

# Nur'ainul Miftahul Huda

## Sequential\_Kalman\_Filter\_and\_ARIMA\_Modeling\_of\_High\_F...

 Paper 5 FMIPA Oke Universitas Tanjungpura

### Document Details

Submission ID

trn:oid::1:3528207307

Submission Date

Apr 6, 2026, 9:59 AM GMT+7

Download Date

Apr 6, 2026, 10:01 AM GMT+7

File Name

Sequential\_Kalman\_Filter\_and\_ARIMA\_Modeling\_of\_High\_Frequency\_Wind\_Speed\_with\_IMR\_Verifi....pdf

File Size

1.2 MB

16 Pages

7,368 Words

38,611 Characters

# 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Filtered from the Report

- Bibliography
- Quoted Text

## Match Groups

-  **33 Not Cited or Quoted 8%**  
Matches with neither in-text citation nor quotation marks
-  **1 Missing Quotations 0%**  
Matches that are still very similar to source material
-  **0 Missing Citation 0%**  
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**  
Matches with in-text citation present, but no quotation marks

## Top Sources

- 8%  Internet sources
- 4%  Publications
- 4%  Submitted works (Student Papers)

## Match Groups

- 33 Not Cited or Quoted 8%**  
Matches with neither in-text citation nor quotation marks
- 1 Missing Quotations 0%**  
Matches that are still very similar to source material
- 0 Missing Citation 0%**  
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**  
Matches with in-text citation present, but no quotation marks

## Top Sources

- 8% Internet sources
- 4% Publications
- 4% Submitted works (Student Papers)

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

|    |                                                                                     |     |
|----|-------------------------------------------------------------------------------------|-----|
| 1  | Internet                                                                            |     |
|    | ejournal.uin-malang.ac.id                                                           | 3%  |
| 2  | Internet                                                                            |     |
|    | ojs3.unpatti.ac.id                                                                  | 1%  |
| 3  | Internet                                                                            |     |
|    | jims-a.org                                                                          | <1% |
| 4  | Student papers                                                                      |     |
|    | Associatie K.U.Leuven                                                               | <1% |
| 5  | Internet                                                                            |     |
|    | acikbilim.yok.gov.tr                                                                | <1% |
| 6  | Internet                                                                            |     |
|    | www.mdpi.com                                                                        | <1% |
| 7  | Student papers                                                                      |     |
|    | University of Kent at Canterbury                                                    | <1% |
| 8  | Internet                                                                            |     |
|    | plastik.me                                                                          | <1% |
| 9  | Student papers                                                                      |     |
|    | RMIT University                                                                     | <1% |
| 10 | Publication                                                                         |     |
|    | J. Bert Keats, Norma Faris Hubele. "Statistical Process Control in Automated Man... | <1% |

|    |                |                              |     |
|----|----------------|------------------------------|-----|
| 11 | Internet       | openknowledge.worldbank.org  | <1% |
| 12 | Internet       | journal.ummat.ac.id          | <1% |
| 13 | Internet       | www.ecostat.kerala.gov.in    | <1% |
| 14 | Student papers | University of the Free State | <1% |
| 15 | Internet       | backoffice.biblio.ugent.be   | <1% |
| 16 | Internet       | romanpub.com                 | <1% |



# Sequential Kalman Filter and ARIMA Modeling of High-Frequency Wind Speed with IMR Verification

Nurfitri Imro'ah<sup>1\*</sup>, Nur'ainul Miftahul Huda<sup>2</sup>, Kartika Sari<sup>3</sup>, and Rahmi Hidayati<sup>3</sup>

<sup>1</sup>Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Tanjungpura, Indonesia

<sup>2</sup>Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Tanjungpura, Indonesia

<sup>3</sup>Departement of Computer System Engineering, Faculty of Mathematics and Natural Sciences, Universitas Tanjungpura, Indonesia

## Abstract

This study uses a dataset of 3,742 observations collected at five-minute intervals to conduct a comprehensive analysis of high-frequency wind speed fluctuations in Pontianak, Indonesia. The IoT-based ESP32 microcontrollers were integrated with a Wi-Fi Sensor Network (WSN) designed to collect data. After identifying strong stochastic noise, positive skewness, and high kurtosis in the initial descriptive study, a sequential modeling framework was designed to address these issues. This framework included Kalman Filter denoising, followed by Autoregressive Integrated Moving Average (ARIMA) modeling. When it came to pre-modeling signal purification and improving time-series stationarity, installing the Kalman Filter proved vital. The results of the comparison show that this two-stage technique performs much better than the solitary ARIMA model. The Root Mean Square Error (RMSE) was reduced from 4.043 to 0.522, representing an improvement of 87%. To further validate the model, Statistical Process Control (SPC) with Individual Moving Range (IMR) charts was used. These graphs confirmed that residuals remained within strict control limits  $3\sigma$ . This research demonstrates that sequential adaptive filtering is an effective and robust method to characterize wind speed and has the potential to improve early warning systems in equatorial areas prone to extreme weather.

**Keywords:** accuracy; control chart; residuals; weather.

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

## 1 Introduction

A high degree of stochasticity and rapid temporal fluctuations are exhibited by the wind speed dynamics in the equatorial region, particularly in West Kalimantan [1]. It is due to complex local atmospheric interactions in this region. However, this low temporal resolution does not allow for the recording of transient phenomena such as wind gusts and short-duration turbulence. It is still the case that the majority of conventional meteorological research relies on daily or hourly average data. Aviation safety, shipping operations, and coastal disaster mitigation systems all rely heavily on these phenomena. To acquire a more granular understanding of local wind behavior,

\*Corresponding author. E-mail: [nurfitriimroah@math.untan.ac.id](mailto:nurfitriimroah@math.untan.ac.id)

it is imperative to conduct high-frequency measurements, particularly at minute intervals [2]. These high-resolution datasets provide a more reliable foundation for real-time monitoring and short-term predictive modeling in dynamic tropical environments, enabling the detection of microvariations that are often masked by time averaging.

High-frequency datasets obtained from Internet of Things sensors and wireless sensor networks often contain substantial stochastic noise, even though high temporal resolution provides more detailed dynamic information. Interference from instruments or random, non-representative atmospheric oscillations could be the source of this noise. When it comes to time series modeling, significant noise at minute intervals can obscure the underlying signal and make it difficult to estimate model parameters, such as those in the Autoregressive Integrated Moving Average (ARIMA) model [3]. If the noise variance is too high, conventional ARIMA models often struggle to capture predictable patterns. This, in turn, leads to unstable residuals and low forecast accuracy. In the Kalimantan region, most local meteorological studies still tend to ignore the adaptive signal-cleaning stage before modeling. It leaves a substantial research need to integrate signal processing techniques to improve the reliability of predictive models on fluctuating sensor data. Consequently, a robust preprocessing mechanism is required to 'purify' the sensor data, ensuring that the subsequent stochastic modeling is based on the actual physical trend rather than measurement artifacts.

The purpose of this research is to propose a two-stage, sequential modeling framework that uses the Kalman Filter as a preprocessing (denoising) stage for high-frequency wind speed data. In contrast to hybrid techniques, which directly incorporate mathematical structures, this framework separates the operations of signal purification and stochastic modeling into two sequential procedures that remain interconnected. The Kalman Filter is an ideal state estimator that operates recursively to separate the actual wind dynamics from stochastic noise. It is accomplished by accounting for the uncertainty in sensor measurements. To prevent the time series structure from becoming unstable before moving on to the ARIMA parameter identification stage, this sequential technique is designed to achieve this. As a result of this, the model can capture predictable patterns with greater precision, without being disrupted by excessive random fluctuations. The raw data from Internet of Things sensors is translated into more precise and consistent estimates of atmospheric conditions through this two-stage approach. This ultimately improves the model's robustness in handling very volatile meteorological data.

A substantial innovative contribution to this research is the use of IMR control charts to evaluate model performance. This contribution complements the other benefits associated with the two-stage sequential strategy proposed. Continuous investigation of the statistical stability of model residuals is enabled by SPC analysis, implemented using IMR control charts. The usual meteorological research, typically relies solely on mean error measures such as MSE, RMSE, or MAE to quantify error. It is in contrast to the current situation. Conventional error measurements often fail to detect mean shifts or changes in variability in modeling results, which are typically unnoticeable. Both types of change are undesirable. By incorporating these control charts, it is possible to visually and statistically assess the consistency and reliability of the sequential Kalman-ARIMA framework. It helps ensure that modeling errors remain within acceptable control limits. This novel, more comprehensive method for high-resolution wind speed analysis and monitoring is made possible by integrating an Internet of Things (IoT) data-gathering system, recursive signal filters, and stability evaluation based on SPC.

The analysis of wind speed has been the topic of a significant amount of research due to the crucial role it plays in optimizing renewable energy and mitigating risks linked to weather. This field has used various models, ranging from the more conventional statistical approaches to the more cutting-edge machine learning strategies. ARIMA models, for example, have been utilized by [4], [5], and [6] for short-term prediction. Additionally, Grey Models ([7]), Random Forest ([8], [9], and [10]), and LSTM networks ([11], [12], and [13]) have demonstrated their effectiveness in capturing both complex and straightforward temporal patterns. Artificial neural

networks (ANNs) have also been widely deployed, with contributions by [14], [15], and [16]. These contributions have demonstrated good predicted accuracy in a variety of geographical scenarios. In addition, comparative studies conducted by [17], [18], [19], and [20] have highlighted the significance of selecting models depending on the resolution of the data and the features of the climate.

However, the majority of these studies rely on data on wind speed that is collected daily or monthly, which may ignore the rapid swings that occur in metropolitan environments located in equatorial regions. When this is applied to tropical places like Pontianak, Indonesia, where the weather conditions are highly unpredictable and can change within minutes, it creates a gap in the study that has been done previously. The present investigation fills this void by employing time series modeling on high-frequency wind speed data collected locally at intervals of five minutes, incorporating Kalman Filter smoothing to reduce noise, and validating model stability by utilizing IMR control charts, a combination rarely investigated in regional climatology studies.

## 2 Methods

### 2.1 Autoregressive Integrated Moving Average (ARIMA)

ARIMA models are typically employed to evaluate time series data for enhanced comprehension and prediction. The ARIMA model is derived from a generic version of the Autoregressive Moving Average (ARMA) model. This model is called ARIMA  $(p, d, q)$ , where  $p$  represents the autoregressive components,  $d$  indicates the integrated components, and  $q$  signifies the moving average components, with  $p$ ,  $d$ , and  $q$  being non-negative integers. Given the assumption that the sequence of random variables  $Z_t$  follows the ARIMA process, the definition of  $Z_t$  can be written as [21]:

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t$$

where  $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ ,  $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ ,  $B$  is Backshift Operator, and  $a_t$  is error at time  $t$ .

The subsequent list ultimately delineates the processes necessary to implement the ARIMA model [22].

1. Ascertain the orders of  $p$ ,  $d$ , and  $q$ . It is the initial phase of the order-choosing procedure. Analyzing time series plots, Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) can be employed to achieve this objective.
2. Estimate the parameters of an ARIMA model using a methodology such as the least squares method.
3. The diagnostic test utilizes the residuals obtained from the ARIMA model to ascertain whether the assumptions of white noise (normal distributed and independent residuals) have been fulfilled.

### 2.2 Kalman Filter

The Kalman Filter is a recursive linear estimator that was developed to minimize the mean-square error of a system's state by processing noisy measurements [23]. The recursive structure of this technique, unlike static filtering techniques, enables dynamic adaptation to system changes. As a result, it is well-suited to high-frequency wind speed data, which often exhibit stochastic variations [24].

The implementation in this study follows a state-space representation [25]:

$$x_{t+1} = A_t x_t + F_t u_t + G_t w_t$$

and

$$z_t = H_t x_t + v_t$$

where  $x_t$  is condition vector at time  $t$ ;  $u_t$  is deterministic input variables at time  $t$ ;  $w_t$  is a noise vector in a system model is derived from generating random integers by a computational program;  $z_t$  is measurement variables at time  $t$ ;  $H$  is the coefficient matrix of the measurement model possesses dimensions  $1 \times m$ . The Kalman Filter is comprised of the stages that may be used [26].

1. Initial value initialization

$$\hat{x}_0 = \bar{x}_0$$

and

$$P_0 = P_{x_0}$$

2. As part of the prediction stage, the computation of estimated values and error covariance is carried out. Error covariance is a value that offers information about the accuracy of the system's prediction and estimation updates. This information may be used to improve the system further.

$$\hat{x}_{t+1}^- = A_t \hat{x}_t + w_t$$

and

$$P_{t+1}^- = A_t P_t A_t^T + Q_t$$

3. The correction phase incorporates  $K_{t+1}$ , the Kalman Gain, which enhances the estimated value from the prediction phase. This value often varies between 0 and 1. A Kalman Gain value approaching 1 indicates greater accuracy in measurement findings.

$$K_{t+1} = P_{t+1}^- H_{t+1}^T (H_{t+1} P_{t+1}^- H_{t+1}^T + R_{t+1})^{-1}$$

and

$$\hat{x}_{t+1} = \hat{x}_{t+1}^- + K_{t+1} (z_{t+1} - H_{t+1} \hat{x}_{t+1}^-)$$

and

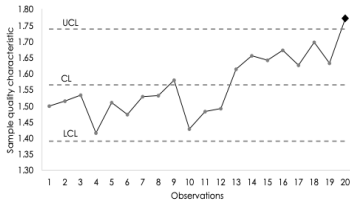
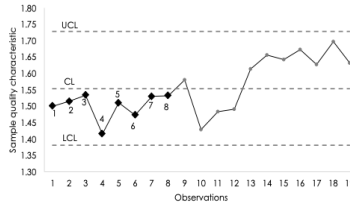
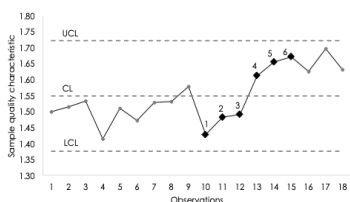
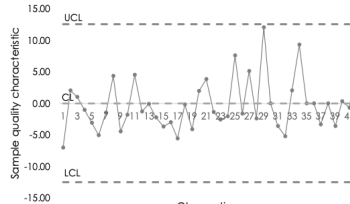
$$P_{t+1} = (1 - K_{t+1} H_{t+1}) P_{t+1}^-$$

The process of the method is divided into two basic stages: (1) The Prediction Stage, which involves the computation of the previous state ( $\hat{x}_{t+1}^-$ ) and error covariance ( $P_{t+1}^-$ ). (2) The Correction Stage, which involves the calculation of the Kalman Gain ( $K_{t+1}$ ) in order to bring the state estimate up to date based on the residual of the actual measurement ( $z_{t+1}$ ). Before applying the ARIMA model to the smoothed series, this iterative technique ensures the series remains statistically stable. For analyzing wind speed, the Kalman Filter is instrumental because it can eliminate high-frequency turbulence noise while preserving the underlying meteorological patterns.

### 2.3 Individual Moving Range (IMR) Chart for Time Series Data

A graphical representation of quality attributes measured or quantified in a sample according to time or observations is the form that statistical quality control takes as a tool [27]. This display may be used to analyze the quality of the sample. A tool that is utilized to control quality is the control chart. These charts are extremely sensitive to changes in the mean of the process, which makes them beneficial for the early detection of potential defects that, if left unchecked, could develop into more serious issues. It is necessary to regulate the production process by removing the specific variables that cause variations in the process. Variations are a part of every process, but the production process must be controlled. Consequently, this will guarantee that the reasons are the only ones responsible for general changes. One of the components of the control chart is referred to as the Centre Line (CL). Regarding the state that is being regulated, it is a representation of the average value of the quality criteria associated with that condition. The two other horizontal lines are referred to by their titles, Upper Control Limit (UCL) and

**Table 1:** An illustration of control chart

| No. | Control Chart                                                                       | Conditions                                                                                | Conclusion     |
|-----|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|----------------|
| 1.  |    | There is a particular point in time when the control restrictions are breached.           | Out of control |
| 2.  |    | A consistent sequence can be found for the eight points positioned below CL.              | Out of control |
| 3.  |   | There is a sequence of six consecutive points that continue to increase regularly.        | Out of control |
| 4.  |  | The control limits that have been specified are not exceeded by the observed data points. | In control     |

149 Lower Control Limit (LCL) (see Table 1). These lines are situated at the top and bottom of the  
150 graphical representation, respectively [28].

Time series control charts are valuable tools for monitoring processes that include accumulating data points over time and displaying autocorrelation [29]. Through the utilization of appropriate time series models and the execution of residual analysis, organizations can successfully detect changes in the behavior of processes. It can guarantee quality and consistency across a wide range of applications. The control chart is based on the fundamental assumptions that the data are independent and that there is no discernible link between them [30]. As a result of the fact that the residual assumption in the time series model is white noise, which is independent and uncorrelated among residuals, the observations utilized in the time series control chart are residuals [31]. This is because only one variable is observed, and that variable is the residual model. Given this, generating a control chart for the time series model may be utilized for any model that assumes the residuals are influenced by white noise. All these conditions are fulfilled by the IMR control chart, which also provides computations as

$$UCL = \bar{x} + 3\left(\frac{|x_i - x_{i-1}|}{d_2}\right); CL = \bar{x}; LCL = \bar{x} - 3\left(\frac{|x_i - x_{i-1}|}{d_2}\right)$$

151 where  $x_i$  is  $i$ -th observations and  $d_2$  is constant. In time series models, residuals are assumed to

adhere to a normal distribution, with a mean of zero and a variance of  $\sigma_{\epsilon}^2$ . Therefore, the value of the  $\bar{x}$  is equal to zero [32]. Within the context of this IMR control chart, residual behavior is evaluated. The derived model has a low level of accuracy if this plot's results demonstrate that it is not under control. A high level of accuracy is possessed by the obtained model, which enables it to be applied to produce predictions for a number of different time periods in the future.

## 2.4 Data Collection and IoT System Architecture

A custom-engineered environmental monitoring system, following a Wireless Sensor Network (WSN) design, was used to collect wind speed data for this investigation. Because of its robust processing capabilities and integrated dual-mode communication, the ESP32 microcontroller was chosen as the focal point for hardware integration [33]. To ensure data accuracy across a wide range of meteorological conditions, a high-precision anemometer and a wind vane with magnetic or ultrasonic sensing were used to measure wind speed and direction.

As part of the data-collection operation, which adhered to a systemized workflow, the sensors conducted continuous environmental sampling. To prevent sensor drift and environmental interference, automated calibration was incorporated into the firmware. The firmware maintained a high level of measurement accuracy thanks to this action. To guarantee a constant connection to the centralized cloud server, a system that used both Wi-Fi and GSM (SIM800L) protocols was implemented. A communication layer enabled by the Internet of Things (IoT) was then used to disseminate the processed data at 5-minute intervals. This backend technology enabled storing structured data in a relational database and receiving real-time data. It also supplied a synchronized, high-resolution time-series dataset that could be used for statistical modeling and analysis when it was later performed. Figure 1 depicts the actual field deployment of the Internet of Things-based weather monitoring station that was installed in Pontianak City. To minimize aerodynamic interference during the data-gathering period, the system was positioned at an appropriate elevation to ensure there were no obstructions to the wind flow.

## 3 Results and Discussion

This study uses wind speed data that was recorded continuously at a frequency of every five minutes, beginning on December 11, 2024, and continuing until December 24, 2024. An autonomous sensor established in the Pontianak City region served as the observation location for the data collection. A digital anemometer-type wind speed sensor equipped with an Internet of Things (IoT)-based data acquisition system is the sensor utilized in the process of data collecting. This equipment can automatically and in real-time capture wind speed data, with an accuracy of up to  $\pm 0.1$  meters per second and a time resolution of 5 minutes. The sensor is installed at a typical meteorological observation height roughly ten meters above the ground. It is executed to guarantee that the data representation is precise and aligned with the conditions of the local environment. This device can capture the wind speed and incorporate a data storage and transfer module that enables the fast processing of additional information. This vast period and data resolution make it possible to conduct a more in-depth investigation of the short-term dynamics of the wind and the micro-fluctuations that are not observable in the data collected daily or monthly.

In the Pontianak City region, the wind speed displayed a variable pattern over the observation period from 11 to 24 December 2024, with a continuous peak tendency occurring during the day. It can be seen by referring to Table 2 and Figure 2, which illustrate the data. The potential for significant local atmospheric dynamics at that period was reflected in the maximum wind speed recorded at 52.95 m/s on 12 December 2024. With a standard deviation of 6.42 m/s and a variance of 41.19 m/s, the average wind speed during the observation time was 4.28 m/s. It indicates extensive data regarding the mean value during the observation period. A



**Figure 1:** The IoT-based weather monitoring system deployed in Pontianak City for high-frequency wind speed data collection

skewness value of 2.72 suggests that the wind speed distribution is positively asymmetric. It shows that low-speed occurrences predominate, with some extreme spikes scattered throughout the distribution. It is supported by the fact that the kurtosis value is 10.70, which indicates a distribution with a high peak and a heavy tail, indicating a frequent occurrence of outliers. This complete analysis was carried out on 3,742 data points with a resolution of 5 minutes, creating a comprehensive picture of the short-term wind variability in tropical regions, including Pontianak, which has dynamic weather features.

The plot of the time series (Figure 2) shows that wind speeds during the observation period (about 18 December to 24 December) changed dynamically and exhibited significant differences. The majority of the wind speed figures were within the 0–10 kilometers per hour range, indicating that wind conditions were mainly light to moderate. However, there were a few abrupt bursts that exceeded 30 kilometers per hour. Given that these increases were random and did not recur regularly, they may be interpreted as gust events or local meteorological disturbances. Because these extreme values are present, the time series pattern becomes uneven and exhibits high volatility. There is no discernible increase or negative trend observable during the entirety of the observation period. Descriptively, this data does not imply a long-term trend, as the average value appears to remain relatively consistent over time.

**Table 2:** Numerical summary

| Count | Mean | St. Deviation | Variance | Kurtosis | Skewness |
|-------|------|---------------|----------|----------|----------|
| 3,742 | 4.28 | 6.42          | 41.19    | 10.70    | 2.72     |

In this research, the Kalman Filter is utilized to smooth the wind speed data gathered from minute-by-minute measurements. Due to local turbulence, rapid pressure changes, and sensor

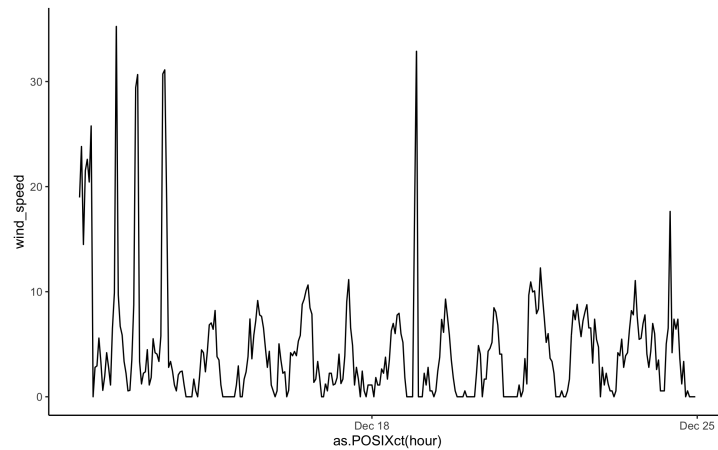


**Figure 2:** Time Series Plot of Wind Speed

interference, wind speed data collected at a high resolution tends to contain significant noise and sharp oscillations. Considering this, it is necessary to have a way of processing data capable of removing these random components without removing any significant information from the real trend. Estimating the real state value of the wind speed being observed is one of the functions of the Kalman Filter, which helps improve the quality of the raw data. For future analysis, this technique makes it possible to obtain more stable and representative data, particularly in time series prediction models such as ARIMA. In addition, it is anticipated that utilizing the state estimate from the Kalman Filter as the input to the model will increase the accuracy of the predictions and reduce the mistakes that the model produces. The implementation of the Kalman Filter involves constructing a state-space model derived from the attributes of the wind speed time series, specifically:

1. The state ( $x_k$ ) represents actual wind speed, which is not witnessed.
2. Measurement ( $z_k$ ) represents actual wind speed measured by device.
3. Transition matrix ( $A$ ) and observation matrix ( $H$ ) are based on linear data patterns.
4. Process and measurement noise covariance values are derived from the first data variability analysis.

### 3.1 ARIMA Modelling

This research examines the application of original data through the ARIMA model (called Data 1) and the ARIMA model applied to data smoothed by the Kalman Filter (called Data 2) to assess the influence of data filtering on the prediction model's quality. The first step in the modeling process is to check whether the dataset, including wind speed, exhibits stable statistical features over time. According to the results of the Augmented Dickey-Fuller (ADF) test, which are provided in Table 3, both Data 1 and Data 2 produce a p-value  $< 0.01$ . It provides strong statistical evidence to reject the null hypothesis ( $H_0$ ), which claims that the data have a unit root. Based on the mathematical analysis, this is a contradiction. Both datasets can be considered stationary as a result of this conclusion. Plots of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) (see Figure 3 and 4) are analyzed in order to determine the order ( $p, d, q$ ) after stationarity has been established successfully. The first set of data yields three primary model possibilities: ARIMA (5,0,0), (5,1,0), and (3,1,1). Even though the Data 1 are influenced by large random fluctuations (noise), the presence of several orders in the data suggests a complex temporal dependence structure. Meanwhile, Data 2 produces a wider set of options, including (1,0,0), (2,1,2), (1,1,1), and (5,0,0). A more distinct autoregressive pattern can be observed as a result, indicating that the recursive filtering process has successfully separated the deterministic signal from the random noise.

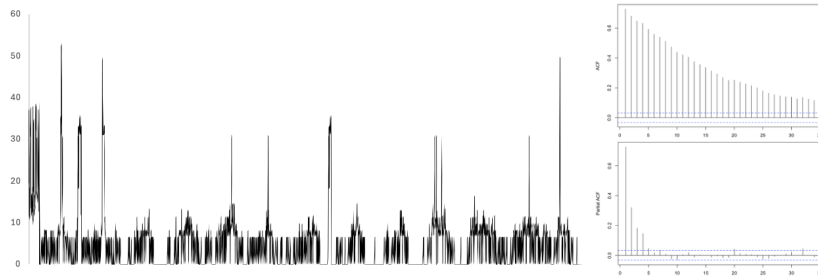


**Figure 3:** Time series, ACF, and PACF plot for wind speed data



**Figure 4:** Time series, ACF, and PACF plot for wind speed data smoothed by the Kalman Filter

**Table 3:** Stationarity Test and Order Identification

|                      | Data   | Result         |
|----------------------|--------|----------------|
| Stationarity Test    | Data 1 | p-value < 0.01 |
|                      | Data 2 | p-value < 0.01 |
| Order Identification | Data 1 | (5,0,0)        |
|                      |        | (5,1,0)        |
|                      |        | (3,1,1)        |
|                      | Data 2 | (1,0,0)        |
|                      |        | (2,1,2)        |
|                      |        | (1,1,1)        |
|                      |        | (5,0,0)        |

### 3.2 Parameter Estimation

In Table 4, which summarizes the parameter estimates for the candidate ARIMA models, it is clear that there is a significant difference in the statistical robustness of Data 1 and Data 2. The ARIMA(5,0,0) model exhibits a long-range temporal dependence, with highly significant autoregressive coefficients ( $\phi_1$  to  $\phi_5$ , p-values < 0.01). However, more sophisticated models indicate parameter inefficiency and inflated standard errors due to the presence of inherent stochastic noise. Alternatively, the Kalman Filter (Data 2) shows a notable signal-purification effect. The ARIMA(5,0,0) model reveals that only the initial lags ( $\phi_1, \phi_2$ ) remain highly significant (p-value < 0.001). It indicates that once high-frequency distortions are suppressed, the autoregressive structure becomes more parsimonious and stable. Further confirmation of the asymptotic stability of the estimates is provided by the fact that the intercept is consistently significant across all of the top-performing models. This implies that the Pontianak equatorial region has a strong long-term mean wind velocity.

**Table 4:** Parameter Estimation for Candidate ARIMA Models

| Data   | Candidate Model | Parameter  | Estimate (Std. Error) | Sig. |
|--------|-----------------|------------|-----------------------|------|
| Data 1 | ARIMA(5,0,0)    | Constant   | 4.2839 (0.5427)       | ***  |
|        |                 | $\phi_1$   | 0.3983 (0.0163)       | ***  |
|        |                 | $\phi_2$   | 0.1896 (0.0175)       | ***  |
|        |                 | $\phi_3$   | 0.1132 (0.0178)       | ***  |
|        |                 | $\phi_4$   | 0.1301 (0.0176)       | ***  |
|        |                 | $\phi_5$   | 0.0475 (0.0165)       | **   |
|        | ARIMA(5,1,0)    | $\phi_1$   | -0.5766 (0.0150)      | ***  |
|        |                 | $\phi_2$   | -0.3726 (0.0188)      | ***  |
|        |                 | $\phi_3$   | -0.2457 (0.0194)      | ***  |
|        |                 | $\phi_4$   | -0.1038 (0.0189)      | ***  |
|        |                 | $\phi_5$   | -0.0466 (0.0165)      | **   |
|        | ARIMA(3,1,1)    | $\phi_1$   | 0.0250 (0.0698)       |      |
|        |                 | $\phi_2$   | -0.0301 (0.0423)      |      |
|        |                 | $\phi_3$   | -0.0318 (0.0311)      |      |
|        |                 | $\theta_1$ | -0.6080 (0.0210)      | ***  |
| Data 2 | ARIMA(1,0,0)    | Constant   | 3.6756 (0.8939)       | ***  |
|        |                 | $\phi_1$   | 0.9906 (0.0024)       | ***  |
|        | ARIMA(2,1,2)    | $\phi_1$   | 0.2992 (0.4239)       |      |
|        |                 | $\phi_2$   | -0.0561 (0.3246)      |      |
|        |                 | $\theta_1$ | -0.3681 (0.4229)      |      |
|        |                 | $\theta_2$ | 0.1014 (0.3205)       |      |
|        | ARIMA(1,1,1)    | $\phi_1$   | -0.3132 (0.2014)      |      |
|        |                 | $\theta_1$ | 0.2446 (0.2052)       |      |
|        | ARIMA(5,0,0)    | Constant   | 3.7365 (0.9122)       | ***  |
|        |                 | $\phi_1$   | 0.9287 (0.0170)       | ***  |
|        |                 | $\phi_2$   | 0.0890 (0.0239)       | ***  |
|        |                 | $\phi_3$   | -0.0112 (0.0239)      |      |
|        |                 | $\phi_4$   | -0.0023 (0.0239)      |      |
|        |                 | $\phi_5$   | -0.0133 (0.0173)      |      |

Note: Standard errors are in parentheses. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

### 3.3 Model Selection and Evaluation

The comparative evaluation of candidate models is presented in Table 5. This evaluation is based on diagnostic performance and accuracy metrics and provides the statistical rationale for selecting the most appropriate forecasting model for each dataset.

Based on the significant parameters estimated in the previous stage, the functional forms of the optimal ARIMA models for both raw and filtered datasets are formulated. Table 6 presents the explicit mathematical equations derived from the ARIMA(5,0,0) structure, which will serve as the basis for the subsequent wind speed forecasting and residual stability analysis.

An indication that long-term stochastic disturbances still influence the raw data fluctuations is that the autoregressive coefficients in Data 1 are more uniformly distributed across the lags ( $\phi_1$  to  $\phi_5$ ). Conversely, in Data 2, there is a notable movement in the weight of the coefficient  $\phi_1$ , which has a value of 0.9287 and is in the dominant position. This physical proof demonstrates that the Kalman Filter successfully extracts the main signal, enabling the model to focus more on dependencies in the immediate past. To ensure that the model continues to provide a reliable reference point for average wind speed, a constant (3.7365 in Data 2) is essential. Because they have reduced the impact of random turbulence recorded in the white noise component ( $e_t$ ), the equations in Data 2 provide a more deterministic, trustworthy representation for use in disaster

**Table 5:** Model Selection Criteria: Diagnostic Checking and Accuracy Metrics

| Data   | Candidate Model | Diagnostic test |        | Accuracy Metrics |              |              | Status |
|--------|-----------------|-----------------|--------|------------------|--------------|--------------|--------|
|        |                 | Normality       | Indep. | AIC              | MAE          | RMSE         |        |
| Data 1 | ARIMA(5,0,0)    | Yes             | Yes    | <b>21088.69</b>  | 2.703        | <b>4.043</b> | Best   |
|        | ARIMA(5,1,0)    | Yes             | No     | 21174.83         | <b>2.559</b> | 4.093        | -      |
|        | ARIMA(3,1,1)    | Yes             | No     | 21160.54         | 2.564        | 4.087        | -      |
| Data 2 | ARIMA(1,0,0)    | Yes             | No     | 5816.30          | 0.096        | 0.523        | -      |
|        | ARIMA(2,1,2)    | Yes             | Yes    | 5811.31          | 0.093        | 0.523        | -      |
|        | ARIMA(1,1,1)    | Yes             | Yes    | 5808.05          | <b>0.092</b> | 0.523        | -      |
|        | ARIMA(5,0,0)    | Yes             | Yes    | <b>5807.22</b>   | 0.097        | <b>0.522</b> | Best   |

Note: Bold values indicate the selected best model for each data group.

**Table 6:** Mathematical Equations of the Best-Fitted ARIMA Models

| Data   | Order   | Mathematical Equation                                                                                |
|--------|---------|------------------------------------------------------------------------------------------------------|
| Data 1 | (5,0,0) | $Y_t = 4.2839 + 0.3983Y_{t-1} + 0.1896Y_{t-2} + 0.1132Y_{t-3} + 0.1301Y_{t-4} + 0.0475Y_{t-5} + e_t$ |
| Data 2 | (5,0,0) | $Y_t = 3.7365 + 0.9287Y_{t-1} + 0.0890Y_{t-2} - 0.0112Y_{t-3} - 0.0023Y_{t-4} - 0.0133Y_{t-5} + e_t$ |

Note:  $Y_t$  represents the wind speed at time  $t$ , and  $e_t$  denotes the white noise residual.

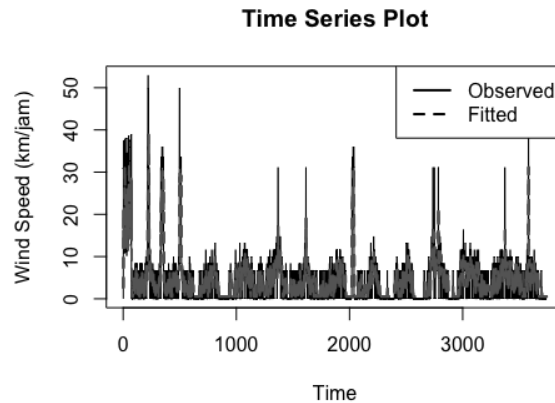
mitigation systems.

### 3.4 Model Verification and Performance Analysis

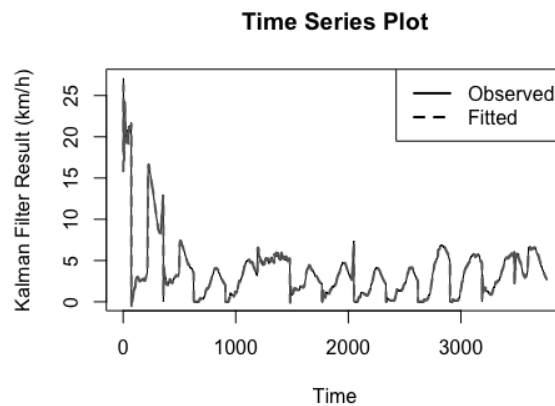
In the final stage of model evaluation, the actual observed values are compared with the model's estimates, and residual stability is assessed. The purpose of this visual analysis is to verify that forecast errors are within the statistical control limits and to ensure that the ARIMA(5,0,0) model can make accurate predictions.

A comparison between the actual data and the model estimates is shown in Figure 5a (for Data 1) and Figure 5b (for Data 2), respectively. In Data 1, the plot shows quite dramatic variations and a significant level of uncertainty (noise). As a result, the estimated line struggles to identify the primary trend amid high-frequency noise. The predicted line in Data 2 demonstrates a very precise tracking of the filtered signal, in contrast to the previous data. It provides a visual demonstration that the Kalman Filter integration has effectively extracted the deterministic component of wind speed. As a result, the ARIMA model generates predictions that are significantly smoother and more accurate than those produced from the raw data.

An Individual Moving Range (IMR) control chart analysis was performed on the model residuals, as depicted in Figure 6, to ensure the model's statistical trustworthiness. A substantial error variance is reflected in the residuals of Data 1, which show a very wide distribution of points with high control limits (UCL and LCL). This pattern is observed. Nevertheless, the control limits are significantly reduced in the vicinity of the Center Line (CL) in the I-MR Chart for Data 2, where the majority of the residuals are stably concentrated. The fact that the residuals in Data 2 remain stable despite several outliers in the early period indicates that the model has successfully captured almost all the systematic information in the data. For wind speed forecasting at the research location, the Kalman Filter-based ARIMA(5,0,0) model is deemed to be valid and reliable. These results demonstrate that the residuals of the filtered model are statistically controlled and consist of white noise.



(a) Raw Data (Data 1)



(b) Filtered Data (Data 2)

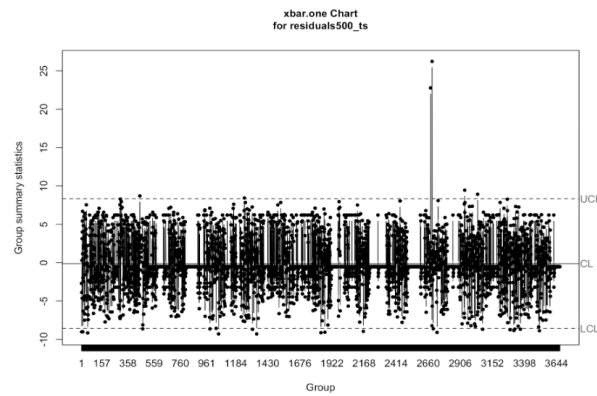
**Figure 5:** Comparison between observed wind speed and ARIMA(5,0,0) model estimations for both datasets.

## 4 Conclusion

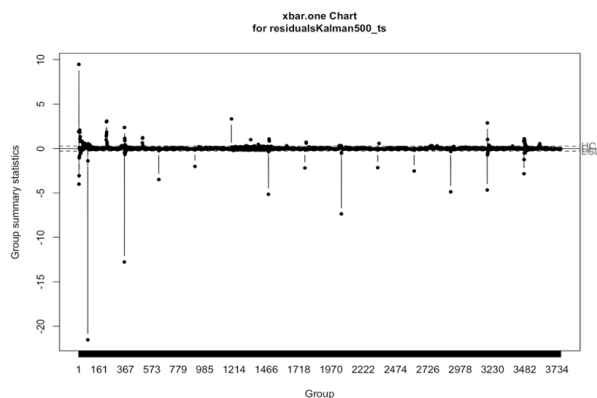
This study concludes that the high-frequency wind speed data in the area under investigation exhibit significant volatility, with the highest recorded velocity of 52.95 m/s and an average of 4.285 m/s. The high skewness and kurtosis indicate that the distribution is not Gaussian and is prone to strong outliers, making typical time series modeling more difficult.

Due to high-frequency stochastic noise and sudden data spikes, the findings reveal that a standalone ARIMA model applied to raw data produces unstable estimations. The incorporation of the Kalman Filter as a pre-processing stage, on the other hand, results in a more parsimonious and dependable ARIMA(5,0,0) model. It is accomplished by reducing random fluctuations, thereby considerably improving the signal quality.

This research uses Individual Moving Range (IMR) control charts as a diagnostic and verification tool, a method rarely used in the meteorological literature. This research introduces a significant methodological innovation. When compared with the raw data, which shows a large number of out of control points, the control charts provide strong evidence of the superior stability of the Kalman-filtered dataset. This is because almost all observations remain within the  $3\sigma$  statistical control limits. The contributions these discoveries make to the development of localized weather-forecasting methodologies, the optimization of early warning systems, and the strategic management of wind-based renewable energy in tropical regions are considerable and



(a) Residuals of Data 1



(b) Residuals of Data 2

**Figure 6:** Residual stability verification using Individual Moving Range (IMR) charts. Note the contrast between the wide variance in Data 1 (a) and the refined, narrowed control limits in Data 2 (b) following the signal purification process, indicating the robustness of the hybrid ARIMA-Kalman approach.

significant.

## CRedit Authorship Contribution Statement

**Nurfitri Imro'ah:** Conceptualization, Methodology, Software, Writing–Original Draft. **Nur'ainul Miftahul Huda:** Data Curation, Formal Analysis, Software. **Kartika Sari:** Investigation, Validation, Visualization. **Rahmi Hidayati:** Data Curation, Validation, Visualization.

## Declaration of Generative AI and AI-assisted technologies

No generative AI or AI-assisted technologies were used during the preparation of this manuscript.

## Declaration of Competing Interest

The authors declare no competing interests.

## Funding and Acknowledgments

This research received no external funding.

## Data and Code Availability

The data and code supporting the findings of this study are available from the corresponding author upon reasonable request and subject to confidentiality agreements.

## References

- [1] M. R. Abdillah et al., "Extreme wind variability and wind map development in western java, indonesia," *International Journal of Disaster Risk Science*, vol. 13, pp. 465–480, 3 Jun. 2022. DOI: [10.1007/s13753-022-00420-7](https://doi.org/10.1007/s13753-022-00420-7).
- [2] G. Gualtieri, "Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 167, p. 112741, Oct. 2022. DOI: [10.1016/j.rser.2022.112741](https://doi.org/10.1016/j.rser.2022.112741).
- [3] J. P. Murcia et al., "Validation of european-scale simulated wind speed and wind generation time series," *Applied Energy*, vol. 305, p. 117794, Jan. 2022. DOI: [10.1016/j.apenergy.2021.117794](https://doi.org/10.1016/j.apenergy.2021.117794).
- [4] H. R. Alsamamra, S. Salah, and J. H. Shoqeir, "Performance analysis of arima model for wind speed forecasting in jerusalem, palestine," *Energy Exploration Exploitation*, vol. 42, pp. 1727–1746, 5 Sep. 2024. DOI: [10.1177/01445987241248201](https://doi.org/10.1177/01445987241248201).
- [5] A. Demirtop and O. Sevlı, "Wind speed prediction using lstm and arima time series analysis models: A case study of gelibolu," *Turkish Journal of Engineering*, vol. 8, pp. 524–536, 3 Jul. 2024. DOI: [10.31127/tuje.1431629](https://doi.org/10.31127/tuje.1431629).
- [6] K. Szostek, D. Mazur, G. Drałus, and J. Kuszniel, "Analysis of the effectiveness of arima, sarima, and svr models in time series forecasting: A case study of wind farm energy production," *Energies*, vol. 17, p. 4803, 19 Sep. 2024. DOI: [10.3390/en17194803](https://doi.org/10.3390/en17194803).
- [7] X. Li, K. Li, S. Shen, and Y. Tian, "Exploring time series models for wind speed forecasting: A comparative analysis," *Energies*, vol. 16, p. 7785, 23 Nov. 2023. DOI: [10.3390/en16237785](https://doi.org/10.3390/en16237785).
- [8] C.-Y. Ho, K.-S. Cheng, and C.-H. Ang, "Utilizing the random forest method for short-term wind speed forecasting in the coastal area of central taiwan," *Energies*, vol. 16, p. 1374, 3 Jan. 2023. DOI: [10.3390/en16031374](https://doi.org/10.3390/en16031374).
- [9] V. K. Saini, R. Kumar, A. S. Al-Sumaiti, S. A., and E. Heydarian-Forushani, "Learning based short term wind speed forecasting models for smart grid applications: An extensive review and case study," *Electric Power Systems Research*, vol. 222, p. 109502, Sep. 2023. DOI: [10.1016/j.epsr.2023.109502](https://doi.org/10.1016/j.epsr.2023.109502).
- [10] S. M. Malakouti, "Estimating the output power and wind speed with ml methods: A case study in texas," *Case Studies in Chemical and Environmental Engineering*, vol. 7, p. 100324, Jun. 2023. DOI: [10.1016/j.cscee.2023.100324](https://doi.org/10.1016/j.cscee.2023.100324).
- [11] C. Huang, H. R. Karimi, P. Mei, D. Yang, and Q. Shi, "Evolving long short-term memory neural network for wind speed forecasting," *Information Sciences*, vol. 632, pp. 390–410, Jun. 2023. DOI: [10.1016/j.ins.2023.03.031](https://doi.org/10.1016/j.ins.2023.03.031).
- [12] J. Wang, H. Zhu, Y. Zhang, F. Cheng, and C. Zhou, "A novel prediction model for wind power based on improved long short-term memory neural network," *Energy*, vol. 265, p. 126283, Feb. 2023. DOI: [10.1016/j.energy.2022.126283](https://doi.org/10.1016/j.energy.2022.126283).
- [13] Y. Cui, Z. Chen, Y. He, X. Xiong, and F. Li, "An algorithm for forecasting day-ahead wind power via novel long short-term memory and wind power ramp events," *Energy*, vol. 263, p. 125888, Jan. 2023. DOI: [10.1016/j.energy.2022.125888](https://doi.org/10.1016/j.energy.2022.125888).

- [14] S. M. Valdivia-Bautista, J. A. Domínguez-Navarro, M. Pérez-Cisneros, C. J. Vega-Gómez, and B. Castillo-Téllez, "Artificial intelligence in wind speed forecasting: A review," *Energies*, vol. 16, p. 2457, 5 Mar. 2023. DOI: [10.3390/en16052457](https://doi.org/10.3390/en16052457).
- [15] J. Duan, H. Zuo, Y. Bai, J. Duan, M. Chang, and B. Chen, "Short-term wind speed forecasting using recurrent neural networks with error correction," *Energy*, vol. 217, p. 119397, Feb. 2021. DOI: [10.1016/j.energy.2020.119397](https://doi.org/10.1016/j.energy.2020.119397).
- [16] A. Barjasteh, S. H. Ghafouri, and M. Hashemi, "A hybrid model based on discrete wavelet transform (dwt) and bidirectional recurrent neural networks for wind speed prediction," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107340, Jan. 2024. DOI: [10.1016/j.engappai.2023.107340](https://doi.org/10.1016/j.engappai.2023.107340).
- [17] Q. B. Phan and T. T. Nguyen, "Enhancing wind speed forecasting accuracy using a two-nested ceemdan-cnn-bilstm model," *ICT Express*, vol. 10, pp. 485–490, 3 Jun. 2024. DOI: [10.1016/j.icte.2023.11.009](https://doi.org/10.1016/j.icte.2023.11.009).
- [18] Y.-M. Zhang and H. Wang, "Multi-head attention-based probabilistic cnn-bilstm for day-ahead wind speed forecasting," *Energy*, vol. 278, p. 127865, Sep. 2023. DOI: [10.1016/j.energy.2023.127865](https://doi.org/10.1016/j.energy.2023.127865).
- [19] N. Imro'ah, N. M. Huda, H. Pratiwi, and M. Y. Ayyash, "Spatio-temporal modeling of fire hotspots using gstar(1;1) model with meteorology based weight matrices," *Hacetatepe Journal of Mathematics and Statistics*, vol. 54, pp. 2525–2542, Dec. 2025. DOI: [10.15672/hujms.1726843](https://doi.org/10.15672/hujms.1726843).
- [20] Y. Li, K. Sun, Q. Yao, and L. Wang, "A dual-optimization wind speed forecasting model based on deep learning and improved dung beetle optimization algorithm," *Energy*, vol. 286, p. 129604, Jan. 2024. DOI: [10.1016/j.energy.2023.129604](https://doi.org/10.1016/j.energy.2023.129604).
- [21] N. Imro'ah and N. M. Huda, "Double intervention analysis on the arima model of covid-19 cases in bali," *Journal of the Indonesian Mathematical Society*, vol. 31, p. 1347, 1 Mar. 2025. DOI: [10.22342/jims.v31i1.1347](https://doi.org/10.22342/jims.v31i1.1347).
- [22] R. Ospina, J. A. M. Gondim, V. Leiva, and C. Castro, "An overview of forecast analysis with arima models during the covid-19 pandemic: Methodology and case study in brazil," *Mathematics*, vol. 11, p. 3069, 14 Jul. 2023. DOI: [10.3390/math11143069](https://doi.org/10.3390/math11143069).
- [23] S. Zhong, B. Peng, J. He, Z. Feng, M. Li, and G. Wang, "Kalman filtering based on dynamic perception of measurement noise," *Mechanical Systems and Signal Processing*, vol. 213, p. 111343, May 2024. DOI: [10.1016/j.ymssp.2024.111343](https://doi.org/10.1016/j.ymssp.2024.111343).
- [24] R. Havangi, "Adaptive robust unscented kalman filter with recursive least square for state of charge estimation of batteries," *Electrical Engineering*, vol. 104, pp. 1001–1017, 2 Apr. 2022. DOI: [10.1007/s00202-021-01358-7](https://doi.org/10.1007/s00202-021-01358-7).
- [25] C.-M. Wen, Z. Yan, Y.-C. Liang, H. Wu, L. Zhou, and Y. Yao, "A control chart-based symbolic conditional transfer entropy method for root cause analysis of process disturbances," *Computers Chemical Engineering*, vol. 164, p. 107902, Aug. 2022. DOI: [10.1016/j.compchemeng.2022.107902](https://doi.org/10.1016/j.compchemeng.2022.107902).
- [26] M. Khodarahmi and V. Maihami, "A review on kalman filter models," *Archives of Computational Methods in Engineering*, vol. 30, pp. 727–747, 1 Jan. 2023. DOI: [10.1007/s11831-022-09815-7](https://doi.org/10.1007/s11831-022-09815-7).
- [27] P. H. Tran, A. A. Nadi, T. H. Nguyen, K. D. Tran, and K. P. Tran, "Application of machine learning in statistical process control charts: A survey and perspective," in 2022, pp. 7–42. DOI: [10.1007/978-3-030-83819-5\\_2](https://doi.org/10.1007/978-3-030-83819-5_2).
- [28] W. Zhang, X. Zhao, Z. Liu, K. Liu, and B. Chen, "Converted state equation kalman filter for nonlinear maneuvering target tracking," *Signal Processing*, vol. 202, p. 108741, Jan. 2023. DOI: [10.1016/j.sigpro.2022.108741](https://doi.org/10.1016/j.sigpro.2022.108741).

- 428 [29] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. S. Rossi, "Unsupervised anomaly detection  
429 for iot-based multivariate time series: Existing solutions, performance analysis and future  
430 directions," *Sensors*, vol. 23, p. 2844, 5 Mar. 2023. DOI: [10.3390/s23052844](https://doi.org/10.3390/s23052844).
- 431 [30] W. Yao, D. Li, and L. Gao, "Fault detection and diagnosis using tree-based ensemble  
432 learning methods and multivariate control charts for centrifugal chillers," *Journal of*  
433 *Building Engineering*, vol. 51, p. 104243, Jul. 2022. DOI: [10.1016/j.jobe.2022.104243](https://doi.org/10.1016/j.jobe.2022.104243).
- 434 [31] T. Umairah, N. Imro'ah, and N. M. Huda, "Arima model verification with outlier factors  
435 using control chart," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, pp. 0579–  
436 0588, 1 Mar. 2024. DOI: [10.30598/barekengvol18iss1pp0579-058](https://doi.org/10.30598/barekengvol18iss1pp0579-058).
- 437 [32] N. Imro'ah and N. M. Huda, "Control chart as verification tools in time series model,"  
438 *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, pp. 995–1002, 3 Sep. 2022.  
439 DOI: [10.30598/barekengvol16iss3pp995-1002](https://doi.org/10.30598/barekengvol16iss3pp995-1002).
- 440 [33] Zhang, Subramanian, Pinelli, Lazarus, Besing, and R. Cortes, "Performance characterization  
441 of a wireless sensors network system (wsns) for measurements of hurricane wind effects on  
442 structures," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 254, p. 105895,  
443 Nov. 2024. DOI: [10.1016/j.jweia.2024.105895](https://doi.org/10.1016/j.jweia.2024.105895).