomon, D. Smith, G. Boer, P. Gonzalez, V. Kharin, W. Merryfield, C. Deser, S. J. Mason *et al.*, “A verification framework for interannual-to-decadal predictions experiments,” *Climate Dynamics*, vol. 40, pp. 245–272, 2013.
- [31] W. A. Müller, C. Appenzeller, F. J. Doblas-Reyes, and M. A. Liniger, “A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes,” *Journal of Climate*, vol. 18, pp. 1513–1523, 2005. [Online]. Available: <https://api.semanticscholar.org/CorpusID:123051255>
- [32] J. S. Whitaker and A. F. Loughé, “The relationship between ensemble spread and ensemble mean skill,” *Monthly Weather Review*, vol. 126, no. 12, pp. 3292 – 3302, 1998. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/126/12/1520-0493_1998_126_3292_trbesa_2.0.co_2.xml
- [33] G. Calmettes, G. B. Drummond, and S. L. Vowler, “Making do with what we have: use your bootstraps,” *Advances in physiology education*, vol. 36, no. 3, pp. 177–180, 2012.

- [34] Y. G. Yatracos, "Do applied statisticians prefer more randomness or less? bootstrap or jackknife?" *Statistics and Computing*, vol. 34, no. 2, p. 83, 2024.
- [35] M. Lewis, *Applied statistics for economists*. Routledge, 2012.
- [36] M. J. Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.
- [37] A. Das, D. P. Reddy, and S. Baidya Roy, "Improving the skill of subseasonal to seasonal (s2s) wind speed forecasts over india using statistical and machine learning methods," *Journal of Geophysical Research: Machine Learning and Computation*, vol. 1, no. 4, p. e2024JH000187, 2024.
- [38] G. Yin, T. Yoshikane, R. Kaneko, and K. Yoshimura, "Improving global subseasonal to seasonal precipitation forecasts using a support vector machine-based method," *Journal of Geophysical Research: Atmospheres*, vol. 128, no. 17, p. e2023JD038929, 2023.
- [39] H. C. Bloomfield, D. J. Brayshaw, P. L. M. Gonzalez, and A. Charlton-Perez, "Sub-seasonal forecasts of demand and wind power and solar power generation for 28 European countries," *Earth System Science Data*, vol. 13, no. 5, pp. 2259–2274, Jan. 2021.
- [40] J. Beverley, M. Newman, and A. Hoell, "Rapid development of systematic enso-related seasonal forecast errors," *Geophysical Research Letters*, vol. 50, no. 10, p. e2022GL102249, 2023.
- [41] C. Dalelane, M. Dobrynin, and K. Fröhlich, "Seasonal forecasts of winter temperature improved by higher-order modes of mean sea level pressure variability in the north atlantic sector," *Geophysical Research Letters*, vol. 47, no. 16, p. e2020GL088717, 2020.

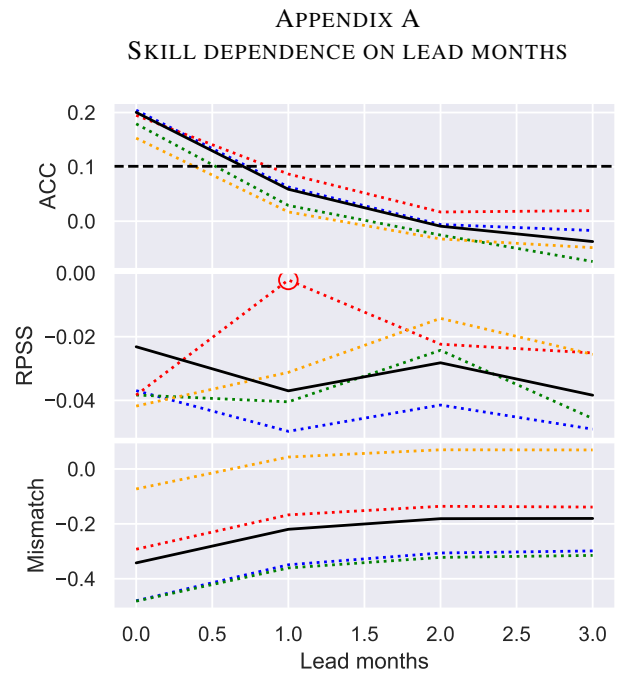


Fig. 4. Skill Metrics plotted against lead months. Different curves represent different skill analysis done over different regions (see colour scheme in the legend of figure 3). The circles highlight the skill scores that are flagged statistically insignificant using the Jackknife method. The dashed horizontal line in the correlation plot (top panel) is the critical threshold above which the values are significant according to the parametric t-test.