ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 04 No. 01. July-September 2025.Page#.3269-3286
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.17011658

Comparing Forecasting Performance of ARFIMA and HAR-RV Models for Realized Variances Using Meteorological Data

Komal Arain
Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan
Naeem Ahmed Qureshi*
Department of Statistics, University of Sindh, Jamshoro, Pakistan
naeem.qureshi@usindh.edu.pk
Velo Suthar
Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan
Hakimzadi Wagan
Department of Agricultural Economics, Sindh Agriculture University Tandojam, Pakistan
Abdul Majid Memon
Department of Statistics, University of Sindh, Jamshoro, Pakistan

ABSTRACT
This study addresses the complex task of modelling and forecasting the realized variances of wind data, with a specific focus on realized volatility across six key atmospheric parameters: wind speed at 30 meters, wind speed at 80 meters, temperature, wind direction, relative humidity, and pressure. Motivated by the global transition toward renewable energy and the need for accurate forecasting tools, this research aims to estimate and to compare the performances of the Heterogeneous Autoregressive (HAR) and Autoregressive Fractionally Integrated Moving Average (ARFIMA) models through in-sample evaluation metrics and to assess the forecasting accuracy of these models through out-of-sample performance. Exploratory analysis revealed long memory behaviour, justifying the use of HAR-RV and ARFIMA models, which were applied to a realized variance series constructed from 2019–2024 data sets. Stationarity and volatility tests confirmed the suitability of both models, with optimal lag structures determined via AIC, BIC, HQC, and log-likelihood values. Forecasts were generated using a rolling window technique and the adequacy of these forecasts was evaluated on standard error metrics such as MSE, RMSE, and MAE. The results showed the superior forecasting capability of the HAR-RV model over the contrasting ARFIMA model. A final 72-hour ahead forecast reinforced HAR-RV's robustness and predictive strength. These findings suggest that HAR-RV modelling is more effective in capturing the dynamics of realized variances, offering valuable insights for operational forecasting and renewable energy integration.

***Keywords**: Comparing Forecasting Performance, Realized Variances, HAR-RV, ARFIMA, Wind Speed, Wind Direction, Temperature, Pressure, Relative Humidity.*

Introduction

The global shift towardrenewable energy has become an imperative to mitigate climate change and achieve sustainable development goals. Among the various renewable energy sources

[1,10,24] wind energy stands out as a crucial component of sustainable power generation, offering a scalable and meteorologically friendly alternative to fossil fuels. However, the inherent variability and stochastic nature of wind present significant challenges for the efficient integration of wind power into electrical grids. Accurate forecasting of wind behaviour is essential for optimizing energy production, ensuring grid stability, and minimizing operational costs. Traditional forecasting models often struggle to capture the complex interplay of meteorological factors that govern wind dynamics. These factors exhibit nonlinear relationships, temporal dependencies, and long-range correlations, necessitating advanced modelling techniques to address these intricacies.

Wind energy's sensitivity to rapidly changing meteorological conditions underscores the need for precise predictions of its driving variables. Various meteorological and atmospheric conditions influence power generation efficiency, wind patterns, and turbine alignment strategies. Conventional forecasting approaches, which often rely on simplistic assumptions, fail to account for the synergistic effects of these factors. Consequently, advancements in time series forecasting have emerged as vital tools to enhance predictive accuracy and support data-driven decision-making in renewable energy systems.

This study aims to address these gaps by applying two advanced statistical frameworks: the Heterogeneous Autoregressive-Realized Variance (hereinafter referred as HAR-RV) model and the Autoregressive Fractionally Integrated Moving Average (hereinafter referred as ARFIMA) model [11]. The HAR-RV model is particularly renowned for its ability to capture multi-scale dynamics and volatility clustering inherent in wind data. By decomposing the time series into components at different time horizons [19] the HAR-RV model efficiently captures both short-term fluctuations and long-term trends. Its hierarchical structure allows the model to integrate information from various time scales, making it adaptable at reflecting abrupt changes as well as persistent patterns. This adaptability is especially beneficial in forecasting wind behaviour, where sudden gusts and gradual shifts coexist, and it offers computational efficiency suitable for real-time applications.

In contrast, the ARFIMA model enhances traditional ARIMA methodologies by incorporating fractional differencing, a feature that allows it to model long-memory phenomenon and persistent correlations with greater precision [18]. Unlike standard ARIMA models that assume only short-term dependencies, ARFIMA is designed to handle non-stationary time series data where past observations exert a lasting influence over extended periods. This capability is crucial for modelling meteorological conditions that often exhibit seasonal trends and fractional integration properties. ARFIMA's flexibility in adjusting to varying degrees of persistence makes it a powerful tool for capturing the long-range dependencies characteristic of meteorological datasets.

By forecasting and comparing the performance of these models across all critical meteorological parameters, this research provides a comprehensive understanding of their capabilities in simulating wind energy dynamics. The detailed analysis not only highlights the theoretical underpinnings of the HAR-RV and ARFIMA models but also demonstrates their practical applicability in addressing the complex forecasting challenges in wind energy generation. The findings from this study are expected to contribute to improved energy planning and grid management strategies, ultimately supporting the integration of renewable energy systems into modern power networks.

High-frequency data presents unique challenges to empirical research. Statistically, a higher number of observations provide greater degrees of freedom, allowing for more precise estimators; however, it also introduces complexities such as model dimensionality and positive semi-definiteness in covariance matrices [6]. Addressing these challenges requires advanced statistical techniques capable of handling the intricate structure of high-frequency datasets, as well as innovative models that accommodate the nuances of high-dimensional time series data. This research specifically focuses on modelling and forecasting time-varying realized variances (also known as "*realized volatility*") for meteorological data, with a particular emphasis on wind data collected from the Jhimpir Wind Power Plant, located in the Thatta District of Sindh, Pakistan. This power plant is a major renewable energy project in Pakistan. Wind energy is a rapidly growing renewable resource that plays a critical role in addressing Pakistan's energy shortage and reducing reliance on fossil fuels. In addition to meeting the growing demand for sustainable energy, wind power aligns with Pakistan's strategic goals of attracting foreign investment in renewable technologies [12]. By leveraging econometric models on high-frequency meteorological data, this study contributes to the field of renewable energy by enhancing forecasting accuracy and providing actionable insights for energy planning and policy. Through sophisticated analysis of wind power data, this research supports the sustainable development of energy resources in Pakistan and offers a methodological framework for high-frequency meteorological data modelling in general.

1.1 Objectives

The following research objectives were achieved during the present study.

1. To estimate and compare both the HAR-RV and ARFIMA models using meteorological data.
2. To identify the best model through different in-sample evaluation strategies.
3. To forecast the realized variances from the selected model.

1.2 Significance of the Study

This study addresses the complex task of modelling and forecasting high-frequency meteorological data, with a specific focus on realized volatility across six key wind parameters. While global interest in renewable energy continues to grow, there remains a research gap in evaluating the forecasting capabilities of advanced univariate econometric models—specifically the HAR-RV and ARFIMA frameworks. High-frequency wind data, characterized by irregular daily observations and non-uniform time intervals [18] presents challenges such as nonlinearity, temporal dependence, and long-memory behaviour—features that traditional models often fail to capture effectively. To address this problem, the present study systematically compares the forecasting accuracy of HAR-RV and ARFIMA models over multiple time horizons (h = 1 to 10), employing robust evaluation techniques such as the rolling window method [19]. The findings are expected to provide significant value to researchers, renewable energy producers, and financial investors by improving volatility forecasting accuracy which will ultimately support optimized energy production, enhanced grid management, and better risk management strategies. Moreover, by aligning with Pakistan's strategic vision to boost renewable energy development and attract foreign investment in green technologies [1] the study contributes to national efforts in climate resilience and sustainable growth.

Literature Review

In this section the more related and update review of literature is presented while dividing this section into four sub-sections. First sub-section particularly focuses on current scenario of renewable energy in Sindh province of Pakistan. The second sub-section describes the potential of wind energy with respect to different geographical and meteorological factors whereas the third and fourth sub-sections are showing the challenges and barriers of wind energy in Sindh Province and the applications of ARFIMA and HAR-RV models in the fields of Renewable Energy, respectively.

Current landscape of wind energy in Sindh (2022-2025)

The installed wind power capacity in Sindh, Pakistan has witnessed substantial growth, reaching approximately 1841 MW [8] to 1845 MW [12,16] by 2022. This expansion highlights a positive trend in harnessing the region's significant wind resources. This growth is a reflection of the government's commitment to increasing the share of renewable energy in the national grid and the active participation of the private sector in developing wind energy projects.

A considerable number of wind power projects are currently operational in Sindh, primarily concentrated within the Jhimpir and Gharo-Keti Bandar Wind Corridor. These projects vary in installed capacity and have been commissioned over the years. The operational wind energy projects in Sindh between 2022 and 2025 include projects like FFC Energy Limited (49.5 MW, commissioned 2013), Zorlu Enerji Pakistan Limited (56.4 MW, commissioned 2013), Three Gorges First Wind Farm Pakistan Private Limited (49.5 MW, commissioned 2014), Foundation Wind Energy –I Limited (50 MW, commissioned 2015), and Sapphire Wind Power Company Limited (52.8 MW, commissioned 2015), among many others [16]. The commissioning dates range from 2013 to 2022, showcasing a decade of active development in the wind energy sector of Sindh.

While a significant wind power capacity is already operational, ongoing efforts continue to expand wind energy generation in Sindh. Several new projects have been commissioned or are under construction within the 2022-2025 timeframe. For instance, the Metro Wind Power project (with a capacity of 60 MW) was commissioned in June 2022, Indus Wind Energy Ltd (50 MW) and Liberty Wind Power 2 Pvt. Ltd (50 MW) were commissioned in March and May 2022, respectively [16]. Planned projects such as the Sindh Wind Farm-Oracle Power with a substantial capacity of 500 MW are expected to be commissioned by 2025 [15]. These activities indicate continued investment and a positive outlook for the growth of wind energy in the region.

Wind Energy Potential in Sindh: Geographical and Meteorological Factors

Sindh's geographical characteristics play a pivotal role in its suitability for wind energy generation. The province's extensive coastal belt, stretching approximately 250 km, along with the presence of the Gharo-Jhimpir wind corridor spanning about 9700 sq. km [3], provide a naturally advantageous setting for large-scale wind power development [2,3]. The Gharo-Jhimpir corridor, in particular, is recognized for its consistently high wind speeds, averaging between 5 to 12 m/s [3]. The relatively flat terrain and proximity to the coast further enhance the feasibility and economic viability of establishing and operating wind farms in this region.

Meteorological conditions in Sindh are also highly favorable for wind energy production. The province experiences sufficient average wind speeds, for example, recording averages of 9.7 m/s during summer, 7.6 m/s in autumn, 7.4 m/s in spring, and 4.8 m/s in winter. The summer and autumn seasons, which coincide with Pakistan's peak electricity demand, exhibit the highest wind speeds. However, a concerning trend of decreasing wind speeds has been observed in some

areas, such as Thatta, since 2009, which could potentially impact the long-term sustainability and efficiency of wind power projects [17]. Prevailing wind directions are largely influenced by the southwest monsoon system which brings strong winds over Pakistan, particularly, benefiting the Sindh region [3].

Data from the Pakistan Meteorological Department (PMD) plays a crucial role in monitoring and analyzing these wind patterns [2,17]. The estimated potential for wind energy in Sindh is substantial. Studies and reports indicate a gross potential of approximately 43.8 GW, with about 11 GW considered usable [3]. Other estimates suggest an economically viable potential of 11 GW in the Jhimpir-Keti Bandar wind corridor alone, with older studies even estimating a technically exploitable potential of up to 50 GW. A World Bank study conducted in 2021 projected that Sindh could accommodate 10,035 MW of renewable energy by 2030, considering various constraints [8].

Challenges and Barriers to Wind Energy Development in Sindh

Despite the significant potential and progress in wind energy development in Sindh, several challenges and barriers hinder the sector from reaching its full capacity. Infrastructure limitations pose a significant obstacle, particularly the inadequacy of transmission lines and grid capacity to efficiently evacuate the generated wind power [3,4,12]. The rapid increase in power generation in the southern region, including wind energy has outpaced the development of the transmission network, leading to congestion and curtailment of wind power projects [12]. This results in underutilization of installed capacity and financial losses for wind power producers.

Policy and regulatory barriers also impede the smooth development of wind energy projects in Sindh. Inconsistent government policies, frequent changes in regulations, and delays in obtaining necessary approvals and tariff notifications create uncertainty for investors and slow down project implementation [2,3,4,12]. The lack of a consistent and supportive policy framework can deter investment and hinder the long-term growth of the wind energy sector.

HAR-RV model in wind energy forecasting

HAR-RV models represent a class of autoregressive models specifically designed to capture the persistence of volatility observed in financial time series, based on the idea of heterogeneous market participants operating on different time scales [22]. These models are typically constructed using realized volatility measures calculated over various time horizons, such as daily, weekly, and monthly. In the context of wind energy forecasting, HAR-RV models have been adapted to predict wind speed and power, considering the multi-scale nature of wind dynamics influenced by factors operating at different temporal frequencies. Research has explored several methodologies employing HAR-RV models for wind energy forecasting. Some studies directly apply HAR-RV models to wind speed or power time series to capture the inherent autoregressive structure and the impact of past realizations on future values. The findings from these studies suggest that HAR-RV models can be effective in predicting wind energy. For instance, a study introducing the HAR-RV model for wind speed forecasting found that it outperformed both the classical ARIMA model and a multi-layer perception ANN model in terms of RMSE, MAE, and mean absolute percentage error MAPE. This indicates that the HAR model was better at capturing wind speed characteristics such as its asymmetric (non-Gaussian) distribution, non-stationary time series profile, and chaotic dynamics compared to conventional models [9]. In some contexts, HAR models have been shown to effectively capture the long-memory features present in wind data [14].

ARFIMA models in wind energy forecasting

ARFIMA models represent a significant extension of the traditional ARIMA framework [13], allowing for the differencing parameter to take fractional values. This key feature enables ARFIMA models to effectively capture long-memory behavior in time series data [7], a characteristic often observed in wind speed and power fluctuations. The fractional differencing parameter, denoted as 'd', plays a crucial role in modeling these long-range dependencies, where past values of the time series can have a persistent, albeit slowly decaying, influence on current and future values. ARFIMA models are particularly suitable for time series that exhibit slow-decaying autocorrelation functions, indicating the presence of long memory [21]. In the realm of wind energy forecasting, ARFIMA models have been applied to predict both wind speed and power across various forecasting horizons.

Materials and methods

Data collection and description

Data for this study were collected from Jhimpir, a key site in southern Sindh, Pakistan, renowned for its consistent wind speeds and favourable atmospheric conditions [12]. The data spans from January 2019 to 2024, with observations recorded every 10 minutes. This high-frequency sampling provides a detailed time series that captures both intra-day fluctuations and seasonal trends as well. Each 10-minute observation includes six key meteorological variables: 1). Wind speed at 30 meters, reflecting near-surface conditions, 2). Wind speed at 80 meters, capturing higher-altitude profiles, 3). Temperature, which influences air density and atmospheric stability, 4). Wind direction showing the directional angle of the wind's origin, 5). Relative humidity which shows the percentage of moisture in the air, affecting sensor accuracy, and finally 6). Pressure of broader weather systems.

Quality control was ensured through sensor calibration, timestamp verification, and pre-analysis procedures (e.g., missing value imputation and outlier detection), ensuring the dataset's reliability for subsequent modelling. All analyses and visualizations were performed using the R programming language. Different packages in R were used for detailed analysis including estimation and forecasting from the employed models.

Methodology

3.2.1 Data pre-processing

Prior to modelling, the high-frequency time series for each variable (wind speeds at two heights, temperature, wind direction, relative humidity, and pressure) were subjected to extensive pre-processing to ensure data quality and compatibility. Missing observations were imputed using linear interpolation, and extreme outliers were identified through robust statistical methods and adjusted accordingly.

3.2.2 Realized variances

To capture intra-day volatility in the high-frequency meteorological data set, realized variances (RV) were calculated for each of the six observed variables using a block-based approach. The raw 10-minute interval data were partitioned into non-overlapping blocks, each consisting of six consecutive observations, thereby representing one-hour intervals. Within each block, first-order differences between successive observations are computed to quantify short-term fluctuations. For a given meteorological variable X_t, the intra-block realized variances are calculated as:

$$\Delta X_{t,k,j} = X_{t,k,j} - X_{t,k,j-1}, \quad j = 1, \dots, 6$$

The realized variance for day *t* was then obtained by summing the squared intra-block realized variances over all N blocks.

$$RV_t = \sum_{j=1}^{6} r_{t,j}^2$$

This process was applied independently to all the variables, producing six distinct realized variance series. By aggregating high-frequency differences in this manner, the procedure effectively filtered out transient noise while preserving the underlying variability structure of each variable.

3.2.3 ARFIMA model

To model the long-range dependence observed in the realized variability series, the ARFIMA model was employed. The ARFIMA model allows for fractional differencing to capture persistent dependencies in the data. In its simplified form, where short-memory components were set to zero (i.e., $p = q = 0$), the model is expressed as:

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t$$

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ are the autoregressive and moving average lag polynomials, respectively, B is the backshift operator, *d* is the fractional differencing parameter estimated from the data, X_t is the time series at time *t*, and finally ε_t is a stochastic term. The fractional differencing operator *(1−B)d(1-B)^d* is implemented via a binomial expansion, thereby enabling a parsimonious representation of the long-memory behaviour in the series.

3.2.4 HAR-RV model

Originally developed in financial econometrics by *[5]*, HAR-RV for realized variance has proven highly effective at capturing the long-memory and multi-scale characteristics of high-frequency time series data. Although its initial application was in forecasting financial market volatility, the model's straightforward structure and intuitive decomposition of variability render it equally promising for meteorological data analysis—such as for wind speed, temperature, relative humidity, and atmospheric pressure.

In a meteorological context, the realized variance RV_t is computed from high-frequency observations (e.g., 10-minute measurements) by aggregating data into non-overlapping blocks (for instance, six consecutive observations to represent one hour). For each meteorological variable, the intra-day variability is estimated as the sum of squared differences between successive observations, effectively filtering out micro-scale noise while preserving meaningful variability. The HAR-RV model then decomposes the daily realized variance into components that capture different temporal horizons:

$$RV_t = \beta_0 + \beta_1 RV_{t-1}^{(d)} + \beta_2 RV_{t-1}^{(w)} + \beta_3 RV_{t-1}^{(m)} + \varepsilon_t$$

Where RV_trealized volatility (RV) at time t, $RV_{t-1}^{(d)}$ is yesterday's realized volatility, $RV_{t-1}^{(w)}$ is the average realized volatility over the last week, $RV_{t-1}^{(m)}$ showing average realized volatility over the last month, and ε_t is the error term. The HAR model's additive structure is designed to reflect the impact of various market participants who operate on different time scales, thereby capturing both short-term shocks and long-term persistence in variability.

3.2.5 Model estimation and order selection

In this study, both the ARFIMA and HAR models were estimated at various lag values (orders) to determine the optimal model specification for the forecasting purpose. For the ARFIMA model, different combinations of autoregressive (AR) and moving average (MA) terms were considered to capture the temporal dependencies and long-range dependencies inherent in the data. Similarly, for the HAR model, various lag structures such as daily, weekly, and monthly were evaluated to account for multi-scale dependencies in the data. The selection of the optimal model order is based on: Akaike Information Criterion (AIC)**,** Bayesian Information Criterion (BIC)**,** Hannan-Quinn Criterion (HQC), and Log-likelihood (LL) of the model**.** These criteria provide a comprehensive evaluation of model fit while penalizing model complexity to avoid over-fitting. This procedure is essential to ensure the robustness and reliability of the chosen model for subsequent forecasting tasks.

3.2.5 Forecasting realized variances

After selecting the two best models i.e., first from ARFIMA specifications and second from HAR-RV specifications, the forecasts of realized variance were generated independently for all the variables used in this study. To assess the forecasting performance of the proposed models, a rolling window forecasting strategy was adopted. In the present study, 50% of the dataset was initially used to fit the model (training data), and the remaining 50% was used to generate iterative one-step-ahead forecasts. Upon completion, the forecasted values were compared against the actual observations in the latter half of the dataset to evaluate forecasting accuracy. This approach ensures forecasts are made under conditions resembling real-time applications, thereby enhancing the practical relevance and reliability of the results.

The accuracy of the out-of-sample forecasts in the time series context is evaluated using four standard error metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). These metrics provide complementary insights into the performance of forecasting models over time. MAE captures the average magnitude of one-step-ahead forecast errors without considering their direction, offering a clear measure of typical prediction accuracy. MSE amplifies the impact of larger forecast errors through squaring, making it useful for detecting models that occasionally produce substantial deviations. RMSE, being the square root of MSE, not only retains the scale of the original series but also highlights the presence of large errors, which is particularly relevant in time series with volatility. These metrics collectively support a rigorous and balanced assessment of model performance in capturing the temporal dynamics of the data. Let $\hat{y}_i$ denote the forecasted value and y_i the actual observed value at time i, for a total of n forecasts. The metrics are formally defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

Results and discussion

This section particularly focuses on the results obtained by employing the research methodology presented in the previous section. This section is divided into different sub-sections to avoid complexities and to provide better understandings of the results.

Descriptive Statistics

Table 1 shows the summary statistics of realized variances computed for all the variables under study. Wind speed 80 shows values ranging from a minimum of 0 m/s to a maximum of 429.40 m/s. The mean wind speed is 3.41 m/s, and the variance is 80.37, indicating moderate variability. Wind speed 30 ranges from 0.000 m/s to 331.17 m/s, with a mean of 2.78 m/s and the variance is 39.88. Wind direction varies between 0.003° and 554.23°, with a mean of 90.98°. Temperature has a minimum of 0°C and a maximum of 290.55°C, with a very low mean of 0.55°C. The variance is 14.36, indicating low variability. Relative Humidity at 80m (%) ranges from 0.003% to 290.55%, with a mean of 9.86%. The variance is 524.58, indicating substantial variability. The skewness of 5.69 and kurtosis of 44.17 suggest a highly skewed and leptokurtic distribution, possibly due to sensor errors or abnormal readings. Furthermore, pressure at 80m (mbar) ranges from 0.000 mbar to 298.02 mbar with a mean of 0.50 mbar. The variance is 24.40, and the skewness of 29.76 with kurtosis of 1291.62 reflects a severely distorted distribution. The skewness and kurtosis of all the variables show positive skewed and leptokurtic behavior.

Table l Descriptive statistics for several meteorological variables under study

Variable	**Min**	**Max**	**Mean**	**Variance**	**Skewness**	**Kurtosis**
Wind speed 80	0.000	429.395	3.412	80.369	20.944	672.724
Wind speed 30	0.000	331.175	2.783	39.881	18.682	568.710
Wind direction	0.003	554.231	90.976	13592.596	1.790	5.825
Temperature	0.000	290.545	0.550	14.356	38.836	2151.202
Relative Humidity	0.003	290.545	9.856	524.577	5.693	44.168
Pressure	0.000	298.016	0.496	24.402	29.758	1291.623

Time plots of variables

The historical data of realized variance from 2019 to 2024 are plotted in Figure 1 which provides a clear visual evidence of time-varying volatility across all six meteorological variables. These plots for Wind speed (30 and 80) shows distinct periods of elevated variance, particularly during late 2019 and early 2020, with sporadic peaks continuing throughout the rest of the timeline. This pattern indicates bursts of high variability followed by relatively calmer periods. The Wind direction plot exhibits consistently high levels of realized variance across the entire time span, suggesting persistent and frequent fluctuations in direction. Temperature and Pressure both display significant spikes in realized variance concentrated around 2019–2020 with much lower and more stable variance observed from 2021 onward. This implies that these variables experienced more extreme variation during earlier periods, stabilizing in later years. The

Humidity plot shows a cyclical pattern, with high and frequent fluctuations in realized variance that appear to follow seasonal or meteorological cycles, along with occasional dips likely due to data gaps or reduced variability. The visual inspection confirms the presence of heteroskedasticity in all variables which was further confirmed through ACFs and PACFs and through Engle's ARCH and McLeod-Li tests.

ACFs (Autocorrelation Functions) & PACF (Partial Autocorrelation Functions)

The analysis of the Autocorrelation Functions (hereinafter referred as ACFs) and Partial Autocorrelation Function (hereinafter referred as PACFs) for the realized variance series across multiple meteorological variables reveals consistent evidence of long memory and temporal dependence. Figure 2(a) showing wind speed at both 30m and 80m heights, the ACFs exhibit a slow, gradual decay with numerous significant lags extending over hundreds of observations, indicating strong persistence in volatility. The PACFs plots show several significant early lags, which diminish rapidly, suggesting the presence of short-term autoregressive effects contributing directly to the variance dynamics. The 80m height data demonstrates a slightly simpler short-term dependence structure compared to 30m, reflected by fewer and lower magnitude PACFs spikes. In Figure 2(b), the wind direction displays a similarly slow decay in the ACFs with persistent significant autocorrelations over a long lag range which has been identified as a characteristic of long memory processes. The PACFs indicate initial significant autoregressive components that taper off, further supporting the coexistence of short-term AR behaviour and long-term dependence.

Similarly, temperature's realized variance exhibits moderate long memory, with the ACFs decaying slower than expected under white noise assumptions. The PACFs show a few significant short-term lags, consistent with autoregressive influences coupled with persistent

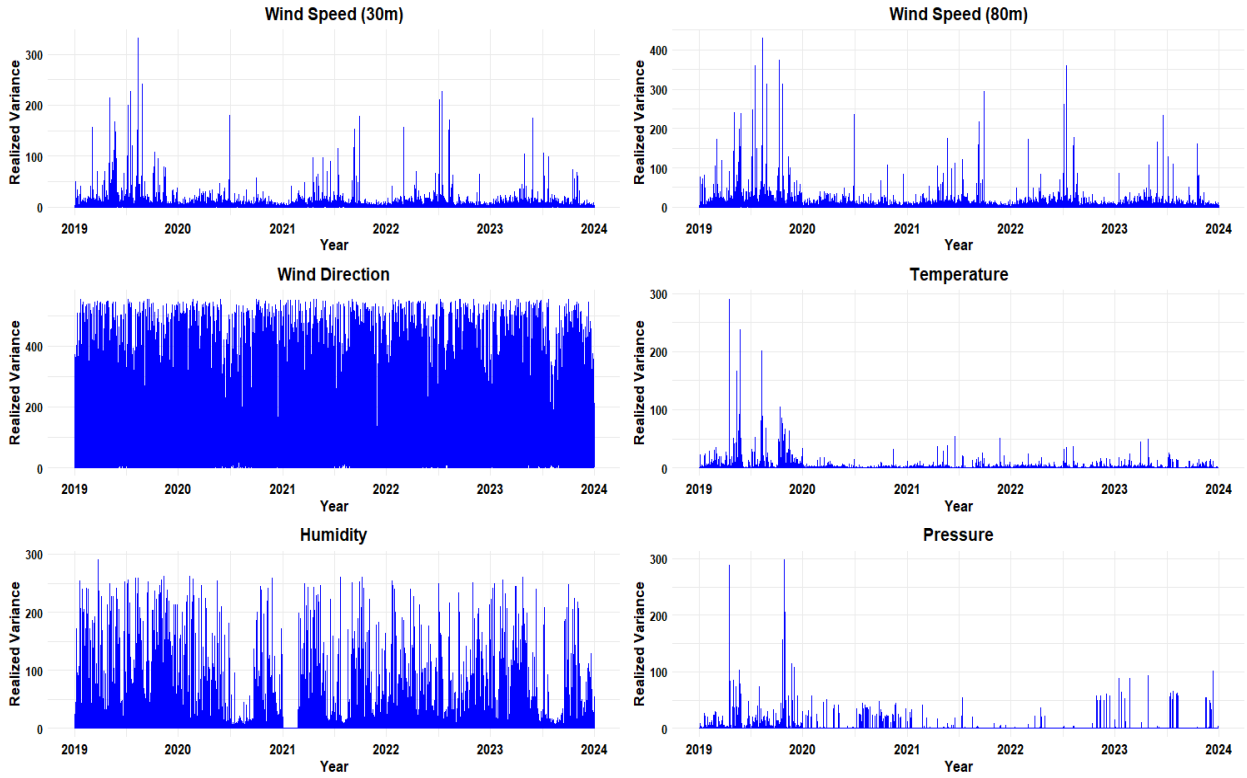

Figure 1 Time Plot of variables under study

variance effects. In Figure 2(c) pressure demonstrates slow ACFs decay with oscillatory patterns and periodic spikes, pointing to long memory combined with cyclical behaviour. The PACFs confirm a short-term AR component through significant initial lags, alongside persistent long-range dependence. Likewise, relative humidity displays the strongest long memory features, as evidenced by an ACF that remains significantly positive well beyond 300 lags. The PACFs highlights pronounced short-term autoregressive effects at the earliest lags, followed by gradual decay, indicating a robust combination of short- and long-term dependence.

Collectively, these findings underscore the importance of models that accommodate fractional integration and multi-scale persistence, such as ARFIMA and HAR-RV, to accurately capture the complex temporal dependence in realized variance for meteorological variables. The persistence patterns observed validates the incorporation of long memory structures in volatility modelling frameworks to improve forecasting and inference in high-frequency meteorological data.

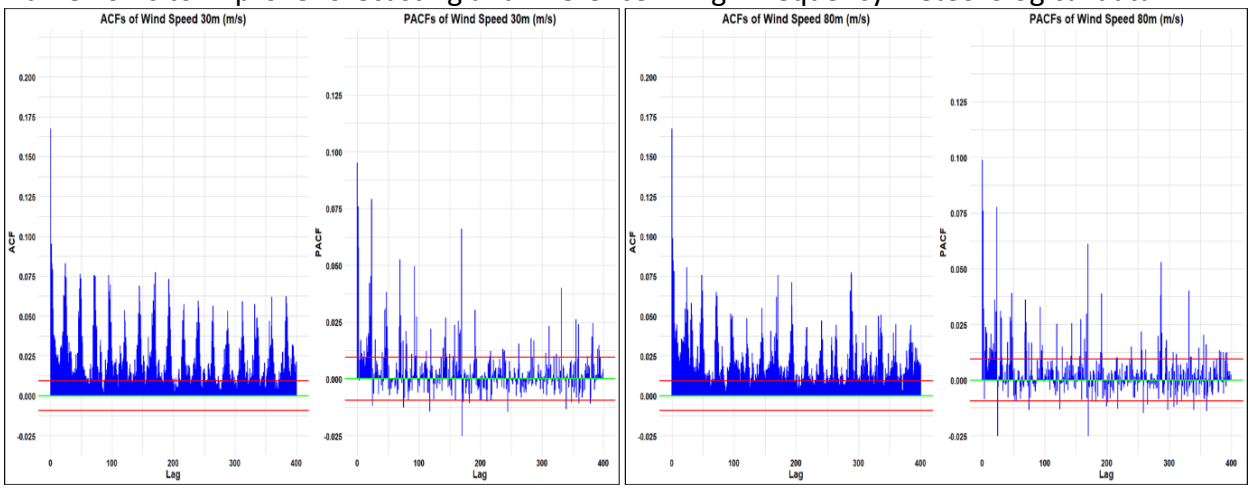

Figure 2(a) ACFs and PACFs of Wind Speed 30 and Wind Speed 80

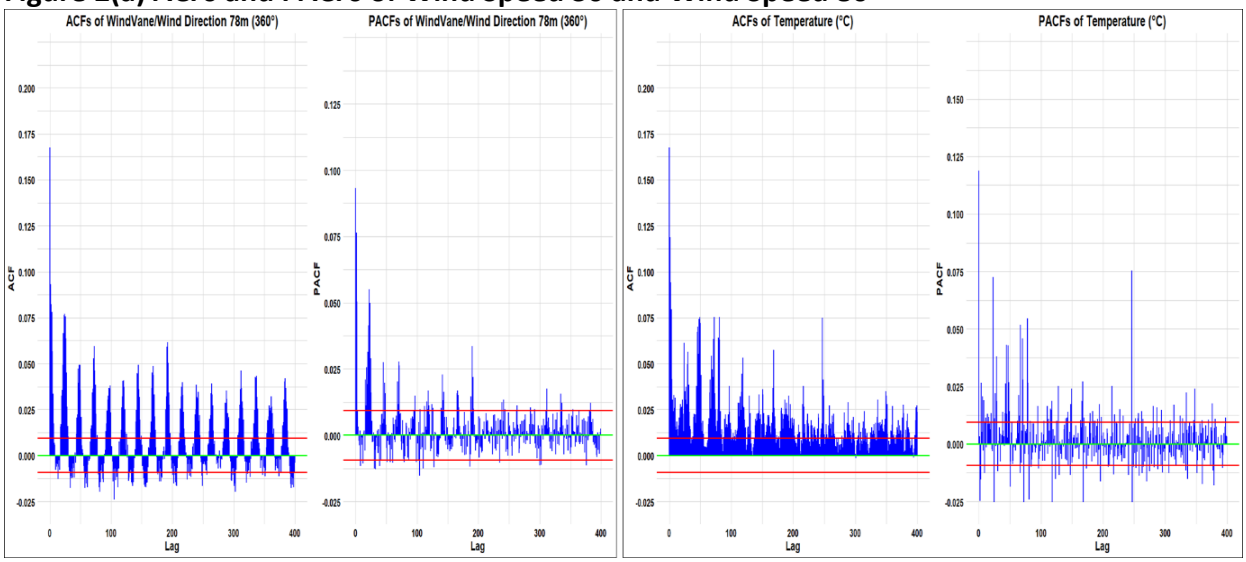

Figure 2(b) ACF and PACF of Wind Direction and Temperature

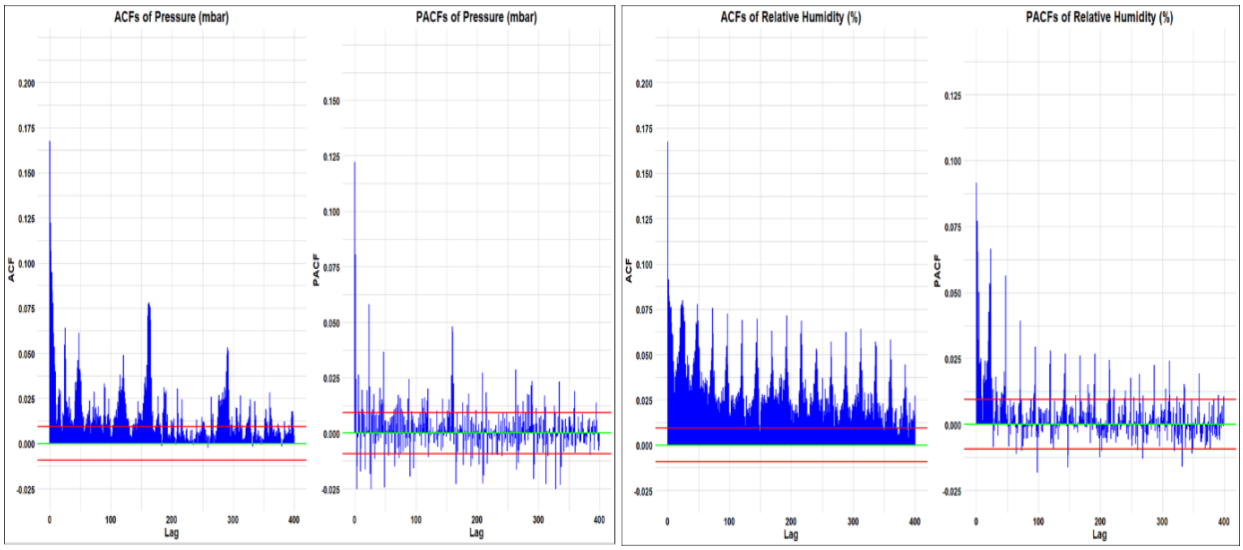

Figure 2(c) ACFs and PACFs of Pressure and Relative Humidity

Testing ARCH effects (volatility clustering)

Table 2 shows the results for the Engle's ARCH and the McLeod-Li tests which have been used to check the presence of heteroskedasticity/ARCH effects (also known as volatility clustering). The results showed that the p-values of test statistics for both tests are very close to 0 for all variables, indicating a strong rejection of the null hypothesis of Homoskedasticity or No ARCH effects at lagged value (m) 50. This confirms that the realized variance series for each variable exhibits significant time-varying volatility and hence HAR-RV and ARFIMA models are used to capture this volatility clustering phenomenon.

Table 2 Engle's ARCH and McLeod-Li results for checking Heteroskedasticity

Variable	**ARCH test P-value (m = 50)**	**McLeod-Li test P-value (m = 50)**
Wind Speed80	0.000	0.000
Wind Speed30	0.000	0.000
Wind Direction78m	0.000	0.000
Temperature	0.000	0.000
Relative Humidity	0.000	0.000
Pressure	0.000	0.000

In-sample ARFIMA model selection

To determine the most suitable ARFIMA(p,d,q) model for the time series data, an in-sample evaluation was performed using various combinations of autoregressive (p) and moving average (q) terms, specifically the orders (1,1), (1,2), (1,3), (2,1), and (2,2). The model selection was based on well-established statistical information criteria — Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan–Quinn Criterion (HQC) as well as the maximized log-likelihood (*LL*) function. Among the models estimated, ARFIMA (1,2) consistently yielded the minimum values for AIC, BIC, and HQ, and the maximum value of the log-likelihood function, indicating a better in-sample fit compared to the alternative specifications. These results suggest that ARFIMA (1, 2) captures the underlying temporal dependence and long-memory structure in the data more effectively than other competing models.

In-sample HAR-RV model selection

To capture the multi-scale dynamics of our hourly wind-power output series, we estimated three Heterogeneous Autoregressive (HAR) models—each designed to reflect fluctuations over distinct horizons—and compared their in-sample performance using AIC, BIC, HQC, and maximized log-likelihood. Because our data are recorded every hour, the three candidate lag structures correspond to meaningful meteorological and operational cycles: the first model, HAR(1, 5, 22), incorporates a one-hour term for very short-run gusts, a five-hour term approximating morning-to-midday boundary-layer developments, and a 22-hour term capturing a complete diurnal cycle; the second, HAR(12, 84, 360), uses a twelve-hour lag to reflect half-day heating-cooling transitions, an 84-hour lag for synoptic patterns recurring roughly every three to four days, and a 360-hour lag for fortnightly shifts in wind regimes; and the third, HAR(24, 168, 720), blends a 24-hour daily cycle, a 168-hour weekly synoptic cycle, and a 720-hour (≈30-days) component to account for slowly evolving seasonal or climatic influences.

Among these estimated HAR models, the HAR (24, 168, 720) specification emerged clearly superior: it delivered the lowest values of AIC, BIC, and HQC while also achieving the highest log-likelihood as shown in the Table l 5. This result confirms that wind-power variability is most effectively modelled by combining daily diurnal patterns (24 hr), recurring weekly weather systems (168 hr), and longer-term variations tied to monthly or seasonal shifts (720 hr). In practice, the 24-hour term captures the predictable day–night cycle of boundary-layer winds; the weekly term reflects the passage of high- and low-pressure systems that typically recur on a seven-day rhythm; and the monthly term approximates gradual changes in large-scale atmospheric circulation and seasonal temperature gradients. By integrating these three scales, HAR (24, 168, and 720) balances parsimony with explanatory power, reproducing both the rapid intraday gustiness and the slow-decaying clustering of wind regimes without the need for a more complex fractional-integration model.

Table 3 Selection of suitable order for ARFIMA model based on different criteria

Variable	**Order**	**Log-Likelihood**	**AIC**	**BIC**	**HQC**
Wind Speed80	(1,1)	-155715.90	311439.81	311474.56	311450.76
Wind Speed80	1,2)	-155710.53	311433.08	311474.55	311449.51
Wind Speed80	(1,3)	-155712.36	311436.72	311488.85	311453.15
Wind Speed80	(2,1)	-155715.27	311440.54	311483.98	311454.23
Wind Speed80	(2,2)	-155710.55	311433.09	311485.22	311449.52
Wind Speed30	(1,1)	-140802.72	281613.44	281648.19	281624.39
Wind Speed30	(1,2)	-140795.35	281600.73	281644.17	281614.41
Wind Speed30	(1,3)	-140857.27	281726.54	281778.67	281742.97
Wind Speed30	(2,1)	-140796.29	281602.58	281646.02	281616.27
Wind Speed30	(2,2)	-140801.64	281615.28	281667.41	281631.71
Wind Direction	(1,1)	-535898.33	1071804.66	1071839.41	1071815.61
Wind Direction	(1,2)	-535882.3	1071776.60	1071822.04	1071794.29
Wind Direction	(1,3)	-535886.01	1071784.02	1071836.15	1071800.45
Wind Direction	(2,1)	-535885.86	1071781.72	1071825.16	1071795.41
Wind Direction	(2,2)	-535883.26	1071778.52	1071830.65	1071794.95
Temperature	(1,1)	-113829.39	227666.78	227701.54	227677.74
Temperature	(1,2)	-113731.64	227510.27	227555.71	227534.96

Temperature	(1,3)	-113777.33	227566.66	227618.79	227583.09
Temperature	(2,1)	-113771.51	227553.02	227596.46	227566.71
Temperature	(2,2)	-113756.62	227525.24	227577.37	227541.67
Relative Humidity	(1,1)	-565394.13	1130796.26	1130831.01	1130807.22
Relative Humidity	(1,2)	-565340.86	1130620.73	1130732.17	1130702.42
Relative Humidity	(1,3)	-565341.59	1130695.19	1130747.32	1130711.62
Relative Humidity	(2,1)	-565350.8	1130711.60	1130755.04	1130725.29
Relative Humidity	(2,2)	-565343.46	1130698.92	1130751.05	1130715.35
Pressure	(1,1)	-124006.31	248020.61	248055.36	248031.56
Pressure	(1,2)	-123954.21	247920.44	247972.56	247936.87
Pressure	(1,3)	-123954.22	247920.45	247972.57	247936.88
Pressure	(2,1)	-123996.84	248003.68	248047.12	248017.37
Pressure	(2,2)	-123961.58	247935.152	247987.28	247951.58

Table 4 Selection of suitable lag order for HAR-RV model based on in-sample evaluation

Variable	**Order**	**LL**	**AIC**	**BIC**	**HQC**
Wind Speed80	(1,5,22)	-155815	311637.2	311671.9	311648.1
Wind Speed30	(1,5,22)	-140924	281855.9	281890.7	281866.9
Wind Direction	(1,5,22)	-268943	537893.0	537927.8	537904.0
Temperature	(1,5,22)	-119789	247585.0	247619.8	247596.0
Relative Humidity	(1,5,22)	-197517	395041.6	395076.4	395052.6
Pressure	(1,5,22)	-134200	268408.0	268442.7	268418.9
Wind Speed80	(12,84,360)	-156202	312412.9	312447.6	312423.8
Wind Speed30	(12,84,360)	-141064	282135.9	282170.6	282146.8
Wind Direction	(12,84,360)	-268146	536300.2	536335.0	536311.2
Temperature	(12,84,360)	-118718	237444.0	237478.7	237455.0
Relative Humidity	(12,84,360)	-196783	393574.7	393609.4	393585.7
Pressure	(12,84,360)	-129666	259339.2	259373.9	259350.1
Wind Speed80	(24, 168, 720)	-154993	309994.3	310029.0	310005.2
Wind Speed30	(24, 168, 720)	-139952	279912.7	279947.4	279923.7
Wind Direction	(24, 168, 720)	-265831	531670.3	531705.0	531681.2
Temperature	(24, 168, 720)	-118159	236326.2	236360.9	236337.1
Relative Humidity	(24, 168, 720)	-194771	389549.8	389584.5	389560.7
Pressure	(24, 168, 720)	-129356	258719.5	258754.1	258730.4

Testing presence of ARCH effects in residuals from selected models

In Table 5, the Engle's ARCH and McLeod-Li tests were conducted to evaluate the presence of ARCH effects in the residuals resulting from the selected models i.e., ARFIMA and HAR-RV al lagged value (m) = 50. As all the p-values of the test-statistics are greater than the conventional level of significance (0.05) resulting the homoskedastic behaviour in residuals.

Table 5 Engle's ARCH and McLeod-Li tests for residuals from ARFIMA HAR model

Variables	**ARCH test (p-value, m = 50)**	**McLeod-Li test (p-value, m = 50)**
Windspeed80	0.081	0.080
Windspeed30	0.076	0.083

Direction78m	0.081	0.079
Temperature	0.077	0.082
Humidity	0.079	0.081
Pressure	0.084	0.077

Out-of-sample forecast evaluation

In Table 6, the comparative analysis of the HAR and ARFIMA models is presented. Based on the most widely used error metrics such as MSE, RMSE and MAE which are used for out-of-sample forecast evaluation, the HAR model consistently outperforms the ARFIMA model in terms of prediction accuracy, as evidenced by lower error values across all metrics. Since wind direction is the most volatile variable so it has the largest value of all error metrics. The selected HAR-RV model will be used to forecast the realized variances of these series.

Table 6 Out-of-sample forecast evaluation based on different error metrics

Variables	**HAR-RV model**			**ARFIMA model**		
	MSE	RMSE	MAE	MSE	RMSE	MAE
Wind Speed30	29.06	5.43	2.39	266.34	16.32	4.56
Wind Speed 80	27.77	5.27	1.91	152.28	12.34	3.61
Wind Direction	13211.74	114.94	83.72	69322.28	263.29	171.53
Temperature	1.54	1.24	0.38	7.34	2.71	0.61
Relative Humidity	403.21	20.08	9.27	1851.34	43.02	18.37
Pressure	6.97	2.64	0.31	33.30	5.77	0.49

4.8 Forecasting realized variances from selected model

To achieve the last objective of the present study, the forecasts for the realized variances were generated using the HAR-RV selected based on in-sample evaluation. In Figure 5, the historical series for Direction78m is marked by considerable volatility, with a sharp peak observed just prior to the 72-hour forecast window. Following this peak, the series exhibits stabilized lower variability. The HAR model projects a relatively flat future trajectory, with tight confidence intervals. Statistically, the narrow-forecast bands indicate high predictive precision, suggesting that the model has effectively captured the underlying post-fluctuation behaviour of the direction series. Humidity displays notable historical volatility, characterized by several abrupt increases followed by a rapid decline. After this stabilization, the HAR model's 72-hour forecast indicates a steady trajectory with relatively narrow confidence intervals. The low forecast variance reflects high model confidence, attributable to the sudden reduction in variability observed immediately before the start of the forecast period. The historical series for Pressure remains relatively stable with minor fluctuations. The HAR model suggests a continuation of this stability over the 72-hour forecast horizon, with a slight upward tendency, accompanied by extremely narrow confidence intervals. Statistically, the tight prediction bands indicate a high degree of certainty, demonstrating that the model is highly effective in forecasting pressure due to the low intrinsic variability of the series. Temperature historically exhibits moderate fluctuations, characterized by peaks and subsequent declines. The 72-hour forecast by the HAR model projects a gradual upward trend.

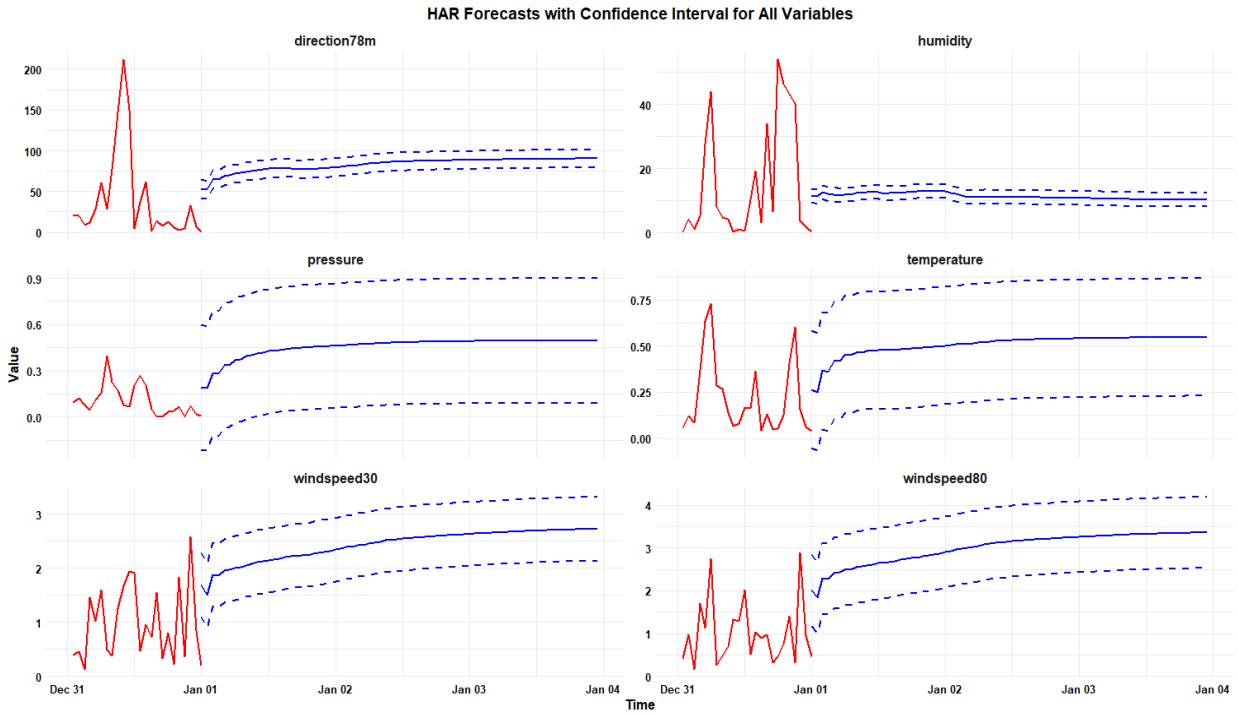

Figure 5 Plot of observed and forecasted time series with 95% confidence intervals

Initially, the confidence intervals are wider, indicating moderate forecast uncertainty; however, they progressively narrow across the forecast horizon. This behaviour suggests that forecast precision improves over time as the model increasingly adapts to the underlying stabilization in temperature dynamics. BothWindspeed30 and Windspeed80 demonstrate considerable historical variability with frequent sharp fluctuations. The HAR model forecasts stabilization in wind speeds over the 72-hour forecast horizon. However, compared to other variables, the confidence intervals for both series are relatively wider, reflecting higher forecast uncertainty. This increased variability is consistent with the inherently dynamic and turbulent nature of wind behaviour, highlighting the challenges in achieving precise wind speed predictions over the forecast period.

Conclusion

This study utilized high-frequency meteorological intra-day data from 2019 to 2024, comprising of six key variables namely Wind Speed 30m, Wind Speed 80m, Temperature, Wind Direction, Relative Humidity, and Pressure to forecast the realized variance using the ARFIMA and HAR-RV models. Initially, the realized variance was constructed from the original time series data. Subsequently, an exploratory analysis was conducted, which included line plots, the ACFs, and the PACFs of the realized variance. The results of this revealed clear evidence of long-memory behavior in the data, suggesting the presence of persistent autocorrelations over time. This empirical characteristic motivated the application of both ARFIMA and HAR-RV models, which are well-suited for capturing long-range dependencies and volatility dynamics in meteorological time series.Prior to model implementation, residual diagnostic tests were conducted on the realized variance series of the meteorological variables. Engle's ARCH and McLeod-Li tests were used to assess the presence of conditional heteroskedasticity (also known as volatility clustering). Both tests returned p-values near to zero, confirming the presence of significant ARCH effects.

These diagnostic results underscore the appropriateness of volatility modeling techniques for the realized variance series, justifying the use of volatility-sensitive models such as ARFIMA and HAR-RV. To determine optimal model specifications, various specifications for these two models were evaluated using model selection criteria including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Criterion (HQC), alongside log-likelihood values and the best fitting model was selected based on these criteria. The ARFIMA model achieved its best performance with an order of (1,2), while the HAR model was most effective with lag structures of (24, 168, 720) hours, corresponding to daily, weekly, and monthly cycles—important periodicities in atmospheric behaviour. Both models captured the residual dynamics effectively; particularly at nominal lag lengths (m = 50) while using intra-day data.. The HAR-RV model consistently outperformed the ARFIMA model across all error measures, demonstrating lower forecast errors and improved residual modelling. Finally, the HAR-RV model proved to be more effective in forecasting realized variance in meteorological data, especially under presence of long memory and volatility clustering, making it a valuable tool for wind power variability modeling and operational forecasting. The 72-hour-ahead forecasts further validated the HAR-RV model's robustness, showing high predictive precision for relatively stable variables such as pressure and wind direction, and moderate but improving accuracy for temperature. Despite wider confidence intervals for wind speeds, the model successfully captured the underlying stabilization trend. These forecasting interpretations highlight the HAR-RV model's strong capability in adapting to diverse atmospheric dynamics, thereby supporting more informed and reliable decision-making in wind energy operations.

References

[1] M. Ashraf, B. Raza, M. Arshad, B. M. Khan, and S. S. H. Zaidi, Performance enhancement of short-term wind speed forecasting model using Real time data, *PLoS ONE***19** (2024). https://doi.org/10.1371/journal.pone.0302664.

[2] S. Aslam, K. Mustafa, M. Raza, S. F. Shah, M. Hussain, and S. Ali, Wind Energy Potential in Pakistan: A Feasibility Study in Sindh Province, *Energies***15** (2022). https://doi.org/10.3390/en15228333.

[3] M. Awais, Wind Energy Potential in Sindh Region, *Innovative Energy Research***11** (2022), 291.

[4] Wind power in Pakistan: How a promising renewable energy source fell out of favour, *Business Recorder* (2023). https://www.brecorder.com/news/amp/40319368.

[5] F. Corsi, A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics***7** (2009), 174–196.

[6] D. Ghanem and A. Smith, What Are the Benefits of High Frequency Data for Fixed Effects Panel Models?,*Journal of the Association of Environmental and Resource Economists***8** (2021), 199–234.

[7] M. Ghofrani and A. Suhreli, Time Series and Renewable Energy Forecasting, in *Dependability Engineering*, Chapter 12, *InTech* (2018). https://doi.org/10.5772/intechopen.71501.

[8] Government of Pakistan, *Updated Nationally Determined Contributions 2021 – Global Energy Monitor: Power Sector Transition in Sindh* (2021). https://www.gem.wiki/Power_Sector_Transition_in_Sindh.

[9] Y. Y. Hong, H. L. Chang, and C. S. Chiu, Recursive wind speed forecasting based on Hammerstein Autoregressive model, *Applied Energy***145** (2015), 191–197.

[10] *IRENA*, *Renewable energy policies in a time of transition*, *International Renewable Energy Agency* (2018).
[11] M. Izzeldin, M. K. Hassan, V. Pappas, and M. Tsionas, Forecasting realised volatility using ARFIMA and HAR models, *Quantitative Finance***19** (2019), 1627–1638. https://doi.org/10.1080/14697688.2019.1600713.
[12] M. H. Khan, Analysis: Taking the wind out of alternative energy's sails, *The Daily Dawn* (2024). https://www.dawn.com/news/1861999.
[13] K. Liu, Y. Chen, and X. Zhang, An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs, *Axioms***6** (2017). https://doi.org/10.3390/axioms6020016.
[14] N. Hu, X. Yin, and Y. Yao, A novel HAR-type realized volatility forecasting model using graph neural network, *International Review of Financial Analysis***98** (2025). https://doi.org/10.1016/j.irfa.2024.103881.
[15] *Sindh Wind Farm–Oracle Power, Sindh, Pakistan*, *Power Technology* (2024). https://www.power-technology.com/data-insights/power-plant-profile-sindh-wind-farm-oracle-power-pakistan.
[16] *Private Power & Infrastructure Board*, *List of Commissioned IPPs & TLP (Operational)* (2025). https://www.ppib.gov.pk/commisioned_ipps.html.
[17] S. Rehman, U. Rehman, M. Zeeshan, and M. Asif, A Feasibility Analysis of Wind Energy Potential and Seasonal Forecasting Trends in Thatta District: A Project to Combat the Energy Crisis in Pakistan, *Energies***18**(1) (2025), 158.
[18] S. Saroha, S. K. Aggarwal, and P. Rana, Wind Power Forecasting, in A. AbouJaoudé (ed.), *Forecasting in Mathematics – Recent Advances, New Perspectives and Applications*, *Intech Open* (2021). https://doi.org/10.5772/intechopen.94550.
[19] P. Soleimani, Forecasting the wind direction by using time series models with long-term memory (case study: Nayer region) (2020). https://www.emerald.com/insight/content/doi/10.1108/ijesm-04-2019-0002/full/html.
[20] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, A Review of Wind Power and Wind Speed Forecasting Methods with Different Time Horizons, *North American Power Symposium* (2010), *IEEE*.
[21] Pakistan's wind power projects to supply electricity to national grid, *Daily Times – Technology Times* (2022). https://metro-power.com/metro-power-company-limited-has-started-producing-50-mega-watt.
[22] N. Ulinnuha, M. Desvina, and M. Muslim, Estimation Parameter d in Autoregressive Fractionally Integrated Moving Average Model in Predicting Wind Speed, *Indonesian Journal of Pure and Applied Mathematics***1** (2019), 61–67.
[23] Y. Wang, Forecasting realized volatility in a changing world: A dynamic model averaging approach (2016). https://www.sciencedirect.com/science/article/abs/pii/S0378426615003647.
[24] *WWEA*, *World Wind Energy Association Annual Report*, *World Wind Energy Association* (2023).