Geoscience Data Journal

WILEY

Geoscience Data Journal RMetS

DATA ARTICLE OPEN ACCESS

Operational Convection-Permitting COSMO/ICON Ensemble Predictions at Observation Sites (CIENS)

Sebastian Lerch[1,2,3] | Benedikt Schulz[2] | Reinhold Hess[4] | Annette Möller[5] | Cristina Primo[4] | Sebastian Trepte[4] | Susanne Theis[4]

[1]Department of Mathematics and Computer Science, Philipps-Universität Marburg, Marburg, Germany | [2]Institute of Statistics, Karlsruhe Institute of Technology, Karlsruhe, Germany | [3]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany | [4]Deutscher Wetterdienst, Offenbach, Germany | [5]Faculty of Business Administration and Economics, Bielefeld University, Bielefeld, Germany

Correspondence: Sebastian Lerch (lerch@mathematik.uni-marburg.de)

Received: 29 September 2025 | **Revised:** 8 December 2025 | **Accepted:** 9 December 2025

Keywords: ensemble prediction | post-processing | probabilistic forecasting | station observations

ABSTRACT

We present the CIENS dataset, which contains ensemble weather forecasts from the operational convection-permitting numerical weather prediction model of the German Weather Service. It comprises forecasts for 55 meteorological variables mapped to the locations of synoptic stations, as well as additional spatially aggregated forecasts from surrounding grid points, available for a subset of these variables. Forecasts are available at hourly lead times from 0 to 21 h for two daily model runs initialised at 00 and 12 UTC, covering the period from December 2010 to June 2023. Additionally, the dataset provides station observations for six key variables at 170 locations across Germany: pressure, temperature, hourly precipitation accumulation, wind speed, wind direction, and wind gusts. Since the forecasts are mapped to the observed locations, the data is delivered in a convenient format for analysis. The CIENS dataset complements the growing collection of benchmark datasets for weather and climate modelling. A key distinguishing feature is its long temporal extent, which encompasses multiple updates to the underlying numerical weather prediction model and thus supports investigations into how forecasting methods can account for such changes. In addition to detailing the design and contents of the CIENS dataset, we outline potential applications in ensemble post-processing, forecast verification, and related research areas. A use case focused on ensemble post-processing illustrates the benefits of incorporating the rich set of available model predictors into machine learning-based forecasting models.

Dataset Details:

Identifier: 10.35097/EOvvQEsgILoXpYTK

Creator: Lerch, S., Schulz, B., Hess, R., Möller, A., Primo, C., Trepte, S., and Theis, S.

Dataset correspondence: lerch@mathematik.uni-marburg.de

Title: Operational convection-permitting COSMO/ICON ensemble predictions at observation sites (CIENS)

Data centre: RADAR4KIT

Publisher: KITOpen

Publication year: 2024

Licence: CC BY 4.0

1 | Introduction

Weather forecasts today are typically based on numerical weather prediction (NWP) models, which use systems of partial differential equations to simulate atmospheric processes. By running NWP models with varying initial conditions and/or model physics, ensemble simulations enable probabilistic forecasting. Despite continual advances (Bauer et al. 2015), NWP ensemble predictions often exhibit systematic errors, making post-processing essential for achieving accurate and reliable forecasts. Here, post-processing refers to methods that leverage past NWP forecast-observation pairs to optimally adjust future forecasts. Over the past decades, a wide variety of post-processing methods have been developed, and these now form a critical component of the forecasting workflow in national and international meteorological services (Vannitsem et al. 2018).

A recent focus of post-processing research has been the application of modern machine learning (ML) methods; see Haupt et al. (2021), Vannitsem et al. (2021) for overviews. For example, post-processing approaches based on random forests (Taillardat et al. 2016), gradient boosting (Messner et al. 2017), and neural networks (Rasp and Lerch 2018) have demonstrated promising results across various applications. The rapid development and growing variety of new methods clearly underscore the need for systematic comparisons and rigorous assessments of the advantages and disadvantages of these approaches. Several studies have undertaken such efforts for univariate (Rasp and Lerch 2018; Schulz and Lerch 2022; Demaeyer et al. 2023) and multivariate (Wilks 2015; Perrone et al. 2020; Lerch et al. 2020; Lakatos et al. 2023) post-processing, using both simulated and real-world data. Nonetheless, there remains a critical need for comprehensive and easily accessible real-world benchmark datasets to enable fair quantitative comparisons and facilitate interdisciplinary research efforts by reducing the time-intensive process of data collection and curation (Dueben et al. 2022)

In recent years, numerous benchmark datasets for weather and climate modelling have been released, including datasets for subseasonal and seasonal weather forecasting (Hwang et al. 2019; Lenkoski et al. 2022; Vitart et al. 2022; Mouatadid et al. 2023) and for data-driven weather and climate prediction (Rasp et al. 2020; Watson-Parris et al. 2022; Rasp et al. 2024). Among these, the WeatherBench 2 dataset (Rasp et al. 2024) provides global gridded NWP forecasts and corresponding reanalysis fields, making it a valuable resource for post-processing research as well (see, e.g., Bülte et al. 2025). Additionally, several recent benchmark datasets explicitly identify post-processing as a key application (Haupt et al. 2021; Ashkboos et al. 2022; Kim et al. 2022). Most closely related to our work is the recent EUPPBench dataset (Demaeyer et al. 2023), which was published as part of the activities within the post-processing working group of the European Meteorological Network (EUMETNET). EUPPBench includes 2 years of forecasts and 20 years of corresponding reforecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), along with station observations across Europe.

In this work, we introduce the CIENS dataset, which provides location-specific ensemble forecasts from the operational convection-permitting NWP model of the German Weather Service (Deutscher Wetterdienst, DWD), along with corresponding observations from 170 stations across Germany. The dataset includes ensemble forecasts of the 20 members for 55 meteorological variables mapped to the locations of the stations, as well as additional spatially aggregated forecasts from surrounding grid points available for some variables. Forecasts are available at hourly lead times from 0 to 21 h, for two daily model runs initialised at 00 UTC and 12 UTC. Spanning December 2010 to June 2023, the dataset also includes station observations of six key variables: pressure, temperature, hourly precipitation accumulation, wind speed, wind direction, and wind gusts. Over this time period, the operational NWP model has undergone significant updates, including changes of resolution, ensemble generation mechanisms, and model physics.

To the best of our knowledge, the CIENS dataset is the largest available archive of pre-processed, analysis-ready weather forecast and observation data specifically designed for station-based post-processing in terms of the temporal extent of the provided data. It complements existing datasets such as EUPPBench and enables addressing a wide range of research questions concerning the development and evaluation of post-processing methods. For instance, adapting post-processing methods to accommodate ongoing changes in the NWP model remains a challenge in operational settings (Hess 2020; Lang et al. 2020; Vannitsem et al. 2021; Primo et al. 2024). The extensive time span of operational ensemble forecasts available in the CIENS dataset makes it ideally suited to address this challenge, among many others. Parts of the CIENS dataset have been used in previous research, primarily focused on wind gust forecasting. For example, Hess (2020), Schulz and Lerch (2022) and Primo et al. (2024) compare various statistical and ML-based post-processing methods, Pantillon et al. (2018) and Eisenstein et al. (2022) investigate meteorological aspects of wind gust forecasts during severe storms, and Arnold et al. (2024) leverage forecast and observation data for methodological advancements in forecast evaluation.

The remainder of this article is structured as follows: Section 2 provides a detailed description of the dataset structure, forecasts, and observations included in the CIENS dataset. Section 4 explores potential applications of the dataset in post-processing research and other areas. The article concludes with a discussion in Section 5. Appendix S1 presents an exemplary use case, where ML-based post-processing methods use different sets of input variables for probabilistic wind gust forecasting are compared.

2 | CIENS Dataset

This section provides an overview of the structure and contents of the CIENS dataset.

2.1 | Data Structure

The CIENS data are provided in four parts, see Table 1 for an overview. This simplifies downloading and handling of the large dataset (with a total size of approximately 370 GB) and was necessitated by technical restrictions of the data repository, where the data is hosted (KITOpen, a central repository service at the Karlsruhe Institute of Technology). To support typical uses in the context of ensemble post-processing (see Section 4;

TABLE 1 | Overview of the different parts of the CIENS dataset. The included NWP forecasts are separated between so-called standard and spatial variables. We refer to standard variables as the meteorological variables taken from the closest grid point, while the spatial variables refer to summary statistics of surrounding grid cells.

Dataset	DOI	Content	Size (GB)
CIENS	10.35097/EOvvQEsgILoXpYTK	'Parent' (or primary) dataset which serves as the official reference and links to the four parts listed below	—
CIENS—Run 00 UTC	10.35097/zzfEJPxDILXwSNPH	NWP forecasts (standard variables) of the model runs initialised at 00 UTC and observational data	75.9
CIENS—Run 00 UTC—Spatial Variables	10.35097/wVDXkDCGnBgFuuGt	NWP forecasts (spatial variables) of the model runs initialised at 00 UTC	109.7
CIENS—Run 12 UTC	10.35097/JKALdQqqLIjGUOBC	NWP forecasts (standard variables) of the model runs initialised at 12 UTC	75.3
CIENS—Run 12 UTC—Spatial Variables	10.35097/rJZCZYljpSReTWNL	NWP forecasts (spatial variables) of the model runs initialised at 12 UTC	109.5

and Section A in Appendix S1), we split the data by initialization times of the model runs (at 00 UTC and 12 UTC), and according to the type of meteorological variables from the NWP forecasts. Specifically, we distinguish between ensemble forecasts taken from the grid point closest to a station location, and spatial forecasts which summarise forecasts from the surrounding 11×11 and 21×21 grid points via their mean value and standard deviation (individually for each ensemble member). Note that while observations are available for all hours of the day, the (complete) observational data are only included in the CIENS—Run 00 UTC dataset to avoid unnecessary duplicates.

The four parts of the CIENS dataset are organised in a similar manner, exemplified by the directory structure of the CIENS—Run 00 UTC data shown in Figure 1. The forecast data is provided as daily netCDF files, with corresponding NWP ensemble predictions for all available observation stations and lead times. Corresponding observation data is provided in yearly netCDF files and can be matched to the forecast data using the provided code.

2.2 | NWP Model Forecasts

The CIENS data set includes model forecasts provided by the ensemble prediction system (EPS) of the DWD from 8 December 2010 to 30 June 2023. During that period, different improvements have been made to the NWP system, partly even resulting in different model names: COSMO-DE-EPS, COSMO-D2-EPS and ICON-D2-EPS. Figure 2 shows an overview of the most relevant changes in the NWP model during the time range considered. The mean verification scores for wind gust forecasts from the NWP model shown alongside the model updates indicate that the model updates can have substantial impacts on the quality of the resulting ensemble forecasts.

Forecast data are provided for the 00 UTC and 12 UTC runs of the ensemble forecast systems which were operational at that time, i.e., of COSMO-DE-EPS (Baldauf et al. 2011; Gebhardt et al. 2011), COSMO-D2-EPS (Baldauf et al. 2018), and ICON-D2-EPS (Reinert et al. 2021). In order to generate long time series, data of COSMO-DE-EPS are used from 8 December 2010 00 UTC until 15 May 2018 12 UTC, when the first data of COSMO-D2-EPS became available. Among other updates, this model change included an increase in the horizontal resolution from 2.8 to 2.2 km and an updated orography. Beginning with 10 February 2021, run 12 UTC, forecast data from the current operational ensemble system ICON-D2-EPS are used. The spatial resolution was kept constant for this model change. There have been numerous additional model updates, which are documented in DWD (2016, 2018, 2025). For example, at the time of writing, four model updates have occurred in the year 2024, and the latest update from 9 July 2024 comprises a revision of the wind gust parameterization and modifications to the radar data assimilation processes. Whether such model updates will have substantial impacts on the forecast quality depends on the specific target variable of interest, along with many other factors such as the location or lead time under consideration.

Forecasts in the CIENS dataset are available for hourly lead times from 0 to 21 h, for each of the 20 members of the ensemble models mentioned above.[1] The forecast model data are interpolated to 170 synoptic observation stations within Germany (see Section 2.3). The interpolation is applied separately for each ensemble member and uses data from the nearest model grid point. Furthermore, medium- and large-scale predictors are derived from the model forecasts including the spatial mean and standard deviation of 11×11 and 21×21 model grid points, respectively, around the locations of the synoptic stations, computed separately for each ensemble member. Those spatial variables constitute supplementary information, that might be of interest for some potential applications, see Hess (2020) and Appendix S1 for examples.

Altogether 55 model variables are available, including near surface parameters such as 2 m-temperature and dew point, wind and wind gusts in 10 m height, total precipitation, cloud coverage, radiation, and many more, but also temperature, relative humidity, wind, vertical velocity, and geopotential on 5 pressure levels from 500 hPa up to 1000 hPa are provided. The complete

```
CIENS -- Run 00 UTC
├── observations/ ........ This directory holds observation data as netCDF files organized by year.
│   ├── obs-2010.nc
│   ├── obs-2011.nc
│   ├── ...
│   └── obs-2023.nc
└── run0/ ........ Directory with NWP forecast data organized in sub-folders named by year and month.
    ├── 201012/ ........ Directory with forecasts for a specific month, as netCDF files organized by day.
    │   ├── grib_2010120800.nc ........ Forecasts are available from 8 December 2010.
    │   ├── grib_2010120900.nc
    │   ├── ...
    │   └── grib_2010123100.nc
    ├── 201101/
    │   ├── grib_2011010100.nc
    │   ├── ...
    │   └── grib_2011013100.nc
    ├── ...
    └── 202306/
        ├── grib_2023060100.nc
        ├── ...
        └── grib_2023063000.nc ........ Forecasts are available until 30 June 2023.
```

FIGURE 1 | Directory structure of the CIENS—Run 00 UTC data.

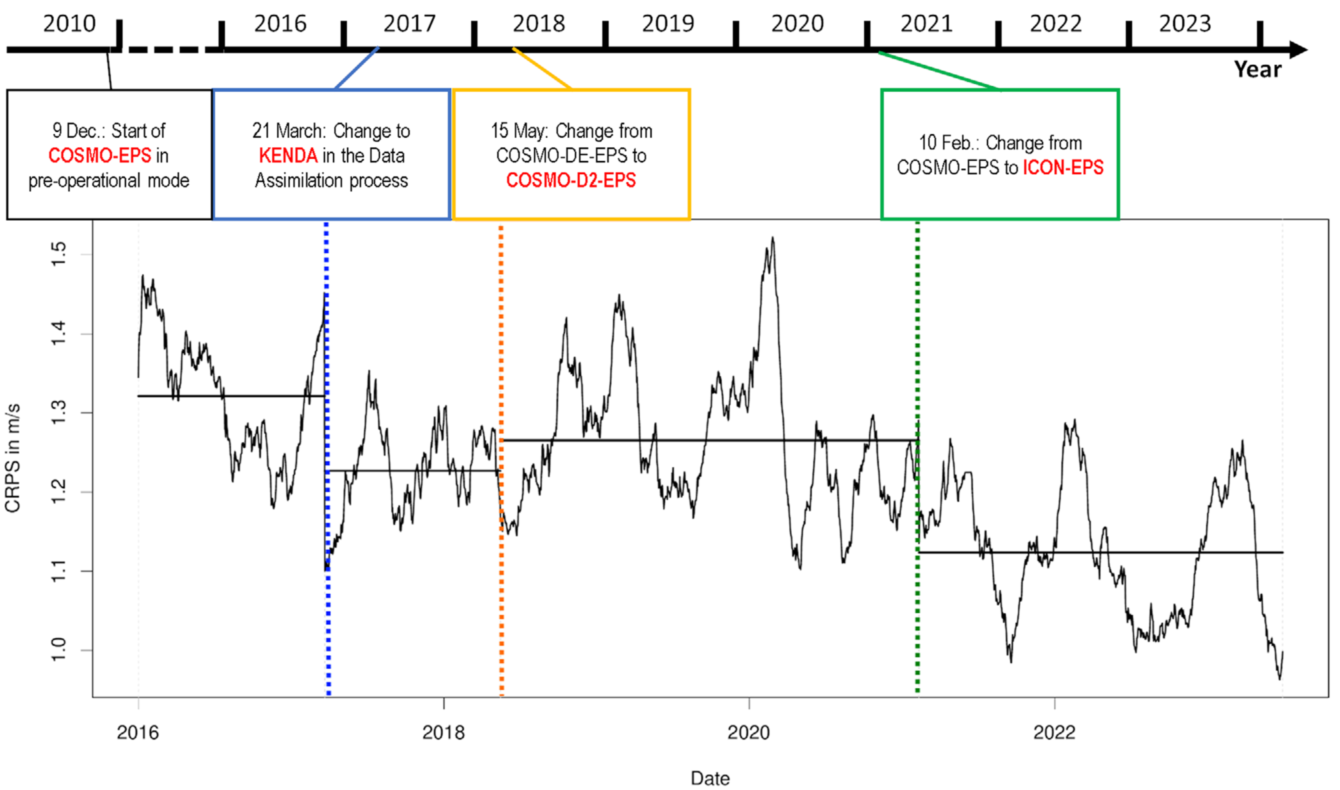

FIGURE 2 | Overview of the most substantial changes in the NWP model underlying the CIENS forecasts indicated by the coloured boxes and lines, along with the temporal evolution of the mean CRPS values (see Section A.2 in Appendix S1) of the raw ensemble forecasts of wind gust with a lead time of 18 h. The CRPS values are averaged over all station locations and smoothed with a 30-day running mean, restricted to the corresponding NWP model version. Horizontal lines indicate the mean CRPS over the corresponding period. The mean CRPS values shown here are restricted to the time period from 2016 until the end of the dataset in June 2023.

list of forecast variables is included in the CIENS repository (Möller et al. 2024).

The zonal and meridional wind components U and V are rotated according to the rotated grid used in the COSMO models. While ICON-D2-EPS uses the same rotated grid as COSMO-D2-EPS, the corresponding U and V forecasts are directed truly geographical and have been rotated according to the rotated grid of the COSMO models to obtain a consistent forecast dataset. Therefore, all U and V forecasts are the zonal and meridional components of the rotated grid of the COSMO model. Since reverting the rotation is not straightforward, this might lead to challenges in comparisons to observations or in using the forecasts in downstream applications.

2.3 | Observations at Station Locations

The CIENS observation data set consists of netCDF files that include 170 European synoptic observation sites distributed within the German domain for the time range from 8 December 2010 to 30 June 2023 (see Figure 3). These station data are part of the synoptic observations distributed via the World Meteorological Observation (WMO) Global Telecommunication System (GTS), available from 2001 onward. The maximum temporal resolution is 1 or 3 h for the standard elements.

The CIENS observation files differ from the original WMO files. An original WMO data file is written in a fixed machine-readable ASCII; however, the CIENS observations are written into netCDF files. WMO files contain all stations and all observation dates for one day with 79 elements. Not all quantities are measured at all stations and may thus be marked with missing values. The standard elements are 2 m-temperature, dew point, precipitation amount, 10 m-wind speed, gust and direction, present weather and cloud cover. In addition, there are cloud heights, sunshine duration, global radiation and many more. However, to enable a large number of observation stations covering all available variables and to minimise missing data and temporal gaps, the CIENS data only includes wind, temperature and precipitation, see Table 2. Note that in contrast to the model predictions of the U and V wind components, the wind direction in the observations is truly geographically directed, which complicates direct comparisons with the forecasts.

The only metadata included is the station identifier (WMO or national identifier), but not the station name nor any geographical information. These can be found in the World Meteorological Organization's official repository named OSCAR/Surface (https://oscar.wmo.int/surface/#/), and have been made available via the Github repository accompanying the dataset.

3 | Dataset Access

The CIENS data are available from the KITOpen repository at https://doi.org/10.35097/EOvvQEsgILoXpYTK under a CC BY 4.0 licence (Schulz et al. 2024). Exemplary code for the R

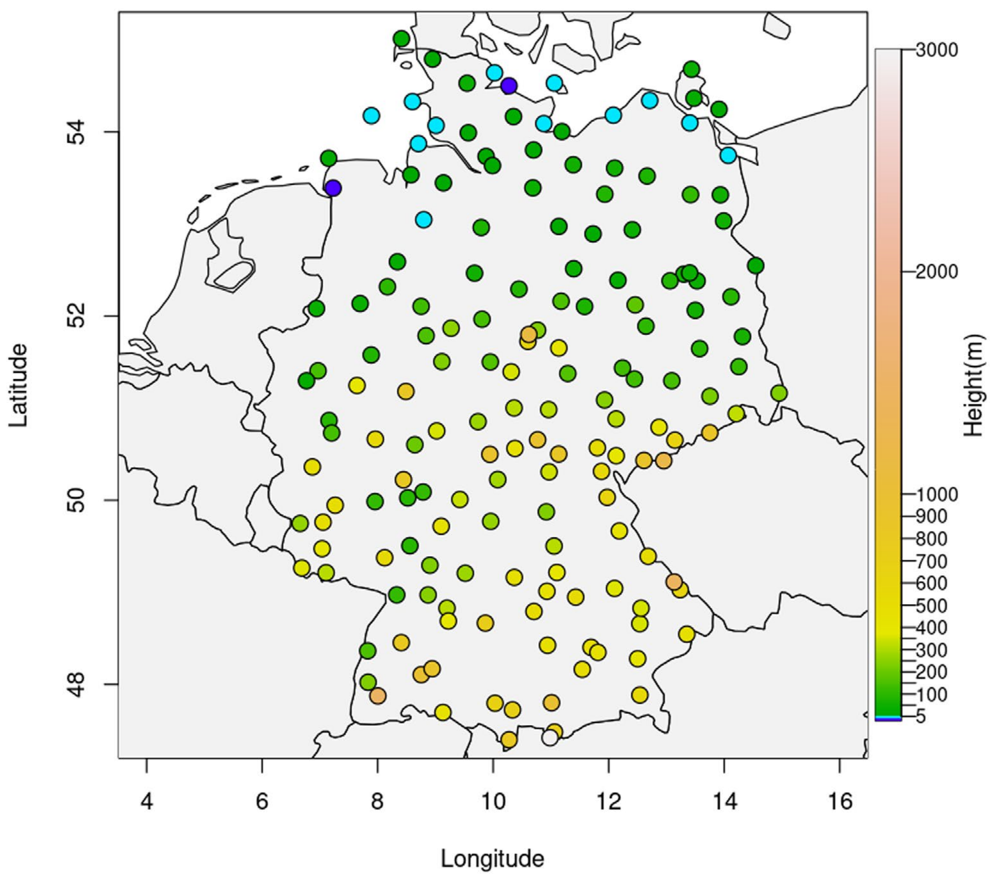

FIGURE 3 | Map of WMO synoptic stations included in CIENS. Colours represent the station altitude (m).

TABLE 2 | Observed variables from the European synoptic stations included in CIENS.

Variable	Name	Unit
wind_speed_of_gust	Wind gusts	m / s
wind_speed	Wind speed	m / s
wind_from_direction	Wind direction	Degree
precipitation_amount	Precipitation amount (hourly)	kg / m^2
air_temperature	Air temperature	K
air_pressure	Air pressure	Pa

programming language and additional documentation, along with all code to reproduce the results and usage examples from Section A in Appendix S1, is available at https://github.com/slerch/CIENS/. Additional location data including the names of the stations, their coordinates, their height and the height of the closest model grid point in the file is also available there.

4 | Potential Uses

The comprehensive collection of forecasts and observations in the CIENS dataset enables researchers to address a wide range of questions. This section aims to present a non-exhaustive list of relevant topics in post-processing research and related fields where the CIENS dataset could serve as a valuable resource.

One primary application of the CIENS dataset in post-processing research is benchmarking both existing and novel methods for the weather variables listed in Table 2 in various settings. In particular, the recent advancements and successes in ML-based post-processing underscore the need for large archives of training data. Given that a key advantage of ML models (and other approaches, such as vine copula-based models, see e.g., Jobst et al. 2023) lies in their ability to effectively leverage information from a wide range of available predictors, the CIENS dataset—with its 55 meteorological variables from the COSMO/ICON model—provides a promising testing ground.

Specifically, it will be interesting to see whether statistical or ML-based post-processing methods can make efficient use of the additionally available spatially aggregated predictions, or whether incorporating information from all ensemble members can provide improvements over methods based on summary statistics alone (Höhlein et al. 2024). Further aspects of model development in post-processing include determining optimal ways to utilise information across multiple lead times (Mlakar et al. 2024) and to effectively combine multiple NWP model runs from different initialization times during model training (Primo et al. 2024). As noted in the introduction, frequent updates to NWP models need to be accounted for by post-processing systems, and thus pose a challenge in operational weather prediction at meteorological services (Vannitsem et al. 2021). Producing a large archive of reforecasts for past dates with an updated model version would be the ideal solution for training post-processing models, but is usually infeasible in terms of the required computational resources in practice, see Hamill (2018) for a detailed discussion. The CIENS dataset, with its extensive archive of operational forecast data encompassing several major updates, allows for detailed investigations of the effects of NWP model changes and the adaptation of post-processing methods, see Section 2.2. Another key research focus in post-processing literature has been on extreme events (e.g., Lerch and Thorarinsdottir 2013; Williams et al. 2014; Pantillon et al. 2018; Friederichs et al. 2018). The large volume of forecast and observation data available for variables such as wind gusts and hourly precipitation accumulation will facilitate comparative studies and targeted model development (Wessel et al. 2024) using the CIENS dataset.

In addition to univariate post-processing of ensemble forecasts for single target variables at specific locations and lead times, many applications require accurate modelling of dependencies across space, time, and variables (Schefzik et al. 2013). Consequently, recent research has increasingly focused on developing multivariate post-processing methods, including new generative ML-based models (Chen et al. 2024; Landry et al. 2025), or vine copula-based methods (Jobst et al. 2024, 2025). The amount of target variables, locations, and lead times in the CIENS dataset provides an opportunity to expand existing benchmarking efforts, particularly through incorporating additional input predictors into multivariate post-processing models.

Beyond ensemble post-processing, the CIENS dataset also supports various other research avenues. For instance, it could serve as a platform for developing new verification methods for probabilistic forecasts. Although substantial progress has been made in both methodology and software tools (for overviews, see e.g., Gneiting and Raftery 2007; Gneiting and Katzfuss 2014; Jordan et al. 2019; Gneiting et al. 2023; Allen 2024), there remains a need for new approaches that address specific challenges such as extremes (Lerch et al. 2017; Allen et al. 2023) and multivariate evaluation (see e.g., Chen et al. 2024, for a discussion from a multivariate post-processing perspective). Additionally, the extensive archive of data allows for a feature-based assessment of forecast quality in both raw and post-processed ensemble predictions; see e.g., Eisenstein et al. (2022) for a study on wind gusts during winter storms. Moreover, the CIENS forecast and observation data could be integrated with other data sources for downstream applications such as hydrological modelling or energy forecasting (potentially in conjunction with post-processing, see Phipps et al. 2022).

In addition to research, the CIENS dataset could serve as a valuable resource for teaching university-level courses in atmospheric sciences, statistics, or computer science and could also be used to run forecasting competitions (Bracher et al. 2024). Finally, the availability of a ready-to-use benchmark dataset alongside open-source software greatly simplifies data collection for student thesis projects.

5 | Discussion and Conclusions

We introduce the CIENS dataset, which encompasses more than 12 years of ensemble predictions from DWD's operational weather prediction model, paired with observations of six meteorological variables at 170 weather stations. The substantial data volume, particularly the wide range of meteorological variables available in the ensemble predictions, makes it a valuable resource for benchmarking existing methods and developing new statistical and ML methods for ensemble post-processing. The dataset is structured to facilitate addressing diverse research questions and allows users to extract relevant subsets with minimal effort. Accompanying code with example pre-processing functionalities and implementations of selected post-processing methods aims at promoting reproducibility and streamlining the dataset's future use.

While a single benchmark dataset cannot capture all aspects relevant to the development of post-processing models, the CIENS dataset offers valuable resources within certain constraints. For example, the operational convection-permitting ensemble prediction system at DWD is limited to forecast lead times of up to 21 h, which may not meet the requirements of all applications, including the growing interest in ML methods for post-processing subseasonal-to-seasonal forecasts (Mayer et al. 2026). Further, the included observation stations were selected to focus on user-relevant variables and to ensure consistent coverage over the dataset's time span with minimal data gaps. This necessitated the exclusion of certain variables, such as solar irradiance and visibility, which have been investigated in recent post-processing research (Schulz et al. 2021; Baran and Lakatos 2023; Horat et al. 2024). Another active area of research involves spatial post-processing methods that utilise two-dimensional gridded forecasts as inputs, often leveraging convolutional neural networks (Grönquist et al. 2021; Veldkamp et al. 2021; Chapman et al. 2022; Li et al. 2022; Horat and Lerch 2024). The CIENS dataset does not include gridded ensemble predictions; instead, in addition to the nearest grid point predictors, it also provides spatial predictors as averages and standard deviations computed over a small set of surrounding grid points for each ensemble member. Consequently, the dataset is less suited for developing spatial post-processing models compared to other available benchmark datasets such as EUPPBench (Demaeyer et al. 2023) or WeatherBench 2 (Rasp et al. 2024). Nevertheless, these two datasets do not include any convection-permitting forecasts so far. Therefore, activities are ongoing to extend EUPPBench by a gridded dataset of COSMO forecasts.

Ultimately, the scientific value of a benchmark dataset is determined by its adoption and use. We believe the CIENS dataset has significant potential as a resource for research projects and teaching across disciplines. Its name reflects this ambition, with the acronym derived from the Latin term *ciens*, which loosely translates to 'to put in motion'.

Acknowledgements

Sebastian Lerch and Benedikt Schulz gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the project C5 'Dynamical feature-based ensemble postprocessing of wind gusts within European winter storms' of the Transregional Collaborative Research Center SFB/TRR 165 'Waves to Weather'. Further, support by DFG Grant Number 395388010 (Sebastian Lerch and Annette Möller) and Grant Number 520017589 (Annette Möller) is gratefully acknowledged. Sebastian Lerch gratefully acknowledges support by the Vector Stiftung through the Young Investigator Group 'Artificial Intelligence for Probabilistic Weather Forecasting'. We further acknowledge support by the German Weather Service (Deutscher Wetterdienst, DWD) through the SPARC-ML project within the extramural research programme, funding reference number 4823EMF01, and open access funding provided by the Open Access Publishing Fund of Philipps-Universität Marburg. We thank Robert Redl for assistance in data handling and the KITOpen team for technical support in making the data publicly available. Further, we thank two anonymous reviewers, whose constructive comments helped to improve an earlier version of this paper. Open Access funding enabled and organized by Projekt DEAL.

Funding

This work was supported by Deutscher Wetterdienst, 4823EMF01, Vector Stiftung, Artificial Intelligence for Probabilistic Weather and Deutsche Forschungsgemeinschaft, 395388010, 520017589, SFB/TRR 165. Philipps-Universität Marburg, Open Access Publishing Fund.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in RADAR4KIT/KITOpe at https://radar.kit.edu/.

Endnotes

[1] Note that before KENDA was introduced in the data assimilation process in March 2021, the 20 members of COSMO-DE-EPS were constructed using five slightly different model configurations applied to the initial and boundary conditions of four different global models. Members 1–5 were based on the IFS model of ECMWF, 6–10 on the formal global model GME of DWD, 11–15 on the GFS (United States National Weather Service) model, and the remaining on GSM (Japan Meteorological Agency).

References

Allen, S. 2024. "Weighted Scoringrules: Emphasizing Particular Outcomes When Evaluating Probabilistic Forecasts." *Journal of Statistical Software* 110: 1–26. https://doi.org/10.18637/jss.v110.i08.

Allen, S., D. Ginsbourger, and J. Ziegel. 2023. "Evaluating Forecasts for High-Impact Events Using Transformed Kernel Scores." *SIAM/ASA Journal on Uncertainty Quantification* 11, no. 3: 906–940. https://doi.org/10.1137/22M1532184.

Arnold, S., G. Gavrilopoulos, B. Schulz, and J. Ziegel. 2024. "Sequential Model Confidence Sets. Preprint." https://arxiv.org/abs/2404.18678.

Ashkboos, S., L. Huang, N. Dryden, et al. 2022. "ENS-10: A Dataset for Post-Processing Ensemble Weather Forecasts." In: Thirty-Sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track. https://openreview.net/forum?id=Z0s5T89qfjc.

Baldauf, M., C. Gebhardt, S. Theis, B. Ritter, and C. Schraff. 2018. "Beschreibung des operationellen Kürzestfristvorhersagemodells COSMOD2 und COSMO-D2-EPS und seiner Ausgabe in die Datenbanken des Deutscher Wetterdienstes (DWD). Technical Report." https://www.dwd.de/DE/leistungen/nwv_cosmo_d2_aenderungen/nwv_cosmo_d2_aenderungen.html.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt. 2011. "Operational Convective-Scale Numerical Weather Prediction With the Cosmo Model: Description and Sensitivities." *Monthly Weather Review* 139: 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1.

Baran, S., and M. Lakatos. 2023. "Statistical Post-Processing of Visibility Ensemble Forecasts." *Meteorological Applications* 30, no. 5: e2157. https://doi.org/10.1002/met.2157.

Bauer, P., A. Thorpe, and G. Brunet. 2015. "The Quiet Revolution of Numerical Weather Prediction." *Nature* 525: 47–55. https://doi.org/10.1038/nature14956.

Bracher, J., N. Koster, F. Krüger, and S. Lerch. 2024. "Learning to Forecast: The Probabilistic Time Series Forecasting Challenge." *American Statistician* 78, no. 1: 115–127. https://doi.org/10.1080/00031305.2023.2199800.

Bülte, C., N. Horat, J. Quinting, and S. Lerch. 2025. "Uncertainty Quantification for Data-Driven Weather Models." *Artificial Intelligence for the Earth Systems*. https://doi.org/10.1175/AIES-D-24-0049.1.

Chapman, W. E., L. D. Monache, S. Alessandrini, et al. 2022. "Probabilistic Predictions From Deterministic Atmospheric River Forecasts With Deep Learning." *Monthly Weather Review* 150, no. 1: 215–234. https://doi.org/10.1175/MWR-D-21-0106.1.

Chen, J., T. Janke, F. Steinke, and S. Lerch. 2024. "Generative Machine Learning Methods for Multivariate Ensemble Postprocessing." *Annals of Applied Statistics* 18, no. 1: 159–183. https://doi.org/10.1214/23-AOAS1784.

Demaeyer, J., J. Bhend, S. Lerch, et al. 2023. "The EUPPBench Postprocessing Benchmark Dataset v1.0." *Earth System Science Data* 15, no. 6: 2635–2653. https://doi.org/10.5194/essd-15-2635-2023.

Dueben, P. D., M. G. Schultz, M. Chantry, D. J. Gagne, D. M. Hall, and A. McGovern. 2022. "Challenges and Benchmark Datasets for Machine Learning in the Atmospheric Sciences: Definition, Status, and Outlook." *Artificial Intelligence for the Earth Systems* 1, no. 3: e210002. https://doi.org/10.1175/AIES-D-21-0002.1.

DWD. 2016. "Modelldokumentation des Regionalmodell Deutschland COSMO-DE (ARCHIV)." https://www.dwd.de/DE/leistungen/nwv_cosmo_de_modelldokumentation/nwv_cosmo_de_modelldokumentation.html.

DWD. 2018. "Modelldokumentation des COSMO-D2 (Regionalmodell Deutschand) (Archiv)." https://www.dwd.de/DE/leistungen/nwv_cosmo_d2_modelldokumentation/nwv_cosmo_d2_modelldokumentation.html.

DWD. 2025. "Icon d2—das regionalmodell." https://www.dwd.de/DE/fachnutzer/forschung_lehre/numerische_wettervorhersage/nwv_aenderungen/_functions/DownloadBox_modellaenderungen/nwv_aenderungen_icon_d2_gesamt.html.

Eisenstein, L., B. Schulz, G. A. Qadir, J. G. Pinto, and P. Knippertz. 2022. "Identification of High-Wind Features Within Extratropical Cyclones Using a Probabilistic Random Forest–Part 1: Method and Case Studies." *Weather and Climate Dynamics* 3, no. 4: 1157–1182. https://doi.org/10.5194/wcd-3-1157-2022.

Friederichs, P., S. Wahl, and S. Buschow. 2018. "Postprocessing for Extreme Events." In *Statistical Postprocessing of Ensemble Forecasts*, 127–154. Elsevier.

Gebhardt, C., S. E. Theis, M. Paulat, and Z. Ben Bouallègue. 2011. "Uncertainties in COSMO-DE Precipitation Forecasts Introduced by Model Perturbations and Variation of Lateral Boundaries." *Atmospheric Research* 100: 168–177. https://doi.org/10.1016/j.atmosres.2010.12.008.

Gneiting, T., and M. Katzfuss. 2014. "Probabilistic Forecasting." *Annual Review of Statistics and Its Application* 1: 125–151. https://doi.org/10.1146/annurevstatistics-062713-085831.

Gneiting, T., and A. E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102, no. 477: 359–378. https://doi.org/10.1198/016214506000001437.

Gneiting, T., D. Wolffram, J. Resin, et al. 2023. "Model Diagnostics and Forecast Evaluation for Quantiles." *Annual Review of Statistics and Its Application* 10, no. 1: 597–621. https://doi.org/10.1146/annurev-statistics-032921-020240.

Grönquist, P., C. Yao, T. Ben-Nun, et al. 2021. "Deep Learning for Post-Processing Ensemble Weather Forecasts." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379, no. 2194: 20200092. https://doi.org/10.1098/rsta.2020.0092.

Hamill, T. M. 2018. "Practical Aspects of Statistical Postprocessing." In *Statistical Postprocessing of Ensemble Forecasts*, 187–217. Elsevier.

Haupt, S. E., W. Chapman, S. V. Adams, et al. 2021. "Towards Implementing Artificial Intelligence Post-Processing in Weather and Climate: Proposed Actions From the Oxford 2019 Workshop." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379, no. 2194: 20200091. https://doi.org/10.1098/rsta.2020.0091.

Hess, R. 2020. "Statistical Postprocessing of Ensemble Forecasts for Severe Weather at Deutscher Wetterdienst." *Nonlinear Processes in Geophysics* 27: 473–487. https://doi.org/10.5194/npg-27-473-2020.

Höhlein, K., B. Schulz, R. Westermann, and S. Lerch. 2024. "Postprocessing of Ensemble Weather Forecasts Using Permutation-Invariant Neural Networks." *Artificial Intelligence for the Earth Systems* 3, no. 1: e230070. https://doi.org/10.1175/AIES-D-23-0070.1.

Horat, N., S. Klerings, and S. Lerch. 2024. "Improving Model Chain Approaches for Probabilistic Solar Energy Forecasting Through Postprocessing and Machine Learning." *Advances in Atmospheric Sciences* 42: 297–312. https://doi.org/10.1007/s00376-024-4219-2.

Horat, N., and S. Lerch. 2024. "Deep Learning for Postprocessing Global Probabilistic Forecasts on Subseasonal Time Scales." *Monthly Weather Review* 152: 667–687. https://doi.org/10.1175/MWR-D-23-0150.1.

Hwang, J., P. Orenstein, J. Cohen, K. Pfeiffer, and L. Mackey. 2019. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19, July. ACM, 2325–2335. "Improving Subseasonal Forecasting in the Western U.S. With Machine Learning." https://doi.org/10.1145/3292500.3330674.

Jobst, D., A. Möller, and J. Groß. 2023. "D-Vine-Copula-Based Postprocessing of Wind Speed Ensemble Forecasts." *Quarterly Journal of the Royal Meteorological Society* 149, no. 755: 2575–2597. https://doi.org/10.1002/qj.4521.

Jobst, D., A. Möller, and J. Groß. 2024. "Gradient-Boosted Generalized Linear Models for Conditional Vine Copulas." *Environmetrics* 35: e2887. https://doi.org/10.1002/env.2887.

Jobst, D., A. Möller, and J. Groß. 2025. "D-Vine Generalized Additive Model Copula-Based Quantile Regression With Application to Ensemble Postprocessing." *Journal of the Royal Statistical Society: Series C: Applied Statistics* 74, no. 4: 994–1020. https://doi.org/10.1093/jrsssc/qlaf011.

Jordan, A., F. Krüger, and S. Lerch. 2019. "Evaluating Probabilistic Forecasts With scoringRules." *Journal of Statistical Software* 90, no. 12: 1–37. https://doi.org/10.18637/jss.v090.i12.

Kim, T., N. Ho, D. Kim, and S. Y. Yun. 2022. "Benchmark Dataset for Precipitation Forecasting by Post-Processing the Numerical Weather Prediction." Working Paper. https://arxiv.org/abs/2206.15241.

Lakatos, M., S. Lerch, S. Hemri, and S. Baran. 2023. "Comparison of Multivariate Post-Processing Methods Using Global ECMWF Ensemble Forecasts." *Quarterly Journal of the Royal Meteorological Society* 149, no. 752: 856–877. https://doi.org/10.1002/qj.4436.

Landry, D., C. Monteleoni, and A. Charantonis. 2025. "Generating Ensembles of Spatially Coherent In Situ Forecasts Using Flow Matching." *Quarterly Journal of the Royal Meteorological Society*: e70055. https://doi.org/10.1002/qj.70055.

Lang, M. N., S. Lerch, G. J. Mayr, T. Simon, R. Stauffer, and A. Zeileis. 2020. "Remember the Past: A Comparison of Time-Adaptive Training Schemes for Non-Homogeneous Regression." *Nonlinear Processes in Geophysics* 27, no. 1: 23–34. https://doi.org/10.5194/npg-27-23-2020.

Lenkoski, A., E. W. Kolstad, and T. L. Thorarinsdottir. 2022. "A Benchmarking Dataset for Seasonal Weather Forecasts." NR-notat. https://nr.brage.unit.no/nr-xmlui/bitstream/handle/11250/2976154/manual.pdf.

Lerch, S., S. Baran, A. Möller, et al. 2020. "Simulation-Based Comparison of Multivariate Ensemble Post-Processing Methods." *Nonlinear Processes in Geophysics* 27, no. 2: 349–371. https://doi.org/10.5194/npg-27-349-2020.

Lerch, S., and T. L. Thorarinsdottir. 2013. "Comparison of Non-Homogeneous Regression Models for Probabilistic Wind Speed Forecasting." *Tellus Series A: Dynamic Meteorology and Oceanography* 65, no. 1: 21206. https://doi.org/10.3402/tellusa.v65i0.21206.

Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting. 2017. "Forecaster's Dilemma: Extreme Events and Forecast Evaluation." *Statistical Science* 32, no. 1: 106–127. https://doi.org/10.1214/16-STS588.

Li, W., B. Pan, J. Xia, and Q. Duan. 2022. "Convolutional Neural Network-Based Statistical Post-Processing of Ensemble Precipitation Forecasts." *Journal of Hydrology* 605: 127301. https://doi.org/10.1016/j.jhydrol.2021.127301.

Mayer, K. J., S. Lerch, C. de Burgh-Day, and M. C. Arcodia. 2026. "Chapter 15—Machine Learning for Subseasonal-to-Seasonal Prediction." In *Sub-Seasonal to Seasonal Prediction (Second Edition)*, edited by A. W. Robertson and F. Vitart, 2nd ed., 539–589. Elsevier.

Messner, J. W., G. J. Mayr, and A. Zeileis. 2017. "Nonhomogeneous Boosting for Predictor Selection in Ensemble Postprocessing." *Monthly Weather Review* 145, no. 1: 137–147. https://doi.org/10.1175/mwr-d-16-0088.1.

Mlakar, P., J. Merše, and J. Faganeli Pucer. 2024. "Ensemble Weather Forecast Post-Processing With a Flexible Probabilistic Neural Network Approach." *Quarterly Journal of the Royal Meteorological Society* 150, no. 764: 4156–4177. https://doi.org/10.1002/qj.4809.

Möller, A., B. Schulz, and S. Lerch. 2024. Replication Material and Code for "Operational Convection-Permitting Cosmo/icon Ensemble Predictions at Observation Sites (ciens)." https://github.com/slerch/CIENS/.

Mouatadid, S., P. Orenstein, G. E. Flaspohler, et al. 2023. "SubseasonalclimateUSA: A Dataset for Subseasonal Forecasting and Benchmarking." Thirty-Seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track.

Pantillon, F., S. Lerch, P. Knippertz, and U. Corsmeier. 2018. "Forecasting Wind Gusts in Winter Storms Using a Calibrated Convection-Permitting Ensemble." *Quarterly Journal of the Royal Meteorological Society* 144, no. 715: 1864–1881. https://doi.org/10.1002/qj.3380.

Perrone, E., I. Schicker, and M. N. Lang. 2020. "A Case Study of Empirical Copula Methods for the Statistical Correction of Forecasts of the ALADIN-LAEF System." *Meteorologische Zeitschrift* 29, no. 4: 277–288. https://doi.org/10.1127/metz/2020/1034.

Phipps, K., S. Lerch, M. Andersson, R. Mikut, V. Hagenmeyer, and N. Ludwig. 2022. "Evaluating Ensemble Post-Processing for Wind Power Forecasts." *Wind Energy* 25, no. 8: 1379–1405. https://doi.org/10.1002/we.2736.

Primo, C., B. Schulz, S. Lerch, and R. Hess. 2024. "Comparison of Model Output Statistics and Neural Networks to Postprocess Wind Gusts." Preprint. https://arxiv.org/abs/2401.11896.

Rasp, S., P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey. 2020. "Weatherbench: A Benchmark Data Set for Data-Driven Weather Forecasting." *Journal of Advances in Modeling Earth Systems* 12, no. 11: e2020MS002203. https://doi.org/10.1029/2020MS002203.

Rasp, S., S. Hoyer, A. Merose, et al. 2024. "WeatherBench 2: A Benchmark for the Next Generation of Data-Driven Global Weather Models." *Journal of Advances in Modeling Earth Systems* 16, no. 6: e2023MS004019. https://doi.org/10.1029/2023MS004019.

Rasp, S., and S. Lerch. 2018. "Neural Networks for Postprocessing Ensemble Weather Forecasts." *Monthly Weather Review* 146, no. 11: 3885–3900. https://doi.org/10.1175/MWR-D-18-0187.1.

Reinert, D., F. Prill, H. Frank, M. Denhard, M. Baldauf, and C. Schraff. 2021. "DWD Database Reference for the Global and Regional Icon and ICON-EPS Forecasting System, DWD." https://www.dwd.de/DWD/forschung/nwv/fepub/icon_database_main.pdf.

Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting. 2013. "Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling." *Statistical Science* 28, no. 4: 616–640. https://doi.org/10.1214/13-STS443.

Schulz, B., M. El Ayari, S. Lerch, and S. Baran. 2021. "Post-Processing Numerical Weather Prediction Ensembles for Probabilistic Solar Irradiance Forecasting." *Solar Energy* 220: 1016–1031. https://doi.org/10.1016/j.solener.2021.03.023.

Schulz, B., and S. Lerch. 2022. "Machine Learning Methods for Postprocessing Ensemble Forecasts of Wind Gusts: A Systematic Comparison." *Monthly Weather Review* 150, no. 1: 235–257. https://doi.org/10.1175/MWR-D-21-0010.1.

Schulz, B., S. Lerch, C. Primo, S. Theis, A. Möller, and R. Hess. 2024. "Dataset: Operational Convection-Permitting COSMO/ICON Ensemble Predictions at Observation Sites (CIENS)." 10.35097/EOvvQEsgILoXpYTK.

Taillardat, M., O. Mestre, M. Zamo, and P. Naveau. 2016. "Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics." *Monthly Weather Review* 144, no. 6: 2375–2393. https://doi.org/10.1175/MWR-D-15-0260.1.

Vannitsem, S., J. B. Bremnes, J. Demaeyer, et al. 2021. "Statistical Postprocessing for Weather Forecasts: Review, Challenges, and Avenues in a Big Data World." *Bulletin of the American Meteorological Society* 102, no. 3: E681–E699. https://doi.org/10.1175/BAMS-D-20-0047.1.

Vannitsem, S., D. S. Wilks, and J. W. Messner. 2018. *Statistical Postprocessing of Ensemble Forecasts*. Elsevier.

Veldkamp, S., K. Whan, S. Dirksen, and M. Schmeits. 2021. "Statistical Postprocessing of Wind Speed Forecasts Using Convolutional Neural Networks." *Monthly Weather Review* 149, no. 4: 1141–1152. https://doi.org/10.1175/MWR-D-20-0219.1.

Vitart, F., A. W. Robertson, A. Spring, et al. 2022. "Outcomes of the Wmo Prize Challenge to Improve Subseasonal to Seasonal Predictions Using Artificial Intelligence." *Bulletin of the American Meteorological Society* 103, no. 12: E2878–E2886. https://doi.org/10.1175/bams-d-22-0046.1.

Watson-Parris, D., Y. Rao, D. Olivié, et al. 2022. "Climatebench v1.0: A Benchmark for Data-Driven Climate Projections." *Journal of Advances in Modeling Earth Systems* 14, no. 10: e2021MS002954. https://doi.org/10.1029/2021MS002954.

Wessel, J. B., C. A. T. Ferro, G. R. Evans, and F. Kwasniok. 2024. "Improving Probabilistic Forecasts of Extreme Wind Speeds by Training Statistical Post-Processing Models With Weighted Scoring Rules." Preprint. https://arxiv.org/abs/2407.15900.

Wilks, D. S. 2015. "Multivariate Ensemble Model Output Statistics Using Empirical Copulas." *Quarterly Journal of the Royal Meteorological Society* 141, no. 688: 945–952. https://doi.org/10.1002/qj.2414.

Williams, R. M., C. A. T. Ferro, and F. Kwasniok. 2014. "A Comparison of Ensemble Post-Processing Methods for Extreme Events." *Quarterly Journal of the Royal Meteorological Society* 140, no. 680: 1112–1120. https://doi.org/10.1002/qj.2198.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information