Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 5 (2026) 610

African Scientific Reports

Comparative evaluation of ARIMA and ETS models for short-term wind speed forecasting across twelve meteorological stations in Pakistan using the expanding-window method

Muhammad Aslam[a,*], Naeem Ullah Khan[a], Atta Ullah[b]

[a]*Department of Mathematics and Statistics, Institute of Business Management, Karachi, Pakistan*
[b]*Department of Mathematics, COMSATS University Islamabad, Attock Campus, Attock, Pakistan*

Abstract

This article presents a statistical forecasting framework for wind speed using two well-known classical forecasting models, the autoregressive integrated moving average (ARIMA) and the exponential smoothing state space (ETS) models with the naïve model as a benchmark for comparison. Twelve meteorological stations across Pakistan were selected based on data availability and analyzed using a comprehensive and rigorous expanding-window data-splitting method. The expanding window method iteratively extends the training data over ten successive splits, with each split used to generate seven-day-ahead wind speed predictions. This approach allows the forecasting performance of the models to be evaluated over multiple forecast origins while progressively incorporating newly available observations into the training data. Root mean square error (RMSE) and mean absolute percentage error (MAPE) were used to examine the predictive behavior of all models and to evaluate their consistency and responsiveness over time. Model comparisons were further assessed by the Diebold-Mariano (DM) test on forecast errors for each model. The ARIMA model performed better (based on RMSE) for short-term wind speed forecasting at seven stations in Pakistan while the ETS model performed better at two stations and performed equally at three stations covered by this study. The DM test also supports the RMSE- and MAPE-based results in favor of the ARIMA model. The findings demonstrate the generally stronger performance of the ARIMA model.

DOI: 10.46481/asr.2026.5.3.610

Keywords: Wind speed, ARIMA, ETS, Expanding-window validation, Forecast accuracy

Article History:
Received: 22 July 2026
Received in revised form: 29 August 2026
Accepted for publication: 29 September 2026
Available online: 29 October 2026

1. Introduction

Today, wind power and other forms of renewable energy are the fastest-growing energy sources in the world. The renewable energy produced by wind speed has numerous advantages over other sources of energy. Accurate statistical modelling and forecasting of wind speed are challenging due to the stochastic and non-stationary behaviour of wind speed. A variety of statistical and time series models have been used to model and forecast wind speed. Autoregressive Integrated Moving Average (ARIMA) model, and its counterpart seasonal autoregressive integrated moving average (SARIMA) and the Exponential Smoothing State Space (ETS) framework, represent foundational models that continue to serve as benchmarks against more complex machine learning and hybrid approaches.

Pakistan has diverse geographical locations that are suitable for this renewable energy. The offshore areas of Sindh and Balochistan, and windy inland areas of Punjab and Khyber Pakhtunkhwa are favourable sites to utilise for renewable energy. Despite the

*Corresponding Author Tel. No.: +92-343-519-7632.
Email address: dr.aslam@iobm.edu.pk (Muhammad Aslam)

diverse geographical locations and literature on wind speed forecasting, comparative evaluations of wind energy modeling specifically for Pakistani meteorological contexts remain scarce. Furthermore, conventional train-test splitting strategies do not accommodate/update the new observations while processing. It often fails to replicate realistic deployment conditions as new observations arrive continuously. The expanding window method, a form of time series cross-validation, overcomes this limitation of the conventional data split method while using ARIMA and ETS models. This method progressively incorporates new data into the training set while running ARIMA and ETS models, to update forecasts. Using the expanding-window method for short-term wind forecasting, the classical time series models: ARIMA and ETS are used to forecast the wind speed in different stations of Pakistan.

The section-by-section organisation of the article is presented here. Section 2 incorporates the most relevant and advanced literature related to wind speed time series forecasting. Section 3 begins with the description of the data and their characteristics along with the methodology and its formulation. Section 4 presents the results and comparisons. Section 5 summarises the paper with key points and recommendations for future research.

2. Literature review

In recent years, wind energy potential in Pakistan has been gaining considerable focus in academia. Accurate wind speed forecasting is fundamental to the effective operation of wind farms, integration with the grid and investment in renewable energy. Wind speed in a given area is affected by various phenomena, such as large-scale atmospheric circulation patterns, local topography, sea–land thermal contrasts and seasonal cycles. These multi-scale drivers lead to non-stationarity and intermittency in wind-speed time series and make the use of conventional linear models difficult. However, statistical time series models have been found to be competitive with more complex machine-learning models, particularly when data are limited and interpretability is desired [1]. The present study emphasizes short-term forecasting, which is of special importance in unit commitment, economic dispatch and energy market bidding.

One of the initial detailed analyses of wind energy potential in Pakistan was carried out by Bhutto *et al.* [2] who found that major wind energy resources are available in the coastline of Sindh and Balochistan plateau. Box *et al.* [3] provide a full account of the techniques of model identification, estimation and the checking of the models. In wind energy applications, the seasonal ARIMA (SARIMA) extensions in Ref. [4] have been used to consider the periodicity of wind speed both diurnal and annual. Erdem and Shi [5] used ARIMA models to wind speed data collected at a turbine site and concluded that the accuracy of the models become satisfactory for short forecast horizons; however, the models perform poorly for longer time horizons. Cadenas and Rivera [6] showed the usefulness of ARIMA model in wind speed prediction with 24 hours ahead in Mexico and obtained good MAPE values. Recently, hybrid ARIMA models with neural network residual correction have been suggested to overcome the linearity limitation of the ARIMA approach by Zhang [7].

On the other hand, the exponential smoothing techniques give exponentially decreasing weights to the previous observations and have been in use for time series forecasting for a long time. The ETS framework introduced by Hyndman *et al.* [8] and expanded in Ref. [9] combines the three components of error, trend and seasonality into a single state space model and enables model selection through information criteria. Empirical studies demonstrate the ETS models have competitive forecasting performance against ARIMA. Simple exponential smoothing models tend to perform better than more involved models in a variety of time series, as shown in Ref. [10] in the M3 Competition. Wang et al. [11] used ETS models for renewable energy forecasting, and observed that simpler error structures often are more general when training data is scarce. ETS models may have some benefits when it comes to the adaptive weighting due to level shifts and smooth trends in wind speed series.

In time series forecasting, the expanding-window evaluation (also called rolling-origin validation) is preferred over traditional cross-validation, because it maintains the time order of the data and avoids data leakage in Refs. [12, 13]. González-Enrique *et al.* [14] obtained good results in NO_2 prediction by implementing a time series cross-validation (LSTM-CVT) scheme for rolling window. This approach was further used by [15] for PM10 forecasting in Metropolitan Lima, with a rolling window of 7 days forecasting for a ten days period. Similarly, García Crespi *et al.* [16] demonstrated the capability of the rolling-origin validation method for modifying the model ranking in multi-step PM10 forecasting.

This approach gives multiple performance estimates, and facilitates checking the stability of the model with time and minimises the variance of the aggregate performance measures. The out-of-sample testing of the time series is examined in great detail in Ref. [13], where they outline the advantages of the so-called rolling window or expanding window testing schemes for realistic performance evaluation of the time series. Hu *et al.* [17] were interested in wind speed forecasting, and tested machine learning models using a sliding window approach, showing the added value of multiple evaluation periods compared to one train-test split. Similar discussions were presented in more detail by Adejumo *et al.* [18], Echi *et al.* [19], Moyi *et al.* [20], Afere *et al.* [21], and Thorat *et al.* [22].

3. Materials and methods

Wind speed data were obtained for twelve meteorological stations distributed across Pakistan and selected to represent the country's diverse climatic and topographic zones. The stations span several provinces including Sindh, Punjab, Khyber Pakhtunkhwa and

Table 1: Descriptive statistics of the wind-speed data.

Station	n	Mean	SD	Min	Q1	Median	Q3	Max
Bahawalpur	20758	3.624	1.673	0.0004	2.514	3.374	4.409	15.003
Chakri	22184	2.469	1.513	0.212	1.419	2.102	3.086	15.944
Gwadar	15731	3.917	2.359	0.098	2.110	3.354	5.274	14.181
Haripur	19808	3.058	1.390	0.236	1.962	2.814	4.163	9.956
Peshawar	17268	2.818	1.333	0.317	1.861	2.611	3.520	12.651
Quaid Abad	17258	2.986	1.589	0.233	1.901	2.688	3.688	15.015
Quetta	19596	3.340	1.796	0.226	1.971	2.958	4.486	13.106
Sadiq Abad	20375	3.414	1.630	0.181	2.262	3.129	4.247	14.457
Sanghar	18288	4.846	2.390	0.364	3.118	4.361	6.325	15.400
Sujawal	20908	5.577	2.337	0.304	3.877	5.230	7.103	15.183
TandoGA	21550	5.078	2.421	0.388	3.225	4.574	6.683	14.804
Umerkot	18107	4.617	2.357	0.312	2.890	4.120	6.046	14.219

Balochistan, capturing coastal, semi-arid, arid, and highland wind regimes. Station selection was guided by data availability, record completeness, and geographic representativeness to ensure that findings can inform national-level wind resource assessment.

Two models, ARIMA and ETS, were fitted to forecast the wind speed data along with naïve model as a benchmark while using expanding-window splitting. Forecasting performance was evaluated for each split using root mean square error (RMSE) and mean absolute percentage error (MAPE) across all stations. A model is considered to be a good performer at station level for each split when it has a lower value of RMSE and MAPE. For each station, lower error values in at least 6 of the 10 splits is considered to be better performer. The Diebold–Mariano (DM) test was applied to the paired out-of-sample forecast errors generated during the expanding-window evaluation. The square error loss function ($d_t = e^2_{ARIMA,t} - e^2_{ETS,t}$) was used where negative loss favor ARIMA and positive loss favor ETS. The two tailed DM test was run for seven-day-ahead forecast horizon by specifying equal predictive accuracy under the null hypothesis against unequal predictive accuracy with 5% level of significance. ARIMA was the better-performing model according to the predefined majority-of-splits criterion and the DM test.

Moreover, for each expanding-window split, the selected ARIMA and ETS model orders were noted along with their residual diagnostics. The Ljung–Box test was used to further assess the residual autocorrelation in both models. The Ljung–Box test states the null hypothesis that there is no autocorrelation (white noise) against the alternative hypothesis that there is autocorrelation. The test indicated statistically significant residual autocorrelation for both models (p-value < 0.05), suggesting that some serial dependence remained unexplained by the fitted models at each split. The selected models' (ARIMA and ETS) specification along with diagnostic results are presented in Table A.17.

3.1. ARIMA model

The $ARIMA(p, d, q)$ model for a time series z_t is defined by Eq. (1):

$$\gamma_p(\beta)(1-\beta)^d z_t = \vartheta_q(\beta)\varepsilon_t, \tag{1}$$

where $\gamma_p(\beta) = 1 - \gamma_1\beta^1 - \gamma_2\beta^2 -, \ldots, -\gamma_p\beta^p$, is the autoregressive polynomial of order p, $\vartheta_q(\beta) = 1 + \vartheta_1\beta^1 + \vartheta_2\beta^2 +, \ldots, +\vartheta_q\beta^q$, is the moving average polynomial of order q and β is the backshift operator and equal to $\beta z_t = z_{t-1}$. Also, β^d denotes differencing of order d and ε_t is white noise.

The orders (p, d, q) were selected automatically for each station and each training expanding window using the *auto.arima()* function in the "*forecast*" package in R. The auto.arima() function employs a stepwise search using the Akaike information criterion (AIC) following unit root tests (KPSS and ADF) to determine the differencing order d.

3.2. ETS model

The ETS(E, T, S) framework specifies the error (E), trend (T), and seasonal (S) components of a time series, with each component classified as none (N), additive (A), or multiplicative (M) where applicable. In this empirical study, the `ets()` function in the R forecast package was used with model = "ZZZ", allowing the error, trend, and seasonal components to be selected automatically based on AICc for each expanding training sample.

The hourly series data was assigned a frequency of 720, corresponding to a seasonal period of 720 observations (30 days). The seasonal frequency was therefore distinct from the seven-day forecasting horizon of 168 hourly observations.

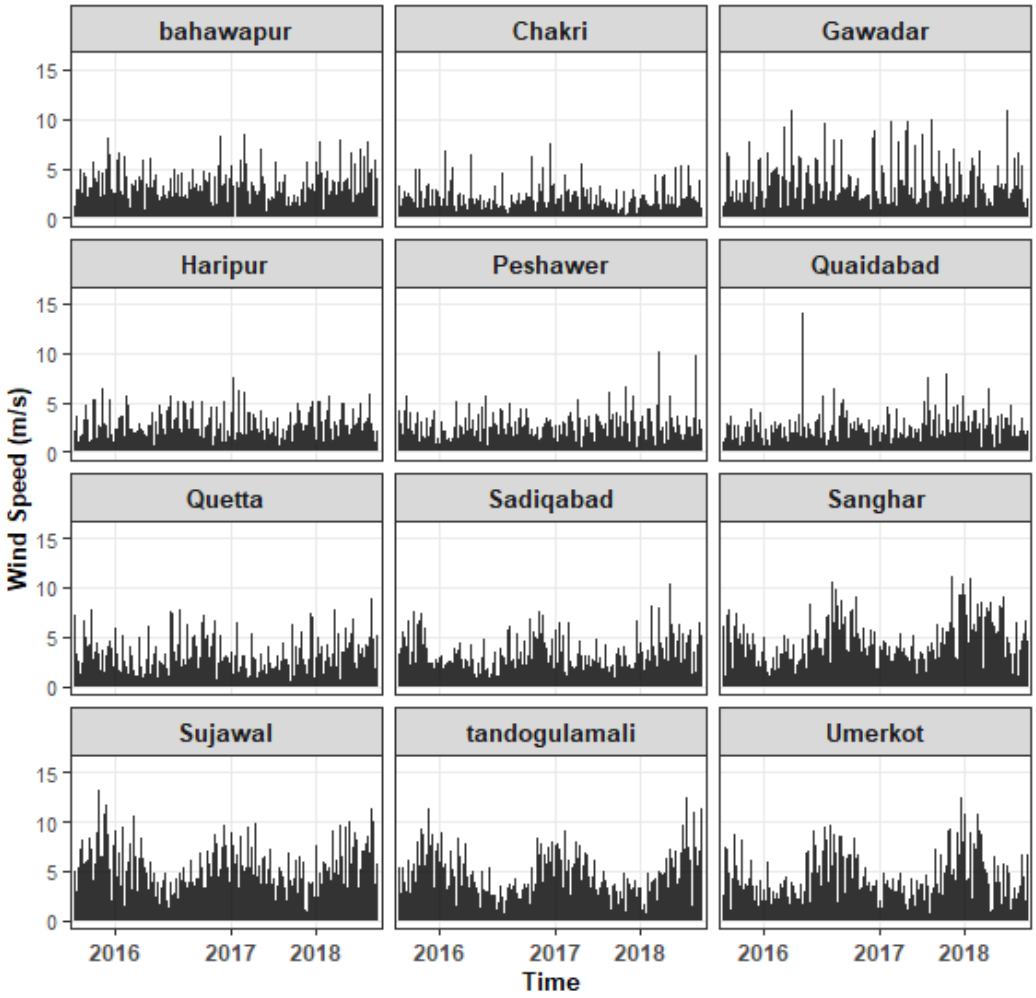


Figure 1: Station-wise wind-speed plots.

3.3. Expanding-window cross-validation

To evaluate the out-of-sample forecasting performance of both models under realistic conditions, an expanding-window cross-validation scheme was implemented. For each station, the full time series was partitioned into ten successive training–test splits. In the first split, a minimum number of observations constituted the initial training set. In each subsequent split, the training window was expanded by a fixed number of observations, while the test set always comprised the immediately following seven days. This design ensures temporal ordering is strictly preserved and that no future information contaminates the training data.

Let T denote the total length of the series. For split k ($k = 1, 2, \ldots, 10$), the training set comprises observations $z_1, z_2, z_3, \ldots, z_{n^*}$, where n^* increases monotonically with k, and the test set comprises $z_{(n^*+1)}, \ldots, z_{(n^*+7)}$. Both ARIMA and ETS models were re-estimated from scratch for each training window, allowing the model orders and parameters to adapt to the growing information set. Seven-step-ahead point forecasts were generated from each estimated model for each split.

The expanding-window evaluation used an initial training window of 20758 hourly observations (for example, Bahawalpur station). Since the objective was 7-day-ahead forecasting, each forecast horizon comprised 168 hourly observations (7×24). After each forecast period, the training window was expanded by 168 observations. The ten forecast blocks were non-overlapping. The details are given in the supplementary tables (Table A.16). Similarly, the same procedure was performed for other stations.

3.4. Performance evaluation metrics

Forecasting accuracy was assessed using two complementary metrics. The root mean square error (RMSE) penalizes larger errors more heavily due to the squared term and is expressed in the original units of wind speed. The mean absolute percentage error (MAPE) provides a scale-free measure of relative forecast accuracy. The $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2}$ and $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{O_i - E_i}{O_i}\right| \times 100$, where O_i is the i^{th} actual value and E_i is i^{th} the predicted value from the model. Both metrics

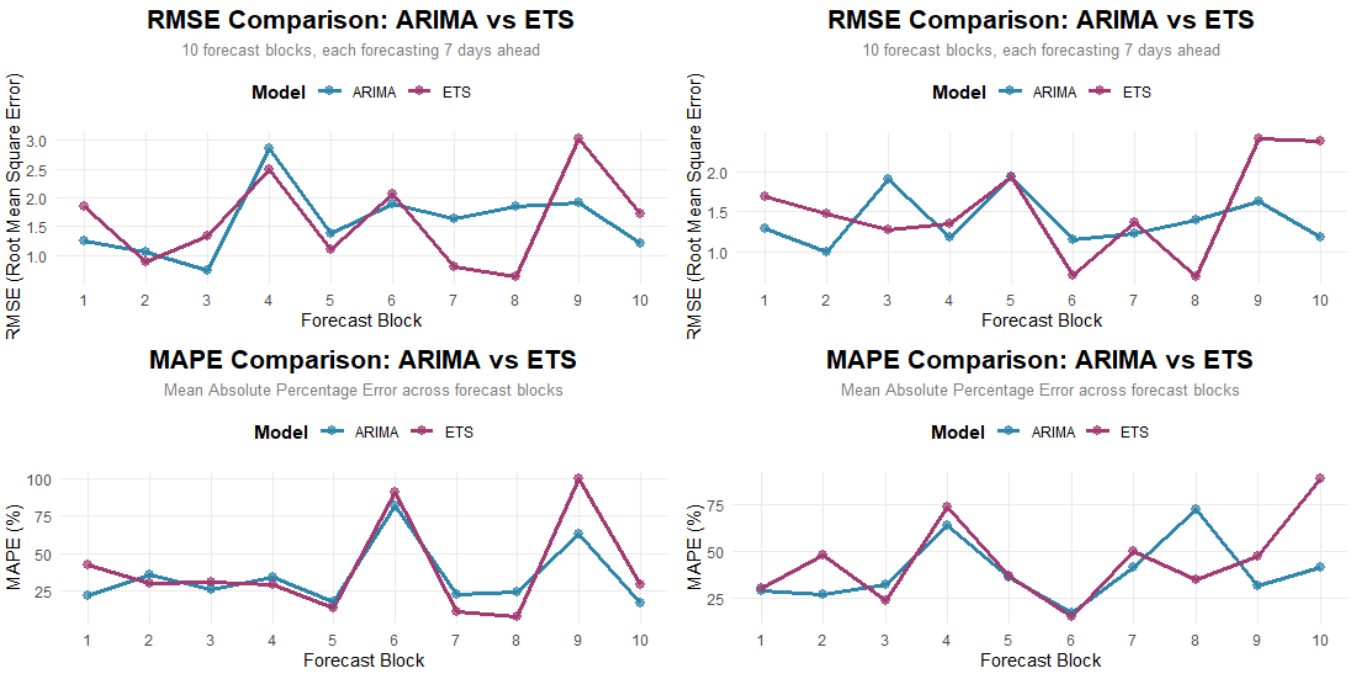


Figure 2: RMSE and MAPE for Bahawalpur (left) and Haripur (right).

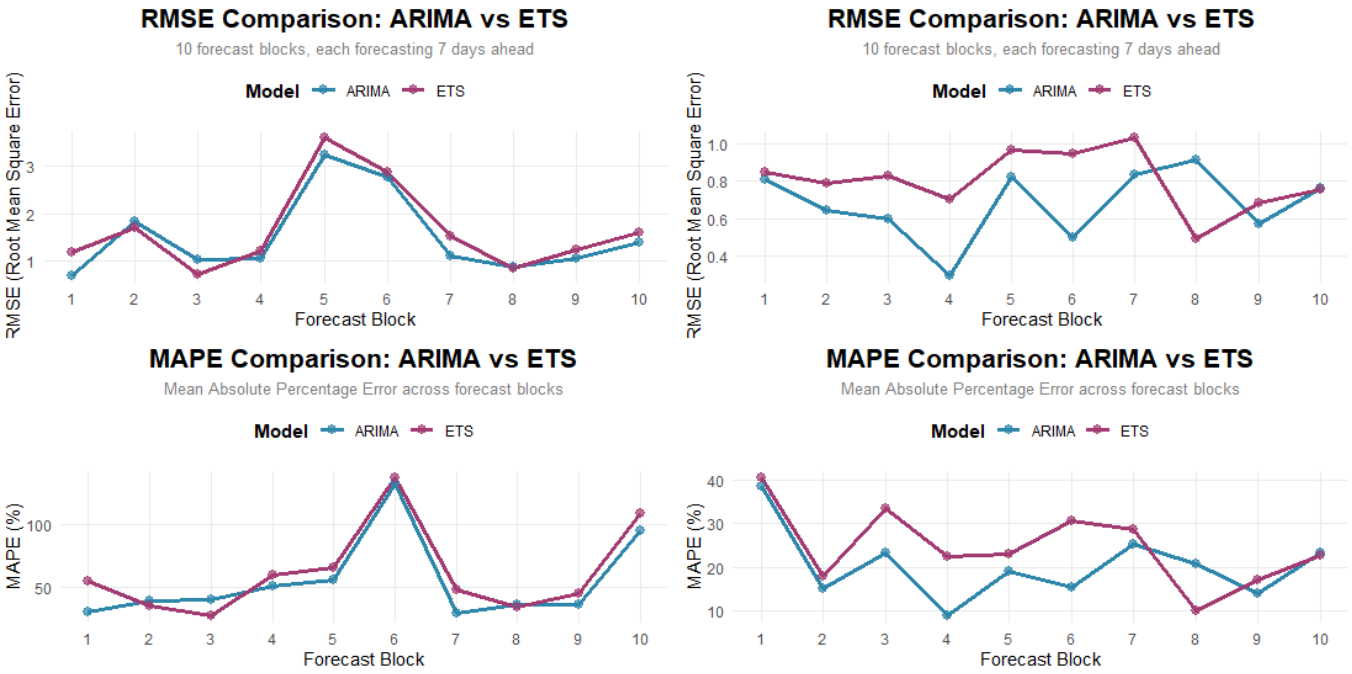


Figure 3: RMSE and MAPE for Peshawar (left) and Quaid Abad (right).

were computed for each of the ten splits for each station and each model, yielding a matrix of performance values that allows assessment of within-split variation across the seven forecast steps (within-prediction-horizon performance) as well as across-split variation reflecting model stability over time.

3.5. Software implementation

All analysis was conducted in R (version 4.5.2) using the "*forecast*" package (Hyndman et al., 2023) for model fitting and prediction, the "*tseries*" package for unit root testing, and base R functions for performance metric computation. Reproducibility

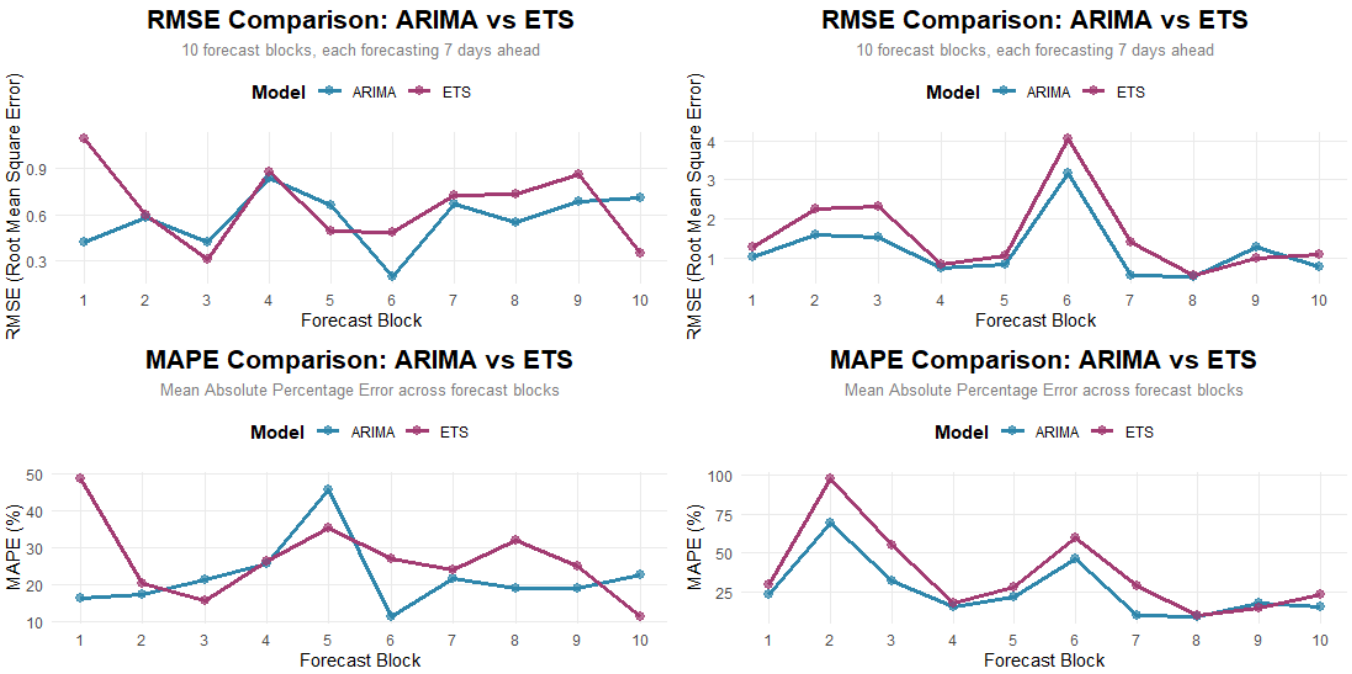


Figure 4: RMSE and MAPE for Chakri (left) and Sadiq Abad (right).

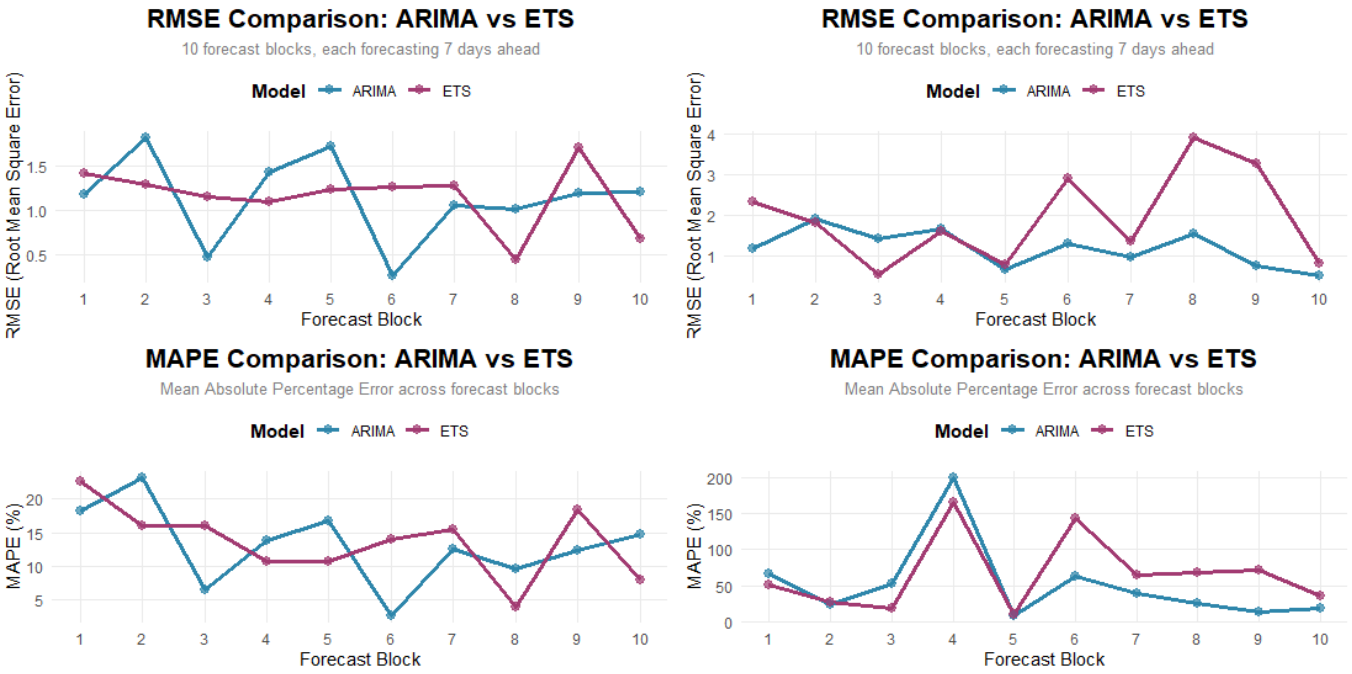


Figure 5: RMSE and MAPE for Sanghar (left) and Gwadar (right).

was supported by setting random seeds where applicable and documenting all parameter choices. Visualization of expanding window performance trajectories was performed using the "*ggplot2"* package.

4. Results and discussion

The main source of data is the NASA POWER database, which is reputed to have good, consistent, and long-term atmospheric records. This study uses hourly wind-speed data measured at the standard height of 20 meters above ground level over a 2.37-year period (13/4/2016 to 31/8/2018). The diurnal cycle of wind speeds at the hourly resolution is especially taken when hourly resolution

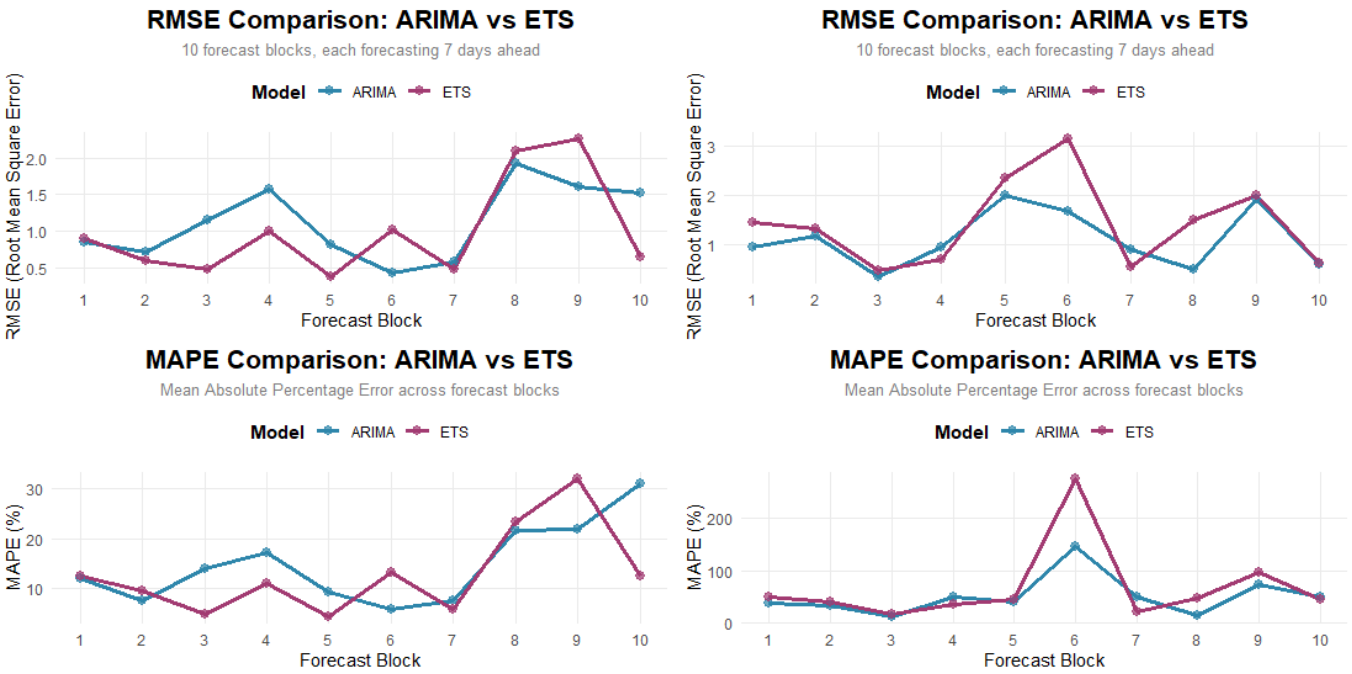


Figure 6: RMSE and MAPE for Umerkot (left) and Quetta (right).

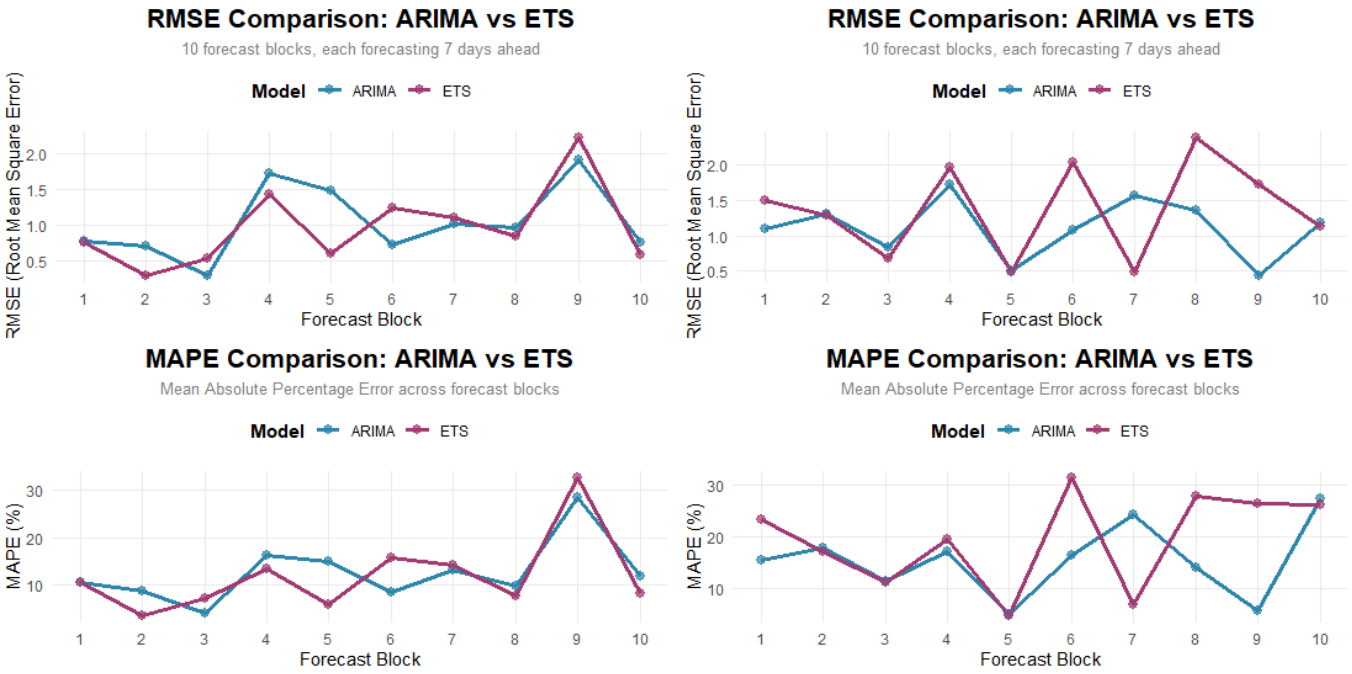


Figure 7: RMSE and MAPE for TandoGA (left) and Sujawal (right).

is chosen, as in coastal and valley sites where the sea breeze and mountain valley systems of winds are very strong and work on hourly scale. This 20 m height is a compromise between the standard meteorological measurement height of 10 m and the rising heights of the modern wind turbine hubs which are standard nearing between 80 and 120 m. No interpolation or other missing-value replacement procedure was used, because no missing observations were identified.

In particular, high wind-speed observations were retained because they may represent genuine meteorological events rather than erroneous measurements. Descriptive statistics including max, min, mean, standard deviation (SD), median, first quartile (Q1) and third quartile (Q3) were computed for each station series to characterize the distributional properties of the wind speed data and presented in Table 1. Similarly, the graphical presentation of the data for each station is presented in Figure 1. We further demonstrate

Table 2: Forecasting performance for Bahawalpur and Sujawal using RMSE and MAPE.

Bahawalpur						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	1.259	21.49	1.857	42.03	1.857	42.028
2	1.058	35.93	0.892	29.74	0.892	29.737
3	0.744	25.87	1.346	30.56	1.346	30.557
4	2.859	34.32	2.498	29.22	2.498	29.216
5	1.378	17.59	1.105	12.97	1.105	12.974
6	1.892	82.21	2.069	91.01	2.069	91.006
7	1.633	22.2	0.808	10.53	0.808	10.529
8	1.851	24.16	0.621	7.19	0.621	7.187
9	1.919	63.01	3.025	99.74	3.025	99.741
10	1.211	16.43	1.723	28.65	1.723	28.651
Sujawal						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	1.168	18.27	1.417	22.63	1.505	23.39
2	1.811	23.18	1.287	15.94	1.283	17.14
3	0.464	6.36	1.145	16.02	0.678	11.41
4	1.422	13.84	1.094	10.6	1.965	19.61
5	1.719	16.76	1.233	10.75	0.484	4.834
6	0.252	2.47	1.253	13.93	2.048	31.29
7	1.045	12.5	1.267	15.56	0.494	7.129
8	1.009	9.53	0.438	3.85	2.384	27.92
9	1.192	12.27	1.707	18.52	1.735	26.34
10	1.204	14.78	0.673	7.94	1.134	26.1

Table 3: Forecasting performance for Chakri and Gwadar using RMSE and MAPE.

Chakri						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	0.424	16.13	1.091	48.48	1.052	46.657
2	0.583	17.33	0.598	20.09	0.583	19.102
3	0.423	21.2	0.313	15.66	0.296	14.591
4	0.836	25.43	0.874	26.07	0.901	25.431
5	0.662	45.54	0.496	35.21	0.445	29.05
6	0.196	11.06	0.486	26.93	0.723	42.27
7	0.667	21.4	0.724	23.81	0.822	28.536
8	0.55	19.01	0.733	31.84	0.716	31.871
9	0.687	18.94	0.859	24.81	0.895	26.9
10	0.713	22.66	0.352	11.34	0.338	9.966
Gwadar						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	1.167	66.77	2.331	51.28	2.437	52.4
2	1.89	24.26	1.815	26.77	1.732	23.72
3	1.422	52.79	0.532	18.36	1.448	57.69
4	1.655	199.05	1.6	165.3	1.6	164.6
5	0.676	7.7	0.79	10.38	1.89	26.28
6	1.292	63.24	2.899	143.6	3.351	165.9
7	0.971	38.7	1.372	64.49	0.772	32.77
8	1.535	24.78	3.887	67.37	3.454	57.95
9	0.762	12.64	3.261	70.88	2.294	50.25
10	0.5	19.35	0.806	35.6	0.828	47.87

the results of the models. The most convenient and intuitive way of presentation of the result was used to depict the performance of the models across each station. The graphical presentation shows the performance of both models using RMSE and MAPE for each

Table 4: Forecasting performance for Peshawar and Quaid Abad using RMSE and MAPE.

Peshawar						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	0.679	30.55	1.192	54.75	1.241	57.216
2	1.84	39.26	1.696	35.04	1.693	34.95
3	1.027	40.3	0.717	27.31	0.701	26.631
4	1.045	51.14	1.21	59.93	1.223	60.603
5	3.247	56.11	3.6	66.32	3.614	66.779
6	2.778	132.67	2.879	137.52	2.833	135.344
7	1.102	30.07	1.521	48.35	1.537	49.035
8	0.869	36.57	0.848	34.39	0.848	34.392
9	1.061	36.32	1.232	45.78	1.23	45.676
10	1.398	95.89	1.602	109.37	1.651	112.88
Quaid Abad						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	0.805	38.63	0.844	40.44	0.813	38.81
2	0.644	15.15	0.786	17.99	0.793	18.22
3	0.599	23.34	0.829	33.59	0.795	32.04
4	0.292	8.79	0.702	22.43	0.742	23.73
5	0.823	18.93	0.967	22.96	0.973	23.15
6	0.497	15.35	0.945	30.59	0.934	30.16
7	0.831	25.35	1.03	28.77	1.046	28.91
8	0.915	20.74	0.492	9.98	0.504	10.29
9	0.57	13.84	0.686	16.94	0.662	16.11
10	0.762	23.35	0.758	22.81	0.765	22.95

Table 5: Forecasting performance for Sanghar and Haripur using RMSE and MAPE.

Sanghar						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	1.097	15.63	1.505	23.39	1.417	22.629
2	1.316	17.94	1.283	17.14	1.288	15.944
3	0.846	11.54	0.678	11.41	1.145	16.019
4	1.728	17.2	1.965	19.61	1.094	10.602
5	0.512	5.09	0.484	4.83	1.233	10.755
6	1.082	16.55	2.048	31.29	1.253	13.93
7	1.575	24.31	0.494	7.13	1.267	15.563
8	1.363	14.2	2.384	27.92	0.438	3.847
9	0.43	5.96	1.735	26.34	1.707	18.521
10	1.186	27.45	1.134	26.1	0.673	7.935
Haripur						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	1.297	28.5	1.681	30.28	1.689	30.481
2	1.002	26.84	1.478	47.7	1.434	46.238
3	1.895	32.1	1.275	23.49	1.268	23.39
4	1.182	63.72	1.36	73.08	1.299	69.674
5	1.934	35.67	1.925	36.59	1.899	36.518
6	1.159	16.84	0.712	14.73	0.711	14.743
7	1.238	41.32	1.362	49.71	1.322	47.952
8	1.4	71.93	0.691	34.95	0.69	34.894
9	1.621	31.1	2.401	47.19	2.335	45.838
10	1.183	41.04	2.37	88.27	2.397	89.29

section defined through expanding window method. The exact and concise results of both models for each station have also been presented in Tables.

Table 6: Forecasting performance for Quetta and Sadiq Abad using RMSE and MAPE.

Quetta						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	0.96	39.28	1.447	49.53	1.447	49.529
2	1.191	33.59	1.33	41.65	1.33	41.643
3	0.352	11.61	0.492	18.04	0.492	18.041
4	0.96	49.74	0.7	36.01	0.7	36.007
5	2.01	39.57	2.356	46.4	2.356	46.396
6	1.672	145.78	3.133	275.6	3.133	275.6
7	0.918	49.47	0.55	22.95	0.55	22.948
8	0.499	14.36	1.502	47.4	1.502	47.4
9	1.918	74.54	1.986	96.66	1.986	96.663
10	0.61	51.11	0.633	44.33	0.633	44.328
Sadiq Abad						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	1.033	23.38	1.29	29.28	1.29	29.283
2	1.599	68.81	2.251	97.42	2.251	97.425
3	1.519	31.91	2.305	55.12	2.305	55.124
4	0.733	15.47	0.839	17.81	0.839	17.811
5	0.823	21.15	1.044	27.66	1.044	27.663
6	3.175	46.14	4.048	59.63	4.048	59.636
7	0.552	9.54	1.395	28.81	1.394	28.803
8	0.506	8.46	0.541	9.2	0.541	9.204
9	1.27	17.52	1.005	13.96	1.005	13.961
10	0.766	15.05	1.078	23.27	1.078	23.265

Table 7: Forecasting performance for TandoGA and Umerkot using RMSE and MAPE.

TandoGA						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	0.78	10.53	0.772	10.52	0.772	10.519
2	0.706	8.59	0.293	3.29	0.293	3.293
3	0.304	3.83	0.539	7.15	0.539	7.149
4	1.732	16.11	1.437	13.24	1.437	13.243
5	1.492	14.92	0.61	5.7	0.61	5.703
6	0.727	8.48	1.254	15.8	1.254	15.797
7	1.031	13.15	1.108	14.11	1.108	14.106
8	0.973	9.77	0.845	7.6	0.845	7.602
9	1.922	28.54	2.225	32.63	2.225	32.625
10	0.763	11.75	0.589	7.99	0.59	7.994
Umerkot						
Split	A RMSE	A MAPE	E RMSE	E MAPE	N RMSE	N MAPE
1	0.848	11.97	0.896	12.5	0.896	12.504
2	0.718	7.56	0.597	9.48	0.597	9.479
3	1.16	13.87	0.487	4.79	0.487	4.787
4	1.579	17.29	0.999	11.03	0.999	11.028
5	0.826	9.15	0.374	4.21	0.374	4.207
6	0.436	5.69	1.02	13.21	1.02	13.214
7	0.585	7.56	0.48	5.8	0.48	5.8
8	1.925	21.53	2.09	23.44	2.09	23.439
9	1.607	21.92	2.26	31.99	2.261	31.992
10	1.526	30.99	0.651	12.59	0.652	12.59

Figures 2–7 show the graphical comparison of the ARIMA and ETS models based on RMSE and MAPE against ten forecast blocks for each station. The performance of the ARIMA and ETS varies in each of the ten forecast blocks. This performance of the

Table 8: 95% coverage and mean interval width over ten splits for ARIMA, ETS and naïve models at each station.

Station	Model	Coverage (%)	Width
Bahawalpur	ARIMA	88.57	5.223
	ETS	91.43	6.771
	Naive	91.43	6.772
Chakri	ARIMA	100	5.031
	ETS	100	6.039
	Naive	100	7.368
Gwadar	ARIMA	100	6.882
	ETS	100	10.456
	Naive	100	8.015
Haripur	ARIMA	94.29	4.768
	ETS	94.29	7.317
	Naive	95.71	7.598
Peshawar	ARIMA	88.57	5.795
	ETS	94.29	7.175
	Naive	95.71	7.387
Quaid Abad	ARIMA	100	4.973
	ETS	100	6.503
	Naive	100	6.852
Quetta	ARIMA	97.14	5.631
	ETS	98.57	7.347
	Naive	98.57	7.347
Sadiq Abad	ARIMA	88.57	4.869
	ETS	90	6.302
	Naive	90	6.303
Sanghar	ARIMA	98.57	4.789
	ETS	100	5.75
	Naive	100	5.75
Sujawal	ARIMA	97.14	4.748
	ETS	94.29	5.508
	Naive	94.29	5.508
TandoGA	ARIMA	98.57	4.445
	ETS	98.57	5.095
	Naive	98.57	5.095
Umerkot	ARIMA	95.71	4.61
	ETS	95.71	5.452
	Naive	95.71	5.452

models is further clarified by numerical values and presented in tables.

Tables 2–7 show the forecasting performance for each of the ten rolling splits (groups) based on RMSE and MAPE of both models for all twelve stations. Based on the average values of RMSE and MAPE, ARIMA and ETS models perform equally for Bahawalpur, Sanghar and Sujawal stations. At Chakri, Gwadar, Haripur, Peshawar, Quaid Abad, Quetta and Sadiq Abad stations, the ARIMA model performs better than the ETS model. Table 7 presents the results for TandoGA and Umerkot stations, where ETS performs better than ARIMA model. Based on RMSE and MAPE, out of the twelve stations analyzed, ARIMA showed stronger forecasting performance at seven, making it the more consistent model overall. ETS, on the other hand, gave better results at two stations, and at the three stations, both models performed at a similar level.

The naïve forecasting method was included as a benchmark to compare the predictive performance of the two models. The RMSE and MAPE of the Naïve model for each station are also given in Tables 2–7. Based on RMSE, ARIMA performs better than the naïve model at a similar number of stations as ETS model. These patterns point to the general strength of ARIMA in handling the autocorrelated nature of wind speed data, while also showing that ETS can still hold its own at certain locations, likely due to differences in local wind behavior or seasonal patterns unique to those sites.

Interval coverage and width are summarized in Table 8, and the Diebold–Mariano results are shown in Table 9.

The 95% CI coverage and mean interval width of ARIMA, ETS and Naïve models for all stations are given in Table 8. This shows that ARIMA has narrower prediction intervals than ETS model across all twelve stations. However, the empirical coverage

Table 9: Diebold–Mariano test statistics and p-values.

Stations	DM	p-value
Quaid Abad	-3.5655	0.000665
Bahawalpur	-0.77492	0.441
Chakri	-1.9524	0.05496
Gwadar	-3.4965	0.000829
Haripur	-1.8748	0.06506
Peshawar	-4.3503	4.60E-05
Quetta	-3.5939	0.000607
Sadiq Abad	-5.0624	3.29E-06
Sanghar	0.25391	0.8003
Sujawal	-3.3632	0.001261
TandoGA	0.76189	0.4487
Umerkot	0.60957	0.5442

of six stations have higher values of ETS than ARIMA model. While five stations have equal coverage values of both models by differentiating three stations has 100% coverage value with two stations having less than 100% coverage but equal value.

Overall, these results suggest that although ARIMA provides sharper prediction intervals, ETS generally offers better coverage and hence more reliable uncertainty quantification. The further detail results of each model for all twelve stations are given in Tables A.10–A.15. In Table 9, the Diebold-Mariano test results and corresponding p-values for all twelve stations are given. If the test statistic (DM) has negative value, indicating lower forecasting loss of ARIMA than ETS and the statistical significance of these differences were assessed by corresponding p-value.

The DM result shows negative at nine stations, indicating that the ARIMA forecasting loss is lower than ETS model at these stations while the the p-values indicate that the forecasting differences are significant at six stations. This indicates that the ARIMA has superior performance than the ETS model. The present evaluation indicates that ARIMA provides the lowest average RMSE and MAPE for the study data than ETS and Naïve models. The DM test also suggests the better performance of the ARIMA model. Moreover, the Ljung-Box test rejected residual whiteness for both models at every station. To check the size of the dependency, we performed residual autocorrelation over lag 168. The maximum autocorrelation of each split was calculated and then the average maximum autocorrelation was reported against each station. The squared autocorrelation, used to describe the fraction of residual variance associated with serial dependence, is reported which are presented in Table A.17. Across all 12 stations, the ARIMA residual autocorrelation explained between 0.3% and 2.5% of residual variance (mean 1.3%), compared with 4.8% to 6.3% for ETS (mean 5.4%). The magnitude of the remaining dependence is modest for ARIMA and it is substantially smaller than for ETS, reinforcing the model's superior out-of-sample forecast accuracy.

Supporting tables

Tables A.10–A.15 report split-level coverage and width. Table A.16 lists the training and forecast partitions; Table A.17 presents the residual diagnostics and model orders.

5. Conclusion

In this study, ARIMA and ETS models were compared for forecasting short-term wind speed at twelve meteorological stations across Pakistan. The naïve model is also included as a benchmark for comparison. An expanding window approach with ten splits was applied, and forecasts were generated seven days ahead. Model accuracy was evaluated through RMSE and MAPE, with a model classified as the better performer at a station if it recorded lower error values in more than half of the ten splits. Also, the model accuracy was evaluated by the DM test applied to the forecast errors of each model for all twelve stations. ARIMA appears to be the better choice for short-term wind speed forecasting at seven stations based on RMSE and MAPE. While the DM statistics and corresponding p-values show ARIMA has lower forecasting loss at nine stations and the forecast-loss difference is statistically significant at six stations. These results provide statistical support for the superior forecasting performance of ARIMA at the stations where significant differences were observed.

That said, the fact that ETS matched or exceeded ARIMA's performance at a few stations suggests that a single model selection may not always be ideal. Choosing the model on a station-by-station basis could yield better forecasting outcomes in practice. Future research might look into combining both models into a hybrid framework, or incorporating additional weather-related variables to push forecasting accuracy even further.

Data availability

The data used in this study is available at https://power.larc.nasa.gov/data-access-viewer/.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Funding

The authors received no specific funding from any public, commercial, or not-for-profit organisation for this research.

References

[1] J. Zeng & W. Qiao, "Short-term wind power prediction using a wavelet support vector machine", IEEE Transactions on Sustainable Energy **3** (2012) 255. https://doi.org/10.1109/TSTE.2011.2180029.
[2] A. W. Bhutto, A. A. Bazmi & G. Zahedi, "Greener energy: Issues and challenges for Pakistan—wind power prospective", Renewable and Sustainable Energy Reviews **20** (2013) 519. https://doi.org/10.1016/j.rser.2012.12.010.
[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel & G. M. Ljung, *Time series analysis: Forecasting and control*, 5th ed., Wiley, 2015. Available online: https://www.wiley.com/en-us/time-series-analysis-forecasting-and-control-5th-edition-p-9781118675021.
[4] H. Liu, H. Q. Tian & Y. F. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction", Applied Energy **98** (2012) 415. https://doi.org/10.1016/j.apenergy.2012.04.001.
[5] E. Erdem & J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction", Applied Energy **88** (2011) 1405. https://doi.org/10.1016/j.apenergy.2010.10.031.
[6] E. Cadenas & W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model", Renewable Energy **35** (2010) 2732. https://doi.org/10.1016/j.renene.2010.04.022.
[7] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing **50** (2003) 159. https://doi.org/10.1016/S0925-2312(01)00702-0.
[8] R. J. Hyndman, A. B. Koehler, R. D. Snyder & S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods", International Journal of Forecasting **18** (2002) 439. https://doi.org/10.1016/S0169-2070(01)00110-8.
[9] R. J. Hyndman & Y. Khandakar, "Automatic time series forecasting: the forecast package for R", Journal of Statistical Software **27** (2008) 1. https://doi.org/10.18637/jss.v027.i03.
[10] S. Makridakis & M. Hibon, "The M3-Competition: Results, conclusions and implications", International Journal of Forecasting **16** (2000) 451. https://doi.org/10.1016/S0169-2070(00)00057-1.
[11] J. Wang, S. Qin, Q. Zhou & H. Jiang, "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China", Renewable Energy **76** (2015) 91. https://doi.org/10.1016/j.renene.2014.11.011.
[12] C. Bergmeir & J. M. Benítez, "On the use of cross-validation for time series predictor evaluation", Information Sciences **191** (2012) 192. https://doi.org/10.1016/j.ins.2011.12.028.
[13] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: An analysis and review", International Journal of Forecasting **16** (2000) 437. https://doi.org/10.1016/S0169-2070(00)00065-0.
[14] J. González-Enrique, I. J. Turias, J. J. Ruiz-Aguilar, J. A. Moscoso-López & L. Franco, "Spatial and meteorological relevance in NO_2 estimations: A case study in the Bay of Algeciras (Spain)", Stochastic Environmental Research and Risk Assessment **33** (2019) 801. https://doi.org/10.1007/s00477-018-01644-0.
[15] M. A. Solis Teran, F. Leite Coelho da Silva, E. A. Torres Armas, N. Carbo-Bustinza & J. L. López-Gonzales, "Modeling air pollution in Metropolitan Lima: A statistical and artificial neural network approach", Environments **12** (2025) 196. https://doi.org/10.3390/environments12060196.
[16] F. García Crespi, E. Y. Funes & M. A. Simon, "Rolling-origin validation reverses model rankings in multi-step PM10 forecasting: XGBoost, SARIMA, and persistence", arXiv preprint arXiv:2603.20315 (2026). https://doi.org/10.48550/arXiv.2603.20315.
[17] J. Hu, J. Wang & G. Zeng, "A hybrid forecasting approach applied to wind speed time series", Renewable Energy **60** (2013) 185. https://doi.org/10.1016/j.renene.2013.05.012.
[18] T. J. Adejumo, A. I. Okegbade, J. I. Idowu, O. J. Oladapo & O. O. Oladejo, "Time series analysis of Nigeria external (foreign) reserves", African Scientific Reports **2** (2023) 54. https://doi.org/10.46481/asr.2023.2.1.54.
[19] I. M. Echi, E. V. Tikyaa, E. J. Eweh, A. A. Tyovenda & T. Igbawua, "Investigation of the spatial distribution of deterministic chaos in some meteorological variables across Nigeria", African Scientific Reports **4** (2025) 280. https://doi.org/10.46481/asr.2025.4.2.280.
[20] U. Moyi, K. Bello, O. J. Ibidoja & G. Muhammad, "Modelling patients waiting and service time by ARIMA model: A case of Federal University Gusau clinic", Recent Advances in Natural Sciences **1** (2023) 7. https://doi.org/10.61298/rans.2023.1.1.7.
[21] A. E. Afere, Y. B. Usman, E. Alih, E. V. Aguda & L. Enesi, "Time series modelling of inflation and consumer price index trends in Nigeria", Recent Advances in Natural Sciences **3** (2025) 225. https://doi.org/10.61298/rans.2025.3.2.225.
[22] A. Thorat, R. Metri & C. Bhattar, "Reliable wind forecasting for grid stability using seasonal ARIMA", Journal of Renewable Energies **29** (2026) 263. https://doi.org/10.54966/jreen.v29i1.1496.

Appendix A. Supporting tables

Table A.10: 95% interval coverage and mean interval width for Bahawalpur and Chakri.

Bahawalpur						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	85.71	71.43	71.43	5.213	6.772	6.772
2	100	100	100	5.213	6.772	6.772
3	100	100	100	5.213	6.771	6.772
4	42.86	100	100	5.214	6.772	6.772
5	100	100	100	5.228	6.771	6.772
6	57.14	57.14	57.14	5.228	6.771	6.771
7	100	100	100	5.229	6.772	6.772
8	100	100	100	5.23	6.771	6.772
9	100	85.71	85.71	5.231	6.771	6.771
10	100	100	100	5.231	6.771	6.771
Chakri						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	5.034	9.199	7.372
2	100	100	100	5.033	6.375	7.371
3	100	100	100	5.033	5.45	7.37
4	100	100	100	5.032	5.612	7.369
5	100	100	100	5.031	5.013	7.368
6	100	100	100	5.031	3.154	7.368
7	100	100	100	5.03	5	7.367
8	100	100	100	5.029	7.239	7.366
9	100	100	100	5.029	5.154	7.365
10	100	100	100	5.028	8.195	7.364

Table A.11: 95% interval coverage and mean interval width for Gwadar and Haripur.

Gwadar						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	6.884	10.453	8.018
2	100	100	100	6.884	10.45	8.017
3	100	100	100	6.885	10.454	8.017
4	100	100	100	6.883	10.456	8.016
5	100	100	100	6.882	10.454	8.016
6	100	100	100	6.881	10.447	8.015
7	100	100	100	6.88	10.454	8.014
8	100	100	100	6.88	10.453	8.013
9	100	100	100	6.879	10.453	8.012
10	100	100	100	6.878	10.482	8.012
Haripur						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	85.71	100	100	4.767	7.316	7.594
2	100	100	100	4.767	7.313	7.597
3	85.71	100	100	4.767	7.313	7.597
4	100	85.71	85.71	4.767	7.313	7.597
5	71.43	85.71	85.71	4.767	7.316	7.597
6	100	100	100	4.769	7.321	7.6
7	100	100	100	4.769	7.321	7.599
8	100	100	100	4.769	7.321	7.599
9	100	85.71	85.71	4.769	7.32	7.599
10	100	85.71	100	4.769	7.319	7.598

Table A.12: 95% interval coverage and mean interval width for Peshawar and Quaid Abad.

Peshawar						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	5.793	7.178	7.389
2	85.71	100	100	5.792	7.176	7.388
3	100	100	100	5.793	7.175	7.388
4	100	100	100	5.793	7.175	7.387
5	42.86	57.14	57.14	5.792	7.174	7.386
6	57.14	85.71	100	5.797	7.174	7.386
7	100	100	100	5.798	7.177	7.386
8	100	100	100	5.796	7.174	7.386
9	100	100	100	5.796	7.176	7.385
10	100	100	100	5.796	7.176	7.385
Quaid Abad						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	4.977	6.506	6.857
2	100	100	100	4.976	6.507	6.856
3	100	100	100	4.975	6.504	6.854
4	100	100	100	4.975	6.504	6.853
5	100	100	100	4.974	6.503	6.852
6	100	100	100	4.973	6.502	6.851
7	100	100	100	4.972	6.501	6.85
8	100	100	100	4.972	6.502	6.849
9	100	100	100	4.971	6.501	6.848
10	100	100	100	4.97	6.5	6.847

Table A.13: 95% interval coverage and mean interval width for Quetta and Sadiq Abad.

Quetta						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	5.755	7.348	7.349
2	100	100	100	5.367	7.348	7.348
3	100	100	100	5.503	7.349	7.349
4	100	100	100	5.399	7.348	7.348
5	100	100	100	5.755	7.347	7.347
6	100	85.71	85.71	5.753	7.347	7.347
7	100	100	100	5.639	7.346	7.347
8	100	100	100	5.751	7.346	7.346
9	71.43	100	100	5.637	7.345	7.345
10	100	100	100	5.753	7.346	7.346
Sadiq Abad						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	4.869	6.303	6.304
2	100	100	100	4.869	6.302	6.303
3	71.43	71.43	71.43	4.869	6.302	6.302
4	100	100	100	4.868	6.302	6.303
5	100	100	100	4.868	6.301	6.302
6	14.29	28.57	28.57	4.868	6.301	6.301
7	100	100	100	4.87	6.303	6.303
8	100	100	100	4.869	6.303	6.303
9	100	100	100	4.868	6.302	6.302
10	100	100	100	4.868	6.302	6.303

Table A.14: 95% interval coverage and mean interval width for Sanghar and Sujawal.

Sanghar						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	4.789	5.752	5.752
2	100	100	100	4.787	5.751	5.752
3	100	100	100	4.79	5.751	5.751
4	100	100	100	4.789	5.751	5.751
5	85.71	100	100	4.789	5.751	5.751
6	100	100	100	4.79	5.75	5.75
7	100	100	100	4.789	5.749	5.75
8	100	100	100	4.789	5.749	5.75
9	100	100	100	4.789	5.749	5.749
10	100	100	100	4.789	5.748	5.748
Sujawal						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	4.808	5.509	5.509
2	100	100	100	4.645	5.508	5.509
3	100	100	100	4.808	5.508	5.508
4	85.71	85.71	85.71	4.823	5.508	5.508
5	100	100	100	4.563	5.509	5.509
6	100	100	100	4.797	5.508	5.508
7	100	100	100	4.808	5.508	5.508
8	85.71	57.14	57.14	4.807	5.507	5.508
9	100	100	100	4.824	5.509	5.509
10	100	100	100	4.592	5.508	5.508

Table A.15: 95% interval coverage and mean interval width for TandoGA and Umerkot.

TandoGA						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	4.444	5.096	5.096
2	100	100	100	4.444	5.096	5.096
3	100	100	100	4.444	5.095	5.095
4	100	100	100	4.443	5.094	5.095
5	100	100	100	4.444	5.094	5.094
6	100	100	100	4.444	5.094	5.094
7	100	100	100	4.444	5.094	5.094
8	100	100	100	4.453	5.094	5.094
9	85.71	85.71	85.71	4.444	5.094	5.095
10	100	100	100	4.445	5.094	5.095
Umerkot						
Split	A cov. (%)	E cov. (%)	N cov. (%)	A width	E width	N width
1	100	100	100	4.61	5.454	5.454
2	100	100	100	4.61	5.453	5.453
3	100	100	100	4.61	5.453	5.453
4	100	100	100	4.61	5.453	5.453
5	100	100	100	4.611	5.453	5.453
6	100	100	100	4.611	5.452	5.452
7	100	100	100	4.61	5.451	5.452
8	71.43	71.43	71.43	4.609	5.451	5.451
9	85.71	85.71	85.71	4.61	5.451	5.452
10	100	100	100	4.611	5.452	5.452

Table A.16: Training and forecasting split data.

Split	Training observations	Training size	Forecast observations	Forecast size
1	1–19077	19077	19078–19245	168
2	1–19245	19245	19246–19413	168
3	1–19413	19413	19414–19581	168
4	1–19581	19581	19582–19749	168
5	1–19749	19749	19750–19917	168
6	1–19917	19917	19918–20085	168
7	1–20085	20085	20086–20253	168
8	1–20253	20253	20254–20421	168
9	1–20421	20421	20422–20589	168
10	1–20589	20589	20590–20757	168

Table A.17: Residual diagnostic results and model specifications.

Station	ARIMA ACF	ARIMA var. (%)	ETS ACF	ETS var. (%)	ARIMA	ETS
Peshawar	0.1017	1.0336	0.2514	6.3226	(0,1,2)	(A,N,N)
Bahawalpur	0.0963	0.9282	0.2341	5.4780	(2,1,2)	(A,N,N)
Chakri	0.0783	0.6130	0.2313	5.3487	(0,1,5)	(A,N,N)
Gwadar	0.1301	1.6920	0.2221	4.9333	(2,1,4)	(A,N,N)
Haripur	0.1340	1.7949	0.2289	5.2378	(0,1,3)	(A,N,N)
Quaid Abad	0.0576	0.3322	0.2466	6.0821	(0,1,5)	(A,N,N)
Quetta	0.1469	2.1581	0.2230	4.9715	(1,1,1)	(A,N,N)
Sadiq Abad	0.0796	0.6333	0.2289	5.2380	(1,1,2)	(A,N,N)
Sanghar	0.0874	0.7646	0.2387	5.6974	(2,1,3)	(A,N,N)
Sujawal	0.1163	1.3522	0.2194	4.8152	(4,1,4)	(A,N,N)
TandoGA	0.1387	1.9241	0.2314	5.3568	(5,1,2)	(A,N,N)
Umerkot	0.1592	2.5359	0.2341	5.4826	(2,1,1)	(A,N,N)