Assessment of Reanalysis Wind Speeds Datasets Against the HadISD Dataset Over South Asia

Muhammad Abid Khan, Koji Dairaku, and Saurabh Kelkar

Abstract

Investigating the characteristics of wind speed holds immense significance within meteorology and various applied sciences. The wind energy industry is experiencing a period of expansion and progress, offering opportunities for establishing offshore wind farms and expanding current land-based wind farms. To comprehend the fundamental aspects of the wind field for the expansion of wind farms, it is crucial to possess reliable long-term data on wind speed. Reanalysis datasets are one of the primary sources for long-term consistent gridded datasets; however, these datasets are fallible to uncertainties. Thus, the reliability of such products is necessary before assessing the potential for wind farm expansion. This study compares daily 10 m wind speed (m/s) from three-reanalysis products (ERA5, JRA55, and NCEP) and their ensemble mean with HadISD wind observations based on the 99% threshold based on the valid dataset available for South Asia during 1974–2004. Statistical metrics such as standard deviation (STD), mean absolute error (MAE), percent bias (PBIAS), and correlation coefficient (CC) are applied to the assessment of the reanalysis products. Results demonstrated that the JRA55 reanalysis product demonstrates the highest level of accuracy among the other products. However, significant disparities were detected between the simulated wind speeds and the actual observations. The JRA55 reanalysis product accurately reflected the changes seen in the Subcontinent observations and was subsequently succeeded by the ERA5 product. The results estimations of JRA55 show exceptional agreement with wind speed observations, as evidenced by its high CC of 0.676, a PBIAS of 21.543, and a low MAE of 1.014. On the other hand, when comparing it to JRA55, ERA5 shows larger MAE and PBIAS, and lower CC. Compared to JRA55 and ERA5, NCEP exhibited a tendency to overestimate, deviating from the expected values. Based on this comparison, this research seeks to assess and distinguish the reliability of reanalysis products when compared to the observed dataset, emphasizing the importance of exercising caution when relying on reanalysis data to evaluate and predict winds in the South Asia region.

Keywords

HadISD · Reliability assessment · Gridded datasets · Statistical metrics · Subcontinent wind

1 Introduction

An important atmospheric parameter, wind speed, is subject to fluctuations over a variety of spatial and temporal dimensions, from extensive regional airflows to intricate turbulent eddies (Khan et al. 2018). Meteorologists, climate scientists, applied disciplines, and engineers alike must be able to understand the nuances of wind behavior and its spatiotemporal fluctuations (Khan and Dairaku 2024; Molina et al. 2021). While wind speed is not as extensively studied as temperature and precipitation due to the limited availability of reliable measurement data, several studies have examined wind speed variability across different continents, and specifically in South Asia. In South Asia, wind regimes are influenced by diverse regional factors such as the Indian Ocean, the Himalayas, and the monsoon system. The coastal areas of South Asia, including the Arabian Sea and the Bay of Bengal, experience higher wind speeds due to the proximity to these bodies of

M. A. Khan (✉) · K. Dairaku · S. Kelkar
Graduate School of Science and Technology, Department of Engineering Mechanics and Energy, University of Tsukuba, Ibaraki, Japan
e-mail: khan.muhammad.sa@alumni.tsukuba.ac.jp

M. A. Khan
Center for Environmental Remote Sensing, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

Z. Zhang et al. (eds.), *Dynamic Evolution of Atmospheric, Ecological, and Hydrological Systems in Circum-Mediterranean Regions*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-86777-4_39

water. The Himalayas play a significant role in shaping wind patterns, causing variations in wind speed and direction across the region. The monsoon system, characterized by seasonal wind changes, further impacts wind variability in South Asia, with peak wind speeds often occurring during the monsoon season (Wenxin et al. 2021). However, detailed assessments of wind speed patterns and cycles specific to South Asia require further research and analysis.

Wind studies rely on various data sources. Wind speed measurements from meteorological stations, although scattered in time and space, provide valuable information. Gridded wind data is available through reanalysis products and climate models, which offer physically consistent atmospheric fields. Reanalysis databases assimilate observations from different sources to generate long-term climate datasets (Wenxin et al. 2021). They serve as valuable resources for studying wind speed trends, validating simulations, and providing boundary conditions for regional climate models. While wind speed observations are sparse, reanalysis products like ERA5, JRA55, and NCEP from ECMWF, JMA, and NCEP offer continuous wind fields at high resolution, respectively. Reanalysis datasets are one of the primary sources for long-term consistent gridded datasets; however, these datasets are fallible to uncertainties. This study aims to analyze wind behavior assessment in South Asia regions using the HadISD dataset, which provides sub-daily surface wind measurements. Statistical computations and frequency distributions of wind speed are examined at daily scales, further highlighting monthly climatology cycles. To fill major scientific gaps, to assess and distinguish the reliability of reanalysis products through the 99% threshold based on the valid dataset HadISD of wind speed and three global reanalysis products and their ensemble mean utilized the current research domain and period is 1974–2004. Thus, the reliability of such products is necessary before assessing the potential for wind farm expansion. The benefit of this research is to enhance the accuracy of wind resource assessments and improve our understanding of regional climate dynamics in South Asia, ultimately contributing to informed decision-making in renewable energy and climate-related sectors. The paper is unfolded into five sections. Sections 2, 3, 4, and 5 are data and methods, results, discussion, and conclusions, respectively.

2 Data and Methods

A meteorological database known as HadISD, maintained by the Met Office Hadley Centre, is utilized in this study as a source of meteorological observations. HadISD has a comprehensive global sub-daily dataset derived from NOAA's NCDC ISD dataset, and it is maintained and distributed by the United Kingdom Met Office Hadley Centre. A quality-controlled dataset of hourly data obtained from weather stations around the world provides hourly estimates of meteorological variables such as wind speed. Rigorous quality control procedures, such as duplicate checks, distributional gap checks, and identification of outlier values, are applied to ensure data accuracy and reliability. The database encompasses 8139 stations globally, with a significant concentration in South Asia covering the period from 1931 to 2020 (Li et al. 2019). Furthermore, details are mentioned for observation datasets and global reanalysis products in the tabulated Table 1. All datasets are regridded by the bilinear interpolation method at 0.1°.

For this study, stations were selected based on strict criteria, requiring valid values for at least 99% of daily time steps within the period of 1974–2004. This selection criterion, more stringent than previous studies (Molina et al. 2021; Li et al. 2019; Dunn et al. 2022), enhances the reliability of the results at the expense of reduced spatial representation for analyzing spatial variability in detail. This approach resulted in 15 daily averaged wind speed series (Molina et al. 2021; Dunn et al. 2022; Zeng et al. 2018) at a height of 10 m, derived from the initially available stations across the South. It should be noted that the distribution of meteorological stations across South Asia is not uniform, with the highest density observed in Afghanistan, the south region of Pakistan, Nepal, Bhutan, Bangladesh, the Himalayas, and the Karakoram region. Attributable to the complex orographic, land, and coastal features of the Subcontinent, the representativeness of these observations for the entire continent is somewhat limited. However, despite this spatial limitation, the HadISD database remains one of the substantial collections of obtainable stations for South Asia (Li et al. 2019). This study will primarily focus on analyzing reanalysis data at grid points that align with the geographic locations of these observations.

To assess the overall accuracy of the reanalysis outcome, employ five statistical metrics by a direct route to compare the average daily wind speed (1974–2004) between the reanalysis datasets and weather stations. These metrics include standard deviation (SD), mean absolute error (MAE), percent bias (PBIAS), correlation coefficient (CC), and climatological mean. The location of the station is also evaluated for its spatial variability of the annual mean wind cycle. By utilizing these metrics, the differences between the reanalysis data and the actual observations can be evaluated comprehensively. The data period is 1974–2004 due to the valid values availability of observation data. It was necessary to use the three-reanalysis of climate datasets ensemble (3RE) to understand the robustness and accuracy of the results when compared to individual datasets. The study area map is present in Fig. 1 (Khan et al. 2024).

Table 1 Outline the main characteristics of the observation datasets and three global reanalysis products evaluated in this academic work

Name	HadISD	ERA5	NECP	JRA55
Full name	Hadley Integrated Surface Database (HadISD)	ECMWF Fifth Major Global Reanalysis (ERA5)	National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP)	Japanese 55-year Reanalysis (JRA55)
Institution	Met Office Hadley Centre	ECMWF	NCEP/NCAR	JMA
Duration covered	1974–2004	1980 to present	1948 to present	1978 to present
Assimilation system	Surface	4D-VAR	3D-VAR	4D-VAR
Grid spacing		0.25° latitude × 0.25° longitude	1.875° latitude × 2° longitude	1.25° latitude × 1.25° longitude
Variable levels	10 m	10 and 100 m	10 m	10 m

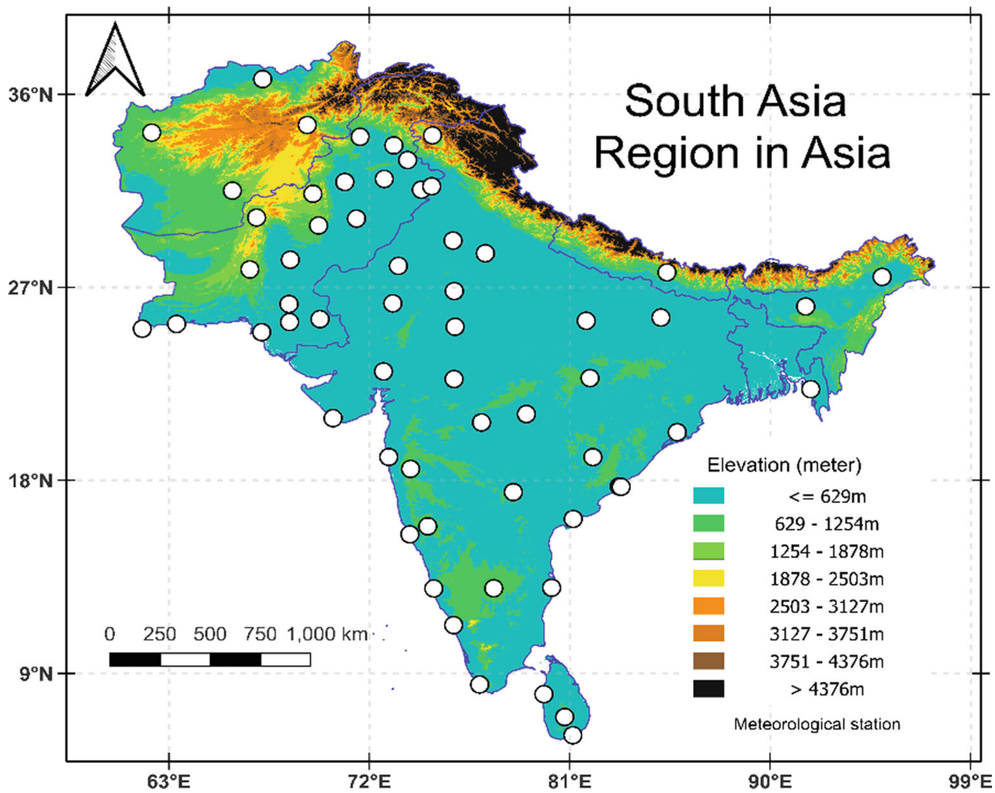

Fig. 1 Topographic map of South Asia and HadISD meteorological available station

3 Results

3.1 Daily Assessment of Reanalysis Datasets

The present study conducted the daily assessment of reanalysis datasets in the South Asia region and the purpose was to assess the accuracy and reliability in capturing real-world conditions based on five statistical metrics. When evaluating and predicting winds across the Subcontinent region, it is imperative to exercise caution when relying on reanalysis data. Statistical analysis results are demonstrated in Fig. 2. Further, the regional monthly variations in wind characteristics across the continent are illustrated in Fig. 2f for all datasets.

As shown in Fig. 2, JRA55 demonstrates the closest performance to the HadISD dataset compared to other products. For instance, JRA55 exhibits the lowest standard deviation

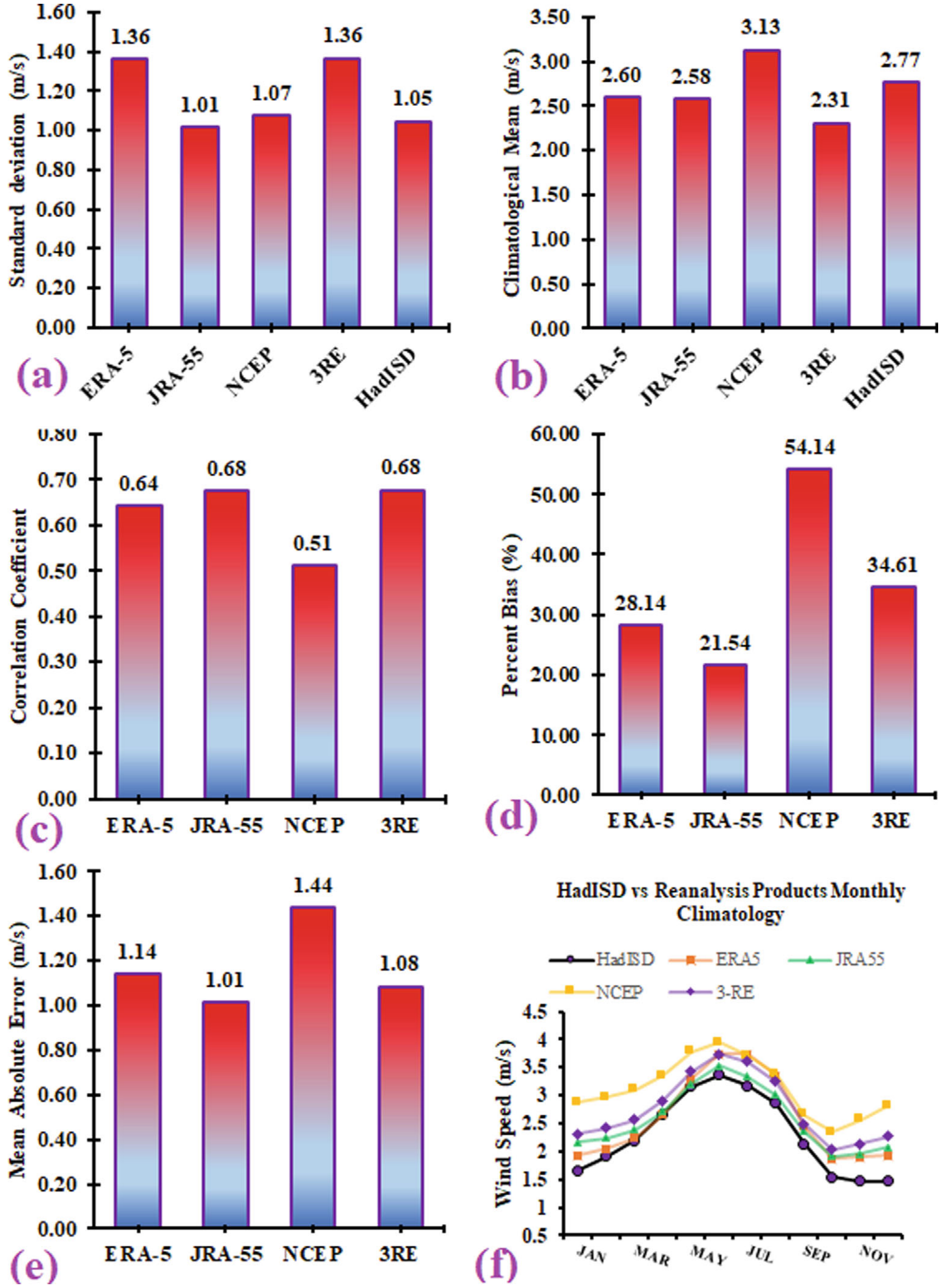

Fig. 2 Statistical daily assessment of reanalysis and observation datasets like **a** standard deviation (SD), **b** climatological mean (CM), **c** correlation coefficient (CC), **d** percent bias (PBIAS), and **e** mean absolute error (MAE). Plus, **f** monthly mean variations (1974–2004) across all datasets

(SD = 1.02 m/s), a strong positive linear relationship (correlation coefficient, CC = 0.68), signifying a small average difference between reanalysis and observational data. The lowest mean absolute error (MAE = 1.01 m/s), exhibits the closest agreement with a PBIAS of 21.54%, while other products indicate overestimation. Also, climatological mean (CM) JRA55 and ERA5 closely match HadISD with mean wind speeds of 2.58 m/s and 2.60 m/s, respectively. On the other hand, NCEP indicates a slightly higher mean wind speed (3.13 m/s), while the three-reanalysis ensemble (3RE) displays a lower value (2.31 m/s). Among the products evaluated in this study, JRA55 has the closest relationship with HadISD based on the analysis of multiple metrics.

3.2 Monthly Regional Climate Dynamics

In the phase throughout all datasets, monthly mean variations (1974–2004) were found in the analysis of wind speed data. A slight decrease in wind speed is usually observed during the end of the year (e.g., November–December), while the lowest wind speeds are usually observed in winter (e.g., January–February) and gradually increase toward summer (e.g., June–July) shown in Fig. 2f. A better understanding of regional climate dynamics can be gained from these findings provide valuable insights into seasonal patterns of wind speeds.

4 Discussion

This study, focusing on the comparison of reanalysis products (ERA5, JRA55, NCEP, and 3RE) with HadISD observational data, provides critical insights into the utility and limitations of these products for wind resource assessments and regional climate studies in South Asia. The comprehensive evaluation, guided by various statistical metrics, enhances the foundation for informed decision-making in the fields of regional wind modeling and climate research. Several important implications arise from the results of this evaluation for the assessment of wind resources and climate studies in South Asia. There are strong correlations between JRA55 reanalysis results and observational data across a variety of metrics, making it a promising reanalysis product. The monthly variation in wind speed data across all datasets indicates a consistent pattern of lower wind speeds during winter months, increasing through summer, followed by a minor decline toward the end of the year, providing valuable insights into seasonal wind dynamics in South Asia. Observations from previous studies (Khan et al. 2018; Khan and Dairaku 2024; Molina et al. 2021; Khan et al. 2025; Wenxin et al. 2021) are supported by the findings of this study. The resulting data is likely to be useful for applications requiring accurate wind speed measurements due to its low SD, high correlation, and small MAE and PBIAS values. Study objectives should, however, be considered when selecting a reanalysis product. JRA55 followed by ERA5 may be a good choice for researchers seeking high accuracy with precision, provided they weigh the trade-offs associated with the performance metrics. However, when considering both its strengths and limitations, NCEP could still have value for certain applications, despite its lower performance.

5 Conclusions

The study concludes with a comprehensive assessment of the performance of multiple reanalysis products representing wind characteristics in South Asia. Atop several metrics, JRA55 performed best. However, the specific requirements of the study should determine the choice of reanalysis product. By enhancing wind resource assessments and understanding regional climate dynamics, these findings contribute to the advancement of wind resource assessments. We compare reanalysis products with observed data to evaluate and compare their reliability. This demonstrates the need to be cautious when evaluating and predicting winds in the Subcontinent region based on reanalysis data.

A holistic understanding of reanalysis product performance requires additional metrics and spatial–temporal analyses to be considered. The evaluation presented here is only a part of the broader research.

References

Dunn R, Azorin-Molina C, Menne MJ, Zeng Z, Casey NW, Shen C (2022) Reduction in reversal of global stilling arising from correction to encoding of calm periods. Environ Res Commun 4(6):061003

Khan MA, Çamur H, Kassem Y (2018) Modeling predictive assessment of wind energy potential as a power generation sources at some selected locations in Pakistan. Model Earth Syst Environ 5(2):555–569

Khan MA, Dairaku K (2024) Assessment of the potential of renewable energy with bias correction due to climate change over South Asia using global atlas dataset. In: Chenchouni H et al (eds) Recent advancements from aquifers to skies in hydrogeology, geoecology, and atmospheric sciences. MedGU 2022. Advances in science, technology & innovation. Springer, Cham. https://doi.org/10.1007/978-3-031-47079-0_67

Khan MA, Dairaku K, Kelkar S (2024) Evaluation of wind speed accuracy enhancement in South Asia through terrain-modified wind speed (Wt) adjustments of high-resolution regional climate modeling. Earth Syst Environ 8:1777–1794. https://doi.org/10.1007/s41748-024-00453-6

Khan MA, Dairaku K, Kelkar S (2025) Assessing wind power generation potential over South Asia using wind speed observation and reanalysis datasets. Stoch Environ Res Risk Assess 39:1179–1207. https://doi.org/10.1007/s00477-025-02918-0

Li D, Feng J, Xu Z, Yin B, Shi H, Qi J (2019) Statistical bias correction for simulated wind speeds over CORDEX-East Asia. Earth Space Sci 6(2):200–211

Molina MO, Gutiérrez C, Sánchez E (2021) Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. Int J Climatol 41(10):4864–4878

Wenxin F, Liu Y, Chappell A, Dong L, Xu R, Ekström M, Fu T, Zeng Z (2021) Evaluation of global reanalysis land surface wind speed trends to support wind energy development using in situ observations. J Appl Meteorol Climatol 60(1):33–50

Zeng Z, Piao S, Li L, Ciais P, Li Y, Cai X, Yang L, Liu M, Wood EF (2018) Global terrestrial stilling: does Earth's greening play a role? Environ Res Lett 13(12):124013