earth

MDPI

Article

A Spectral Analysis-Driven SARIMAX Framework with Fourier Terms for Monthly Dust Concentration Forecasting

Ommolbanin Bazrafshan [1], Hossein Zamani [2], Behnoush Farokhzadeh [3] and Tommaso Caloiero [4,*]

[1] Department of Natural Resources, Faculty of Agricultural and Natural Resources, University of Hormozgan, Bandar Abbas 79161-93145, Iran; o.bazrafshan@hormozgan.ac.ir
[2] Department of Statistics, Faculty of Science, University of Hormozgan, Bandar Abbas 79161-93145, Iran; zamani.huni@hormozgan.ac.ir
[3] Department of Nature Engineering, Faculty of Natural Resources and Environment, Malayer University, Malayer 84621-65741, Iran; b.farokhzadeh@malayeru.ac.ir
[4] National Research Council of Italy, Research Institute for Geo-Hydrological Protection (CNR-IRPI), 87036 Rende, Italy
* Correspondence: tommaso.caloiero@cnr.it; Tel.: +39-0984-841-464

Abstract

This study aimed to forecast monthly $PM_{2.5}$ concentrations in Zabol, one of the world's most dust-prone regions, using four time series models: SARIMA, SARIMAX enhanced with Fourier terms (selected based on spectral peak analysis), TBATS, and a novel hybrid ensemble. Spectral analysis identified a dominant annual cycle (frequency 0.083), which justified the inclusion of two Fourier harmonics in the SARIMAX model. Results demonstrated that the hybrid model, which optimally combined forecasts from the three individual models (with weights $\omega_2 = 0.628$ for SARIMAX, $\omega_3 = 0.263$ for TBATS, and $\omega_1 = 0.109$ for SARIMA), outperformed all others across all evaluation metrics, achieving the lowest AIC (1835.04), BIC (1842.08), RMSE (9.42 $\mu g/m^3$), and MAE (7.43 $\mu g/m^3$). It was also the only model exhibiting no significant residual autocorrelation (Ljung–Box p-value = 0.882). Forecast uncertainty bands were constant across the prediction horizon, with widths of approximately ± 11.39 $\mu g/m^3$ for the 80% confidence interval and ± 22.25 $\mu g/m^3$ for the 95% confidence interval, reflecting fixed absolute uncertainty in the multi-step forecasts. The proposed hybrid framework provides a robust foundation for early warning systems and public health management in dust-affected arid regions.

Keywords: hybrid time series models; deterministic seasonality; dust concentration forecast; early warning systems

check for updates

Academic Editors: Charles Jones and Hesham M. El-Askary

Received: 31 July 2025
Revised: 30 September 2025
Accepted: 7 October 2025
Published: 10 October 2025

Citation: Bazrafshan, O.; Zamani, H.; Farokhzadeh, B.; Caloiero, T. A Spectral Analysis-Driven SARIMAX Framework with Fourier Terms for Monthly Dust Concentration Forecasting. *Earth* **2025**, *6*, 123. https://doi.org/10.3390/earth6040123

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ambient air pollution, particularly elevated levels of fine particulate matter ($PM_{2.5}$ and PM_{10}), poses a significant threat to environmental sustainability and public health worldwide. Prolonged exposure to $PM_{2.5}$ is strongly linked to higher rates of cardiovascular and respiratory morbidity and mortality [1]. In arid and semi-arid regions, frequent dust storms further degrade air quality and amplify the health risks associated with fine particulate matter [2]. With millions of people residing in dust-prone zones across the Middle East and Central Asia, protecting public health requires timely and accurate forecasting of airborne dust concentrations [3]. In this regional context, operational dust-forecasting capabilities are critical for risk mitigation and environmental management [4].

Despite significant advancements in atmospheric monitoring, operational forecasting of dust concentrations remains a complex challenge due to the nonlinear behavior

and pronounced temporal variability of particulate matter. Persistent dust events and fluctuating pollutant loads in arid and urban regions demand robust predictive tools [5]. Classical time series models such as ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal Autoregressive Integrated Moving Average) have been widely utilized to forecast environmental variables, including solar radiation [6], temperature [7], and both $PM_{2.5}$ and PM_{10} concentrations [8–10]. These models are generally well-suited for capturing short-term trends and seasonal cycles in air quality data, requiring limited input parameters and enabling computational efficiency [11,12]. Nevertheless, they often fail to fully capture persistent or deterministic seasonal patterns, particularly when complex cyclic components exist in the data, as evidenced by residual autocorrelation at seasonal lags, which can compromise forecast reliability [13].

To address these limitations, extensions such as SARIMAX models—by incorporating exogenous regressors (e.g., meteorological variables, calendar/seasonal indicators, or lagged external indices)—can better capture systematic cyclic behavior and thereby improve forecast performance [7,11].

Models like TBATS (Box–Cox transformation, ARMA errors, Trend, and Seasonal components). further expand this capability by accommodating multiple or non-integer seasonality, making them attractive for complex time series [14]. However, even these flexible models may be limited in spatial generalization and can underperform during rapid, non-stationary changes such as extreme dust events or abrupt meteorological shifts [15,16].

Recent research has thus shifted toward hybrid approaches, integrating external regressors, spectral decomposition, or nonlinear machine learning techniques to overcome the limitations of classical statistical models [9,17,18]. For example, Doreswamy et al. [12] and Díaz-Robles et al. [19] have shown that ARIMA-based hybrid frameworks with machine learning components can effectively capture both linear and nonlinear dependencies in $PM_{2.5}$ and PM_{10} data. The widespread use of SARIMA and SARIMAX for air quality prediction has been documented in various regions [20,21], but these models often struggle to capture persistent deterministic and recurring seasonal cycles fully. To address this limitation, spectral decomposition techniques, such as the inclusion of Fourier terms in SARIMAX, have been increasingly adopted [11,22].

Recent studies further confirm the value of SARIMAX–Fourier formulations for environmental applications: Necula et al. [23] integrated Fourier-based SARIMAX within hybrid pollution-forecasting frameworks, Avila et al. [24] demonstrated its effectiveness for multi-seasonal environmental series. These contributions underscore the growing relevance of SARIMAX–Fourier models for environmental and air-quality forecasting.

Recent advances in air quality forecasting have demonstrated the high effectiveness of hybrid modeling approaches for predicting $PM_{2.5}$ concentrations, particularly in complex urban environments. By combining statistical models like ARIMA and SARIMA with advanced machine learning and deep learning methods, including BiLSTM, SALSTM, LSSVM, and ANN, researchers have achieved notable improvements in predictive accuracy. These hybrids leverage both linear and nonlinear modeling strengths, capturing intricate temporal patterns in pollutant data [23,25,26]. For example, Du et al. [25] showed substantial gains with a SARIMA–deep learning hybrid, while Albeladi et al. [26] highlighted the effectiveness of LSTM–ARIMA integration. Incorporating spatial features, as Liu et al. [27] did for Wuhan, further enhances accuracy in highly variable pollution settings. Necula et al. [23] developed a SARIMA–BiLSTM hybrid, achieving robust results, and Solis Teran et al. [28] confirmed the operational value of hybrid frameworks in Lima, Peru.

Complementing these methods, the TBATS model represents a major advancement for forecasting time series with complex or non-integer seasonality, such as those found in environmental applications. TBATS utilizes trigonometric terms within its exponential smooth-

ing framework to flexibly model overlapping or evolving seasonality [14,29,30]. Naim et al. [29] and Andreozzi et al. [30] showed that TBATS outperforms traditional approaches, particularly in multi-seasonal contexts. Such hybrid approaches highlight the importance of combining linear and nonlinear modeling strategies, informed by the nature of the data, to achieve accurate forecasts in hydrological and environmental applications [18,19]. Together, these approaches illustrate the importance of combining statistical rigor with domain knowledge and spectral diagnostics to develop robust forecasting tools for environmental health applications.

Despite substantial progress in time series modeling for air quality forecasting, notable research gaps persist, particularly in arid regions such as Zabol in southeastern Iran, where dust dynamics are exceptionally complex and environmental extremes are intensified by the desiccation of the Hamoun wetlands. Zabol has become internationally recognized for its frequent and severe dust storms, with annual $PM_{2.5}$ and PM_{10} concentrations regularly surpassing both national and global air quality standards. The region's characteristic "120-day wind" phenomenon, driven by hyper-arid climatic conditions, results in recurrent, acute dust episodes that substantially elevate risks of respiratory and cardiovascular diseases among the local population [31–35].

While several classical time series models have been applied to air quality forecasting, there is a notable lack of studies systematically evaluating advanced hybrid models, particularly those integrating spectral analysis and deterministic seasonal components, for dust prediction in such challenging environments. Addressing this gap, the present study develops and assesses a SARIMAX model augmented with Fourier terms, guided by spectral diagnostics, for monthly $PM_{2.5}$ forecasting in Zabol. By comparing this approach with traditional SARIMA and TBATS models, the research aims to offer both methodological advancement and actionable insights to support early warning systems and dust management strategies in one of the world's most severely dust-affected regions.

2. Study Area and Datasets

Zabol, situated in southeastern Iran, is widely documented as a wind-exposed, dust-prone area within the Sistan Basin [32,33,36]. The area's hyper-arid climate and low annual precipitation, combined with the progressive desiccation of the Hamoun wetlands, have dramatically increased the frequency and severity of dust storms in recent decades [37,38]. This process has been further exacerbated by recurring droughts, significant land use changes such as agricultural expansion and upstream water diversion, all of which have accelerated wetland loss and intensified dust emission events, leading to pronounced seasonality in particulate pollution [37]. Zabol is strongly affected by the "120-day wind" (Levar), a persistent summer wind system that transports large volumes of mineral dust from the desiccated Hamoun basin into urban and rural environments [33]. These conditions, together with the region's flat and sparsely vegetated topography, create optimal circumstances for dust mobilization and atmospheric transport [32].

Zabol has been reported to exhibit elevated annual average $PM_{2.5}$ and PM_{10} levels in Iran [31]. Given these unique climatic and geographical features, Zabol represents an ideal natural laboratory for investigating the dynamics of particulate matter pollution and the performance of advanced forecasting models in extreme dust conditions. For this study, a monthly time series dataset of fine particulate matter ($PM_{2.5}$; aerodynamic diameter < 2.5 μm) was collected from the primary air quality monitoring station in Zabol. The dataset spans 21 years from 2004 to 2025, yielding 264 monthly observations from government-operated environmental monitoring stations. The geographical location of Zabol and the monitoring site is illustrated in Figure 1.

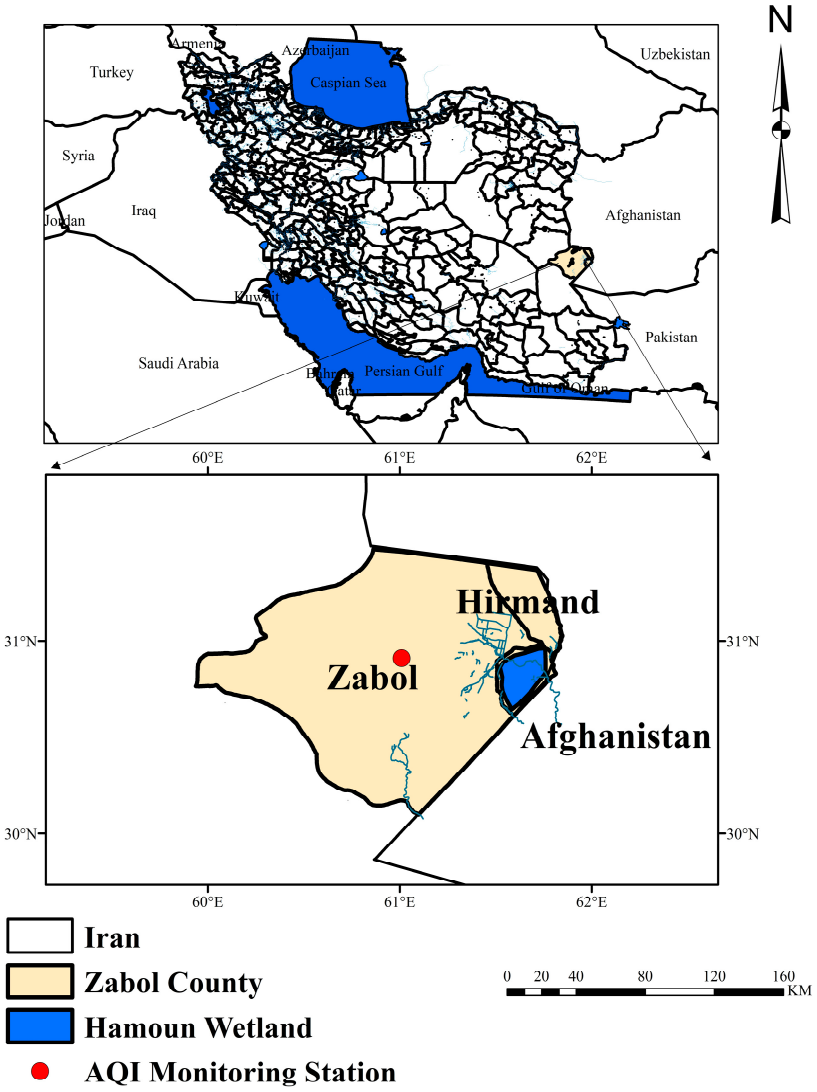

Figure 1. Location of the study area.

3. Methodology

This study aims to develop a reliable and interpretable framework for forecasting monthly concentrations of fine particulate matter ($PM_{2.5}$) in Zabol. The research applies and compares four time series modeling approaches: the classical Seasonal ARIMA (SARIMA) model, the TBATS model, and a modified SARIMAX model augmented with Fourier terms and a hybrid model. The methodology consists of a sequence of analytical stages and a comparative evaluation of model performance based on information criteria and forecast accuracy metrics. All computations and visualizations were performed in R (v4.4.2) using the forecast, stats, and ggplot2 packages.

3.1. Exploratory Data Analysis and Stationarity Assessment

An essential first step in time series modeling is to evaluate the statistical characteristics and stationarity of the data, particularly for ARIMA-based approaches [13]. Exploratory data analysis involved visualizing the time series to identify recurring seasonal patterns and any long-term trends, which are important for determining the suitability of seasonal models [22,39]. To formally test for stationarity, the Augmented Dickey–Fuller (ADF) test was applied, with the optimal lag order selected by AIC [40–42]. A statistically significant result ($p < 0.05$) confirms stationarity in the original series. The ACF and PACF plots were used to diagnose serial dependencies and identify the presence of seasonal

cycles. Distinct spikes at seasonal lags (e.g., lag 12) in the ACF confirm deterministic seasonality, while the PACF helped specify the order of autoregressive terms [43,44]. The absence of persistent autocorrelation at non-seasonal lags suggested suitability for seasonal ARIMA or its extensions.

3.2. Spectral Analysis and Fourier Term Extraction

Spectral analysis was employed to identify and quantify the dominant periodic components within the monthly $PM_{2.5}$ time series. For a discrete stationary time series Y_t with the autocovariance function $\gamma(h)$ the spectral density is defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h} \quad (1)$$

where $-\pi \leq \omega \leq \pi$. The spectral peak is a local maximum of the spectral density estimate. The dominant frequency is $\omega^* = \arg\max_{\omega} f(\omega)$. If the dominant peak is at frequency $f^* = f(\omega^*)$ then the corresponding period is $T = 1/f^*$.

The periodogram, an established nonparametric estimator of spectral density, was utilized to transform the series into the frequency domain, revealing latent periodic structures that may not be readily apparent in time-domain diagnostics such as ACF and PACF plots [39]. Such spectral techniques are commonly applied in environmental time series analysis, particularly for identifying regular seasonal cycles caused by natural processes [44,45]. The periodogram was computed to detect dominant seasonal frequencies, with particular attention to identifying annual cycles (e.g., frequency 1/12 for monthly data). Similar approaches have been utilized in ground-level $PM_{2.5}$ studies to empirically model annual seasonal components [46,47]. To effectively represent deterministic seasonality, two Fourier term pairs, sine and cosine, were generated and included in the SARIMAX model as exogenous regressors. These terms are formally defined as:

$$S_{k,t} = sin\left(\frac{2\pi kt}{s}\right) \text{ and } C_{k,t} = cos\left(\frac{2\pi kt}{s}\right), \quad (2)$$

where $S_{k,t}$ and $C_{k,t}$ represent sine and cosine components at a frequency harmonic k, t denotes the time index (month), s = 12 represents the seasonal period (12 months). k = 1, 2 is the number of harmonics selected based on the spectral peak magnitude.

The inclusion of Fourier terms enables capturing deterministic seasonal variation without excessively increasing model complexity. Prior studies have demonstrated that Fourier-enhanced SARIMAX models yield improved forecast accuracy and residual whiteness when modeling environmental time series containing fixed seasonal cycles [14,22,23].

3.3. Model Formulation and Estimation

To forecast monthly $PM_{2.5}$ concentrations in Zabol, four time series models were specified and estimated according to the statistical characteristics of the dataset and the identified seasonality.

3.3.1. SARIMA Model

A Seasonal ARIMA (SARIMA) model was used to represent both the short-term autocorrelation and the seasonal structure in the monthly $PM_{2.5}$ series. The general SARIMA formulation is:

$$SARIMA(p,d,q)(P,D,Q)_S, \quad (3)$$

where p, d, q are the orders of the non-seasonal autoregressive (AR), differencing (I), and moving average (MA) components; P, D, Q are the orders of the seasonal AR, I, and MA components; s is the seasonal period (here, s = 12 for monthly data).

Using backshift notation, the model is expressed as:

$$\Phi\left(B^S\right)\phi(B)(1-B)^d\left(1-B^S\right)^D y_t = \Theta\left(B^S\right)\theta(B)\varepsilon_t, \quad (4)$$

Here, B is the lag operator, and $\theta(B)$ are non-seasonal AR and MA polynomials; $\Phi\left(B^S\right)$ and $\Theta\left(B^S\right)$ are seasonal AR and MA polynomials; $(1-B)^d\left(1-B^S\right)^D$ denote non-seasonal and seasonal differencing, respectively [39,45]. Model identification was based on ACF and PACF diagnostics, with parameter estimation guided by minimizing RMSE and selecting the lowest Akaike Information Criterion (AIC). The selected model was [12]:

$$SARIMA(1,0,2)(2,0,2), \quad (5)$$

This configuration effectively captures short-term dynamics and annual seasonal variations present in the $PM_{2.5}$ series.

3.3.2. SARIMAX Model with Fourier Terms

The SARIMAX model extends the SARIMA framework by allowing the inclusion of exogenous regressors, such as meteorological variables or lagged external series [48]. In this study, deterministic Fourier terms (sine and cosine pairs) were incorporated as exogenous regressors to parsimoniously capture fixed seasonal cycles, following standard forecasting practice [22] and trigonometric seasonal modeling [14]. This approach enables the model to represent smooth and repeatable periodicity with only a few parameters, avoiding the need for numerous seasonal lags or dummy variables for each season.

The general SARIMAX formulation with Fourier terms is expressed as:

$$y_t = \mu + \sum_{k=1}^{K}\left[a_k sin\left(\frac{2\pi kt}{s}\right) + y_k cos\left(\frac{2\pi kt}{s}\right)\right]\Phi\left(B^S\right)\phi(B)(1-B)^d\left(1-B^S\right)^D y_t = \Theta\left(B^S\right)\theta(B)\varepsilon_t \quad (6)$$

where y_t is the observed $PM_{2.5}$ concentration at time t, μ is the model intercept, a_k and y_k are the Fourier coefficients for sine and cosine terms, s = 12 denotes the seasonal cycle length (months), K is the number of harmonics (here, K = 2), $\phi(B)$, $\theta(B)$, and $= \Theta\left(B^S\right)$ are the non-seasonal and seasonal AR and MA polynomials, and ε_t is white noise error.

To capture extreme values in the time series, we constructed two binary indicator variables. The high-extreme indicator and the low-extreme indicator, which are denoted by $I_{\{y_t > Q_{0.95}\}}$, $I_{\{y_t < Q_{0.05}\}}$ respectively. These dummy variables were included as exogenous regressors in a SARIMAX model to explicitly account for the influence of unusually large and unusually small observations on the conditional mean dynamics. These dummy variables are defined as:

$$I_{\{y_t > Q_{0.95}\}} = \begin{cases} 0, y_t < Q_{0.95} \\ 1, y_t \geq Q_{0.95} \end{cases}, I_{\{y_t < Q_{0.05}\}} = \begin{cases} 0, y_t \geq Q_{0.05} \\ 1, y_t < Q_{0.05} \end{cases} \quad (7)$$

3.3.3. TBATS Model

The TBATS model is an advanced forecasting framework specifically designed to handle complex, overlapping, and non-integer seasonal patterns in time series data [14,22]. TBATS stands for Trigonometric seasonality, Box–Cox transformation, ARMA errors, Trend, and Seasonal components. This approach models seasonal effects using a Fourier (trigonometric) series, automatically applies Box–Cox transformations to stabilize variance, and

incorporates an ARMA structure for modeling residual autocorrelation, as well as trend and damping components where appropriate [18]. In practice, information criteria (notably AIC) were used to guide model selection—covering transformation, trend, damping, and ARMA terms—following general forecasting guidance [49] and evidence from TBATS applications [29,30], to achieve parsimonious yet effective models.

For this study, the TBATS model was implemented using the TBATS function from the forecast package in R [50]. The algorithm automatically determines the optimal Box–Cox λ parameter for variance stabilization, inclusion or exclusion of trend and damping components, ARMA order for error modeling, as well as seasonal periods and the number of harmonics for each detected cycle [14].

TBATS models seasonal components using a truncated Fourier series, which enables flexible modeling of single or multiple seasonal cycles, even when the seasonal periods are non-integer or overlapping. For the i-th seasonal period (l_i), the seasonal component is expressed as:

$$Y_{s,i}(t) = \sum_{k=1}^{K} \left[\beta_{i,k} sin\left(\frac{2\pi kt}{l_i}\right) + y_{i,k} cos\left(\frac{2\pi kt}{l_i}\right) \right], \quad (8)$$

where k_i is the number of Fourier terms for this seasonality. The overall TBATS model integrates these seasonal components with the trend and error terms as follows:

$$Y_t = T_t \sum_{i=1}^{M} Y_{S,i}(t) + \varepsilon_t, \quad (9)$$

where T denotes the trend term [51].

3.4. Model Evaluation

The forecasting performance of the four time-series models—SARIMA (1,0,2) (2,0,2) [12], SARIMAX (0,0,0) (2,0,0) [12] with two Fourier pairs, TBATS and hybrid model was systematically evaluated through multiple criteria and diagnostic tests.

3.4.1. Information Criteria

Model parsimony and in-sample fit were compared using AIC and BIC:

$$AIC = -2\ln(L) + 2k, \quad (10)$$

$$BIC = -2ln(L) + kln(n), \quad (11)$$

where L denotes the model's likelihood, k is the number of estimated parameters, and n is the sample size. Lower values indicate superior model performance [43,52].

3.4.2. Forecast Error Metrics

Predictive accuracy was assessed using RMSE, MAE, and MAPE, which are widely employed in time-series forecasting studies.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}, \quad (12)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|, \quad (13)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| x100\%, \quad (14)$$

These metrics are standard in time-series forecasting and allow a comprehensive assessment of models' performance [53]. Prediction intervals at 80% and 95% confidence

levels were calculated to quantify uncertainty in out-of-sample forecasts, consistent with best practices in air quality forecasting.

3.5. Residual Diagnostic: Ljung–Box Test

To confirm the adequacy of the fitted models, the Ljung–Box test was applied to residuals:

$$Q^* = n(n+2)\sum_{i=1}^{k}(n-l)^{-1}r_1, \tag{15}$$

where r_1 is the residual autocorrelation at lag l. Under the null hypothesis of no residual autocorrelation, Q^* follows a chi-square distribution with $l - p - q$ degrees of freedom, where p and q are the orders in ARIMA(p,q). A non-significant test result ($p > 0.05$) suggests residuals' autocorrelations behave like white noise [43,54].

3.6. Hybrid Model

To improve forecasting accuracy, a hybrid framework was developed by combining three fitted time-series models—SARIMA, SARIMAX, and TBATS—into a single weighted ensemble. At each time point, the hybrid forecast was defined as a convex combination of the individual model outputs formulated as:

$$\hat{y}_{Hybrid} = \omega_1\hat{y}_{SARIMA} + \omega_2\hat{y}_{SARIMAX} + \omega_3\hat{y}_{TBATS} \tag{16}$$

where $\hat{y}_{SARIMA}$, $\hat{y}_{SARIMAX}$ and $\hat{y}_{TBATS}$ denote the out-of-sample predictions from the respective models and (ω_1, ω_2, ω_3) are nonnegative weights that sum to one. The weights were estimated by minimizing the RMSE function using the optim function in R with the L-BFGS-B algorithm, which allows box constraints to enforce ($0 \leq \omega_i \leq 1$) and the simplex constraint ($\omega_1 + \omega_2 + \omega_3 = 1$). This combination strategy follows general forecasting guidance [49] and established practice on forecast combinations [22], leveraging the complementary strengths of SARIMA (parsimonious seasonal dynamics), SARIMAX (seasonality with exogenous regressors), and TBATS (complex and possibly non-integer seasonalities). In our setting, the ensemble typically reduced forecast error relative to individual models by averaging model-specific biases and variances [22].

4. Results

4.1. Exploratory Analysis and Seasonal Decomposition

The time series plot (Figure 2) exhibits a clear seasonal pattern without any discernible long-term trend, indicating stationarity in both mean and variance. This observation was formally confirmed by the Augmented Dickey–Fuller (ADF) test, which yielded a Dickey–Fuller statistic of -12.101 with a p-value of 0.01 at lag order 6. Since the p-value is less than 0.05, the null hypothesis of a unit root was rejected, confirming that the time series is stationary and does not require differencing or further transformation before modeling.

Figure 3 displays the periodogram of monthly $PM_{2.5}$ concentrations at Zabol station. The results reveal a dominant peak at a frequency of 0.083, corresponding to 12 months, indicating a strong annual seasonal pattern in the data. This finding aligns with previous observations of seasonal air pollution variations in the region. The absence of other significant peaks in the periodogram suggests no secondary cycles (such as semi-annual or short-term seasonal patterns) are present in the data. These results confirm the necessity of incorporating seasonal components, such as Fourier terms or SARIMA models, in subsequent analyses.

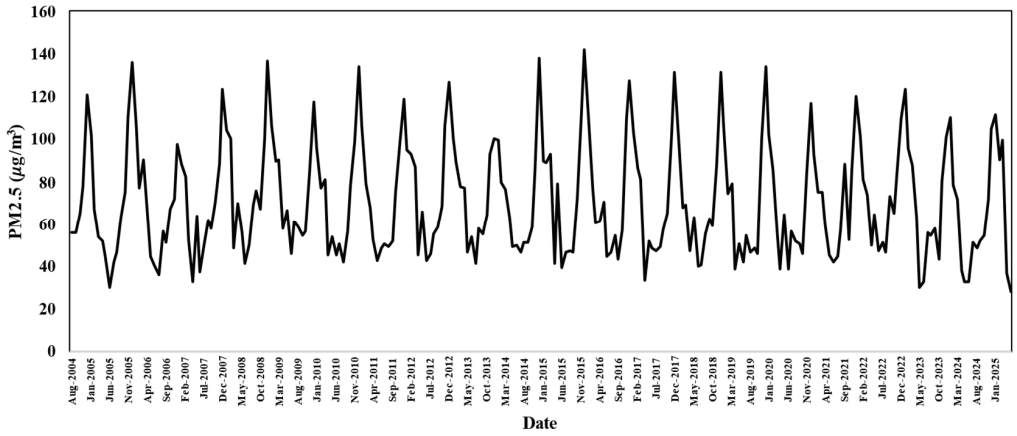

Figure 2. Monthly time series of $PM_{2.5}$ concentrations in Zabol, illustrating pronounced annual seasonality and absence of long-term trend.

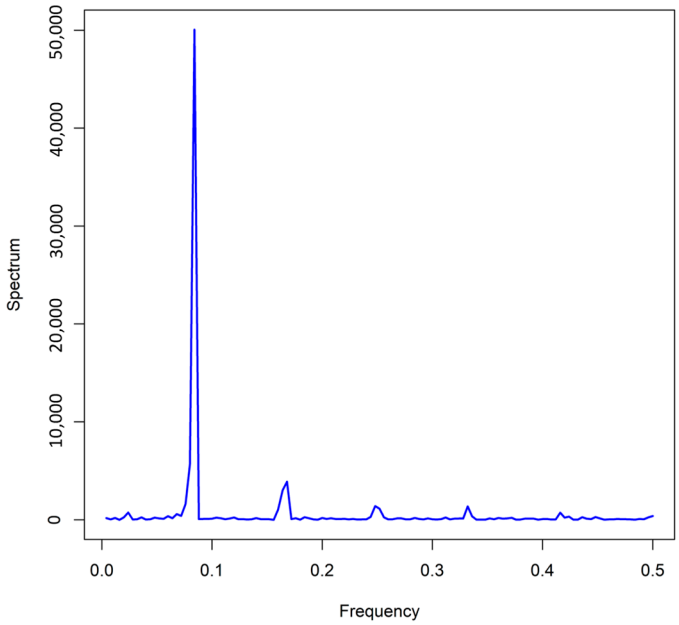

Figure 3. Periodogram of the monthly $PM_{2.5}$ concentrations, indicating a dominant spectral peak at a frequency of 0.083 (corresponding to 12 months).

Figure 4 displays the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the $PM_{2.5}$ time series. The ACF plot clearly shows significant spikes at lag 12 and its multiples (24, 36, . . .), indicating a strong seasonal pattern with a 12-month periodicity. This pattern confirms that $PM_{2.5}$ concentrations are influenced by seasonal factors (such as climatic variations or periodic human activities). In contrast, the ACF fluctuations at non-seasonal lags (1 to 11) decay rapidly, suggesting the absence of long-term dependencies or trends in the data. The PACF plot reveals a significant spike at lag 12, which shows a notable decrease by lag 24. This behavior suggests that the seasonal component likely has a low order. Additionally, the slight increase at lag 1 in the PACF may indicate a non-seasonal autoregressive (AR) component. Based on the ACF and PACF analysis, we propose a SARIMA model structure to capture both the seasonal and non-seasonal components of the $PM_{2.5}$ time series. For the non-seasonal component (p,d,q), the rapid decay in the ACF and the significant spike at lag 1 in the PACF suggest either an AR(1) or MA(1) process, leading to our recommendation of $p = 1$ (non-seasonal AR term) and q = 1 (non-seasonal MA term). Regarding the seasonal component (P,D,Q) [12], the significant spike at lag 12 in the ACF with gradual decay indicates a seasonal AR(1) (P = 1) process,

while the PACF's spike at lag 12 without subsequent spikes suggests a seasonal MA(1) (Q = 1) component. Consequently, our final model proposal is SARIMA (1,0,2) (2,0,2) [12], which effectively accounts for both the observed short-term dependencies through its non-seasonal terms and the strong annual seasonality through its seasonal components. This model structure is particularly suitable given the clear 12-month periodicity evident in both plots and the relatively simple non-seasonal pattern. The inclusion of first-order seasonal AR and MA terms (P = 1, Q = 1) appropriately captures the persistent seasonal effects while maintaining model parsimony.

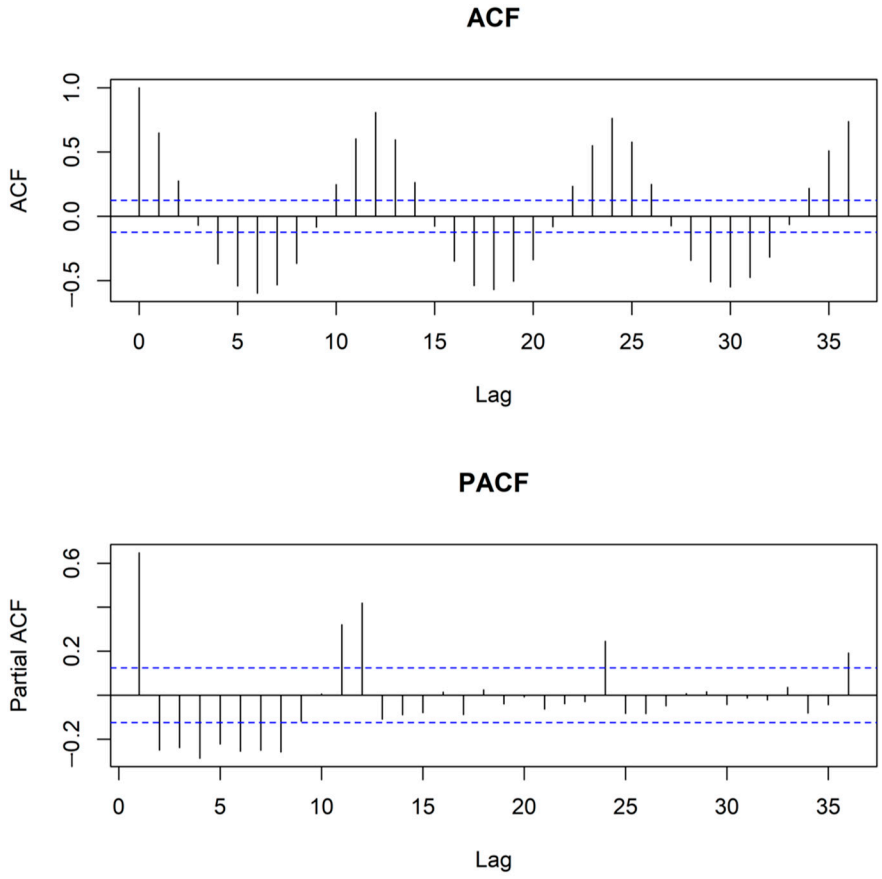

Figure 4. ACF and PACF plots for the $PM_{2.5}$ time series, illustrating prominent annual periodicity (spikes at lag 12).

4.2. Model Estimation and Parameter Results

The estimated coefficients and standard errors for each fitted model are reported in Tables 1–3. For the SARIMA (1,0,2) (2,0,2) [12] model (Table 1), the estimated autoregressive and moving average terms were statistically significant, indicating strong persistence and seasonal dependence in the $PM_{2.5}$ series. The model provided a parsimonious yet robust representation of both short-term and annual cycles.

Table 1. Estimated parameters for SARIMA (1,0,2) (2,0,2) [12].

Coefficient	AR1	MA1	MA2	SAR1	SAR2	SMA1	SMA2	Mean
Estimates	−0.861	1.045	0.127	1.175	−0.183	−0.978	0.220	68.21
S.E.	0.124	0.136	0.072	0.504	0.501	0.493	0.384	7.073

Table 2. Estimated parameters for SARIMAX (0,0,0) (2,0,0) [12] with two Fourier pairs.

Coefficient	SAR1	SAR2	Intercept	S1-12	C1-12	S2-12	C2-12	δ_1	δ_2
Estimates	0.144	0.154	69.57	6.44	−26.16	−4.51	6.37	26.69	−18.61
S.E.	0.065	0.066	0.881	1.23	1.25	1.24	1.22	3.58	3.11

Table 3. Estimated parameters for TBATS.

Coefficient	β	γ_1	γ_2
Estimates	0.131	0.325	0.347
S.E.	0.062	0.059	0.061

The estimated parameters for the SARIMA (1,0,2) (2,0,2) [12] model reveal a complex and persistent structure in the $PM_{2.5}$ time series. The significant negative AR(1) coefficient (−0.861) indicates a strong mean-reverting pattern, where an increase in concentration in one month is typically followed by a decrease in the next. The significant MA terms further capture these short-term dynamics; the strong positive MA(1) coefficient (1.045) suggests that random shocks, such as sudden dust storms, have an immediate and pronounced impact on $PM_{2.5}$ levels, while the weaker MA(2) term shows this influence weakly persists for a second month. The model's seasonal components are particularly dominant. The powerful and significant SAR(1) coefficient (1.175) confirms a robust annual cycle, demonstrating that concentrations in a given month are heavily dependent on those from the same month in the previous year. This is complemented by the significant negative SMA(1) coefficient (−0.978), which effectively models and dampens seasonal shocks. Finally, the estimated mean value of 68.21 $\mu g/m^3$ signifies a chronically high background level of air pollution, far exceeding WHO guidelines. Collectively, these parameters affirm the model's ability to parsimoniously yet robustly represent both the short-term autoregressive moving-average dynamics and the dominant seasonal forces governing dust concentrations in Zabol.

Table 2 shows the results of the SARIMAX (0,0,0) (2,0,0) [12] model with two Fourier pairs and confirms that the seasonality is overwhelmingly deterministic. This is evidenced by the highly significant and large-magnitude Fourier coefficients (e.g., −26.16 and 6.44 for the first harmonic), which dominate the model, in stark contrast to the weak and modest seasonal autoregressive parameters (0.144 and 0.154). The analysis confirms that employing Fourier terms is the primary driver for capturing the dominant 12-month periodicity and achieving predictive accuracy, while the included exogenous variables also demonstrate a strong and significant influence with their substantial coefficients (26.69 and −18.61).

Finally, the TBATS model (Table 3) captured complex seasonal components through its trigonometric and trend-related terms, all of which were significant. This highlights TBATS as a flexible alternative for representing overlapping or evolving seasonalities.

Collectively, the results indicate that SARIMA offers a simple and statistically adequate structure, SARIMAX leverages deterministic cycles via Fourier components, and TBATS provides enhanced flexibility for more complex seasonal dynamics.

Table 4 presents the weight estimates for the hybrid ensemble model, which combines forecasts from SARIMA, SARIMAX, and TBATS models. The weights were optimized by minimizing the Root Mean Square Error (RMSE), resulting in a configuration where SARIMAX receives the highest weight (ω_2 = 0.628), followed by TBATS (ω_3 = 0.263), and SARIMA (ω_1 = 0.109). The significant magnitude of the SARIMAX weight underscores its dominant role, consistent with earlier findings that highlighted its effectiveness in modeling deterministic seasonality and incorporating exogenous variables. Meanwhile, the substantial weight assigned to TBATS suggests it captures complex seasonal patterns or nonlinear-

ities not fully addressed by the other models. All weights are statistically significant, as each estimate is multiple times larger than its standard error, confirming their robustness.

Table 4. Estimated parameters for the Hybrid model.

Coefficient	ω_1	ω_2	ω_3
Estimates	0.109	0.628	0.263
S.E.	0.031	0.044	0.041

The hybrid approach successfully leverages the distinct strengths of each model: SARIMAX provides a strong baseline with its ability to handle external regressors and deterministic seasonality, TBATS adds flexibility for intricate seasonal behaviors, and SARIMA contributes a parsimonious seasonal ARIMA structure. By integrating these components, the ensemble mitigates individual model biases and variances, resulting in improved forecast accuracy and robustness. This strategy demonstrates the value of model combination in enhancing predictive performance beyond what any single model can achieve.

4.3. Model Validation and Diagnostic Checks

Table 5 and Figure 5 present a comprehensive performance evaluation of four forecasting models: SARIMA, SARIMAX, TBATS, and a Hybrid ensemble. The results clearly demonstrate that the Hybrid model achieves the best overall performance, with the lowest AIC (1835.04) and BIC (1842.08) values, indicating superior goodness-of-fit while maintaining optimal parsimony. Furthermore, it delivers the smallest RMSE (9.42) and MAE (7.43), highlighting its exceptional predictive accuracy and ability to minimize absolute errors. While its MAPE (11.67%) is slightly higher than that of the SARIMAX model, this marginal difference is outweighed by its advantages in other critical metrics.

Table 5. Model comparison using statistical criteria (AIC, BIC, RMSE, MAE, MAPE).

Criteria	AIC	BIC	RMSE (μg/m^3)	MAE (μg/m^3)	MAPE
SARIMA	1946.27	1977.96	10.852	8.590	13.860
SARIMAX	1866.45	1901.67	9.69	7.52	11.64
TBATS	2574.202	2623.502	10.295	8.123	13.601
Hybrid	1835.04	1842.08	9.42	7.43	11.67

The SARIMAX model remains the strongest individual model, outperforming both the standalone SARIMA and TBATS models across all criteria, including AIC (1866.45), BIC (1901.67), RMSE (9.69), MAE (7.52), and MAPE (11.64%). This confirms its robustness in leveraging exogenous variables and capturing key data patterns. In contrast, the TBATS model shows the weakest statistical fit with the highest AIC and BIC values, though it performs reasonably in point forecasting accuracy. The SARIMA model serves as a decent baseline but is consistently outperformed by its more sophisticated counterparts. Ultimately, the analysis confirms that the Hybrid model's strategy of combining individual forecasts yields the most robust and accurate predictions, making it the recommended choice for this forecasting task.

Table 6 presents the Ljung–Box test results for residual autocorrelation across the four models, providing crucial insights into their statistical adequacy. The Hybrid model demonstrates the best performance with a test statistic of Q* = 15.311 and a high p-value of 0.882, indicating no significant residual autocorrelation and confirming its ability to effectively capture temporal dependencies through its ensemble approach. Similarly, the SARIMAX model shows excellent results (Q* = 13.071, p = 0.93), validating its strong capacity to model patterns using exogenous variables. The SARIMA model also performs acceptably (Q* = 18.556, p = 0.354), showing no concerning autocorrelation in its residuals.

In stark contrast, the TBATS model exhibits significant residual autocorrelation (Q* = 37.397, $p = 0.039$), revealing substantial limitations in accounting for all temporal patterns. These diagnostic results complement the earlier performance metrics, confirming that both the Hybrid and SARIMAX models not only excel in predictive accuracy but also provide the most statistically adequate residual structures, with the Hybrid model emerging as the most comprehensive solution.

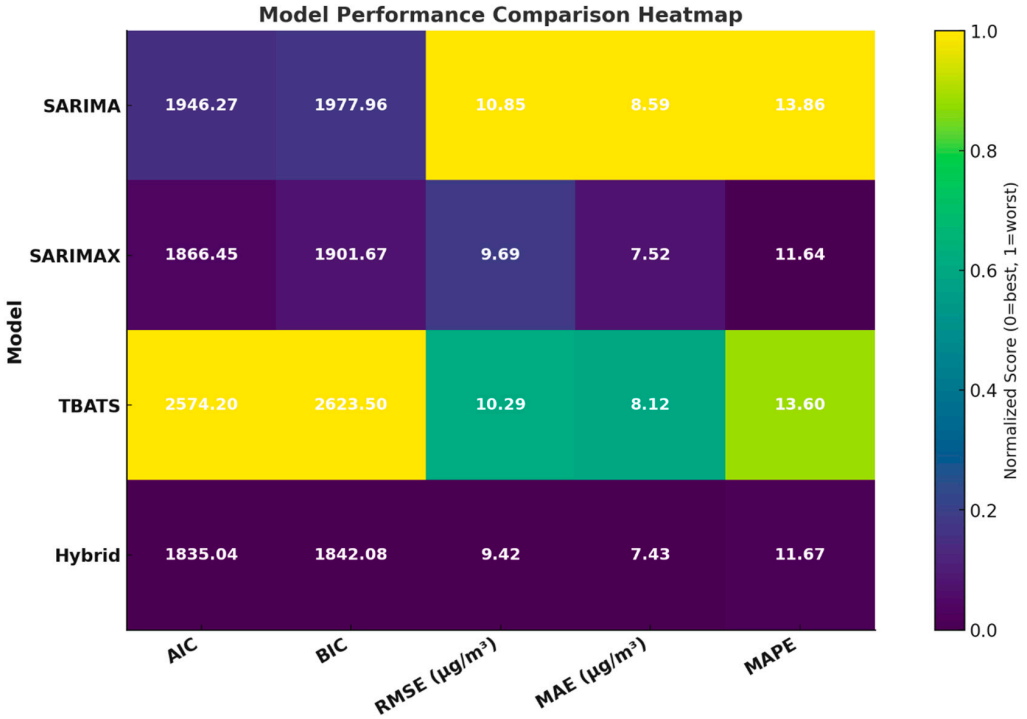

Figure 5. Comparative Performance Evaluation of SARIMA, SARIMAX, TBATS, and Hybrid Models for $PM_{2.5}$ Forecasting.

Table 6. Ljung–Box test results for model residuals.

Model/Statistics	Q*	df	*p*-Value
SARIMA	18.556	17	0.354
SARIMAX	13.071	22	0.93
TBATS	37.397	24	0.039
Hybrid	15.311	21	0.882

Figure 6 displays the residual diagnostics for the SARIMAX (0,0,0) (2,0,0) [12] model, which includes three key components: the residual time series plot (top), the autocorrelation function (ACF) of residuals (bottom left), and a histogram of residuals with a density overlay (bottom right). The residual time series plot shows that the residuals fluctuate randomly around zero without any discernible systematic patterns or trends, indicating that the model has effectively captured the underlying structure of the data. The ACF plot reveals no significant autocorrelations at any lag, as all bars remain within the confidence bounds, suggesting that the model has adequately accounted for temporal dependencies, including seasonal effects. Additionally, the histogram of residuals exhibits a roughly symmetric distribution that aligns well with the normal density curve, supporting the assumption of normality. Collectively, these diagnostics confirm that the SARIMAX (0,0,0) (2,0,0) [12] model is well-specified, with residuals behaving as white noise. For further validation, statistical tests such as the Ljung–Box test (for autocorrelation) or the Shapiro–Wilk test

(for normality) could be conducted, but the visual evidence strongly suggests the model's adequacy for forecasting purposes.

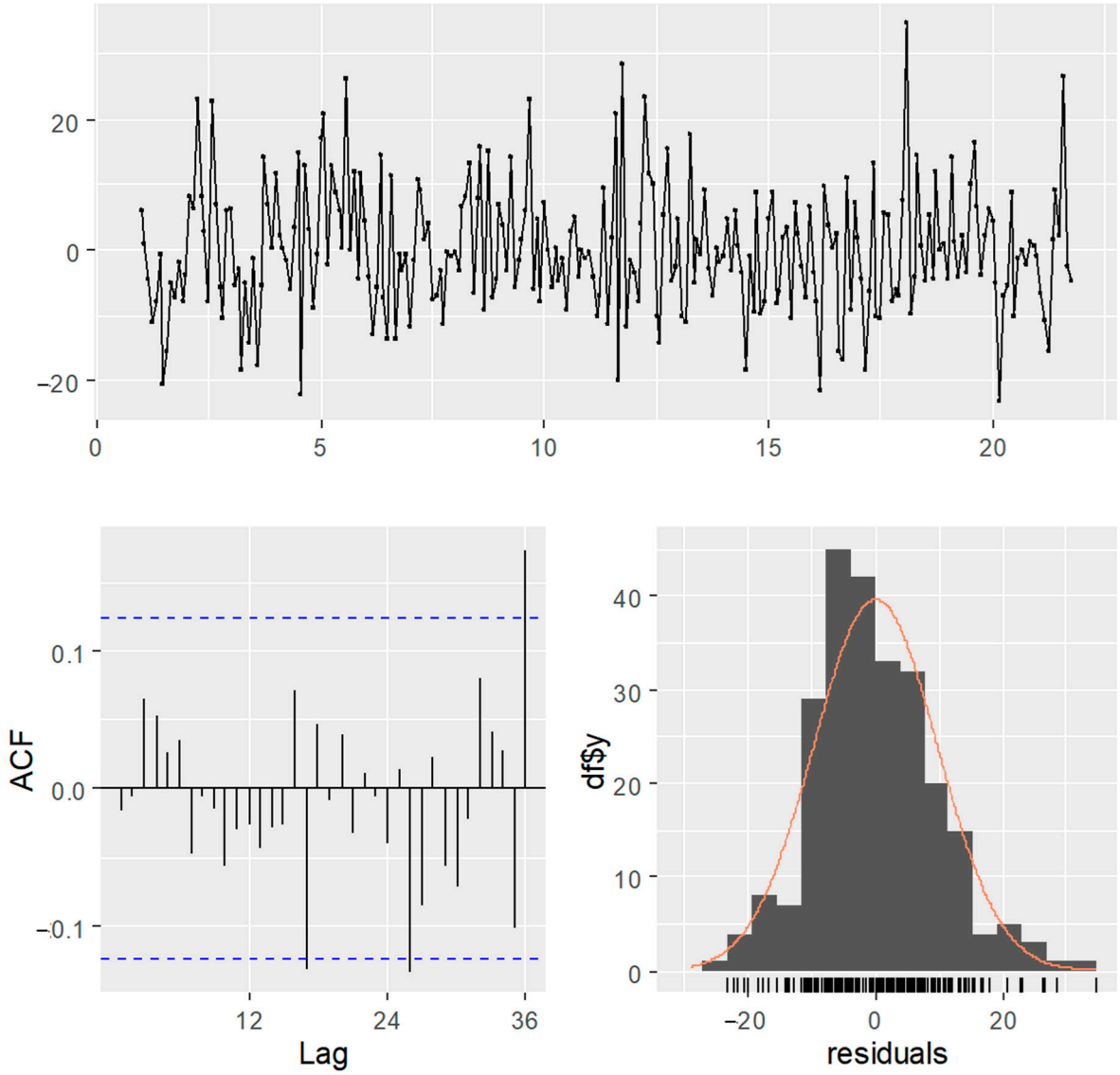

Figure 6. Residual diagnostics for the SARIMAX (0,0,0) (2,0,0) [12] model: (**top**) residual time series; (**bottom left**) ACF of residuals; (**bottom right**) histogram of residuals with density overlay.

Collectively, these results indicate that the SARIMAX model with Fourier terms provides the most robust fit for the monthly $PM_{2.5}$ series in Zabol, outperforming both the classical SARIMA and TBATS models in terms of statistical fit and residual diagnostics.

As shown in Figure 7a–d, the comparison between observed and forecasted $PM_{2.5}$ concentrations at Zabol station from 2004 to 2025 is presented using different models.

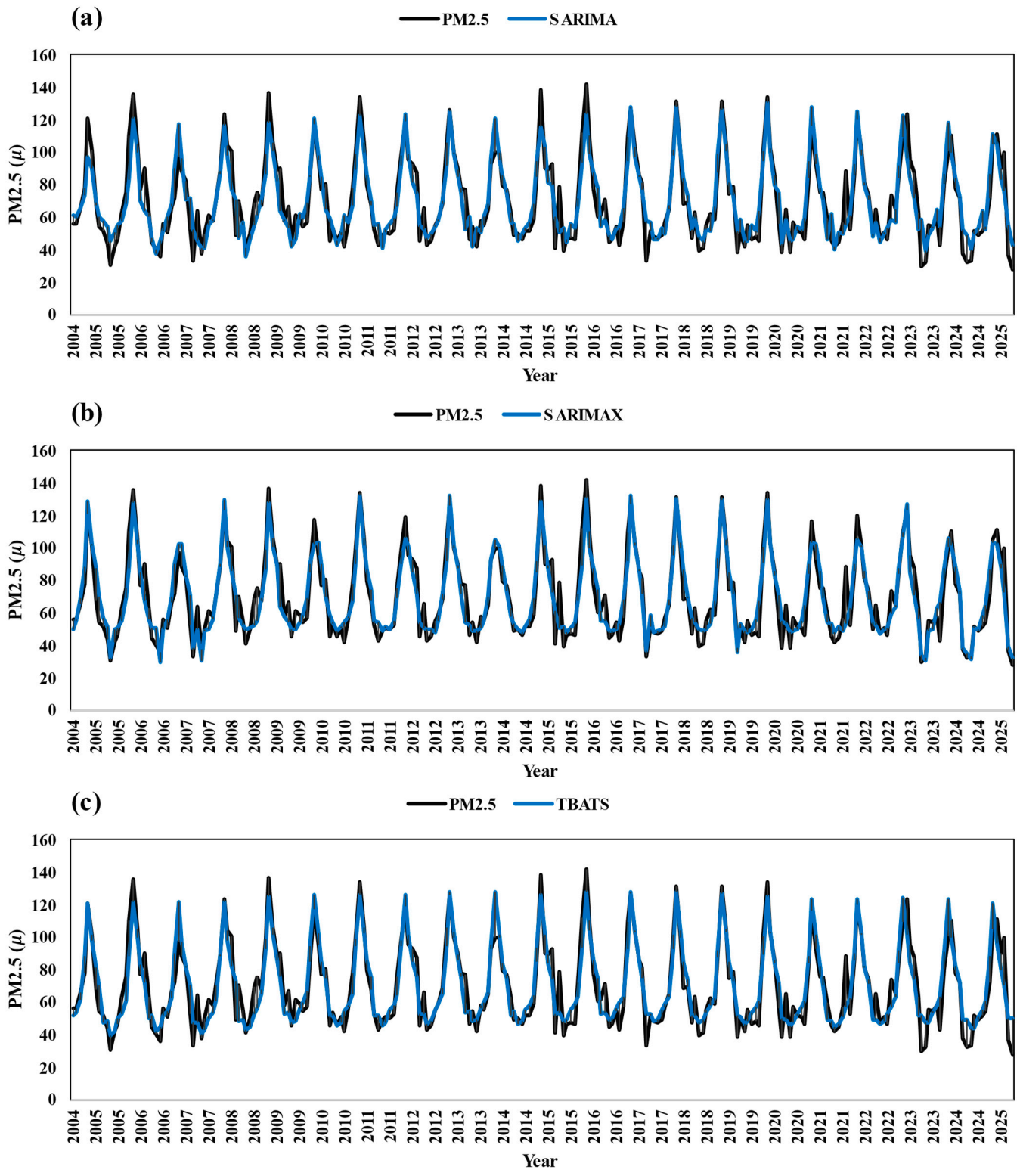

Figure 7. *Cont.*

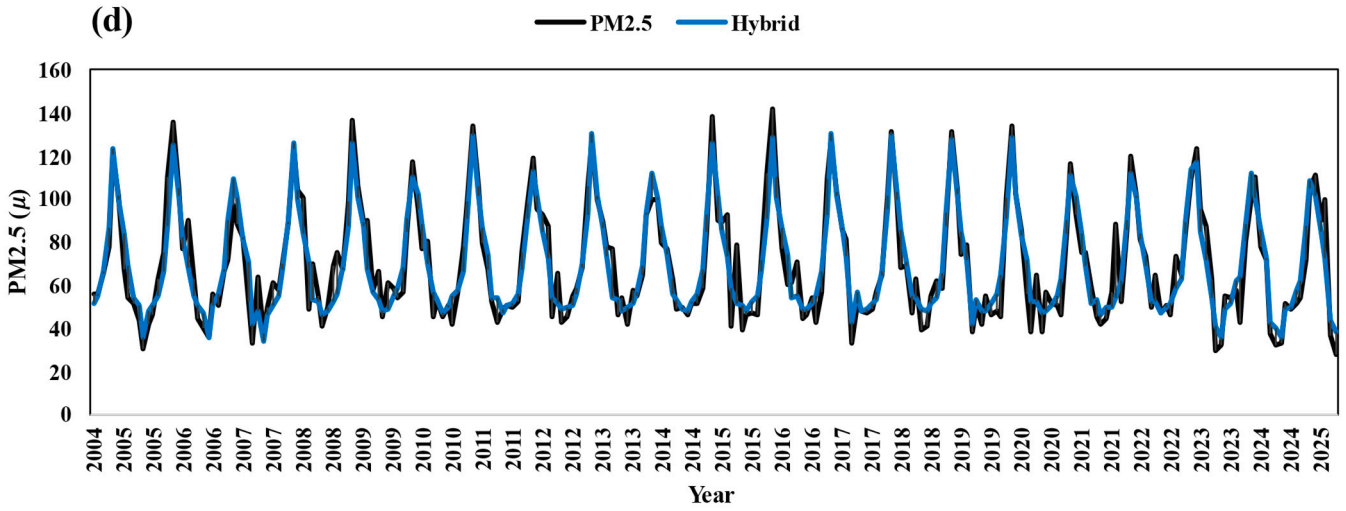

Figure 7. Comparison of Observed and Forecasted $PM_{2.5}$ Concentrations (2004–2025) Using SARIMA (**a**), SARIMAX (**b**), and TBATS (**c**) and Hybrid (**d**) Models.

4.4. Out-of-Sample Forecasting with the Hybrid Model

Figure 8 displays the observed and forecasted monthly $PM_{2.5}$ concentrations in Zabol from 2004 to 2027, generated by the Hybrid model. As clearly shown, the model has successfully captured the seasonal pattern of the historical data. Both the short-term and long-term forecasts (up to 2027) follow the anticipated trends with reasonable accuracy, and the forecasted peak values for the upcoming periods (such as 2026–2027) appear to be more conservative and lower compared to the past observed peaks (like in 2014–2015). Overall, the model demonstrates stable and reliable performance not only for short-term forecasting and seasonal planning but also for the long-term outlook up to 2027. The figure effectively demonstrates the model's strength in capturing and continuing the seasonal variation pattern.

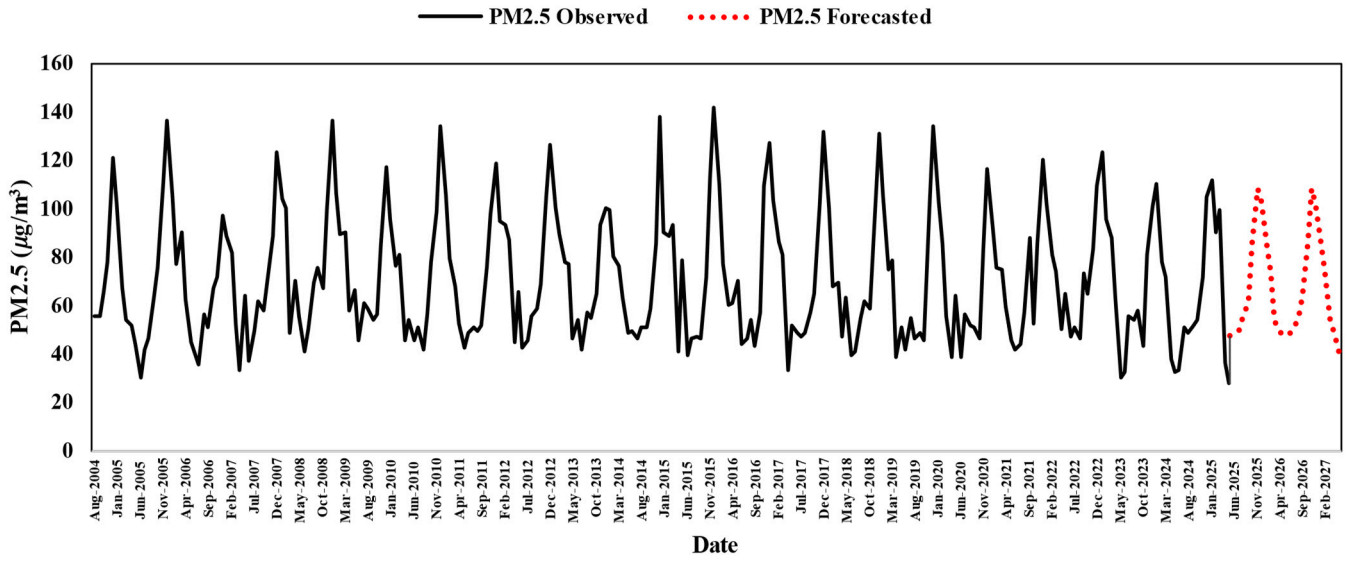

Figure 8. Observed and forecasted for monthly $PM_{2.5}$ concentrations in Zabol, generated by the Hybrid model.

Figure 9a displays the $PM_{2.5}$ concentration forecasts generated by the hybrid model with an 80% confidence interval. The red line represents the mean predicted values for each month, while the blue shaded band around this line indicates the 80% confidence interval range defined by the lower bound (LB_80) and upper bound (UB_80).

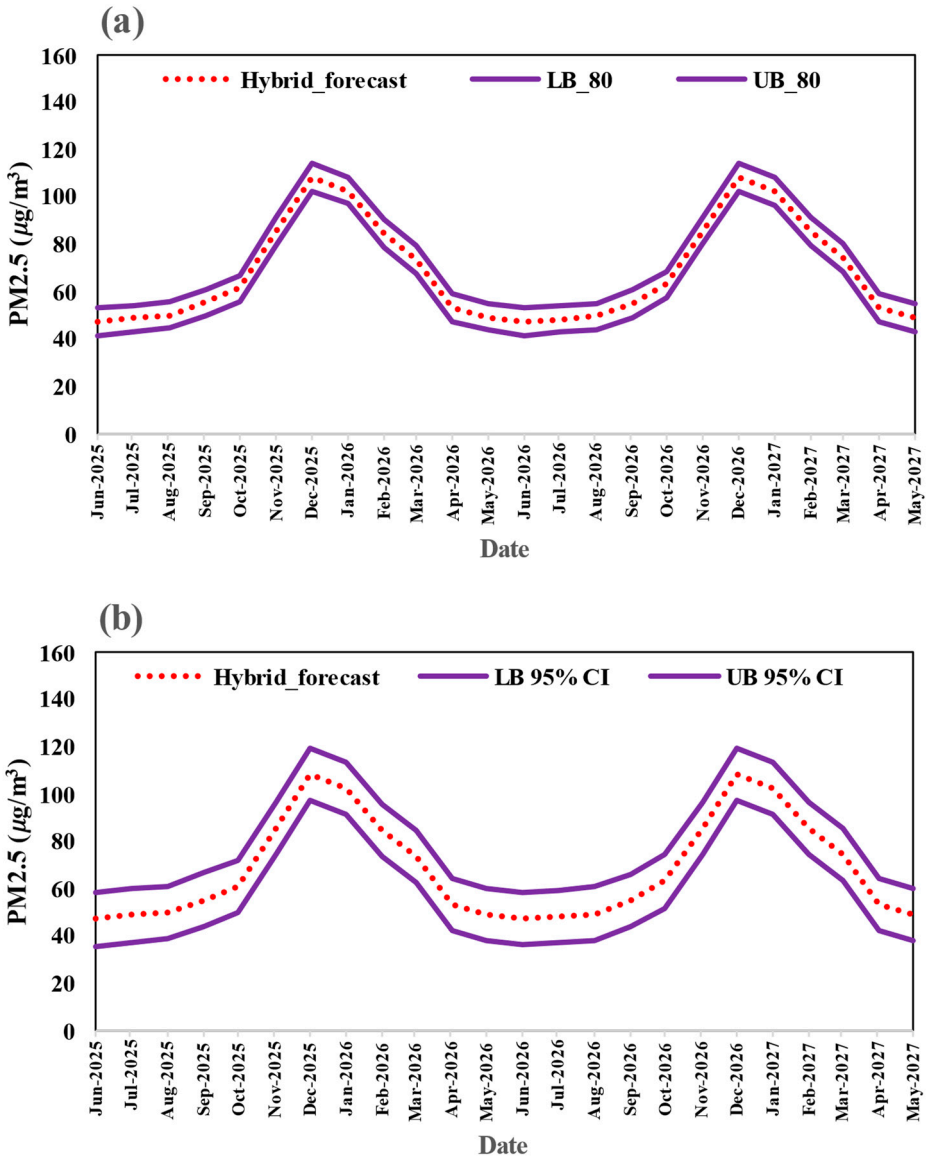

Figure 9. $PM_{2.5}$ Concentration Forecasting Using Hybrid Model with (**a**) 80% and (**b**) 95% Confidence Intervals.

Contrary to typical forecasting models, the width or difference between these bounds remains constant throughout the entire forecast period from June 2025 to May 2027. Calculations based on the data show that the difference between the upper and lower bounds is precisely 11.386 μg/m^3 for all months. For example, in June 2025, the forecast range is from 41.55 to 52.93 μg/m^3, in December 2026 from 102.62 to 114.01 μg/m^3, and in May 2027 from 43.49 to 54.88 μg/m^3. This complete stability in the confidence band width, despite significant variations in the central predicted values ranging from approximately 47 to 108 μg/m^3, indicates that the hybrid model has considered a constant absolute uncertainty for the entire two-year forecast horizon. Therefore, unlike many forecasting models where uncertainty increases with longer time horizons, this model demonstrates the same level of accuracy and uncertainty throughout the entire forecast period.

Figure 9b displays the 95% confidence interval forecasts generated by the hybrid model.

Contrary to initial expectations, the width of this confidence interval remains constant throughout the entire forecast period from June 2025 to May 2027. The difference between the upper and lower bounds is consistently 22.252 μg/m^3 across all months, with no observable expansion in the interval width over time. For example, in June 2025, the forecast

range spans from 36.11 to 58.37 μg/m^3, in December 2026 from 97.18 to 119.43 μg/m^3, and in May 2027 from 38.06 to 60.31 μg/m^3. This consistent interval width, despite significant fluctuations in the central predicted values, indicates that the model maintains a stable level of uncertainty throughout the entire two-year forecast horizon. The 95% confidence interval provides a wider range of potential outcomes compared to the 80% interval, but unlike typical forecasting models, this range does not expand with longer prediction horizons. Therefore, both confidence levels demonstrate consistent performance across short-term and long-term forecasts, suggesting the model's reliability remains constant throughout the prediction period.

Overall, the Hybrid model demonstrates reliable predictive performance for seasonal dust episodes in Zabol, as evidenced by the close alignment between forecasted values and historical observations, along with reasonably tight confidence bounds. This underscores the suitability of the Hybrid modeling approach for operational air quality management and public health planning in regions characterized by pronounced seasonal variability in $PM_{2.5}$ levels.

5. Discussion

The present study aims to predict $PM_{2.5}$ concentrations at the Zabol station using SARIMA, SARIMAX, TBATS, and a Hybrid model on a monthly scale from 2004 to 2025. The selection of these classical models, along with the Hybrid ensemble approach, over more complex machine learning-based methods such as LSTM or XGBoost, is primarily due to the limitations of observational data in this region and the need for models with higher interpretability and manageable parameter complexity [55]. Classical time series models, particularly in cases where data exhibit clear seasonal patterns—such as $PM_{2.5}$ in Zabol—demonstrate acceptable performance even with limited data volume. The Hybrid model further enhances predictive robustness by combining the strengths of individual models.

Preliminary data analysis revealed a strong seasonal pattern without a long-term trend, which was confirmed by the ADF test. The periodogram also identified an annual cycle with a frequency of 0.083 (a 12-month period), reinforcing the necessity of incorporating seasonal components in the modeling. ACF/PACF analysis further emphasized that seasonal dependencies at lags 12, 24, etc., are dominant.

According to the findings of Rashki et al. [56,57] and Behrooz et al. [15], the seasonal pattern of $PM_{2.5}$ concentration in Zabol is primarily influenced by the 120-day winds of Sistan, with the highest levels of these particles observed during summer (June to September). This significant increase in particle concentration stems from the key role of these strong winds in transporting dust from adjacent arid regions, particularly the deserts of Afghanistan and the Sistan Plain, to the Zabol area. This process not only leads to worsening air pollution but also significantly impacts air quality degradation and increases health risks for residents during this period. Such findings highlight the necessity of considering regional climatic factors and wind patterns in air pollution management and environmental planning. Furthermore, recurrent droughts and the desiccation of Hamoun Wetland (resulting from reduced water inflow from the Hirmand River) during spring and summer exacerbate local dust emissions by decreasing soil moisture [58].

Although statistical time-series models do not explicitly account for causal drivers such as emission sources or meteorological dynamics, they remain widely adopted and validated for short- to medium-term air quality forecasting [8,19,59]. In highly seasonal environments like Zabol, where the "120-day wind" imposes a strong deterministic cycle, the Hybrid model—integrating SARIMA, SARIMAX, and TBATS—proves capable of reproducing recurrent dust episodes with improved accuracy. As such, the forecasts generated here

should not be interpreted as substitutes for process-based or emission-driven models, but rather as complementary tools that provide operational value for early-warning and public health protection in data-scarce regions.

The evaluation of model accuracy and goodness-of-fit indicated that the Hybrid model, with its weighted combination of SARIMA, SARIMAX, and TBATS, outperformed all individual models. It achieved the lowest AIC (1835.04) and BIC (1842.08), along with the smallest RMSE (9.42) and MAE (7.43), demonstrating its superior predictive accuracy and robustness. The Ljung–Box test further confirmed the Hybrid model's dominance in residual diagnostics (Q = 15.311, $p = 0.882$), unequivocally verifying its capacity to eliminate temporal dependencies. In contrast, SARIMAX showed strong performance but with minor residual autocorrelation ($p = 0.93$), while TBATS exhibited significant limitations ($p = 0.039$). This empirical evidence corroborates established theoretical frameworks [14,22] that advocate for model parsimony and ensemble techniques when confronting linearly structured environmental time series.

This superiority aligns with broader findings in the field. In this context, Chai et al. [60] demonstrated that integrating satellite-based predictors with machine learning models substantially improves $PM_{2.5}$ estimation accuracy in China, further supporting the broader potential of hybrid frameworks for air quality applications. Our results reinforce this paradigm, showing that even without satellite data, a hybrid approach combining classical time-series models can significantly enhance prediction robustness in data-scarce regions.

The TBATS model exhibited significant residual autocorrelation (Ljung–Box $p = 0.039$), indicating its inability to fully capture the data's temporal dependencies. This limitation may arise from the model's reliance on trigonometric seasonality, which struggles to adapt to the rigid, meteorologically driven annual cycle dominated by the 120-day Sistan winds. Additionally, TBATS may not adequately account for abrupt dust events causing short-term shocks. Alternative configurations (e.g., increasing harmonics or enabling ARMA errors) were tested but did not yield improved results, suggesting inherent incompatibility with this dataset's structure [51].

Out-of-sample forecast analysis using the Hybrid model demonstrates its effectiveness in capturing the seasonal patterns of $PM_{2.5}$ concentrations in Zabol [58]. Contrary to conventional expectations, as the forecast horizon extends to 2026–2027, the model's uncertainty does not progressively increase but rather maintains constant levels of uncertainty throughout the entire forecast period. The Hybrid model demonstrates constant absolute uncertainty over the two-year forecast horizon, with the 80% confidence interval consistently maintaining a fixed width of 11.386 $\mu g/m^3$ and the 95% confidence interval maintaining a constant width of 22.252 $\mu g/m^3$. This consistent pattern persists even when $PM_{2.5}$ concentrations exceed the 100 $\mu g/m^3$ threshold, indicating the model's ability to maintain the same level of accuracy for extreme values. Although the Hybrid model has successfully maintained consistent uncertainty levels, it is worth noting that this characteristic may stem from inherent limitations in accounting for cumulative error propagation in multi-step predictions and structural deviations caused by long-term climatic variations [59,61,62]. This finding suggests that while the model demonstrates consistent performance for both short-term and long-term predictions, it may face challenges in fully capturing the inherent uncertainty in long-term forecasts.

Unlike many conventional time series models, the hybrid model developed in this study does not exhibit cumulative error dependency in multi-step-ahead predictions. As observed in the results, the confidence interval bandwidth remains completely stable throughout the entire forecast period from June 2025 to May 2027. This stability indicates that the model possesses high resistance to error propagation and accumulation, maintaining the same level of accuracy even over long-term forecast horizons [63,64].

Beyond the specific case of Zabol, the proposed Hybrid framework has potential for broader applications. Similar arid and dust-prone regions such as the Middle East, Central Asia, and North Africa face persistent dust episodes driven by deterministic seasonal cycles but often lack dense monitoring networks. In such settings, integrating spectral diagnostics with classical time-series forecasting and ensemble techniques can provide a cost-effective and interpretable early-warning tool. Moreover, this methodological structure can be extended to other environmental variables with strong cyclic behavior, such as ozone concentrations, temperature anomalies, and rainfall extremes, thereby enhancing its transferability and practical relevance [65,66].

6. Conclusions

This study developed and evaluated a suite of time series models, including SARIMA, SARIMAX, TBATS, and a novel Hybrid ensemble, for forecasting monthly $PM_{2.5}$ concentrations in Zabol, Iran, using data from 2004 to 2025. The key finding was that the Hybrid model, which intelligently combines the forecasts of the three individual models through an optimized weighting scheme, demonstrated superior performance across all evaluation metrics, achieving the lowest AIC, BIC, RMSE, and MAE. This result underscores the significant advantage of leveraging ensemble techniques to integrate the strengths of diverse modeling approaches.

The research revealed several critical insights about dust forecasting in arid regions. First, it demonstrates that a hybrid approach, rather than relying on a single complex model, can most effectively capture the physical processes governing dust transport, as evidenced by the Hybrid model's optimal weights. Second, the study highlights how regional climatic factors—particularly the Sistan wind regime and Hamoun wetland desiccation—create a strong, deterministic seasonal pattern that was best captured by the Fourier terms in the SARIMAX model.

Several limitations warrant consideration. Dependence on limited observational data presents additional challenges. These limitations suggest that while the Hybrid model provides a solid foundation, it could be further enhanced by complementing it with other approaches for comprehensive dust forecasting. Additionally, the absence of meteorological variables such as relative humidity, temperature anomalies, and wind speed fluctuations represents a further limitation. The inclusion of these factors through reanalysis of datasets or remote sensing could enhance predictive performance.

Future research directions should focus on advancing hybrid modeling frameworks by integrating the interpretability of classical models with the pattern recognition capabilities of machine learning. Incorporating additional data sources like satellite observations and meteorological reanalysis could enhance prediction accuracy, particularly for extreme events. The development of more sophisticated ensemble forecasting systems would help quantify and communicate prediction uncertainty more effectively.

From a practical perspective, these findings support several policy recommendations. The demonstrated Hybrid forecasting approach could form the basis for early warning systems to protect public health during dust events. In particular, forecasts can be linked to health-relevant thresholds, where $PM_{2.5}$ levels above 35 $\mu g/m^3$ correspond to the WHO 24 h guideline, and values exceeding 100 $\mu g/m^3$—frequently observed during the 120-day wind season—represent a critical local risk threshold. Incorporating these benchmarks into operational use would enable timely alerts, targeted protective measures, and prioritization of vulnerable populations. The methodology therefore provides environmental managers with a powerful tool for anticipating air quality degradation and implementing mitigation strategies.

This research advances dust forecasting capabilities in several important ways. It provides a validated hybrid modeling framework adaptable to similar arid environments, offers new insights into the synergistic value of combining model types for seasonal forecasting, and establishes clear pathways for future improvements through integrated modeling approaches. The findings contribute to both the methodological development of environmental forecasting and the practical management of air quality in dust-affected regions.

Author Contributions: Conceptualization, H.Z.; methodology H.Z., O.B. and T.C.; software, H.Z. and O.B.; validation, H.Z., O.B. and T.C.; formal analysis, H.Z., O.B., B.F. and T.C.; investigation, H.Z., O.B., B.F. and T.C.; resources, H.Z.; data curation, O.B. and T.C.; writing—original draft preparation, B.F.; writing—review and editing, O.B. and T.C.; visualization, O.B. and B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets that support the findings of this study are subject to local data-sharing restrictions and therefore cannot be made publicly available. However, they are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pope, C.A.; Ezzati, M.; Cannon, J.B.; Allen, R.T.; Jerrett, M.; Burnett, R.T. Mortality risk and $PM_{2.5}$ air pollution in the USA: An analysis of a national prospective cohort. *Air Qual. Atmos. Health* **2018**, *11*, 245–252. [CrossRef]
2. Al-Hemoud, A.; Gasana, J.; Al-Dabbous, A.; Alajeel, A.; Al-Shatti, A.; Behbehani, W.; Malak, M. Exposure levels of air pollution ($PM_{2.5}$) and associated health risk in Kuwait. *Environ. Res.* **2019**, *179*, 108730. [CrossRef] [PubMed]
3. World Bank Group. *Sand and Dust Storms in the Middle East and North Africa Region: Sources, Costs, and Solutions*; World Bank: Washington, DC, USA, 2019.
4. Karami, S.; Kaskaoutis, D.G.; Kashani, S.S.; Rahnama, M.; Rashki, A. Evaluation of nine operational models in forecasting different types of synoptic dust events in the Middle East. *Geosciences* **2021**, *11*, 458. [CrossRef]
5. Yi, Z.; Zeng, Z.; Wang, Y.; Li, W.; Zhang, B.; Gui, H.; Zhang, X. Improving short-term forecasting of surface dust concentration in Northern China: Integrating machine learning with multi-numerical models. *Atmos. Pollut. Res.* **2025**, *16*, 102480. [CrossRef]
6. Alsharif, M.H.; Younes, M.K.; Kim, J. Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. *Symmetry* **2019**, *11*, 240. [CrossRef]
7. Elshewey, A.M.; Shams, M.Y.; Elhady, A.M.; Shohieb, S.M.; Abdelhamid, A.A.; Ibrahim, A.; Tarek, Z. A novel WD-SARIMAX model for temperature forecasting using daily Delhi climate dataset. *Sustainability* **2023**, *15*, 757. [CrossRef]
8. Abhilash, M.S.K.; Thakur, A.; Gupta, D.; Sreevidya, B. Time series analysis of air pollution in Bengaluru using ARIMA model. In *Ambient Communications and Computer Systems*; *Advances in Intelligent Systems and Computing*; Springer: Singapore, 2018; Volume 696, pp. 413–426.
9. Barthwal, A.; Acharya, D.; Lohani, D. Prediction and analysis of particulate matter ($PM_{2.5}$ and PM_{10}) concentrations using machine learning techniques. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 1323–1338. [CrossRef]
10. Sharma, D.; Thapar, S.; Sachdeva, K. Enhancing particulate matter prediction in Delhi: Insights from statistical and machine learning models. *Environ. Monit. Assess.* **2025**, *197*, 723. [CrossRef]
11. Alharbi, F.R.; Csala, D. A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach. *Inventions* **2022**, *7*, 94. [CrossRef]
12. Doreswamy; Harishkumar, K.S.; Km, Y.; Gad, I. Forecasting Air Pollution Particulate Matter ($PM_{2.5}$) Using Machine Learning Regression Models. *Procedia Comput. Sci.* **2020**, *171*, 2057–2066. [CrossRef]
13. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2015.
14. De Livera, A.M.; Hyndman, R.J.; Snyder, R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **2011**, *106*, 1513–1527. [CrossRef]
15. Behrooz, R.D.; Mohammadpour, K.; Broomandi, P.; Kosmopoulos, P.G.; Gholami, H.; Kaskaoutis, D.G. Long-term (2012–2020) PM_{10} concentrations and increasing trends in the Sistan Basin: The role of Levar wind and synoptic meteorology. *Atmos. Pollut. Res.* **2022**, *13*, 101460. [CrossRef]

16. Chen, G.; Li, S.; Knibbs, L.K.; Hamm, N.A.S.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M.J.; Guo, Y. A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52–60. [CrossRef]
17. Al-qaness, M.A.A.; Dahou, A.; Ewees, A.A.; Abualigah, L.; Huai, J.; Abd Elaziz, M.; Helmi, A.M. ResInformer: Residual Transformer-Based Artificial Time-Series Forecasting Model for PM2.5 Concentration in Three Major Chinese Cities. *Mathematics* **2023**, *11*, 476. [CrossRef]
18. Gjika, E.; Ferrja, A.; Kamberi, A. A study on the efficiency of hybrid models in forecasting precipitations and water inflow: Albania case study. *Adv. Sci. Technol. Eng. Syst. J.* **2019**, *4*, 302–310. [CrossRef]
19. Díaz-Robles, L.; Ortega, J.; Fu, J.; Reed, G.; Chow, J.; Watson, J.; Moncada-Herrera, J. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos. Environ.* **2008**, *42*, 8331–8340. [CrossRef]
20. Gopu, P.; Panda, R.R.; Nagwani, N.K. Time series analysis using ARIMA model for air pollution prediction in Hyderabad City of India. In *Soft Computing and Signal Processing*; *Advances in Intelligent Systems and Computing*; Springer: Singapore, 2021; pp. 47–56.
21. Gupta, V.; Gharekhan, D.; Samal, D.R. Machine Learning Based $PM_{2.5}$ and PM_{10} Concentration Modeling for Delhi City. *J. Indian Soc. Remote Sens.* **2025**, *53*, 81–99. [CrossRef]
22. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018.
23. Necula, S.C.; Hauer, I.; Fotache, D.; Hurbean, L. Advanced Hybrid Models for Air Pollution Forecasting: Combining SARIMA and BiLSTM Architectures. *Electronics* **2025**, *14*, 549. [CrossRef]
24. Avila, M.L.; Alonso, A.M.; Peña, D. Modelling time series with multiple seasonalities: An application to hourly NO_2 pollution levels. *Stoch. Environ. Res. Risk Assess.* **2025**, *39*, 2063–2093. [CrossRef]
25. Du, M.; Zhang, Y.; Sun, L.; Wang, Y.; Zhou, X. A New Hybrid Model and Its Application for Forecasting of Daily $PM_{2.5}$ Concentrations. *SSRN* **2023**. [CrossRef]
26. Albeladi, K.; Zafar, B.; Mueen, A. Time Series Forecasting Using LSTM and ARIMA. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 313–320. [CrossRef]
27. Liu, Z.; Ge, C.; Zheng, K.; Bao, S.; Cui, Y.; Yuan, Y.; Zhang, Y. Forecasting Daily $PM_{2.5}$ Concentrations in Wuhan with a Spatial-Autocorrelation-Based Long Short-Term Memory Model. *Atmos. Environ.* **2024**, *331*, 120605. [CrossRef]
28. Solis Teran, M.A.; Leite Coelho da Silva, F.; Torres Armas, E.A.; Carbo-Bustinza, N.; López-Gonzales, J.L. Modeling Air Pollution in Metropolitan Lima: A Statistical and Artificial Neural Network Approach. *Environments* **2025**, *12*, 196. [CrossRef]
29. Naim, I.; Mahara, T.; Idrisi, A.R. Effective Short-Term Forecasting for Daily Time Series with Complex Seasonal Patterns. *Procedia Comput. Sci.* **2018**, *132*, 1832–1841. [CrossRef]
30. Andreozzi, L.; Teresa, B.M.; Magnano, L. Time Series Models for Different Seasonal Patterns. In Proceedings of the 34th International Symposium on Forecasting (ISF 2014), Rotterdam, The Netherlands, 29 June–2 July 2014.
31. Miri, A.; Ahmadi, H.; Ekhtesasi, M.R.; Panjehkeh, N.; Ghanbari, A. Environmental and Socio-Economic Impacts of Dust Storms in Sistan Region, Iran. *Int. J. Environ. Stud.* **2009**, *66*, 343–355. [CrossRef]
32. Miri, A.; Maleki, S.; Middleton, N. An Investigation into Climatic and Terrestrial Drivers of Dust Storms in the Sistan Region of Iran in the Early Twenty-First Century. *Sci. Total Environ.* **2021**, *757*, 143952. [CrossRef] [PubMed]
33. Alizadeh-Choobari, O.; Zawar-Reza, P.; Sturman, A. The "Wind of 120 Days" and Dust Storm Activity over the Sistan Basin. *Atmos. Res.* **2014**, *143*, 328–341. [CrossRef]
34. Javan, S.; Rahdar, S.; Miri, M.; Djahed, B.; Kazemian, H.; Fakhri, Y.; Eslami, H.; Fallahzadeh, R.A.; Gholizadeh, A.; Taghavi, M. Modeling of the PM_{10} Pollutant Health Effects in a Semi-Arid Area: A Case Study in Zabol, Iran. *Model. Earth Syst. Environ.* **2021**, *7*, 455–463. [CrossRef]
35. Vaezi, A.; Shahbazi, R.; Sheikh, M.; Lak, R.; Ahmadi, N.; Kaskaoutis, D.G.; Tagaris, E. Environmental Pollution and Human Health Risks Associated with Atmospheric Dust in Zabol City, Iran. *Air Qual. Atmos. Health* **2024**, *17*, 2491–2513. [CrossRef]
36. Rashki, A.; Eriksson, P.G.; Rautenbach, C.d.W.; Kaskaoutis, D.G.; Grote, W.; Dykstra, J. Assessment of Chemical and Mineralogical Characteristics of Airborne Dust in the Sistan Region, Iran. *Chemosphere* **2013**, *90*, 227–236. [CrossRef]
37. Maleki, S.; Koupaei, S.S.; Soffianian, A.; Saatchi, S.; Pourmanafi, S.; Rahdari, V. Human and Climate Effects on the Hamoun Wetlands. *Weather Clim. Soc.* **2019**, *11*, 609–622. [CrossRef]
38. Miri, A.; Shirmohammadi, E.; Sorooshian, A. Influence of Meteorological Factors and Air Pollutants on Bacterial Concentration across Two Urban Areas in the Sistan Region of Iran. *Urban Clim.* **2023**, *51*, 101650. [CrossRef]
39. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications*, 3rd ed.; Springer: New York, NY, USA, 2000.
40. Dickey, D.A.; Fuller, W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.
41. Said, S.E.; Dickey, D.A. Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* **1984**, *71*, 599–607. [CrossRef]

42. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2005.
43. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: New York, NY, USA, 2002.
44. Wei, W.W.S. *Time Series Analysis: Univariate and Multivariate Methods*; Addison Wesley: Boston, MA, USA, 2006.
45. Ripley, B.D. *Modern Applied Statistics with S*; Springer: New York, NY, USA, 2002.
46. Kumar, V.; Senarathna, D.; Gurajala, S.; Olsen, W.; Sur, S.; Mondal, S.; Dhaniyala, S. Spectral Analysis Approach for Assessing the Accuracy of Low-Cost Air Quality Sensor Network Data. *Atmos. Meas. Tech.* **2023**, *16*, 5415–5427. [CrossRef]
47. Liu, Y.; Adamu, A.; Tan, J. Spatial Characterization of Periodic Behaviors of Ground $PM_{2.5}$ Concentration across the Yangtze River Delta and the North China Plain during 2014–2024: A New Insight on Driving Processes of Regional Air Pollution. *Environ. Res.* **2025**, *277*, 121648. [CrossRef]
48. Vagropoulos, S.I.; Chouliaras, G.I.; Kardakos, E.G.; Simoglou, C.K.; Bakirtzis, A.G. Comparison of SARIMAX, SARIMA, Modified SARIMA and ANN-Based Models for Short-Term PV Generation Forecasting. In Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 4–8 April 2016; pp. 1–6.
49. NHS England and NHS Improvement. *Advanced Forecasting Techniques*; NHS England: London, UK, 2020.
50. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [CrossRef]
51. Zhao, D.; Zhang, H. The Research on TBATS and ELM Models for Prediction of Human Brucellosis Cases in Mainland China: A Time Series Study. *BMC Infect. Dis.* **2022**, *22*, 934. [CrossRef]
52. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. *Akaike Information Criterion Statistics*; D. Reidel Publishing: Dordrecht, The Netherlands, 1986.
53. Stock, J.H.; Watson, M.W. *Introduction to Econometrics*, 2nd ed.; Pearson: Boston, MA, USA, 2007.
54. Ljung, G.M.; Box, G.E.P. On a Measure of Lack of Fit in Time Series Models. *Biometrika* **1978**, *65*, 297–303. [CrossRef]
55. Shan, S.; Ni, H.; Chen, G.; Lin, X.; Li, J. A Machine Learning Framework for Enhancing Short-Term Water Demand Forecasting Using Attention-BiLSTM Networks Integrated with XGBoost Residual Correction. *Water* **2023**, *15*, 3605. [CrossRef]
56. Rashki, A.; Kaskaoutis, D.; Rautenbach, C.D.; Eriksson, P. Changes of Permanent Lake Surfaces, and Their Consequences for Dust Aerosols and Air Quality: The Hamoun Lakes of the Sistan Area, Iran. In *Atmospheric Aerosols—Regional Characteristics—Chemistry and Physics*; IntechOpen: London, UK, 2012.
57. Rashki, A.; Rautenbach, C.D.; Eriksson, P.G.; Kaskaoutis, D.G.; Gupta, P. Temporal Changes of Particulate Concentration in the Ambient Air over the City of Zahedan, Iran. *Air Qual. Atmos. Health* **2013**, *6*, 123–135. [CrossRef]
58. Shahriar, A.; Mozaffari, G.A.; Poudineh, S. The Impact of Drought Periods and Wind Erosion on the Physical Development of Desert Cities (Case Study: Zabol–Iran). *Desert* **2018**, *23*, 199–209.
59. Pozza, S.A.; Lima, E.P.; Comin, T.T.; Gimenes, M.L.; Coury, J.R. Time Series Analysis of $PM_{2.5}$ and $PM_{10-2.5}$ Mass Concentration in the City of Sao Carlos, Brazil. *Int. J. Environ. Pollut.* **2010**, *41*, 90–108. [CrossRef]
60. Chai, J.; Song, J.; Xu, Y.; Zhang, L.; Guo, B. Enhancing the applicability of satellite remote sensing for $PM_{2.5}$ estimation using machine learning models in China. *J. Sens.* **2022**, *2022*, 7148682. [CrossRef]
61. Kumari, S.; Muthulakshmi, P. SARIMA Model: An Efficient Machine Learning Technique for Weather Forecasting. *Procedia Comput. Sci.* **2024**, *235*, 656–670. [CrossRef]
62. Gocheva-Ilieva, S.; Ivanov, A. Assaying SARIMA and Generalised Regularised Regression for Particulate Matter PM_{10} Modelling and Forecasting. *Int. J. Environ. Pollut.* **2019**, *66*, 41–62. [CrossRef]
63. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
64. Petropoulos, F.; Apiletti, D.; Assimakopoulos, V.; Babai, M.Z.; Barrow, D.K.; Taieb, S.B.; Bergmeir, C.; Bessa, R.J.; Bijak, J.; Boylan, J.E.; et al. Forecasting: Theory and practice. *Int. J. Forecast.* **2022**, *38*, 705–871. [CrossRef]
65. Perera, S.; Maneja, R.H.; Allali, M.; Rakovski, C.; Linstead, E.; Struppa, D.; Qasem, A.; El-Askary, H. Time series decomposition of land surface temperature for long-term trend forecasting and impact on nesting sea turtle habitats in the Arabian Gulf. *J. Appl. Meteorol. Climatol.* **2024**, *63*, 1479–1497. [CrossRef]
66. Xin, N.; Su, J.; Hasan, M.M. MMformer with Adaptive Transferable Attention: Advancing Multivariate Time Series Forecasting for Environmental Applications. *arXiv* **2025**, arXiv:2504.14050. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.