https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

SgriWind-CN: A Sub-Daily Gridded High-Resolution Near-Surface Wind Speed Dataset for China, Generated by an AI-Driven Integration of In-Situ Observations

Xueru Bai[1], Mingyuan Zhang[1], Jinlin Zha[1, *], Anning Huang[2], Xinsheng Zhu[3], Jian Tang[4], Deming Zhao[5], and Jian Wu[1]

[1]Key Laboratory of Atmospheric Environment and Processes in the Boundary Layer over the Low−Latitude Plateau Region, Department of Atmospheric Science, Yunnan University, Kunming, China

[2]School of Atmospheric Sciences, Nanjing University, Nanjing, China

[3]Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of the People's Republic of China, Nanjing, China

[4]National Meteorological Center of China Meteorological Administration, Beijing, China

[5]Key Laboratory of Regional Climate and Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Correspondence to: Jinlin Zha (zhajl@ynu.edu.cn)

Abstract. Reliable near-surface wind speed (NSWS) data underpin a wide range of research in hydrology, climatology, wind energy, and environmental change. However, estimates of long-term change in NSWS remain highly uncertain, largely because of incomplete temporal coverage, uneven station distribution, and inhomogeneities in observational records. Here, we release the SgriWind-CN dataset, a new sub-daily gridded NSWS dataset over China, generated from 6-hourly observations collected at 2,384 meteorological stations during 1980–2023. After rigorous quality control and sub-daily homogenization, we apply a deep learning model, PConvUNet, to reconstruct a sub-daily, gridded wind speed dataset. The resulting product effectively reproduces the spatial distribution of NSWS and performs well in data-sparse areas. Quantitative evaluation shows that PConvUNet substantially outperforms conventional interpolation methods: it best preserves the original statistical distribution, as reflected in the lowest Kullback–Leibler divergence, and reduces the mean absolute error on the test set by at least 65.3% relative to alternative approaches. The model also exhibits an advantage in capturing extreme wind events. A reverse interpolation validation test further confirms the product's robustness, yielding a national mean bias of less than 0.05 $m \cdot s^{-1}$. Beyond NSWS, the dataset also includes surface roughness, land cover type, elevation, latitude, and longitude to support diverse user applications and independent evaluation. This study aims to provide a robust, high-resolution data foundation for advancing understanding of wind variability and its broader impacts across Earth system sciences. SgriWind-CN is publicly available at https://doi.org/10.5281/zenodo.22812484 (Bai et al., 2026).

Keywords: Near-surface wind speed, Homogenization, Sub-daily, PConvUNet, China

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

1 Introduction

Near-surface wind speed (NSWS) is a fundamental variable that characterizes atmospheric dynamics and climate change, and its variability is closely tied to boundary-layer physics, land-surface processes, and ecological functioning (He et al., 2010; Chang et al., 2012; Wu et al., 2018). By modulating mass and energy exchanges in the lower atmosphere, NSWS strongly influences aerosol and dust dispersion and transport, affecting visibility, air quality, and regional weather conditions (Chen et al., 2020; Zhang et al., 2020; Ashkenazy and Yizhaq, 2023). In the context of hydrological cycles, NSWS plays a critical role in regulating evapotranspiration at both global and regional scales, with direct implications for the occurrence and severity of drought events (McVicar et al., 2012; Wang and Dickinson, 2012; Liu et al., 2014; Yang et al., 2022; Wang et al., 2025). As a key renewable energy source, NSWS is foundational to wind resource assessment, optimal turbine siting, and efficient energy exploitation (Karnauskas et al., 2018; Zeng et al., 2019). Given its broad relevance, NSWS data have been widely applied in diverse fields, including extreme wind event assessment, wind power generation estimation, air pollutant dispersion modelling, and evapotranspiration quantification (Dupré et al., 2020; Tuinenburg et al., 2020; Dayal et al., 2021; Bai and Wang, 2023). These applications impose increasingly stringent requirements on data accuracy, continuity, and spatiotemporal resolution, underscoring the need for a reliable, high-quality NSWS dataset with fine spatiotemporal resolution.

One of the most notable features of NSWS trends is the long-term decline over the six decades, commonly referred to as "terrestrial stilling" (Vautard et al., 2010; McVicar et al., 2012; Azorin-Molina et al., 2014). A similar decreasing trend has been documented across China (Lin et al., 2013; Wu et al., 2017), but growing evidence suggests that this declining trend may have reversed in recent decades (Li, Y et al., 2018; Zha et al., 2021; Zhang and Wang, 2020). For instance, Yang et al. (2021) found that NSWS reversed over China around 2014, shifting from a significant decline to a considerable recovery. Liu et al. (2024) applied higher weights to sparsely distributed stations and reported a 31% larger increasing trend compared with the unweighted estimates. In contrast, Li et al. (2022) suggested the annual mean NSWS showed a continuous reduction over China, with no statistically significant recovery. Overall, the long-term trend in NSWS, especially the recent recovery, remains highly uncertain and lacks consensus. This uncertainty stems from multiple factors, including differences in station network density, spatial weighting strategies, and, most critically, inhomogeneities introduced by non-climate factors.

NSWS is highly sensitive to changes in surface roughness and surrounding built environments. Non-climatic factors, such as instrument replacement and station relocation, can introduce numerous inhomogeneity breakpoints (IBs) into the wind speed series. In China, more than 600 stations have been relocated (Cheng et al., 2024). An analysis of daily mean NSWS revealed that over 59% of stations are affected by observation automation and station relocation during 1970–2017 (Si et al., 2018; Zhang and Wang, 2023). A recent study pointed out that IBs linked to non-climate factors can induce artificial trends in NSWS (Yan et al., 2026). Hence, rigorous homogenization correction of raw observations is a prerequisite for generating robust findings and reliable interdisciplinary research.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



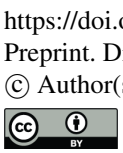
Beyond the issue of inhomogeneity, the limited capacity of existing daily or monthly mean NSWS datasets to capture regional-scale spatiotemporal patterns poses another major challenge. Station observations are temporally discontinuous and spatially heterogeneous, with poor representativeness in regions of sparse network coverage, making it difficult to depict the full spatiotemporal variability of NSWS (Liu et al., 2019; Liu et al., 2024). Gridded reanalysis products offer spatially and temporally consistent fields based on numerical models, yet they are subject to uncertainties arising from input data quality. Furthermore, reanalyses tend to smooth out smaller-scale observational variability and fail to represent wind speed variations over complex terrain (Rose and Apt, 2016; Wohland et al., 2019; Hersbach et al., 2020). Given these limitations, the development of high-resolution gridded NSWS data based on observations is of great value.

In recent years, considerable efforts have been devoted to the gridded reconstruction of surface observational data (Han et al., 2023; Zhang et al., 2024; Wang et al., 2026). Existing approaches broadly fall into two categories: spatial interpolation based solely on station observations, and data fusion that incorporates background fields from reanalysis or numerical models. Classical interpolation methods, such as inverse distance weighting (Shepard, 1968) and Kriging (Krige, 1951), rely on spatial autocorrelation and distance–decay relationships to estimate values at unobserved locations. Data fusion techniques, including optimal interpolation (Gandin, 1965) and Bayesian model averaging (Raftery et al., 2005), combine station data with background fields to produce spatially continuous products. However, these methods often introduce large errors in regions with complex terrain and sparse stations and make them difficult to use (Bishop et al., 2001; Hofstra et al., 2008; Baez-Villanueva et al., 2020).

Accompanied by the rapid advancement of artificial intelligence (AI), machine learning algorithms and big data analysis technologies are expected to be applied in data reconstruction over the station-sparse regions. These include conventional machine learning models such as random forests (Zhou et al., 2026) and support vector machines (Zhang et al., 2017), as well as generative models such as generative adversarial networks (Zhu et al., 2020) and diffusion models (Sun et al., 2025). Some studies have demonstrated that AI-based methods offer advantages in fusing multi-source information and modelling nonlinear relationships, showing promise for wind speed prediction and gridded reconstruction in regions with complex terrain and sparse stations (Houndekindo and Ouarda, 2025; Long et al., 2025). Building on these methods, some researchers have made progress in gridded NSWS reconstruction (Zhou, L. et al., 2022; Wang et al., 2026). However, the existing studies focus on daily or monthly mean NSWS, and sub-daily gridded homogenized NSWS datasets are unavailable. Given the strong diurnal cycle of NSWS, daily or monthly averages smooth out key information, inevitably obscuring sub-daily variability, thereby hindering refined analyses in wind-related research and applications. There is thus a clear and urgent need for reliable sub-daily gridded NSWS data.

To address the aforementioned challenges, we have developed SgriWind-CN, a sub-daily gridded, homogenized, high-resolution NSWS dataset for China. Specifically, we compiled 6-hourly wind observations at 10 m height from 2,384 stations during 1980–2023 and performed comprehensive homogenization testing for these raw observations. We then constructed a partial convolution-based deep learning model to establish a gridded NSWS dataset. Finally, we conducted a thorough reliability evaluation of the dataset.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

2 Datasets

The 6-hourly (02:00, 08:00, 14:00, and 20:00 BJT) wind speed observations at 10 m height used in this study are obtained from the China Meteorological Administration (CMA). These data span the period 1980–2023 and comprise records from 2,384 stations (Fig. 1). In addition, 6-hourly air temperature and pressure observations are also obtained from the CMA, and sea level pressure is computed for each station using the pressure-altitude formula and station elevation data. To construct the gridded observational dataset, hourly 10-m wind speed data from the fifth-generation ECMWF reanalysis (ERA5) (Hersbach et al., 2023) at 0.25°×0.25° resolution are used for model training and testing. The ERA5 data cover the period 1950–2024 and are cropped to the domain of 18.00–53.75 °N and 72.00–137.75 °E, yielding a grid of 144 × 264. To validate the generalization ability of the model, 6-hourly NCEP FNL Operational Global Analysis data (National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2015) are used as an independent test set. As NCEP FNL is separate from the ERA5 data used to train the model, it provides a test of generalization across data sources. It is worth noting that the spatial resolution of NCEP FNL data with a long time scale is only available with a coarse resolution of 1°×1°, whereas a high resolution of 0.25°×0.25° is only available from 2016 to 2025. To match the spatial resolution of the training set, we therefore select the NCEP FNL data with 0.25°×0.25° resolution. Elevation data are obtained from the Copernicus GLO-30 Digital Elevation Model (European Space Agency, 2024). Based on these data, we derive slope and calculate the surface roughness as $1/\cos(\text{slope})$ (Zhang et al., 2021). The land cover type data, developed by Yang and Xin (2021), is used; it contains nine major land cover types in China from 1990 to 2019.

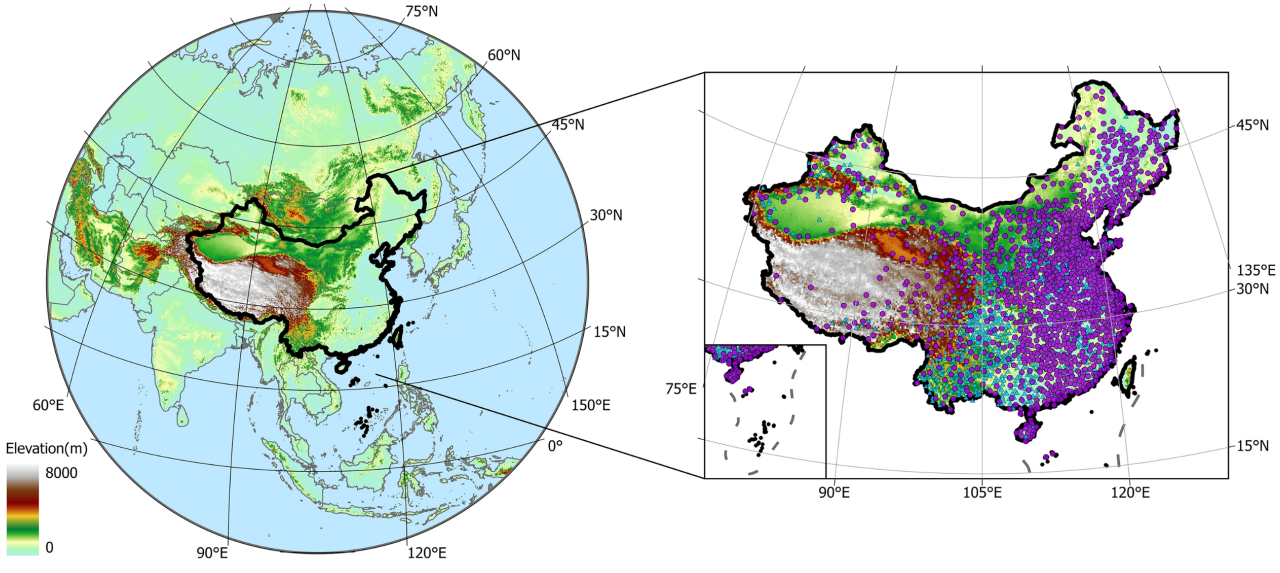


Figure 1. Spatial patterns of stations used in this study. Triangles and circles indicate the locations of all stations. Blue triangles represent excluded stations, while purple circles represent retained stations.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



3 Methods

3.1 Quality control and homogenization

The quality control (QC) and homogenization procedure is divided into three steps: (1) QC; (2) calculation of the reference sequence; and (3) detection and correction of IBs (Fig. S1).

(1) QC

The 6-hourly wind speed observations are required to meet the following QC criteria: (a) stations with an annual missing data rate exceeding 10% are excluded; and (b) for each station, at least three observation times per month are required to have fewer than 11 missing days and no consecutive missing data exceeding 5 days. After applying these criteria, 1,498 stations are retained (purple dots in Fig. 1). The calculated sea level pressure is processed following the same screening procedure.

(2) Calculation of the reference sequence

Because of the strong spatiotemporal variability in NSWS, the traditional approach of using nearby stations as reference sequences for the target station is not applicable. Geostrophic wind is mainly determined by large-scale pressure fields, and compared with wind observations, pressure is less affected by local environments (Alexandersson, 2001). Therefore, geostrophic wind sequences calculated from pressure can better reflect the long-term trends of wind speed (Wan et al., 2010; Zhou, C. et al., 2022). Geostrophic wind is calculated using sea level pressure and latitude-longitude information of triangular regions formed by three stations. Moreover, considering the uncertainty associated with using only three stations, referring to Zhang and Wang (2020), we select all available stations within a 3° radius of each target station and compute geostrophic wind from multiple three-station combinations. Subsequently, the geostrophic wind exhibiting the highest correlation with the observed NSWS at the target station is selected as the reference series. To ensure the robustness of the reference sequence, each station is required to have at least 20 triangles; otherwise, no reference sequence is computed for that station. Finally, 1,433 stations satisfy this criterion and retain a reference sequence.

(3) Detection and correction of IBs

We employ two homogenization methods: the Penalized Maximal F test (PMF) (Wang, 2008a, b) and the Penalized Maximal *t* test (PMT) (Wang et al., 2007; Wang, 2008a). Both methods account for first-order autocorrelation in the time series and apply empirical penalty functions to correct the uneven distribution of false-alarm rate and detection power. Furthermore, we employ the quantile-matching (QM) adjustment (Wang et al., 2010) to correct the mean bias and the distributional differences across the IBs. Based on the reference sequences constructed in step (2), the PMT is applied to the stations with a reference sequence, and the PMF is applied to the remaining stations. Because metadata record files are unavailable, we identify IBs as reliably as possible using two criteria: (1) IBs associated with a mean change of less than 0.1 in the normalized series are excluded (Cao et al., 2016; Li, Z. et al., 2018), as such a small change is insufficient to indicate an inhomogeneity; (2) Because IBs at the same station should be consistent across the four observation times, it is necessary to ensure that the breakpoints at least three observation times remain consistent within a window of six months before and

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



after. Subsequently, based on the identified IBs, the raw wind speed series is adjusted using the QM technique to obtain the homogenized wind speed series.

3.2 Gridding procedure based on PConvUNet deep learning model

Partial convolution (PConv) is initially introduced for image restoration and is effective in reconstructing high-fidelity images under extensive missing-data conditions (Liu et al. 2018). U-shaped neural network (UNet) is a classical encoder-decoder architecture whose defining feature is the use of skip connections, which directly bridge corresponding encoder and decoder layers by concatenating their feature maps, thereby preserving spatial detail and facilitating precise reconstruction of fine-grained structures (Ronneberger et al., 2015). The combination of PConv and UNet yields the PConvUNet architecture, which has shown robust performance in data interpolation and reconstruction, particularly in the domains characterized by sparse observational coverage (Kadow et al., 2020). This hybrid framework has been applied to reconstruct multiple environmental variables, including surface air temperature (Kadow et al., 2020), wind speed (Zhou, L. et al., 2022), surface solar radiation (Jiao et al., 2023), soil moisture (Wei et al., 2024), and sea ice thickness (Ma et al., 2025), consistently producing accurate and physically plausible results. To obtain a high-quality gridded NSWS dataset, we use the PConvUNet deep learning model to convert the homogenized discrete stations data into a gridded dataset at a resolution of 0.25°×0.25°, thereby better meeting the requirements of related research fields (Fig. 2a). Given the performance of the ERA5 reanalysis product in representing spatiotemporal variations in atmospheric circulation and wind fields (Ramon et al., 2019; Hersbach et al., 2020; Zhang and Wang, 2020), we further use its hourly 10 m wind speed at a resolution of 0.25°×0.25° to train the PConvUNet model.

3.2.1 Data preprocessing

The ERA5 and homogenized NSWS data are preprocessed following the procedure described in Section 2.1. The resulting gridded ERA5 wind fields serve as reference fields for generating the PConvUNet training samples. To construct an independent test set, we randomly select one year from each decade of the full time series, yielding eight test years, including 1953, 1967, 1974, 1982, 1991, 2002, 2014, and 2024; these years are withheld and excluded from both the training and validation sets. The hourly data from the remaining 67 years are partitioned into training and validation sets at an 11:1 ratio, yielding 538,406 and 48,946 image samples, respectively. The test set uses the same binary mask as the training set, with regions lacking observations set to 0 and those with observations set to 1.

3.2.2 PConvUNet reconstruction

The data used for gridding are the homogenized NSWS observations at stations. To improve training stability, we first standardize the input data and apply Min-Max normalization to map the data to the [0,1] interval, thereby mitigating the influence of outliers on gradient-based optimization. Spatial interpolation onto the target grid (144×264) is performed using inverse distance weighting, with the interpolation radius constrained to preserve local fidelity and minimize smoothing

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



artifacts induced by aggregation. The preprocessed gridded data, treated as single-channel grayscale images, are then fed into the PConvUNet architecture. Leveraging partial convolutions and dynamic mask updates, the model performs pixel-wise reconstruction of wind speed fields, yielding spatially coherent, gap-free datasets. The partial convolution operation at each spatial location is given in Equation (1).

$$x' = \begin{cases} \boldsymbol{W}^T(\boldsymbol{X} \odot \boldsymbol{M}) \dfrac{\text{sum}(1)}{\text{sum}(\boldsymbol{M})} + b, & \text{if sum}(\boldsymbol{M}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Where $\boldsymbol{W}$ is the weight of the convolution filter, b represents the corresponding bias term, $\boldsymbol{X}$ represents the feature values of the convolution window, $\boldsymbol{M}$ is the corresponding binary mask, $\odot$ indicates the element-wise multiplication, and 1 is a matrix of the same shape as $\boldsymbol{M}$ but with all elements being 1. The output value only depends on the unmasked input. To ensure consistent output magnitude across convolutional windows containing varying numbers of valid inputs, PConvUNet applies the scaling factor defined as $\frac{\text{sum}(1)}{\text{sum}(\boldsymbol{M})}$ to scale. Where $\text{sum}(1)$ is the total number of elements in the window, and $\text{sum}(\boldsymbol{M})$ denotes the count of unmasked positions. Following each partial convolution operation, the binary mask is updated: a given output location is marked as valid if and only if at least one unmasked input contributes to its computation. This mask propagation rule is formalized in Equation (2).

$$m' = \begin{cases} 1, & \text{if sum}(\boldsymbol{M}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The above processing approach enables the model to have a distinct capability to construct contiguous missing regions. As long as the input contains any valid pixels, by applying the partial convolution layer continuously for a sufficient number of times, the mask will eventually become a valid value of 1.

The total loss ($\mathcal{L}_{total}$) function is designed by weighting multiple loss functions, including pixel valid region loss ($\mathcal{L}_{valid}$), pixel hole region loss ($\mathcal{L}_{hole}$), perceptual loss ($\mathcal{L}_{\text{perceptual}}$), original style output loss ($\mathcal{L}_{\text{styleout}}$), synthetic output loss ($\mathcal{L}_{\text{stylecomp}}$), and total variation loss ($\mathcal{L}_{\text{tv}}$). It is formulated as follows:

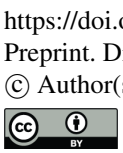
$$\mathcal{L}_{total} = \mathcal{L}_{valid} + 10\mathcal{L}_{hole} + 0.05\mathcal{L}_{perceptual} + 16\left(\mathcal{L}_{style_{out}} + \mathcal{L}_{style_{comp}}\right) + 0.1\mathcal{L}_{tv} \quad (3)$$

Training spans 120 hours and comprises 50,000 optimization steps. The Adam optimizer is employed with an initial learning rate of 6×10^{-4}. The learning rate decay factor is set to 0.3, and decay is triggered at 70% (35,000 iterations) and 90% (45,000 iterations) of the total number of iterations. Detailed processes are shown in Fig. 2b.

3.2.3 Validation Methods

To assess the reliability of different masking strategies, we implement two experimental protocols: (1) random masking scheme and (2) fixed masking scheme. In the random masking scheme, missing regions are stochastically generated during both training and testing according to a predefined masking ratio; consequently, the locations, geometries, and spatial extents of masked grids vary across samples. This scheme emulates the stochastic nature of real-world spatial data gaps and serves

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



to evaluate the model's adaptability, generalization capability, and robustness to unseen missing patterns. In contrast, the fixed masking scheme employs a static binary mask derived from the empirical spatial distribution of observational gaps, i.e., grid cells consistently unobserved across the historical record are designated as permanently masked. The identical mask is applied uniformly to both the training and testing datasets. This design more faithfully reproduces the data deficiencies inherent in observation networks and is therefore better suited for quantifying reconstruction accuracy under realistic, stationary missing data conditions.

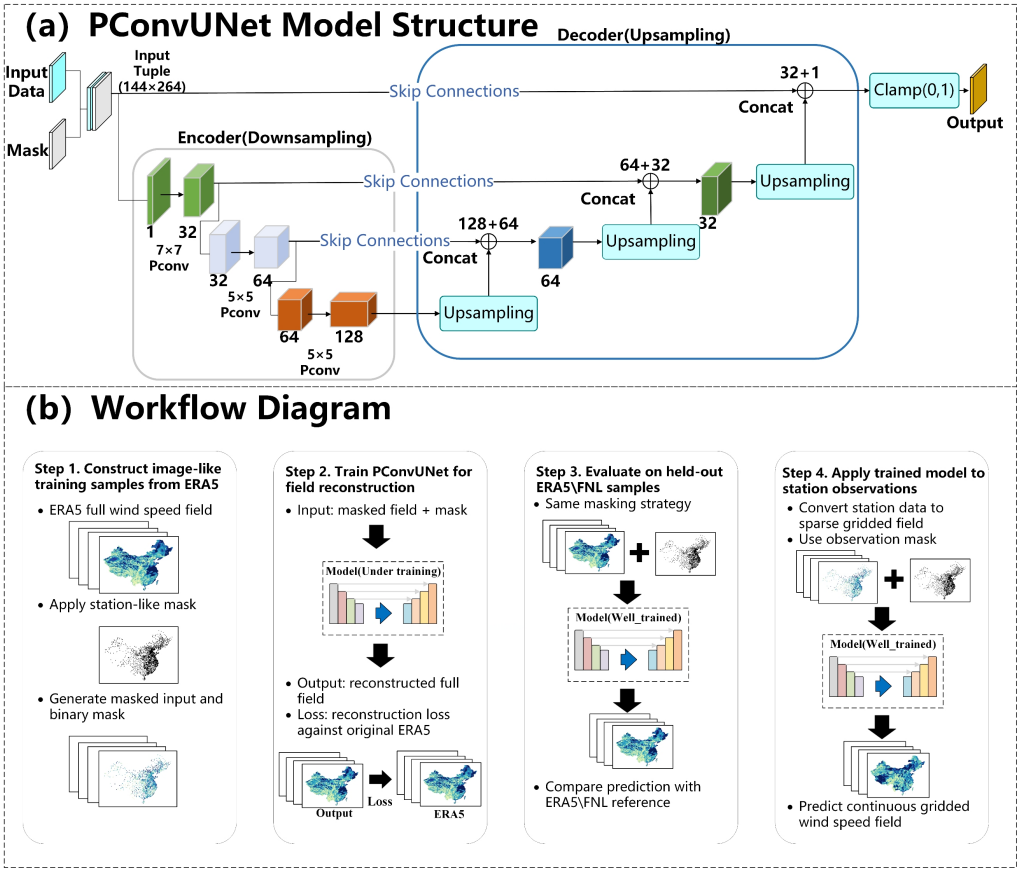


Figure 2. Architectural diagram of the PConvUNet deep learning model. This model is based on the UNet architecture, with standard convolutional layers systematically replaced by partial convolutional (PConv) layers. The encoder and decoder each comprise three hierarchical stages; within the decoder, feature map upsampling is performed via nearest-neighbor interpolation.

3.3 Other methods

The spatially constrained multivariate clustering method is used to divide the study region into 7 sub-regions (Assuncao et al., 2006; Chuan et al., 2024). Linear trend is calculated using least-squares regression, and the statistical significance is evaluated using a two-tailed Student's *t*-test. A piecewise linear regression algorithm is applied to identify a significant turning point of the long-term trend within a given time sequence.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

220 **4 Results**

4.1 Quality control and homogenization of raw wind speed observations

The raw NSWS observations were characterized by substantial missing values and anomalously low wind speeds (Fig. 3). Data completeness varied considerably among the four observation times. Missing data were most severe at 02:00 BJT, when only 30% of the stations had a missing rate below 0.05% (Fig. 3 a1). At the other observation times (08:00, 14:00, and
225 20:00 BJT), by contrast, 98% of stations had a missing rate below 0.05% (Fig. 3, a2–a4). Regionally, Southwest China exhibited the most pronounced data missing, with a mean missing rate exceeding 34.0%, and the longest consecutive missing period in this region exceeded 10 years. The anomalously low wind speeds were manifested primarily as anomalous zero-wind-speed records, which occurred mainly at 02:00 and 08:00 BJT, with the longest consecutive zero-value period exceeding 300 days. These zero-wind-speed records were likewise concentrated in Southwest China, whereas the coastal
230 areas and northeastern China exhibited relatively few such records (Fig. 3, b1–d4).

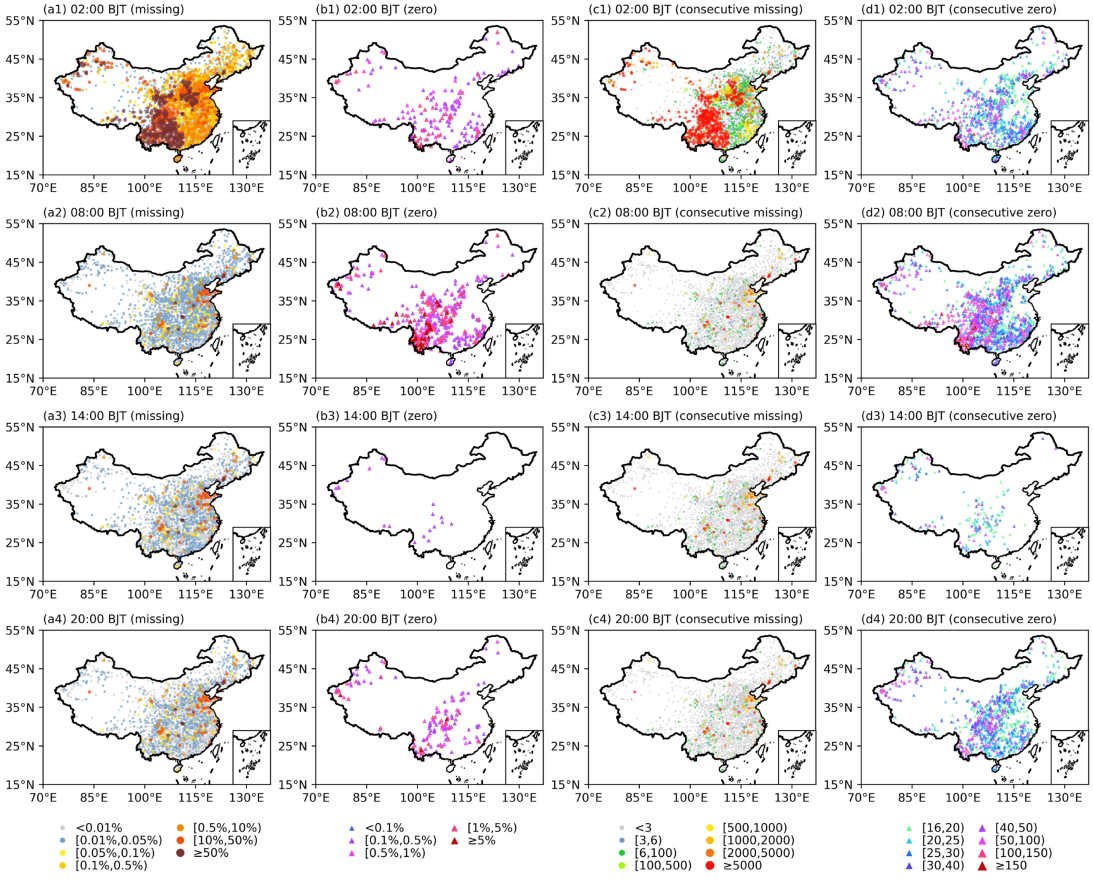


Figure 3. Spatial patterns of raw observation NSWS quality from 1980 to 2023. (a1–a4) Mean missing rate of stations. (b1–b4) Proportion of calendar months in which all NSWS records are zero. (c1–c4) Maximum number of consecutive days with missing NSWS observations. (d1–d4) Longest consecutive days with NSWS zero-value measurements.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

Following the method described in Section 3.1, quality control was applied to the raw station data. Despite excluding 886 stations with frequent or consecutive missing data, the remaining dataset still exhibited inhomogeneities. Ten randomly selected examples displayed varying degrees of inhomogeneities, expressed mainly as abrupt increases or decreases in the mean NSWS after IBs, as well as significant changes in NSWS variability (Fig. 4). For instance, at the station located at 31.17° N, 112.57° E, two apparent abrupt increases were detected around 1992 and 2015, with mean shifts exceeding 1.0 m s^{-1}. At the station located at 43.87°N, 126.53°E, an abrupt change was detected in 2012, with a mean shift even exceeding 3.0 m s^{-1}. Such abrupt changes were fairly common across observation stations. Approximately 40% of stations had at least one analogous IB, with 2,776 breakpoints identified in total. Obviously, the sudden change in NSWS and its prolonged persistence are highly questionable. Hence, the direct use of raw NSWS data could introduce significant uncertainties into the results.



Figure 4. Temporal changes of raw wind speed observations from randomly selected 10 observed stations.

After quality control and homogenization, the long-term changes of the NSWS are shown in Figure 5. Homogenization mainly adjusted the trend of the raw wind speed. At all observation times, the mean value of homogenized NSWS

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

(Hom_NSWS) was higher than that of the raw observation. The mean wind speed increased by 0.2 m s^{-1} after homogenization, with the largest increase found in period P1 (approximately 1980–1990), reaching 0.3 m s^{-1}. The long-term variation in the raw observation was characterized by three distinct phases, identified using a piecewise linear regression algorithm: a significant decline before approximately 1990 (P1), a slowed decline during 1990–2013 (P2), and a significant increase during 2013–2023 (P3). However, when the same turning points derived from the raw observations were applied to the Hom_NSWS, a markedly different trend was observed. During the P1 period, the Hom_NSWS and raw wind speeds were nearly identical, with both trends ≤ −0.22 m s^{-1} $decade^{-1}$ (p<0.05). During the P2 period, the Hom_NSWS showed a stronger decline than raw observations, by about 0.07 m s^{-1} $decade^{-1}$ on average. During the P3 period, the homogenized and raw observations showed trends opposite in sign, and the Hom_NSWS didn't exhibit an increasing trend. The absence of a recent recovery in Hom_NSWS is broadly consistent with the findings of Yan et al. (2026), which highlighted the sensitivity of recent wind speed trends to inhomogeneity records, while differences in study periods, regions, and homogenization methods may lead to some variation in the estimated trend. Overall, homogenization effectively corrected the biases in mean values and the trend of wind speed from inhomogeneities in the observational series. In particular, around 2010, homogenization weakened the increasing trend in the raw observation.

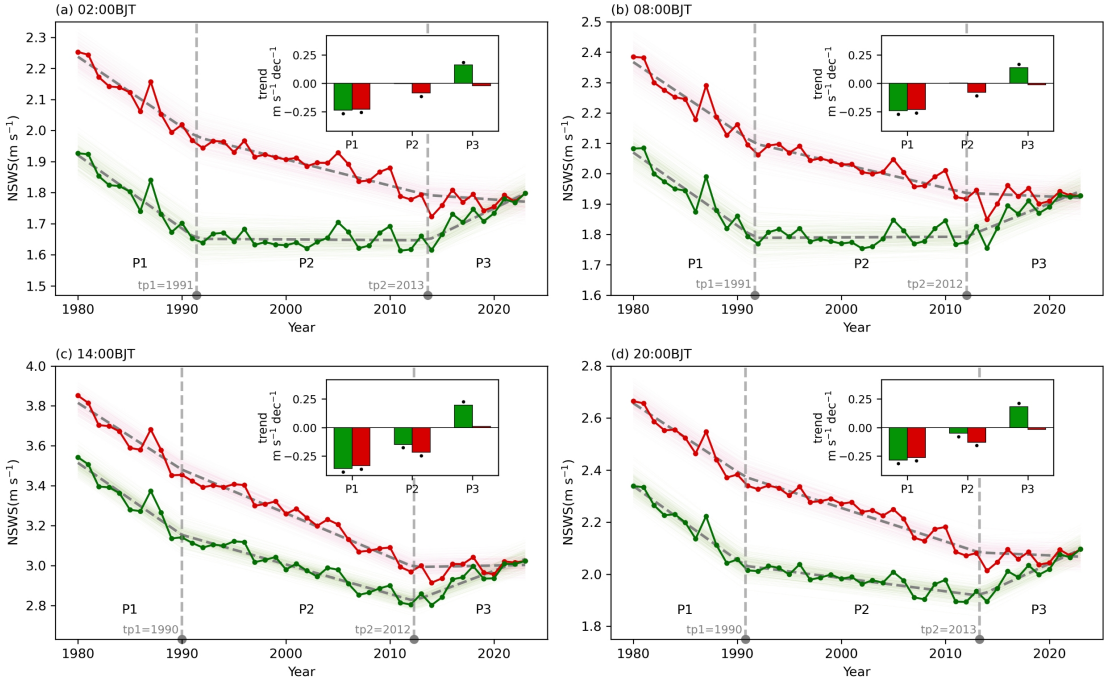


Figure 5. Comparison of Hom_NSWS (red) and raw NSWS (green) from 1980 to 2023. The grey dashed line denotes the piecewise linear regression fit, and the thin solid line represents the piecewise linear regression fit based on the mean NSWS from a randomly selected 40% of the stations (n=300). tp1 and tp2 indicate the statistically identified breakpoints derived from the piecewise linear regression analysis. The insets show the linear trends for the three periods separated by tp1 and tp2. Dots indicate trends that are significant at the 95% confidence level according to a two-tailed Student's t-test.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

4.2 Gridding of homogenized observation data over China based on PConvUNet

Figure 2 illustrates the framework of the PConvUNet model and the workflow used for gridding discrete station observations. Using an eight-year test subset randomly selected from the ERA5 reanalysis dataset (Fig. 6, a1–h1), we masked the grid points without observation stations (Fig. 6, a2–h2) and then applied the trained PConvUNet model to reconstruct wind speed at these points. Relative to the ERA5 reference data (0.25°×0.25°), the fixed-mask PConvUNet consistently captured fine-scale NSWS patterns within the gridded domain across all test years (Fig. 6, a3–h3). Reconstructed NSWSs at all grid cells were significantly positively correlated with the ERA5 reference values (Fig. 7, a1–h1). We further compared the fixed-mask PConvUNet with two alternative approaches: (i) a conventional random-mask strategy and (ii) widely used imputation and interpolation methods, including random forest and bilinear interpolation. By comparison, the alternative methods showed larger biases across the entire domain than the fixed-mask PConvUNet. The random-mask strategy failed to reproduce fine-scale NSWS features, particularly in regions with sparse station coverage (Fig. 6a4–h4). Similarly, both random forest imputation (Fig. 7, a3–h3) and bilinear interpolation (Fig. 7, a4–h4) performed less well in resolving localized wind speed features, especially over western China, where station density was low (Fig. 6, a5–h5, a6–h6), indicating that spatially blurred reconstructions lack structural fidelity. Probability density analysis further supported these results. For all test years, the fixed-mask PConvUNet achieved the lowest Kullback–Leibler (KL) divergence relative to the ERA5 reference values (0.0052±0.0014), indicating the best performance in preserving the distribution characteristics of the data. In contrast, the mean KL divergences of random forest and bilinear interpolation were 0.039±0.007 and 0.074±0.001, respectively, which were higher than those of PConvUNet (Fig. 8). Overall, these results indicate that the fixed-mask PConvUNet not only captured fine-scale spatiotemporal characteristics of wind speed but also effectively preserved the statistical integrity of the original distribution, including its tail behavior and high wind speed characteristics.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

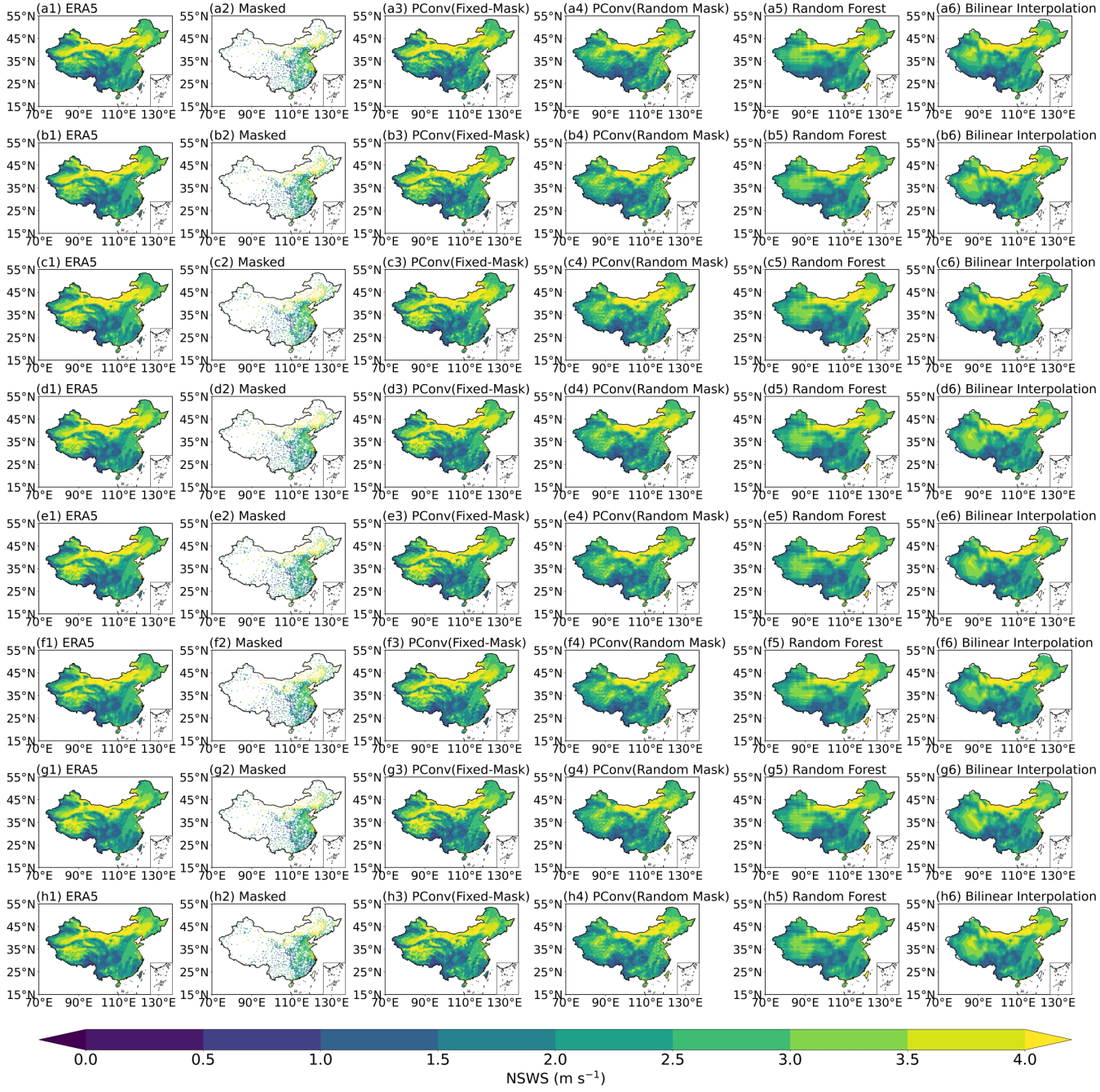


Figure 6. Comparison of wind speed reconstructions from multiple methods against ERA5 reference data. (a1–h1) ERA5 reference fields. (a2–h2) The same ERA5 fields masked to retain only grid cells collocated with discrete stations. (a3–h3) Reconstructions from the PConvUNet model trained with fixed binary masks. (a4–h4) Reconstructions from the PConvUNet model trained with random binary masks. (a5–h5): random forest interpolation. (a6–h6): bilinear interpolation. (a–h) correspond to eight randomly selected test years (1953, 1967, 1974, 1982, 1991, 2002, 2014, 2024), sampled at approximately decadal intervals.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

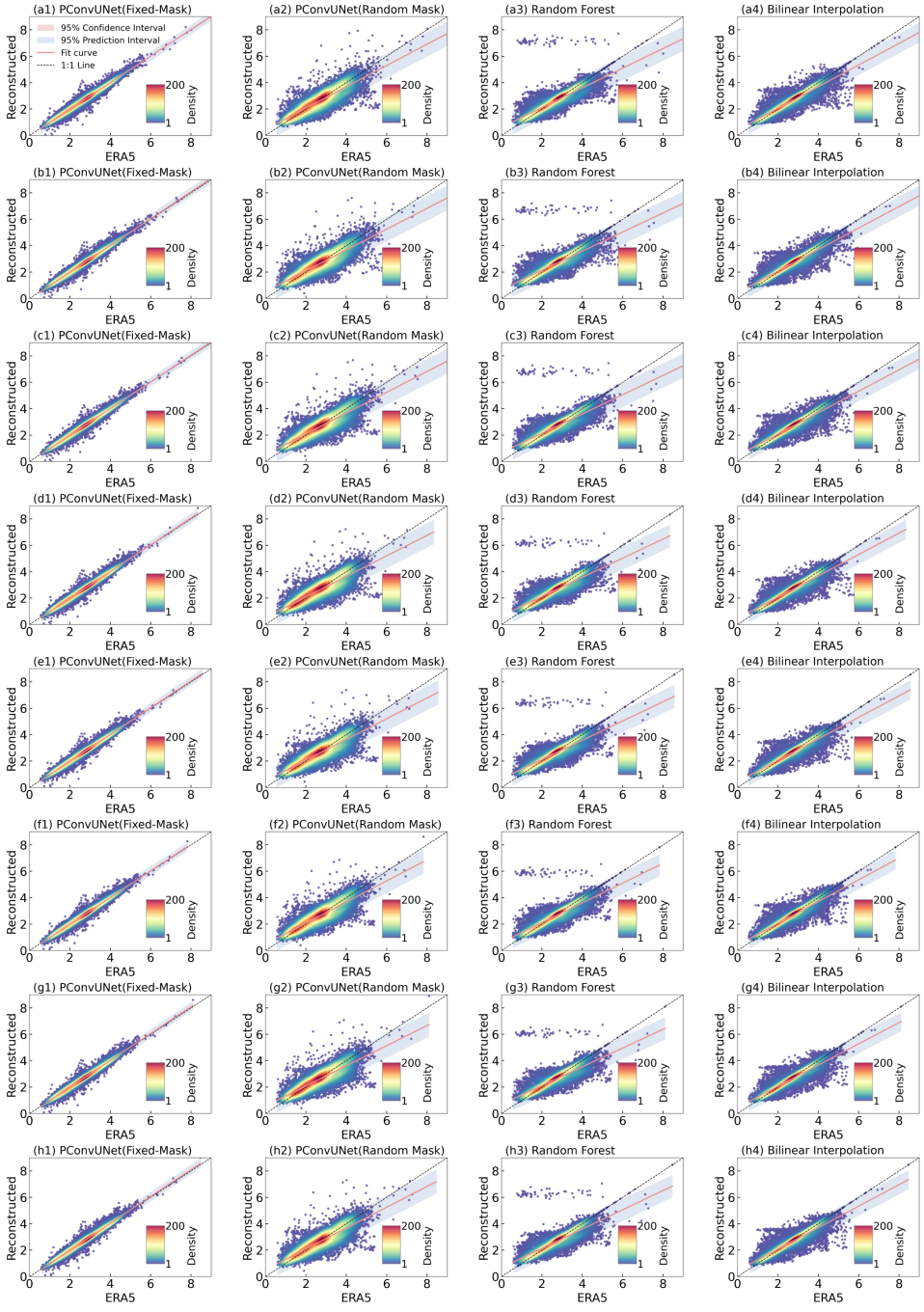


Figure 7. Residual analysis of wind speed reconstructions from four methods relative to ERA5 reanalysis data. (a1–h1) Results for the PConvUNet with fixed binary masks. (a2–h2) Results for the PConvUNet with random binary masks. (a3–h3): random forest interpolation. (a4–h4): bilinear interpolation. (a–h) correspond to eight independently selected test years (1953, 1967, 1974, 1982, 1991, 2002, 2014, 2024), spanning approximately decadal intervals to ensure temporal diversity and robustness.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

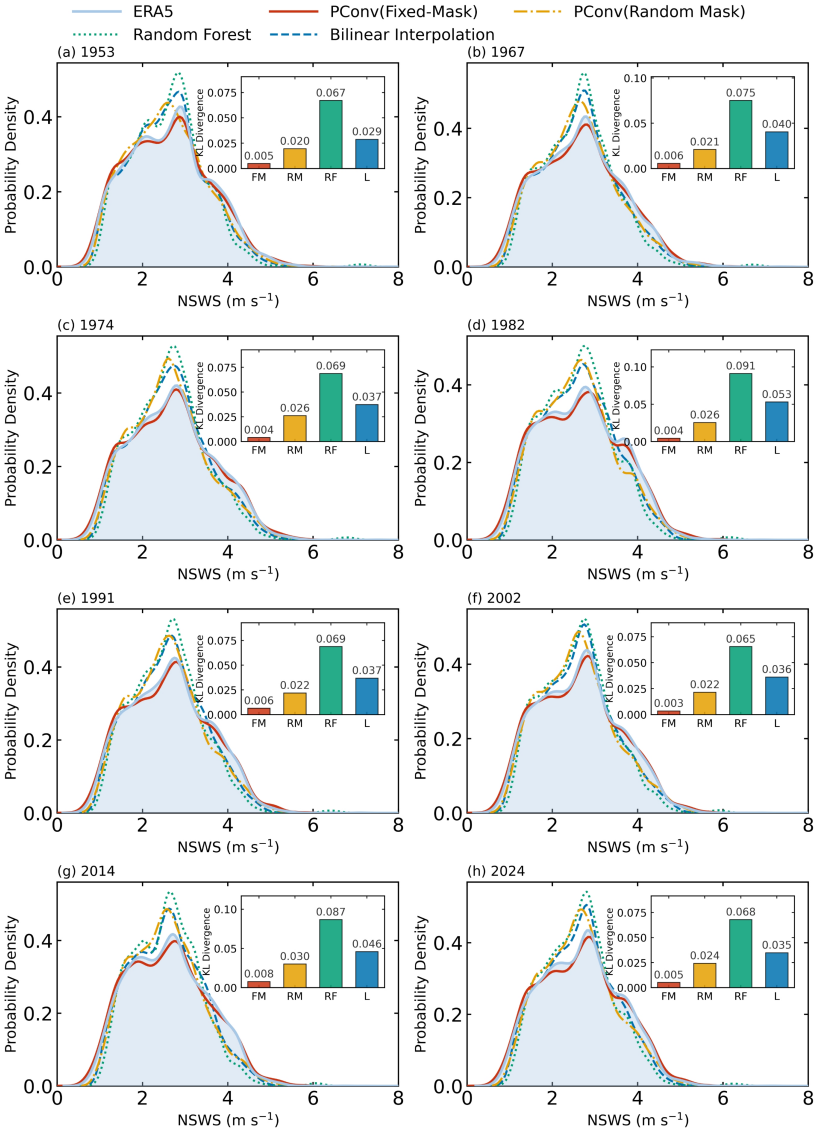


Figure 8. Probability density distributions of wind speed derived from different reconstruction methods. (a–h) correspond to eight randomly selected years. Light blue solid lines denote ERA5 reference data; red solid lines denote outputs from the PConvUNet trained with fixed masking; yellow dashed lines denote outputs from the PConvUNet trained with random masking; green dashed lines denote the outputs from the random forest; dark blue dashed lines denote the outputs from bilinear interpolation. In the insets, bars show the KL divergence between the outputs of each method and the ERA5 reference data.

To assess the model's generalization capability on unseen data, we applied it to the NCEP FNL dataset, which was masked and processed following the same procedure as for the ERA5 test set. Similarly, PConvUNet outperformed the other methods in reconstructing spatial features. In station-sparse regions such as the Tibetan Plateau, it better reproduced the fine-

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



scale features of the FNL dataset (Fig. 9). Moreover, PConvUNet more accurately captured wind events across a range of intensities in the unseen dataset (Fig. S2). Using the FNL analysis as ground truth, PConvUNet achieved the best performance across all five wind speed categories, with a mean absolute deviation of 0.61%, which was 1.16 percentage points lower than the average of the second-best methods (1.77%), corresponding to a relative improvement of 65.5%. The model exhibited modest advantages for calm and light air categories, but showed the strongest performance for wind events ≥8 m s^{-1} (Fig. S2), indicating its significance in reproducing strong wind events.

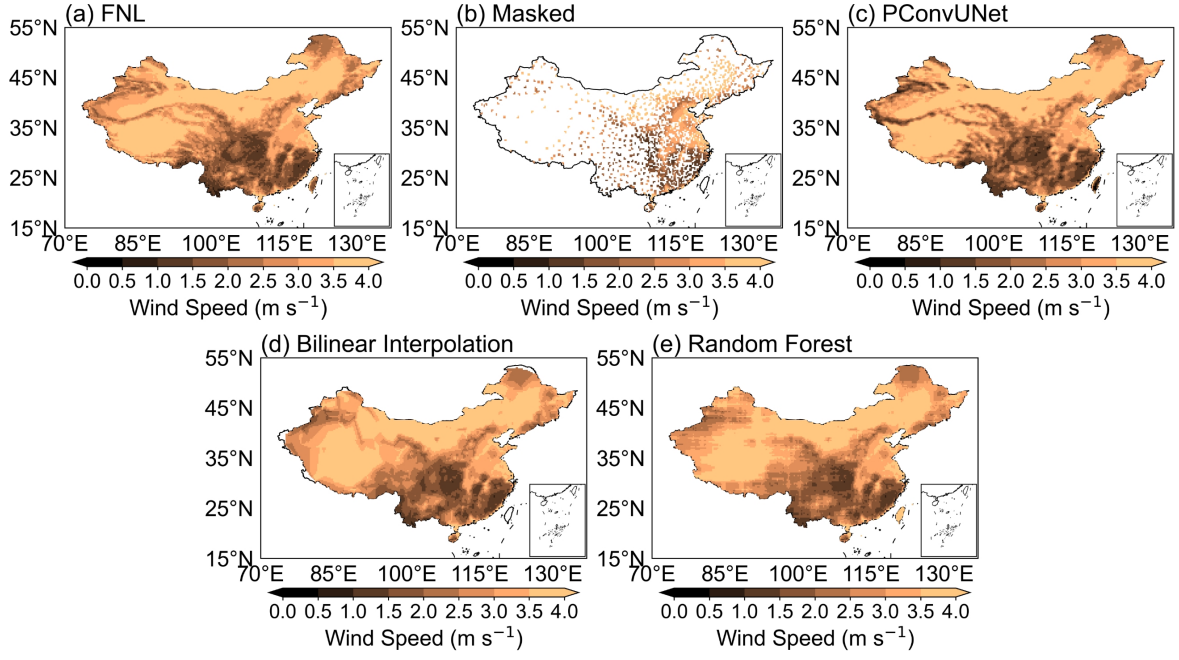


Figure 9. Generalization capability testing of the PConvUNet using the FNL dataset. (a) Spatial distribution of the mean values of the original FNL data, (b) the spatial distribution of FNL data after masking according to regions with observational stations, (c) gridding results generated by the PConvUNet. (d) and (e) are the same as (c), but for bilinear interpolation and random forest, respectively.

We calculated the root mean square error (RMSE) and mean absolute error (MAE) between the reconstructed values and ERA5 true values across seven sub-regions. The error differences were then defined as the RMSE and MAE of bilinear interpolation minus those of the PConvUNet, namely ΔRMSE and ΔMAE. A positive value indicates that PConvUNet achieved a lower reconstruction error than bilinear interpolation. The results show that the PConvUNet consistently outperformed bilinear interpolation over the sparse observation station coverage, such as the Tibetan Plateau and northwest China (Fig. 10a–b). Compared to bilinear interpolation, based on PConvUNet, the RMSE and MAE in the Tibetan Plateau decreased by 0.35 and 0.2 m s^{-1}, respectively; in northwest China, they decreased by 0.56 and 0.38 m s^{-1}, respectively. To quantify the advantage of the PConvUNet across station-sparse regions, we defined a Meteorological Station Area Density

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

Index (MSADI; $\mathrm{MSADI} = \frac{\text{Number of meteorological stations}}{(\text{Regional area / Total area}) \cdot 100}$). Lower (higher) MSADI values corresponded to sparser (denser) observational station coverage within a region. For instance, the Tibetan Plateau and northwest China have large areas with relatively few stations, yielding low MSADI values (3.78 and 3.09, respectively). In contrast, eastern China and southeastern China, characterized by small areas and high station density, exhibited high MSADI values (45.43 and 33.30, respectively). Although both PConvUNet and bilinear interpolation exhibited increasing RMSE and MAE with decreasing MSADI, the errors of the bilinear method in areas with sparse stations were higher than those of the PConvUNet. Meanwhile, the uncertainty ranges of RMSE and MAE for the PConvUNet were also smaller than those of the bilinear interpolation (Fig. 10c–d), indicating that the PConvUNet is less sensitive to the deterioration of MSADI. Consequently, conducting data interpolation with the PConvUNet has higher reliability over areas with sparse stations compared to the traditional bilinear interpolation.

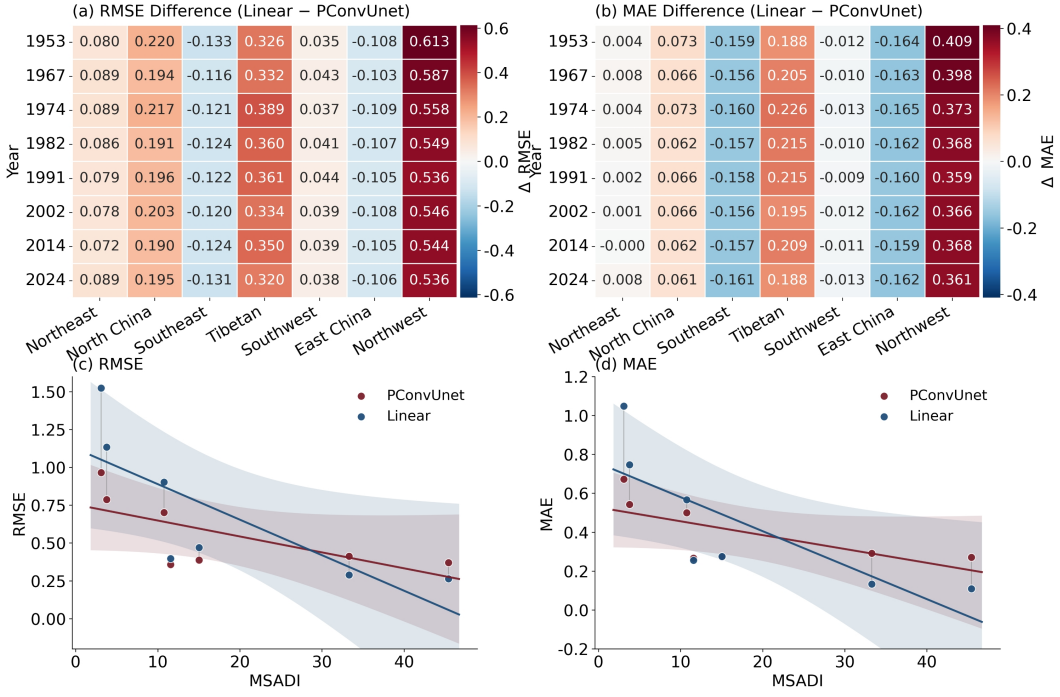


Figure 10. Differences in the RMSE and MAE between interpolated wind speed estimates and ERA5 reference data across all regions (a, b), and the RMSE and MAE of the interpolation data compared with the true values varied with MSADI in seven regions (c, d). Shading indicates the 95% confidence interval of the corresponding linear regression fits.

4.3 Robustness evaluation of the reconstructed gridding NSWS dataset

To evaluate the performance of PConvUNet in reconstructing wind fields over the regions lacking observational stations, we applied the Hom_NSWS dataset as input to generate gridded wind speed estimates. Subsequently, we inverted the original masking scheme by designating the locations of actual meteorological stations as missing (i.e., masked) pixels, thereby

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

constructing a controlled "reverse interpolation" test set. The resulting masked gridded data were then fed back into the PConvUNet model for reconstruction. Importantly, for this experiment, the model was retrained under the inverted mask configuration to ensure fair and robust evaluation. Results show that the spatial patterns of the reverse-interpolated wind speed fields (Fig. 11, a2–d2) closely matched those of the Hom_NSWS across all four observation times (Fig. 11, a1–d1). At the majority of stations, the RMSEs between reverse-interpolated wind speed and Hom_NSWS remain low across all four observation times (< 0.4 m s^{-1}) (Fig. 11, a3–d3). Temporally, the reverse-interpolated sequences also exhibited strong agreement with discrete station observations, with a mean bias of less than 0.05 m s^{-1} across all four times (Fig. 12). Furthermore, Weibull distribution analysis revealed near-identical shape and scale parameters between the reverse-interpolated datasets and original Hom_NSWS datasets, confirming preservation of wind speed distribution characteristics without statistically significant deviation (insets in Fig. 12).

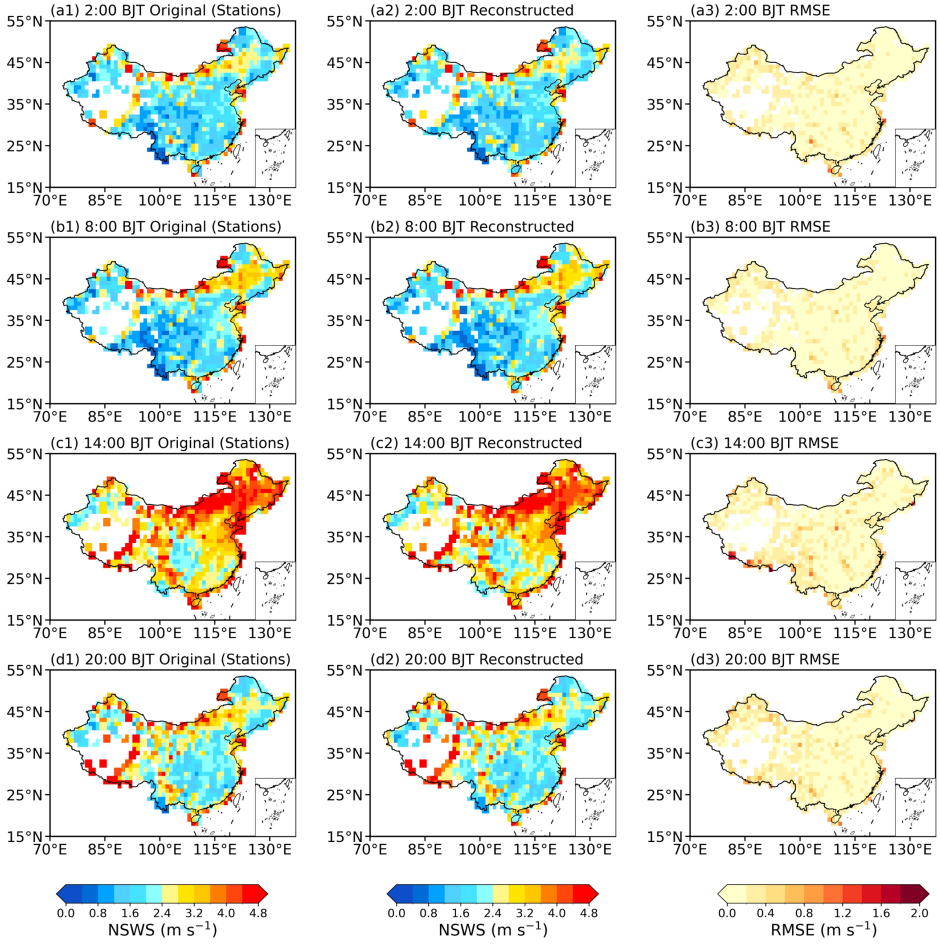


Figure 11. Spatial patterns of the original station-based Hom_NSWS (a1–d1) and the reconstructed gridded Hom_NSWS inferred by inputting the filled data into the PConvUNet model (a2–d2), together with their RMSE (a3–d3), during all four observation times (02:00, 08:00, 14:00, and 20:00 BJT).

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

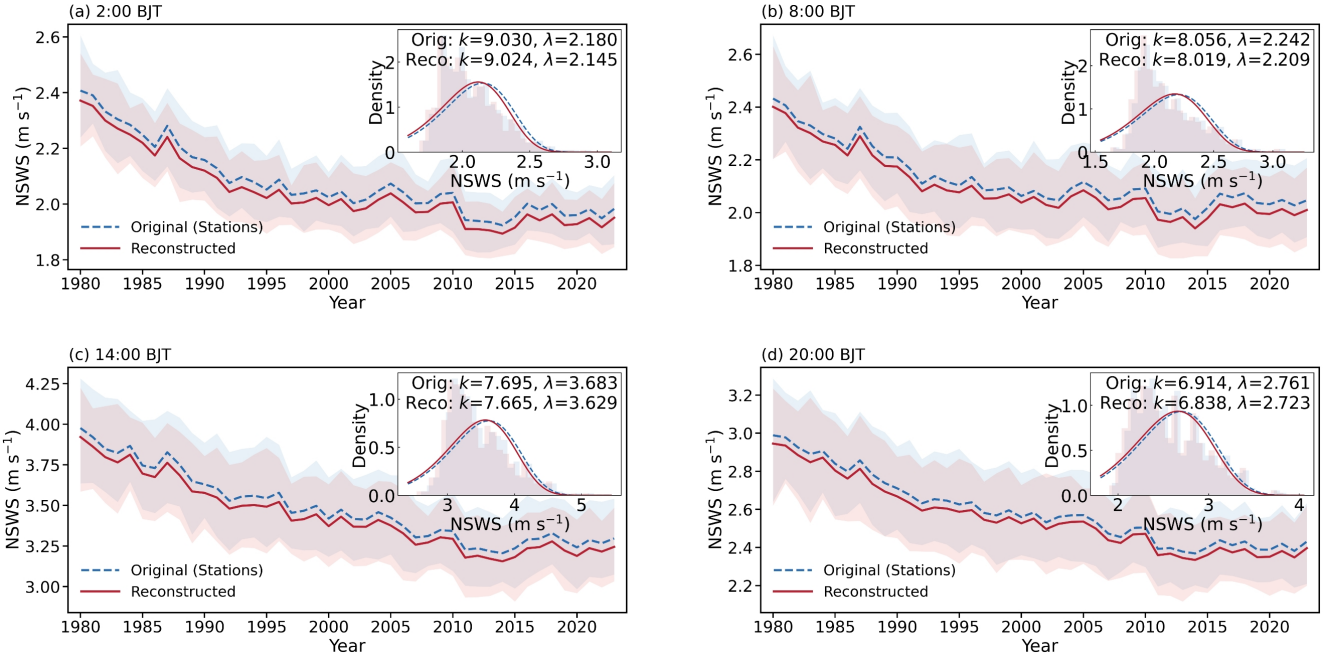


Figure 12. Temporal evolution of reconstructed gridded Hom_NSWS at the stations inferred by inputting the filled data into the PConvUNet model (GH_NSWS) (red lines) and the homogenized NSWS at stations (Hom_NSWS) (blue lines) at 02:00 BJT (a), 08:00 BJT (b), 14:00 BJT (c), and 20:00 BJT (d). Shading area represents the 95% uncertainty interval of the annual average wind speed calculated after 300 tests with 40% randomly selected samples. The insets show the Weibull distributions of the GH-NSWS and Hom_NSWS, and denote shape and scale parameters, respectively.

Across seven regions, the station estimates reconstructed by PConvUNet exhibited broadly consistent error levels and diurnal variations (Fig. 13). Spatially, the RMSEs for the seven regions were consistently concentrated around 0.3 m s^{-1}, and the mean differences were near 0 m s^{-1}. Quantitatively, Northeast China and the Tibetan Plateau showed the highest reconstruction accuracy, with average RMSE across the four observation times of only 0.36 and 0.37 m s^{-1}, respectively. East and Southeast China exhibited the lowest accuracy, with average RMSE of 0.50 and 0.52 m s^{-1}, respectively. This may be attributed to the larger areas of missing data created by masking station locations in these station-dense regions, which made reconstruction more challenging. Temporally, the RMSE at 14:00 BJT was 0.15m s^{-1} higher than at the other observation times. Overall, the error distributions across regions remain non-significant, and the influence of spatial differences on model uncertainty was limited, indicating that the PConvUNet exhibits high robustness in gridding discrete station observations.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

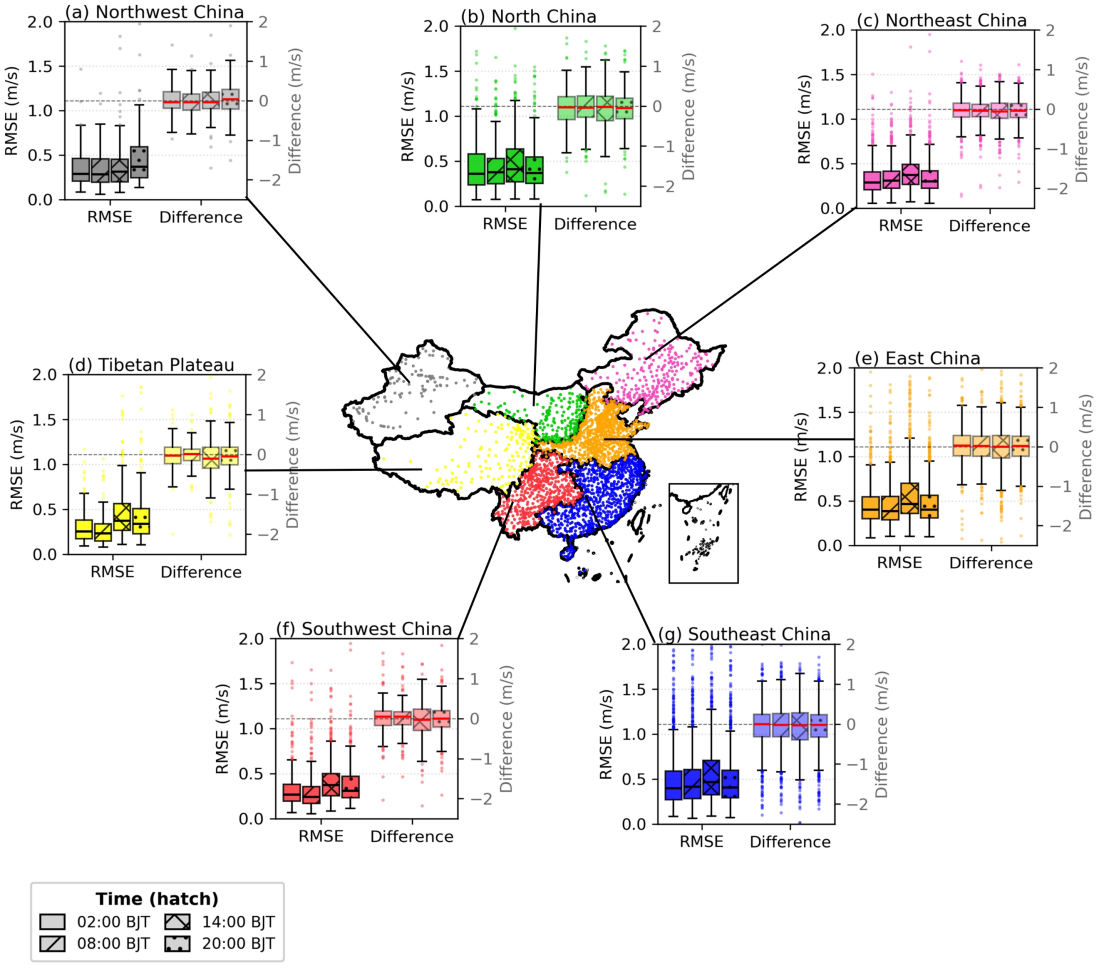


Figure 13. Comparison of RMSE and bias between reconstructed and Hom_NSWS at four observation times across seven regions in China. Each subplot corresponds to a specific region. Four distinct hatching patterns represent different observation times: no hatching indicates 02:00 BJT, slashes denote 08:00 BJT, cross-hatching signifies 14:00 BJT, and dots represent 20:00 BJT. The horizontal dashed line in each subplot indicates the zero-bias line.

5 Conclusions and Discussions

This study develops the SgriWind-CN dataset by integrating NSWS observations over China. Through systematic quality control and homogenization correction, specifically addressing IBs induced by instrument changes, station relocations, and other non-climatic factors, the climatic consistency of wind speed observations is improved. Meanwhile, a deep learning model, PConvUNet, is then applied to fuse the ERA5 reanalysis spatial fields with wind speed observations, achieving high-resolution gridded NSWS data. The main results are summarized as follows.

1. Based on 1,498 station records, the first sub-daily, homogenized NSWS dataset across China was developed at a spatial resolution of 0.25° × 0.25° for 1980–2023.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



2. More than 40% of the stations exhibited IBs during homogenization. After homogenization, the mean NSWS increased by 0.2 m s^{-1}, and the overall declining trend was intensified.

3. Comparative validations and evaluations show that the PConvUNet model markedly outperforms conventional interpolation methods in reconstructing the wind field, especially over the regions with complex terrain and sparse stations, where RMSE and MAE were reduced by 0.45 and 0.29 m s^{-1}, respectively.

4. The wind speeds reconstructed by PConvUNet show strong agreement with the observations in both temporal variation and spatial distribution, with a mean bias of less than 0.05 m s^{-1}. These results demonstrate that the model can accurately capture spatiotemporal characteristics of the station observations.

It should be noted that this study only addresses non-climatic IBs and spatial reconstruction but does not fill temporal gaps in the observational records; consequently, this may lead to a deviation when calculating the daily mean NSWS. Meanwhile, this study employs a partial convolutional neural network for spatial field reconstruction. It can only learn the spatial correlations from reanalysis data and lacks physics-informed constraint capabilities. Future work needs to integrate multi-source data fusion techniques, including reanalysis, satellite-derived winds, and relevant covariates, to reconstruct missing values and enhance coverage in data-sparse regions, thereby generating a more complete, continuous, and physically consistent NSWS record to support both fundamental research and operational applications.

Code and data availability

SgriWind-CN is publicly available at Zenodo: https://doi.org/10.5281/zenodo.22812484 (Bai et al., 2026). It provides gridded 10 m near-surface wind speed data over China at a spatial resolution of 0.25°×0.25°. The dataset also includes surface roughness, land-cover type, elevation, latitude, and longitude. All data are provided in NetCDF format.

Author contributions

Xueru Bai: Formal analysis, Investigation, Methodology, Visualization, Validation. **Mingyuan Zhang**: Formal analysis, Investigation, Methodology, Visualization, Validation. **Jinlin Zha**: Conceptualization, Formal analysis, Investigation, Validation, Writing – Original Draft, Writing – Review and Editing, Validation. **Anning Huang**: Validation. **Xinsheng Zhu**: Validation. **Jian Tang**: Validation, Data curation. **Deming Zhao**: Formal analysis, Validation. **Jian Wu**: Validation, Funding acquisition.

Competing interests

The authors declare no competing interests.

Open Research

All observational data supporting this study's findings are presented in the manuscript. The observed dataset is accessed at CMA Data (CMA, 2003). All figures are prepared using Python (Python, 2023) (version 3.12.4).

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2023YFF0805504), the National Natural Science Foundation of China (42475040, 42375174), the Yunnan Fundamental Research Projects (202401AW070008), and the Major Project of Yunnan Foundational Research Program (202501BC070001).

References

Alexandersson, H.: Homogenization of Climate Data, Difficult but Necessary, in: Detecting and Modelling Regional Climate Change, edited by: India, M. B. and Bonillo, D. L., Springer, Berlin, Heidelberg, 3–12, https://doi.org/10.1007/978-3-662-04313-4_1, 2001.

Ashkenazy, Y. and Yizhaq, H.: The diurnal cycle and temporal trends of surface winds, Earth Planet. Sci. Lett., 601, 117907, https://doi.org/10.1016/j.epsl.2022.117907, 2023.

Assunção, R. M., Neves, M. C., Câmara, G., and Da Costa Freitas, C.: Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees, Int. J. Geogr. Inf. Sci., 20, 797–811, https://doi.org/10.1080/13658810600665111, 2006.

Azorin-Molina, C., Vicente-Serrano, S. M., McVicar, T. R., Jerez, S., Sanchez-Lorenzo, A., López-Moreno, J.-I., Revuelto, J., Trigo, R. M., López-Bustins, J. A., and Espírito-Santo, F.: Homogenization and assessment of observed near-surface wind speed trends over Spain and Portugal, 1961–2011, J. Climate, 27, 3692–3712, https://doi.org/10.1175/JCLI-D-13-00652.1, 2014.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. D., and Thinh, N. X.: RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements, Remote Sens. Environ., 239, 111606, https://doi.org/10.1016/j.rse.2019.111606, 2020.

Bai, P. and Wang, Y.: The importance of heat storage for estimating lake evaporation on different time scales: Insights from a large shallow subtropical lake, Water Resour. Res., 59, e2023WR035123, 2023.

Bai, X., Zhang, M., Zha, J., Huang, A., Zhu, X., Tang, J., Zhao, D., and Wu, J.: A Sub-Daily Gridded High-Resolution Near-Surface Wind Speed Dataset for China, Generated by an AI-Driven Integration of In-Situ Observations, Zenodo [data set], https://doi.org/10.5281/zenodo.22812484, 2026.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Mon. Wea. Rev., 129, 420–436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2, 2001.

Cao, L., Zhu, Y., Tang, G., Yuan, F., and Yan, Z.: Climatic warming in China according to a homogenized data set from 2419 stations, Int. J. Climatol., 36, 4384–4392, https://doi.org/10.1002/joc.4639, 2016.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Chang, Y.-C., Chen, G.-Y., Tseng, R.-S., Centurioni, L. R., and Chu, P. C.: Observed near-surface currents under high wind speeds, J. Geophys. Res. Oceans, 117, https://doi.org/10.1029/2012JC007996, 2012.

Chen, Q., Fan, J., Yin, Y., and Han, B.: Aerosol impacts on mesoscale convective systems forming under different vertical wind shear conditions, J. Geophys. Res. Atmos., 125, e2018JD030027, https://doi.org/10.1029/2018JD030027, 2020.

Cheng, C. S., Zhu, J., Yang, T., Yang, S., Nie, G., and Tang, J.: The decline trends of the observed winds over China mostly attributed to urbanization-induced wind blocking from the anemometers, Urban Clim., 55, 101998, https://doi.org/10.1016/j.uclim.2024.101998, 2024.

Chuan, T., Zha, J., Wu, J., Zhao, D., Fan, W., Jiang, H., and Lyu, Y.: Non-synchronization of the decadal transition in winter near-surface wind speed across northern and southern China, Geophys. Res. Lett., 51, e2024GL110246, https://doi.org/10.1029/2024GL110246, 2024.

Dayal, K. K., Cater, J. E., Kingan, M. J., Bellon, G. D., and Sharma, R. N.: Wind resource assessment and energy potential of selected locations in Fiji, Renew. Energy, 172, 219–237, https://doi.org/10.1016/j.renene.2021.03.034, 2021.

Dupré, A., Drobinski, P., Alonzo, B., Badosa, J., Briard, C., and Plougonven, R.: Sub-hourly forecasting of wind speed and wind energy, Renew. Energy, 145, 2373–2379, https://doi.org/10.1016/j.renene.2019.07.161, 2020.

European Space Agency. (2024). Copernicus Global Digital Elevation Model. OpenTopography [data set]. https://doi.org/10.5069/G9028PQB

Gandin, L. S.: Objective Analysis of Meteorological Fields, Israel Program for Scientific Translations, Jerusalem, 1965.

Han, J., Miao, C., Gou, J., Zheng, H., Zhang, Q., and Guo, X.: A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations, Earth Syst. Sci. Data, 15, 3147–3161, https://doi.org/10.5194/essd-15-3147-2023, 2023.

He, Y., Monahan, A. H., Jones, C. G., Dai, A., Biner, S., Caya, D., and Winger, K.: Probability distributions of land surface wind speeds over North America, J. Geophys. Res. Atmos., 115, 2008JD010708, https://doi.org/10.1029/2008JD010708, 2010.

Hersbach, H., Bell, B., Berrisford, P., et al.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J-N.: ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.adbb2d47, 2023.

Hofstra, N., Haylock, M., New, M., Jones, P., and Frei, C.: Comparison of six methods for the interpolation of daily, European climate data, J. Geophys. Res., 113, D21110, https://doi.org/10.1029/2008JD010100, 2008.

Houndekindo, F. and Ouarda, T. B. M. J.: LSTM and Transformer-based framework for bias correction of ERA5 hourly wind speeds, Energy, 328, 136498, https://doi.org/10.1016/j.energy.2025.136498, 2025.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Jiao, B., Su, Y., Li, Q., Manara, V., and Wild, M.: An integrated and homogenized global surface solar radiation dataset and its reconstruction based on a convolutional neural network approach, Earth Syst. Sci. Data, 15, 4519–4535, https://doi.org/10.5194/essd-15-4519-2023, 2023.

Kadow, C., Hall, D. M., and Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nat. Geosci., 13, 408–413, https://doi.org/10.1038/s41561-020-0582-5, 2020.

Karnauskas, K. B., Lundquist, J. K., and Zhang, L.: Southward shift of the global wind energy resource under high carbon dioxide emissions, Nat. Geosci., 11, 38–43, https://doi.org/10.1038/s41561-017-0029-9, 2018.

Krige, D. G.: A statistical approach to some basic mine valuation problems on the Witwatersrand, J. South. Afr. Inst. Min. Metall., 52, 201–203, https://doi.org/10.2307/3006914, 1951.

Li, X., Li, Q.-P., Ding, Y.-H., and Wang, M.: Near-surface wind speed changes in eastern China during 1970–2019 winter and its possible causes, Adv. Clim. Change Res., 13, 228–239, https://doi.org/10.1016/j.accre.2022.01.003, 2022.

Li, Y., Chen, Y., Li, Z., and Fang, G.: Recent recovery of surface wind speed in northwest China, Int. J. Climatol., 38, 4445–4458, https://doi.org/10.1002/joc.5679, 2018.

Li, Z., Yan, Z., Cao, L., and Jones, P. D.: Further-adjusted long-term temperature series in China based on MASH, Adv. Atmos. Sci., 35, 909–917, https://doi.org/10.1007/s00376-018-7280-x, 2018.

Lin, C., Yang, K., Qin, J., and Fu, R.: Observed coherent trends of surface and upper-air wind speed over China since 1960, J. Climate, 26, 2891–2903, https://doi.org/10.1175/JCLI-D-12-00093.1, 2013.

Liu, F., Sun, F., Liu, W., Wang, T., Wang, H., Wang, X., and Lim, W. H.: On wind speed pattern and energy potential in China, Appl. Energy, 236, 867–876, https://doi.org/10.1016/j.apenergy.2018.12.056, 2019.

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B.: Image inpainting for irregular holes using partial convolutions, in: Computer Vision – ECCV 2018, edited by: Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., Springer, Cham, Switzerland, 85–100, https://doi.org/10.1007/978-3-030-01252-6_6, 2018.

Liu, X., Zhang, X. J., Tang, Q., and Zhang, X. Z.: Effects of surface wind speed decline on modeled hydrological conditions in China, Hydrol. Earth Syst. Sci., 18, 2803–2813, https://doi.org/10.5194/hess-18-2803-2014, 2014.

Liu, Y., Zhou, L., Qin, Y., Azorin-Molina, C., Shen, C., Xu, R., and Zeng, Z.: Impacts of anemometer changes, site relocations and processing methods on wind speed trends in China, Atmos. Meas. Tech., 17, 1123–1131, https://doi.org/10.5194/amt-17-1123-2024, 2024.

Long, Z., Zha, J., Zhang, H., Chuan, T., Lu, W., Yan, Y., et al.: Deep learning with ERA5 variables accurately simulates near-surface wind speed variations across a complex terrain region, Geophys. Res. Lett., 52, e2025GL116108, https://doi.org/10.1029/2025GL116108, 2025.

Ma, Z., Yang, Q., Xu, Y., et al.: Reconstruction of Antarctic sea ice thickness from sparse satellite laser altimetry data via deep learning, Sci. Data, 13, https://doi.org/10.1038/s41597-025-06390-x, 2025.

McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, T. G., Grieser, J., Jhajharia, D., Himri, Y., Mahowald, N. M., Mescherskaya, A. V., Kruger, A. C., Rehman, S., and Dinpashoh, Y.: Global review and

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation, J. Hydrol., 416–417, 182–205, https://doi.org/10.1016/j.jhydrol.2011.10.024, 2012.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce: NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids, NSF National Center for Atmospheric Research, https://doi.org/10.5065/D65Q4T4Z, 2015.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian model averaging to calibrate forecast ensembles, Mon. Weather Rev., 133, 1155–1174, https://doi.org/10.1175/MWR2906.1, 2005.

Ramon, J., Lledó, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J.: What global reanalysis best represents near-surface winds?, Q. J. Roy. Meteor. Soc., 145, 3236–3251, https://doi.org/10.1002/qj.3616, 2019.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by: Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., Springer, Cham, Switzerland, 234–241, https://doi.org/10.1007/978-3-319-319-24574-4_28, 2015.

Rose, S. and Apt, J.: Quantifying sources of uncertainty in reanalysis-derived wind speed, Renew. Energy, 94, 157–165, https://doi.org/10.1016/j.renene.2016.03.028, 2016.

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 1968 23rd ACM National Conference, ACM '68, 517–524, https://doi.org/10.1145/800186.810616, 1968.

Si, P., Luo, C., and Liang, D.: Homogenization of Tianjin monthly near-surface wind speed using RHtestsV4 for 1951–2014, Theor. Appl. Climatol., 132, 1303–1320, https://doi.org/10.1007/s00704-017-2140-7, 2018.

Sun, S., Nai, C., Pan, B., Li, W., Li, L., Li, X., Foufoula-Georgiou, E., and Lin, Y.: Fusion of multi-source precipitation records via coordinate-based generative models, Nat. Commun., 17, 1227, https://doi.org/10.1038/s41467-025-67987-9, 2025.

Tuinenburg, O. A., Theeuwen, J. J. E., and Staal, A.: High-resolution global atmospheric moisture connections from evaporation to precipitation, Earth Syst. Sci. Data, 12, 3177–3188, https://doi.org/10.5194/essd-12-3177-2020, 2020.

Vautard, R., Cattiaux, J., Yiou, P., Thépaut, J.-N., and Ciais, P.: Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness, Nat. Geosci., 3, 756–761, https://doi.org/10.1038/ngeo979, 2010.

Wan, H., Wang, X. L., and Swail, V. R.: Homogenization and trend analysis of Canadian near-surface wind speeds, J. Climate, 23, 1209–1225, https://doi.org/10.1175/2009JCLI3200.1, 2010.

Wang, K. and Dickinson, R. E.: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability, Rev. Geophys., 50, RG2005, https://doi.org/10.1029/2011RG000373, 2012.

Wang, K., Bai, P., and Liu, X.: Three paradoxes related to potential evapotranspiration in a warming climate, Curr. Clim. Change Rep., 11, 6, https://doi.org/10.1007/s40641-025-00203-4, 2025.

Wang, K., Cai, H., He, Y., Li, Y., Xia, H., Yin, C., Yu, H., Zhao, R., Zhang, Z., and Zhou, C.: A homogenized gridded climate dataset for China, Sci. China Earth Sci., https://doi.org/10.1007/s11430-025-1889-0, 2026.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.



Wang, X. L.: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test, J. Appl. Meteor. Climatol., 47, 2423–2444, https://doi.org/10.1175/2008JAMC1741.1, 2008a.

Wang, X. L.: Penalized maximal F test for detecting undocumented mean shift without trend change, J. Atmos. Oceanic Technol., 25, 368–384, https://doi.org/10.1175/2007JTECHA982.1, 2008b.

Wang, X. L., Wen, Q. H., and Wu, Y.: Penalized maximal t test for detecting undocumented mean change in climate data series, J. Appl. Meteor. Climatol., 46, 916–931, https://doi.org/10.1175/JAM2504.1, 2007.

Wang, X. L., Chen, H., Wu, Y., Feng, Y., and Pu, Q.: New techniques for the detection and adjustment of shifts in daily precipitation data series, J. Appl. Meteor. Climatol., 49, 2416–2436, https://doi.org/10.1175/2010JAMC2376.1, 2010.

Wei, Z., Miao, L., Peng, J., Zhao, T., Meng, L., Lu, H., Peng, Z., Cosh, M. H., Fang, B., Lakshmi, V., and Shi, J.: Bridging spatio-temporal discontinuities in global soil moisture mapping by coupling physics in deep learning, Remote Sens. Environ., 313, 114371, https://doi.org/10.1016/j.rse.2024.114371, 2024.

Wohland, J., Omrani, N., Witthaut, D., and Keenlyside, N. S.: Inconsistent wind speed trends in current twentieth century reanalyses, J. Geophys. Res. Atmos., 124, 1931–1940, https://doi.org/10.1029/2018JD030083, 2019.

Wu, J., Zha, J., and Zhao, D.: Evaluating the effects of land use and cover change on the decrease of surface wind speed over China in recent 30 years using a statistical downscaling method, Clim. Dyn., 48, 131–149, https://doi.org/10.1007/s00382-016-3065-z, 2017.

Wu, J., Zha, J., Zhao, D., and Yang, Q.: Changes in terrestrial near-surface wind speed and their possible causes: An overview, Clim. Dyn., 51, 2039–2078, https://doi.org/10.1007/s00382-017-3997-y, 2018.

Yan, Y., Wu, J., Chao, Q., and Sun, Y.: Overestimation of the recent observed near-surface wind speed recovery in China, npj Clim. Atmos. Sci., 9, 49, https://doi.org/10.1038/s41612-026-01322-x, 2026.

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth Syst. Sci. Data, 13, 3907–3925, https://doi.org/10.5194/essd-13-3907-2021, 2021.

Yang, P., Zhang, S., Xia, J., Zhan, C., Cai, W., Wang, W., Luo, X., Chen, N., and Li, J.: Analysis of drought and flood alternation and its driving factors in the Yangtze River Basin under climate change, Atmos. Res., 270, 106087, https://doi.org/10.1016/j.atmosres.2022.106087, 2022.

Yang, Q., Li, M., Zu, Z., and Ma, Z.: Has the stilling of the surface wind speed ended in China?, Sci. China Earth Sci., 64, 1036–1049, https://doi.org/10.1007/s11430-020-9738-4, 2021.

Zeng, Z., Ziegler, A. D., Searchinger, T., Yang, L., Chen, A., Ju, K., Piao, S., Li, L. Z. X., Ciais, P., Chen, D., Liu, J., Azorin-Molina, C., Chappell, A., Medvigy, D., and Wood, E. F.: A reversal in global terrestrial stilling and its implications for wind energy production, Nat. Clim. Change, 9, 979–985, https://doi.org/10.1038/s41558-019-0622-6, 2019.

Zha, J., Shen, C., Zhao, D., Wu, J., and Fan, W.: Slowdown and reversal of terrestrial near-surface wind speed and its future changes over eastern China, Environ. Res. Lett., 16, 034028, https://doi.org/10.1088/1748-9326/abe2cd, 2021.

https://doi.org/10.5194/essd-2026-806
Preprint. Discussion started: 12 November 2026
© Author(s) 2026. CC BY 4.0 License.


Zhang, J., Liu, B., Ren, S., Han, W., Ding, Y., and Peng, S.: A 4 km daily gridded meteorological dataset for China from 2000 to 2020, Sci. Data, 11, 1230, https://doi.org/10.1038/s41597-024-04029-x, 2024.

Zhang, L., Wang, P., Huang, C., Ai, B., and Feng, W.: A method of optimizing terrain rendering using digital terrain analysis, ISPRS Int. J. Geo-Inf., 10, 666, https://doi.org/10.3390/ijgi10100666, 2021.

Zhang, X., Liu, G., Wang, H., and Li, X.: Application of a hybrid interpolation method based on support vector machine in the precipitation spatial interpolation of basins, Water, 9, 760, https://doi.org/10.3390/w9100760, 2017.

Zhang, Y., Gao, L., Cao, L., Yan, Z., and Wu, Y.: Decreasing atmospheric visibility associated with weakening winds from 1980 to 2017 over China, Atmos. Environ., 224, 117314, https://doi.org/10.1016/j.atmosenv.2020.117314, 2020.

Zhang, Z. and Wang, K.: Stilling and recovery of the surface wind speed based on observation, reanalysis, and geostrophic wind theory over China from 1960 to 2017, J. Climate, 33, 3989–4008, https://doi.org/10.1175/JCLI-D-19-0281.1, 2020.

Zhang, Z. and Wang, K.: Homogenization of observed surface wind speed based on geostrophic wind theory over China from 1970 to 2017, J. Climate, 36, 3667–3679, https://doi.org/10.1175/JCLI-D-22-0445.1, 2023.

Zhou, C., Azorin-Molina, C., Engström, E., Minola, L., Wern, L., Hellström, S., Lönn, J., and Chen, D.: HomogWS-se: A century-long homogenized dataset of near-surface wind speed observations since 1925 rescued in Sweden, Earth Syst. Sci. Data, 14, 2167–2177, https://doi.org/10.5194/essd-14-2167-2022, 2022.

Zhou, H., Ning, S., Yang, R., Nazrollozoda, S., and Hu, Z.: A high-resolution dataset revealing the dynamical variations of the relative humidity in China, J. Hydrol., 670, 135163, https://doi.org/10.1016/j.jhydrol.2026.135163, 2026.

Zhou, L., Liu, H., Jiang, X., Ziegler, A. D., Azorin-Molina, C., Liu, J., and Zeng, Z.: An artificial intelligence reconstruction of global gridded surface winds, Sci. Bull., 67, 2060–2063, https://doi.org/10.1016/j.scib.2022.09.022, 2022.

Zhu, D., Cheng, X., Zhang, F., Yao, X., Gao, Y., and Liu, Y.: Spatial interpolation using conditional generative adversarial neural networks, Int. J. Geogr. Inf. Sci., 34, 735–758, https://doi.org/10.1080/13658816.2019.1599122, 2020.