sustainability

Article

Forecasting Accuracy of Traditional Regression, Machine Learning, and Deep Learning: A Study of Environmental Emissions in Saudi Arabia

Suleman Sarwar [1,*], Ghazala Aziz [2] and Daniel Balsalobre-Lorente [3]

1 Department of Finance and Economics, College of Business, University of Jeddah, Jeddah 23445, Saudi Arabia
2 Department of Business Administration, College of Administrative and Financial Sciences, Saudi Electronic University, Jeddah 13316, Saudi Arabia; g.aziz@seu.edu.sa
3 Department of Political Economy and Public Finance, Economics and Business Statistics and Economic Policy, University of Castilla-La Mancha, 13001 Ciudad Real, Spain; daniel.balsalobre@uclm.es
* Correspondence: suleman@uj.edu.sa

Abstract: Currently, the world is facing the problem of climate change and other environmental issues due to higher emissions of greenhouse gases. Saudi Arabia is not an exception due to the dependence of the Saudi economy on fossil fuels, which adds to the problem. However, due to the nonlinear pattern of pollution-creating gases, including nitrogen and sulfur dioxide, it is not effortless to rely on forecasting accuracy. Nevertheless, it is essential to denoise the data to extract the reliable outcomes used by different econometric approaches. Hence, the current paper introduces a hybrid model combining compressed sensor denoising (CSD) with traditional regression, machine learning, and deep learning techniques. Comparing different hybrid models and various denoising techniques revealed that CSD-GAN is the best model for accurately predicting NO_2 and SO_2, as compared with ARIMA, RLS, and SVR. Also, when the comparison is made between predicted and actual NO_2 and SO_2 levels, these are aligned, proving that CSD-GAN is superior in its level and direction of prediction. It can be concluded that the GAN model is the best hybrid model for predicting NO_2 and SO_2 emissions in Saudi Arabia. Hence, this model is recommended to policymakers for predicting environmental externalities and framing policies accordingly.

Keywords: ARIMA; machine learning; deep learning; environment; forecasting; Saudi Arabia

Citation: Sarwar, S.; Aziz, G.; Balsalobre-Lorente, D. Forecasting Accuracy of Traditional Regression, Machine Learning, and Deep Learning: A Study of Environmental Emissions in Saudi Arabia. *Sustainability* **2023**, *15*, 14957. https://doi.org/10.3390/su152014957

Academic Editor: Jin Tao

Received: 4 September 2023
Revised: 9 October 2023
Accepted: 13 October 2023
Published: 17 October 2023

1. Introduction

The need to increase awareness of environmental threats is essential, and people need to be more cautious about their actions relating to environmental issues in light of the recent increase in natural disasters, warming and cooling phases, and other weather patterns. Under the parameters of the present discussion, human civilization and globalization are the key contributors to the ongoing transformation of the global environment. There are many human-caused threats to the environment today. They include pollution, climate change, ozone depletion, acid rain, dwindling natural resources, population growth, improper garbage disposal, deforestation, and biodiversity loss. Unsustainable resource usage is the root cause of almost all of these operations. Large quantities of carbon dioxide and other greenhouse gases are released into the atmosphere when fossil fuels are used for energy in factories and vehicles. The environmental impact of these processes is substantial and adverse [1]. Consequences for the environment and human health on a global scale continue to be very unsettling. Unsafe water, insufficient sanitation and hygiene, air pollution, and global climate change are responsible for about ten percent of all worldwide deaths and disease loads.

Increasing industrial activity and motor vehicle usage in metropolitan areas are causing many health and environmental issues [2,3]. The impacts of air pollution on health are quite complicated since there are several causes, and their personal effects differ. "Nitrogen dioxide (NO_2) and sulfur dioxide (SO_2)" are the most common inorganic gas pollutants. The most significant atmospheric NO_2 foundations are burning fossil fuels and vehicle exhausts [4,5]. Rising SO_2 levels in the atmosphere are mainly caused by the use of sulfur-containing fuels such as coal for residential uses in metropolitan areas [6,7]. Inhaled air pollutants have an undesirable influence on human health by harming the lungs and respiratory system. They are also absorbed by the circulation and circulated throughout the body [8].

Nevertheless, the danger differs from one pollutant to another. "ecosystem" refers to a group working in the construction industry. Nitrogen oxides may render youngsters more vulnerable to respiratory ailments, especially during the winter. However, it is essential to predict the pollutant gasses so governments can take preemptive actions to counter these environmental issues.

Some previous studies used different artificial intelligence models to predict environmental externalities. For example, in order to forecast the values of a response variable used in modeling particulate matter, ref. [9] contrasted ANN with multiple linear regression (MLR), a statistical technique (PM_{10} and $PM_{2.5}$). The effectiveness of ANNs and decision tree models for estimating PM_{10} concentrations was assessed by [10]. Similar to this, ref. [11] compared the Back-propagation neural network (BPNN) and Autoregressive integrated moving average (ARIMA), another statistical time-series forecasting tool, for forecasting carbon monoxide (CO), SPM (suspended particulate matter), and sulfur dioxide (SO_2) in an industrial area. To forecast hourly $PM_{2.5}$ concentrations, ref. [8] developed a new forecasting method based on random forest and ANN approaches. The author of [9] conducted a comparative review of the modeling methodologies for simulating PM_{10} pollution concentrations using "ANN, LASSO, SVR, RF, kNN, and xGBoost". However, the forecasting of environmental pollutants is still lacking accuracy due to the high noise level in the data. Hence, we have attempted to input our valuable input into the existing literature.

To overcome the problems of single models, a few researchers also tried to build hybrid models. In this regard, to estimate PM and SO_2 concentrations [6], a hybrid forecasting model was developed based on a meta-heuristic approach known as the gray wolf optimizer. In addition, ref. [11] developed a hybrid air pollution estimating model that projected NO_2 and PM concentrations in China using fuzzy time series and uncertainty analysis. They also reported an application of SVM-based air pollution modeling. In this work, the authors situated two unique hybrid adaptive predicting models for one-step pollution prediction in Taiyuan, China. Their findings demonstrated that the combined "SVM and ANN" models performed better for air quality prediction than a single statistical learning model. In a similar line, ref. [12] created a unique hybrid-Garch strategy using SVM and ARIMA algorithms to estimate environmental externalities every hour for ten days.

However, initial air pollution data are erratic and noisy, like most time series. Forecasting these sounds and instability is pointless or harmful. Preprocessing the data and removing the interfering evidence from the original time series are required to increase predicting accuracy [13,14]. The denoising procedure has already been used in the forecast, and the results have proved its efficacy. For time-series analysis, ref. [7] developed a novel entropy-based wavelet denoising approach. To forecast the exchange rates, ref. [15] suggested a "Slantlet denoising-based least squares support vector regression (LSSVR)" model. The author of [16] suggested a neural network model based on exponential smoothing denoising for stock market forecasting. The author of [17] introduced a new model for exchange rate forecasting by integrating the "Markov switching" model with the Hodrick–Prescott filter. The author of [18] suggested a hybrid model for predicting water demand that combines the extended Kalman filter with genetic programming. The author of [19] implemented Fourier transform into a fuzzy time-series stock price predicting model. The author of [20] suggested a unique multivariate wavelet denoising-based method for assessing the portfolio value at risk (PVaR). The author of [21] suggested an enhanced wavelet modeling framework for eliminating noise in time-series forecasting.

Exponentially smoothing [22], as well as the Hodrick–Prescott (HP) filter [23], Kalman filter [24], "Fourier transform (FT), discrete cosine transform (DCT) [25], and wavelet transform" [26] are a few examples of denoising techniques that have been studied in the field of data processing. Unfortunately, the denoising techniques mentioned above all have a fatal flaw: they are susceptible to the values of the parameters that control them due to their fixed-basis construction. Recently, a denoising technique called compressed sensing-based denoising (CSD) has gained popularity since it is a more adaptable algorithm based on sparsity [27,28]. However, given an appropriate sparse transform basis, the CSD process may keep most of the information owing to sparsity. In contrast, most other denoising algorithms may lose some information due to their principles. Due to these two factors, this research's unique hybrid forecasting strategy uses CSD as an efficient data-denoising tool.

The novelty of the research under discussion primarily revolves around its pioneering approach to predicting environmental pollutants in Saudi Arabia, specifically NO_2 and SO_2. While preceding studies made commendable attempts, the unique combination of multiple denoising methods with advanced artificial intelligence techniques sets this study apart. This research introduces innovative CSD-AI strategies, namely CSD-SVR and CSD-GAN, which focus on data denoising to extract pertinent information, consequently reducing forecasting errors. Another distinctive feature is the presentation of the CSD-GAN methodology, which fills a gap in previous academic pursuits. Additionally, rather than relying on a solitary algorithm, this study harnesses the potential of four diverse algorithms—ARIMA, RLS, SVR, and GAN—in isolation and hybrid configurations alongside denoising techniques. This multifaceted approach facilitates a holistic evaluation and paves the way for an integrated, sophisticated model for accurately forecasting environmental pollutants. Comparing various AI-based denoising models contributes to developing a critical integrated model for forecasting SO_2 and NO_2.

This study aims to accurately predict NO_2 and SO_2 by combining "compressed sensing-based denoising (CSD)" and artificial intelligence (AI) techniques. However, the comparison is also made between different denoising techniques and single and hybrid models. As far as the structure of the paper is concerned, it has five sections: Section 1 deals with an introduction leading to Section 2, which is a brief literature review of related studies. Section 3 is regarding the methodology and description of data, whereas Section 4 presents the primary analysis results. The last section is about the concluding remarks and policy recommendations.

2. Literature Review

National and international restrictions have grown in response to the recent decline in air quality caused by increased air pollution. The need to know in advance what the future air quality levels will be emphasizes the significance of taking action to avoid air pollution. On a busy highway area, ref. [29] observed the horizontal distributions of air pollutants. They emphasized a poor correlation between distance and the particle air pollutant concentrations. Using monitors close to downtown Shanghai, ref. [30] studied the air pollutants and highlighted the vertical profiles of traffic-emitted pollutants and bimodal distribution patterns. To vertically anticipate the periodic features of air pollutants in the vicinity of viaduct environments, ref. [31] developed a back-propagation neural network.

Using two years' worth of observation data from the "Shanghai roadside station", ref. [32] performed research to estimate the "NO, NO_2, CO, and O_3" air contaminants in the atmosphere. The study found that the air effluence beneath the raised road was worse than that caused by vehicles on the sides of the road. To determine the air quality, they suggested using an LSTM model. Four air pollutants were estimated in the proposed model with a minor estimation error [33]. To forecast surface-level PM concentrations and track the impacts of urban traffic on the air quality in Shanghai, Du et al. suggested a deep learning model named DeepAir [34]. In addition to observing the impacts of the COVID-19 epidemic on the air, ref. [35] forecasted the PM_{10} and SO_2 air contaminants in Sakarya. For the prediction, they employed "recurrent artificial neural networks". They attained correlation levels of 0.88 for SO_2 and 0.67 for PM_{10}. In China, ref. [30] employed the random forest (RM) technique to estimate SO_2 emissions. They contrasted the RM algorithm's performance with that of other machine learning techniques. The author of [36] used CNN to estimate PM10. To improve estimate accuracy, they adopted the Bagging model. According to them, the model's accuracy, based on atmospheric variables, has reached 14.9469.

Using the AUSTAL 2000 model, [37] examined how a cement plant affected the values of air pollutants, including "CO, SO_2, NOx, and PM_{10}". They established unique classifications for each era after collecting emission data for 19 years. In their investigation, Perez et al. employed a neural network and a linear model to forecast air pollutants in Coyhaique, Chile. The research demonstrated that the linear model performed worse than the neural network model. With the neural network model, they attained an estimated accuracy of 0.95 [38]. To study the air quality index in Chennai, India, Refs. [39,40] recommended an approach that integrated support vector regression with long short-term memory. Compared to previous methods, deep learning models provide a value for the AQI that is more precise and accurate. They suggested an innovative recurrent neural network deep-learning model to forecast air pollution concentrations over the next two days. They computed the procedure using a particle swarm optimization technique. Their work aims to forecast the levels of six air pollutants for air quality. The authors of [41] performed a review to observe the features and functions of smart buildings. They described the strategies for accomplishing the objectives of smart buildings. The nine categories of performance metrics that were identified also needed to be improved. To enhance the performance of smart buildings, they looked at nine sets of performance metrics.

A strategy to predict urban air quality was put out by [42]. Their model is run on a dataset of 15 locations in India. They compared the new method's performance to other existing forecasting models. For the estimate of $PM_{2.5}$ concentrations, Chiang et al. suggested a hybrid time-series model that combines the autoencoder, CNN, and GRU approaches. Ecosystem refers to a group working in the construction industry [43]. Du et al. proposed a novel attention encoder–decoder model for multivariate time-series estimation issues. The Bi-LSTM deep learning structure serves as the foundation for the suggested model. The suggested model was evaluated using five multivariate datasets, and it was discovered to accurately predict the outcomes [44]. Du et al. suggested a hybrid multimodal deep learning system that combines "1D CNN and GRU" algorithms on multimodal traffic data to predict short-term traffic flows. The model accurately forecasted a complicated traffic flow [45]. In a different study, ref. [46] suggested a deep learning model for PM that included "one-dimensional CNN and Bi-LSTM modules". The model's accuracy for PM prediction was good. The experimental findings supported the forecast of air pollution.

3. Methodology and Data Setting

3.1. Denoising Methods

In 2004, Donoho first presented the concept of compressed sensing (CS). It provides a new approach to signal sampling that goes against Shannon's theorem. Using convex optimization, CS seeks to recover a sparse signal from a limited set of non-adaptive, linear data [47]. Among the various potential uses of compressed sensing (CS), the CSD method for signal denoising has been proposed [28]. To help with understanding, a sparse depiction is offered first. When signals are represented sparsely, they may be stated concisely in terms of a sound basis, such as the Fourier or wavelet basis. What follows is the corresponding mathematical expression:

$W \epsilon R^n$, and its orthonormal form is $\omega = [\omega_1 \omega_2 \ldots . \omega_n]$:

$$W = \sum_{i=1}^{n} S_i \omega_i \tag{1}$$

In Equation (1), $_i$ is the ith coefficient of W:

$$S_i = (W, \omega_i) \tag{2}$$

As a result, W may be expressed as s, representing the matrix $n \times n$, whose columns are $\omega_1 \ldots \omega_n$. Since coefficient s is sparse in this situation, Equation (1) obtains the spare form of W.

The CSD process has the following three steps:

1. The sparse or approximation sparse representation for the signal W may be written as $s = \omega^T W$ if the signal $W \epsilon R^n$ is sparse under an orthogonal basis.
2. We created an $m \times n$, $(m < n)$ dimensional observation matrix to quantify the sparse coefficients s and produced an observation vector, $Z = \varphi_s$. The transformed basis is unaffected by this observation matrix φ. The whole sensing procedure is as follows:

$$Z = \varphi\omega^T W \tag{3}$$

3. After receiving the compressed sensed signal Z, the recovery of W from Z is carried out as follows:

$$min||\omega^T W||_0,\ s.t = \varphi\omega^T W. \tag{4}$$

The following formula may be used to resolve the NP-hard issue in the equation above:

$$min||\omega^T W||_1,\ s.t = \varphi\omega^T W. \tag{5}$$

For the noise pollution in W, the minimization problem has to be adjusted as follows:

$$min||\omega^T W||_1,\ s.t = ||\varphi\omega^T W - Z||_2 \leq \varepsilon. \tag{6}$$

Equation (6) could be resolved using the "orthogonal matching pursuit (OMP)" technique, a popular and effective strategy for assuring the success of recovery. It could reconstruct the signal W extremely precisely using the compressed detected signal Z due to the potential of the sparse encoding of the signal W in certain transform domains.

CSD provides more flexible parameter settings than the standard denoising methods, such as "Fourier filter and wavelet denoising". The frequency and amplitude thresholds in the frequency domain must be specified for the Fourier filter. Additionally, this strategy may result in information loss. Similarly, the disadvantage of the wavelet denoising method is that it requires frequency thresholds to be defined for different time scales when processing massive volumes of data. In contrast, based on CS theory, CSD may produce a suitable denoising result by choosing an appropriate sparse transform basis and sampling rate.

3.2. Artificial Intelligence (AI)

3.2.1. Least Squares Support Vector Regression

Ref. [48] was the one who initially suggested the Support Vector Machine (SVM). The fundamental concept behind support vector regression (SVR) is to transfer the original data into a high-dimensional feature space, where linear regression is performed. The following formulation represents the regression function:

$$f(x) = \sum_{t=1}^{T} w_t K(x, x_t) + b \tag{7}$$

When w_t and b are the weights arrived at by minimizing the regularized risk function, $K(x,\ x_t)$ is the mapping function, and $f(x)$ is the prediction estimate. As a result, the optimization problem that results from Equation (7) is as follows:

$$\begin{gathered} min\ \frac{1}{2} w^T w + \gamma \sum_{t-1}^{T} (\xi_t + \xi_t^*) \\ s.t\ w^T \varphi(x_t) + b - y_t \leq \varepsilon + \xi_t^*, (i = 1,2,\ldots.T) \\ y_t - \left(w^T \varphi(x_t + b\right) \leq \varepsilon + \xi_t, (i = 1,2,\ldots.T) \end{gathered} \tag{8}$$

where the nonnegative variables ξ_t and ξ_t^* are the slack variables, which indicate the distance between the actual values and the corresponding border values of the $\varepsilon - tube$, and γ is the penalty parameter. The network structure of the relevant algorithm of SVM is reported in Figure 1.

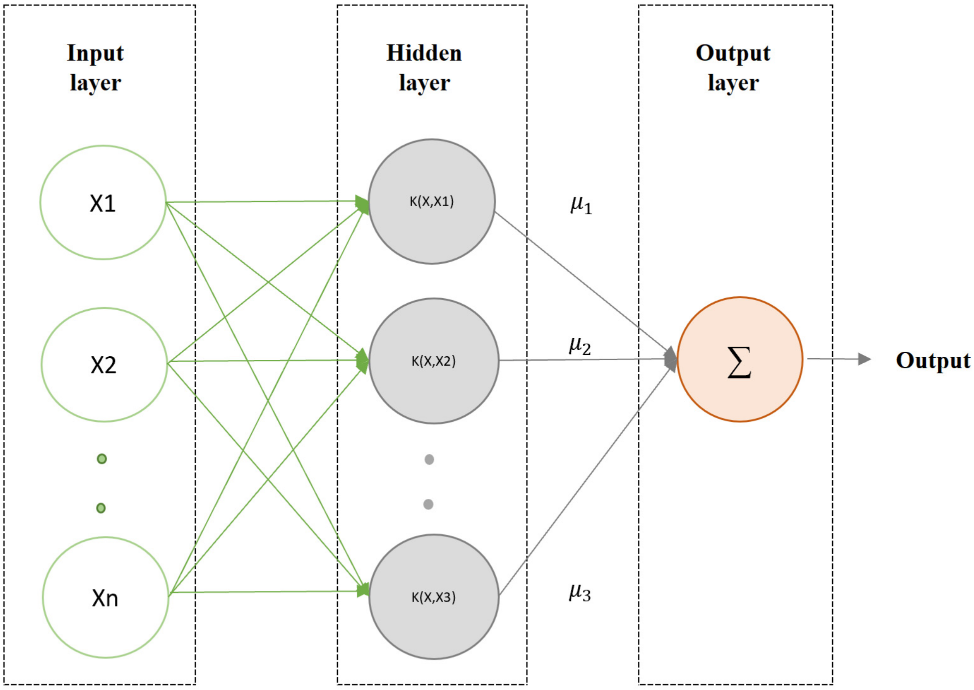

Figure 1. Network structure diagram of SVM.

3.2.2. Generative Adversarial Network (GAN)

Utilizing generative adversarial networks (GANs) in time-series forecasting has reshaped predictive modeling perspectives. GANs operate with two deeply intertwined neural networks: a generator that crafts sequences, and a discriminator that discerns genuine sequences from the generated ones. To effectively harness GANs for forecasting, one must preprocess time-series data to ensure consistent intervals and normalization, optimizing the neural architectures' performance. Sequence prediction is facilitated by introducing lagged versions of the series as input. Within this framework, the generator, when fed with random noise, aims to replicate the dynamics of genuine data. Simultaneously, the discriminator refines its skill in differentiating actual future sequences from the generator's concoctions. Their adversarial interplay iteratively refines the quality of the generator's output. When forecasting, the generator's refined output, grounded in recent observations, is employed, tapping into GANs' prowess at modeling complex data distributions, capturing potential nonlinearities, and intricate patterns for enhanced predictive accuracy. The network structure of the relevant algorithm of SVM is reported in Figure 2.

Gated recurrent units (GRUs) and generative adversarial networks (GANs) hail from different facets of deep learning, yet their integration offers promising avenues in various applications. GRUs, a variant of recurrent neural networks, excel in sequence-based tasks, capturing temporal dependencies through specialized gating mechanisms. On the other hand, GANs consist of a duet of networks—a generator and a discriminator—collaborating in an adversarial setting to produce high-fidelity data mimics. Incorporating GRUs within the GAN architecture can enhance sequential data modeling. Specifically, when GANs target sequence generation tasks, a GRU-based generator or discriminator can be pivotal. The temporal dynamics grasped by GRUs ensure that the generated sequences are plausible regarding individual data points and their sequential structure. Conversely, a GRU-infused discriminator becomes adept at identifying discrepancies in the temporal patterns of generated sequences. This symbiosis marries the generative prowess of GANs with the sequence-savvy nature of GRUs, advancing the state-of-the-art in sequential data generation.

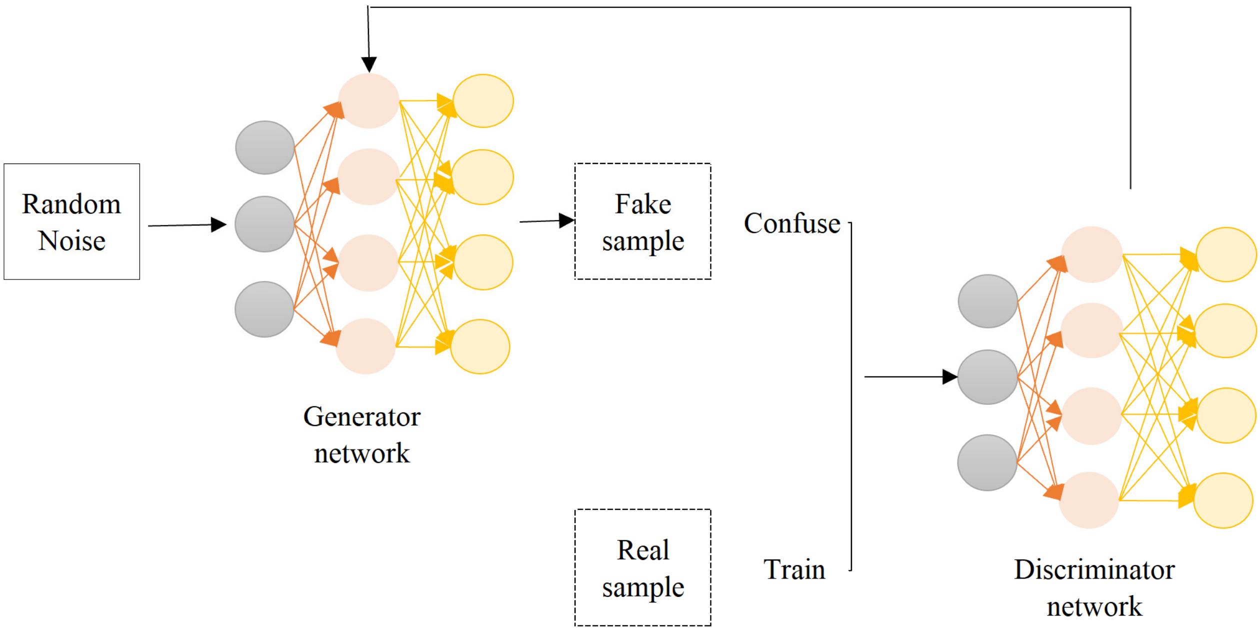

Figure 2. Network structure diagram of GAN.

3.3. AI Forecasted Models Integrated with Compressed Sensing-Based Denoising (CSD-AI)

Based on the previously discussed methodologies, a novel hybrid model, the "CSD-AI" learning paradigm for SO_2 and NO_2 forecasting, is developed, and multi-step-ahead prediction is used. There are numerous techniques for doing this, according to [49]; however, the direct forecasting approach is used in this study. Based on the time series x_t $(t = 1, 2, \ldots T)$, the following equation is utilized to obtain an m-step forward forecast for x_{t+m}.

$$\hat{X}_{t+m} = f(X_t, X_{t-1}, \ldots, X_{t-(l-1)}) \tag{9}$$

$\hat{X}_t$ is the period's forecast value, X_t is the actual value, and l is the lag order. In CSD-AI, there are two steps:

1. The original data X comprises a trend T, and noise X is first represented by an appropriate transform basis; in our instance, a wavelet basis. The sparse coefficients are then sampled using a Gaussian white noise sampling matrix. Ultimately, the cleaned data T may be obtained via the OMP recovery process for more research.
2. After data denoising, a powerful AI approach, such as an SVM or ANN, is used to model the cleaned data T and make predictions for the original X.

3.4. Data Description

The data related to NO_2 and SO_2 were extracted from the King Abdullah Petroleum Studies and Research Center (KAPSARC) from 1 August 2019 to 15 July 2020, from which the training data are from 1 August 2019 to 28 May 2020, and the testing data are from 29 May 2020 to 15 July 2020. The training data are around a third-fourth, and the testing data are about one-fourth of the complete data. In addition, we have used the data from 16 July to 30 August for out-of-sample forecasting. Our emphasis is on the daily frequency data; however, a few of the observations are missing in the daily data, which is simulated by using the Markov chain Monte Carlo (MCMC) algorithm. The units for NO_2 and SO_2 are parts per billion (ppb).

Table 1 reports the data description of the variables, where the skewness is above 0, and kurtosis is above 3, which indicates that the distribution is nonnormal. The findings of the Jarque–Bera statistics are significant and also reject the null hypothesis of normality. For conditional heteroscedasticity, we employed the ARCH-LM test that confirms the presence of conditional heteroscedasticity. In section B, the results of BDS are reported, which are used to confirm the existence of nonlinearity in the data series [50]. The null hypothesis of BDS presents the series as linearly dependent. In our case, all the values of BDS are significant at 1 percent, meaning that the series have nonlinearly dependence. However, in the presence of nonnormality and conditional heteroscedasticity, we have multiple machine learning options that are useful for forecasting, such as SVM and neural networks, which are capable of handling the nonnormality and conditional heteroscedasticity issues and present robust results [51].

Table 1. Data description.

Section A: Descriptive	NO_2	SO_2
Mean	3.974	1.528
Maximum	6.370	4.190
Minimum	0.200	0.190
Std. Dev.	0.388	0.635
Skewness	1.640	1.181
Kurtosis	10.070	5.218
Jarque–Bera	519.074	70.012
Probability	0.000	0.000
ARCH-LM	89.261 ***	210.674 ***
Section B: BDS		
2	0.352 ***	0.140 ***
3	0.394 ***	0.189 ***
4	0.410 ***	0.200 ***
5	0.573 ***	0.342 ***
6	0.499 ***	0.418 ***

Notes: *** represents the level of significance at 1%.

3.5. Performance Evaluation Criteria

The mean absolute percentage error is calculated by taking the average of all of the observed values' absolute deviation values. The value of the arithmetic mean, which is defined as the set of differences that are not cancelled out by one another, is given as a percentage. The chart illustrates the actual predicted inaccuracy in an accurate manner and accurately indicates the extent of data dispersion.

$$MAPE = \frac{100}{n} \sum\nolimits_{i=1}^{n} \left| \frac{T_i - X_i}{T_i} \right| \tag{10}$$

MSE represents the variation between estimators. A lower number indicates a more accurate prognosis. The *MSE* measures how dispersed the data collection is.

$$MSE = \frac{1}{n} \sum\nolimits_{i=1}^{n} (T_i - X_i)^2 \tag{11}$$

The average difference between the values that were anticipated and those that were actually observed may be more properly expressed using a statistic called the root mean square error, which is sometimes referred to as the standard error. When the anticipated and actual values are in perfect agreement with one another, this error is equal to zero. The author of [52] recommends *MSE* and *RMSE* as important criteria for comparisons.

$$RMSE = \sqrt{\frac{1}{n} \sum\nolimits_{i=1}^{n} (T_i - X_i)^2} \tag{12}$$

Since positive and negative variance values do not cancel out, the mean absolute deviation accurately represents the expected error. T_i is the actual values, X_i is the predicted values, and n is the total number of predicted values.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|T_i - X_i| \tag{13}$$

3.6. Benchmark Models

The CSD technique's ability to improve forecast correctness is evaluated first. For this reason, a set of hybrid models is developed by combining CSD with well-known forecasting techniques, such as the most traditional method of robust least squares (RLS) [53] and the most widely used AIs of SVR [54] and GAN [55], and then, by contrasting these hybrid models (CSD-ARIMA, CSD-RLS, CSD-SVR, and CSD-GAN). Two viewpoints may be used to outline the primary justifications for adopting ARIMA, RLS, SVR, and GAN as forecasting models in hybrid model development. On the one hand, RLS is the most common linear regression model, and it has long been employed as a standard in prediction research. On the other hand, SVR and GAN have been widely used as the most common AI approaches, notably for predicting SO_2 and NO_2 [56,57]. Despite their distinct strengths, they can only partially be shown to be superior to each other. As a result, the suggested hybrid framework implements both potent intelligence models (SVR and GAN) as forecasting models.

The benefits of the CSD-AI learning paradigm that was proposed are investigated in the second step. As a consequence of this, in order to produce a set of hybrid benchmarks, an additional five well-known denoising techniques, including exponential smoothing (ES) [58], the Hodrick–Prescott (HP) filter [59], Kalman filter (KF) [60], and wavelet denoising (WD) [61], have been included as preprocessors for the original data. In general, for the proposed CSD-AI models (i.e., CSD-ANN and CSD-RNN), three single benchmarks (i.e., ARIMA, RNN, and ANN), one CSD-based ARIMA hybrid benchmark (i.e., CSD-ARIMA), and a set of hybrid models with an additional five denoising techniques are constructed for comparison. These benchmarks are used to evaluate the performance of the proposed CSD-AI models. The sequence of study is mentioned in Figure 3, which proposes us to focus on ARIMA, RLS, SVR, and GAN.

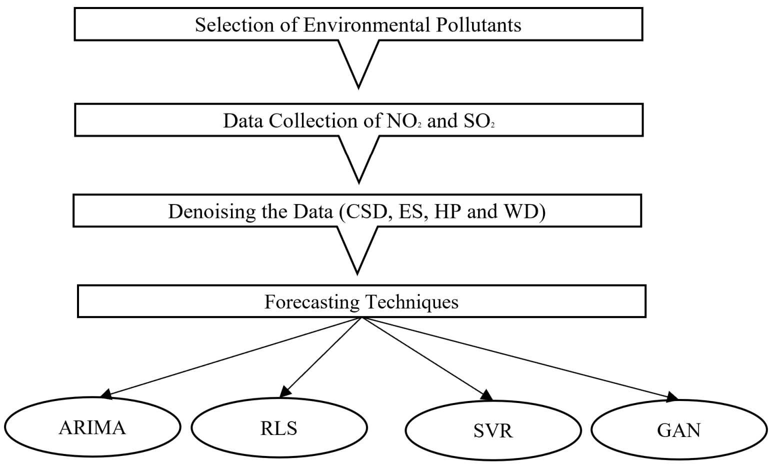

Figure 3. Structure and steps of study.

3.7. Parameter Settings

Research [22,23] on the issue is combined with trial and error to identify the parameters of the denoising approach to be used. CSD uses a Symlet-6 sparse transform basis, a sample size of 500, and 125 iterations of the OMP algorithm. A smoothing factor of 0.2 has been used in ES. In the HP filter, the smoothing value is set to 100. In KF, we achieved a covariance between measurements of 0.25 and a process covariance of 0.0004. Within DCT, 100 is used as the cutoff for the lowest frequency. The frequency thresholds in

WD are determined using the soft threshold method, Symlet 6 as the wavelet basis, and 8 decomposition iterations [62].

The optimal ARIMA model for each training sample is chosen by minimizing the Schwarz criterion (SC) in order to create forecasting models [63]. This investigation employs a feed-forward neural network (FNN) $(I - H - O)$ [64] in ANN, with seven hidden nodes, one output neuron, and I input neurons, where I is the lag order decided upon using autocorrelation and partial correlation tests and is ultimately set at 6. There are 10,000 iterations of each ANN model conducted on the training data. All of the models have been coded in the computer application Matlab R2019a, and all of the programmers have been run on a HP laptop i7.

4. Results and Discussion

The initial stage in the CSD-AI-based learning paradigm that has been presented is to use CSD to denoise the data gathered on the NO_2 concentration, and the related outcome can be shown in Figure 4. The second phase is to make projections based on the cleansed data using a specific and very accurate forecasting program (e.g., SVE and GAN). In addition, a set of benchmark models, which may include single or hybrid forecasting models, is executed to make comparisons.

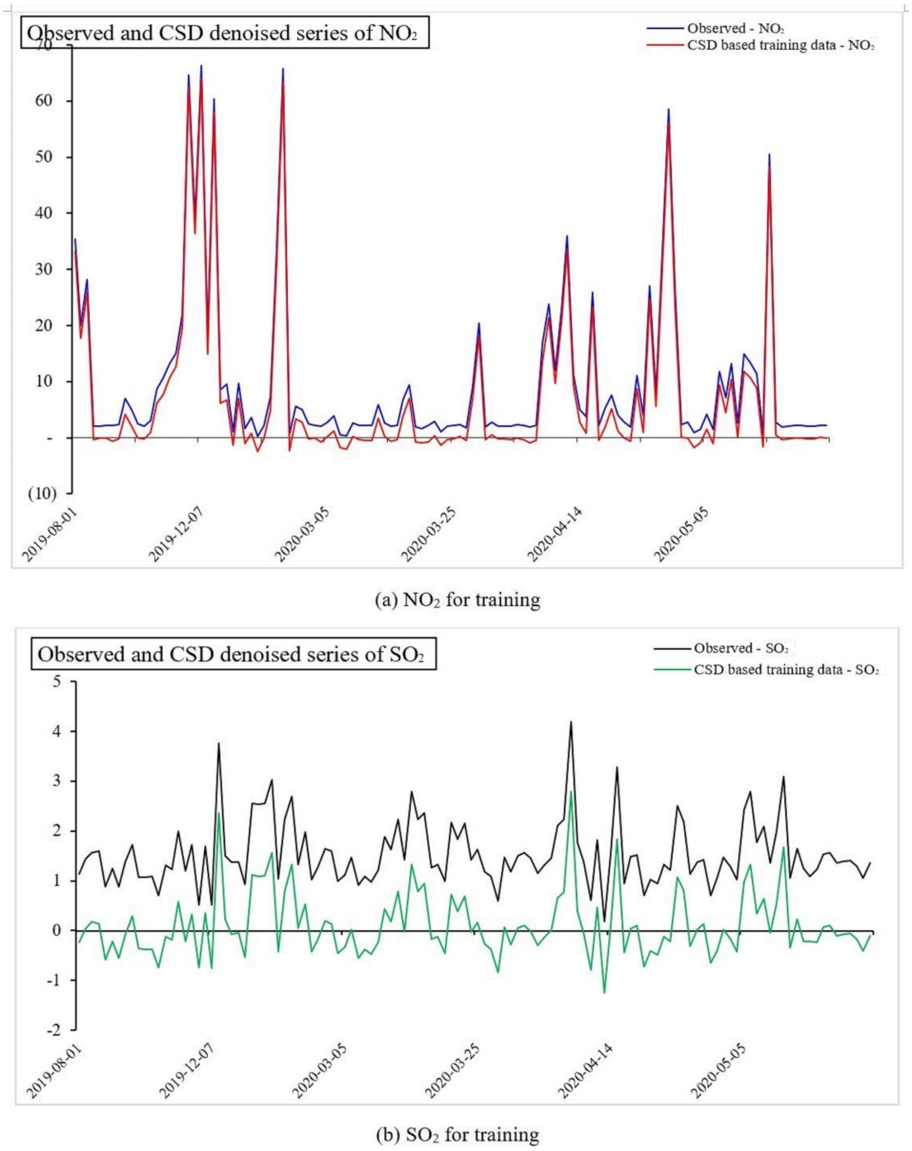

Figure 4. CSD−based denoising series of NO_2 and SO_2 for training.

First, we discuss how CSD helps with better predicting. Figures 5–8 compare the prediction accuracy (in terms of *MAPE*, *MSE*, *RMSE*, and *MAD*) of CSD-based hybrid learning paradigms to the accuracy of their benchmarks that do not use CSD. Also, the results of the Diebold–Mariano (DM) tests were conducted on CSD-based hybrid models as well as single models.

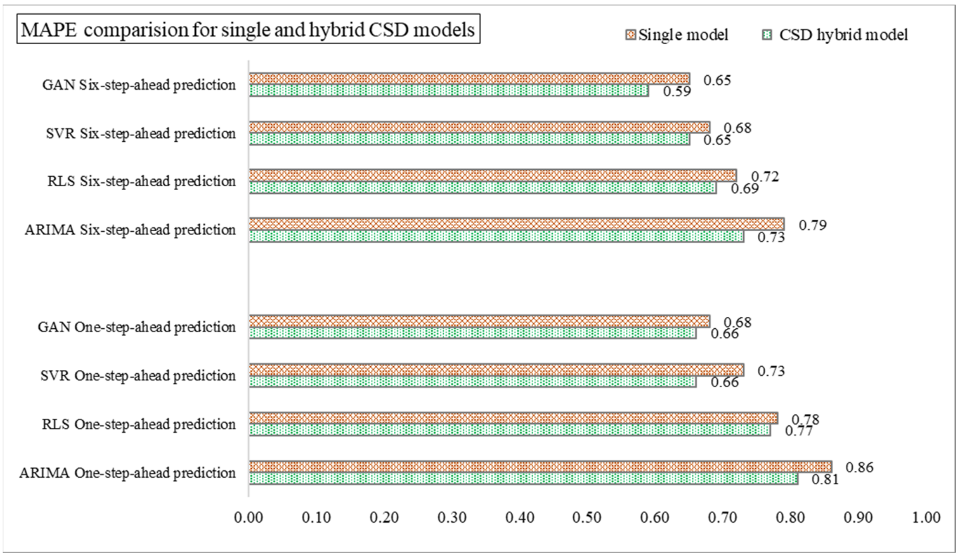

Figure 5. *MAPE* comparison for single benchmark and CSD-based hybrid models.

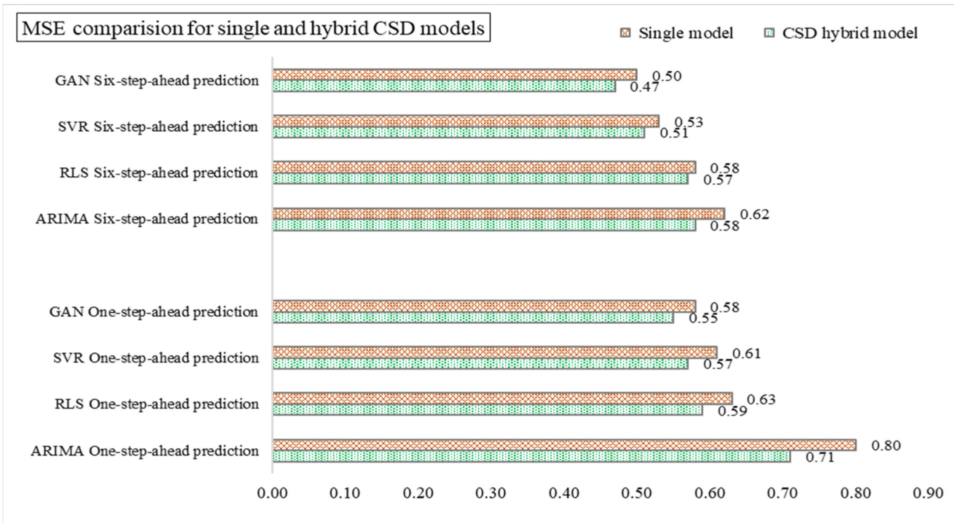

Figure 6. *MSE* comparison for single benchmark and CSD-based hybrid models.

4.1. Implementation of CSD-Based Models in Forecasting (Effectiveness of CSD in Forecasting)

The comparison in terms of *MAPE* for single as well as hybrid CSD models is reported in Figure 5. It can be noted that the CSD hybrid models show better performance as compared to single models. The *MAPE* value of the hybrid models is lower than the *MAPE* of single models. In one-step-ahead prediction, the *MAPEs* of the CSD-based SVR and GAN hybrid models are lower than those of the CSD-based traditional models, such as ARIMA and RLS. The performance of CSD-SVR and CSD-GAN in one-step-ahead prediction is the same. However, in six-step-ahead prediction, CSD-SVR and CSD-GAN outperform the traditional model with lower *MAPE* values. However, the performance of CSD-GAN is more prominent, with the lowest *MAPE* value. This indicates that the directional prediction capability of the hybrid models is superior to the traditional model in both one-step- and six-step-ahead prediction.

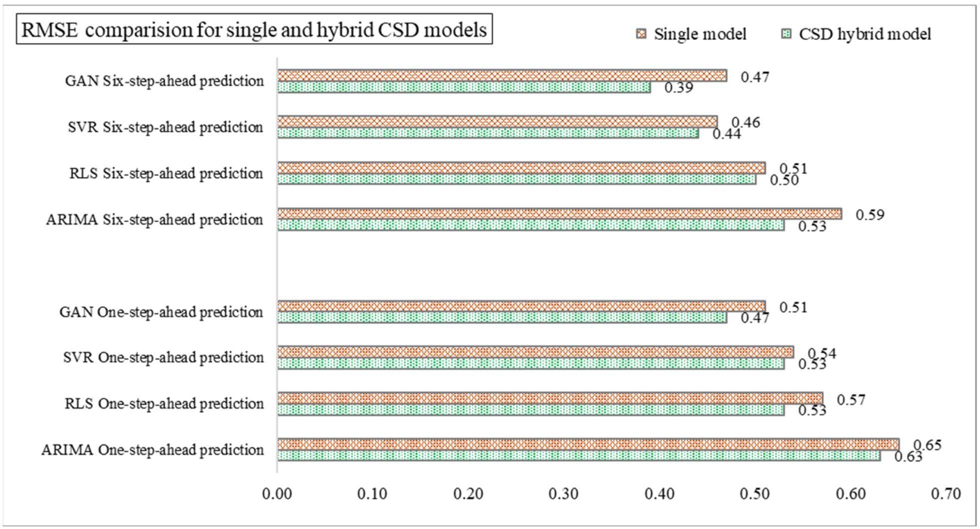

Figure 7. *RMSE* comparison for single benchmark and CSD-based hybrid models.

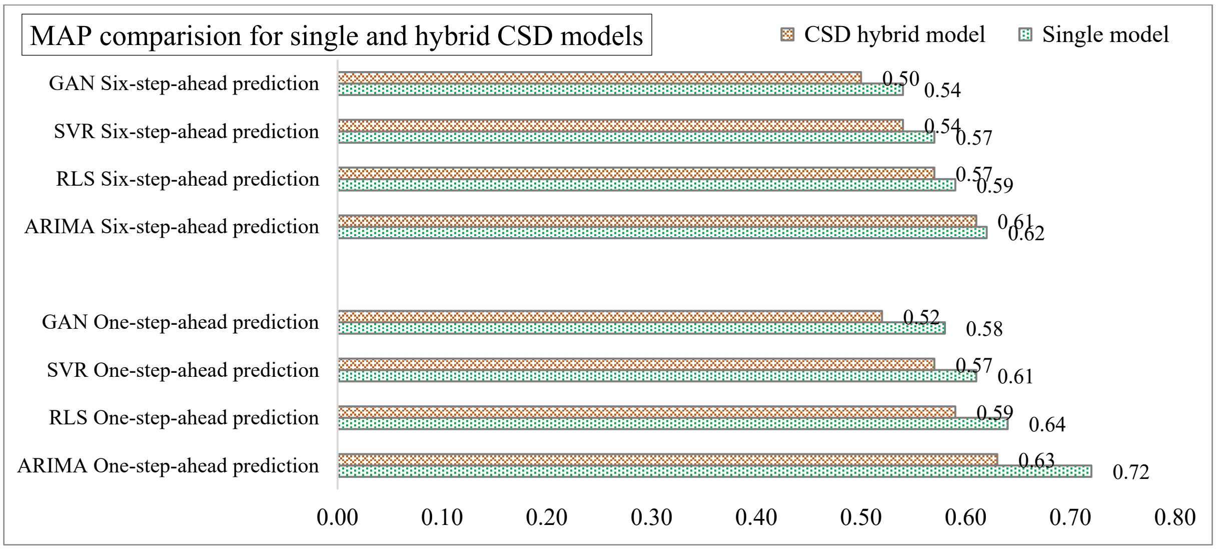

Figure 8. *MAD* comparison for single benchmark and CSD-based hybrid models.

To check the performance of the single and hybrid CSD models (CSD-SVR and CSD-GAN), *MSE* is also used, and the results are reported in Figure 6. In one-step-ahead prediction, the *MSE* values of CSD-SVR and CSD-GAN are lower than the traditional CSD-RLS and CSD-ARIMA models. Likewise, single AI models (SVR and GAN) also show better performance with lower *MSE* values than the *MSE* value of RLS. Similarly, in six-step-ahead prediction, the AI models perform better than the traditional model in both the single (SVR, GAN) and CSD-based models (CSD-SVR, CSD-GAN). The *MSE* values of SVR and GAN are lower than RLS. Also, CSD-SVR and CSD-GAN show lower *MSE* values than CSD-RLS. This proves the superiority of the hybrid models for directional prediction. A comparison of CSD-SVR's and CSD-GAN's performance validates the superiority of the CSD-GAN model with the most negligible *MSE* value.

Aside from *MAPE* and *MSE*, *RMSERMSE* is the third measure to compare the performance of single (ARIMA, RLS, SVR, and GAN) and hybrid (CSD-ARIMA, CSD-RLS, CSD-SVR, and CSD-GAN) models. Figure 7 shows the outcomes when SVR and GAN perform better than RLS with a lower *RMSERMSE* value in one-step-ahead prediction. Furthermore, in the case of the hybrid models, CSD-SVR and CSD-GAN have lower *RMSERMSE* values than CSD-RLS. This means that the AI models are better than single or hybrid models in one-step-ahead prediction. However, CSD-GAN is the best because it has the most negligible *RMSE* value.

Additionally, in six-step-ahead prediction, the AI models outperform traditional models, either single or hybrid. The *RMSE* values of SVR and GAN are lower than the *RMSE* values of RLS and ARIMA. Likewise, the *RMSE* values of CSD-SVR and CSD-GAN are also lower than the *RMSE* value of CSD-RLS. This proved that the prediction levels are much better with hybrid AI models. However, CSD-GAN is superior to CSD-SVR.

The last condition for the performance comparison of single and CSD-based hybrid models is *MAD*, and Figure 8 presents the outcomes. In the case of one-step prediction, the AI models are better than the traditional models with lower *MSE* values. Also, the CSD-based hybrid models (CSD-SVR and CSD-GAN) have superior performance with lower *MAD* values. Single models (SVR, GAN) have a lower *MAD* than RLS in one-step-ahead prediction. Also, CSD-SVR and CSD-GAN show lower *MAD* values compared to CSD-RLS. The same applies to six-step-ahead prediction, where the AI models perform better with lower *MAD* values. However, it is worth noting that the CSD-based hybrid model (CSD-SVR and CSD-GAN) outperforms the single and traditional models. Nevertheless, CSD-GAN has the lowest *MAPE*, showing its superiority in the direction and level of prediction.

Some significant conclusions can be drawn from the results regarding the *MAPE*, *MSE*, *RMSE*, and *MAD*. The first and most prominent conclusion is regarding the superiority of the CSD-based AI models (CSD-ARIMA, CSD-RLS, CSD-SVR, and CSD-GAN) in terms of prediction regarding their level and direction. This means novel hybrid techniques are best when it comes to prediction. Also, CSD-GAN proved to be the best prediction model in one-step and six-step predictions. The reason behind the superior performance of CSD is its ability to lower the noise significantly in NO_2 data, which results in better performances of SVR and GAN.

Additionally, the CSD-based hybrid models proved to be better in their level and direction of prediction when compared to their single models, which also shows the significance of the CSD approach. In addition, both AI models (SVR and GAN) perform better than RLS, which is a traditional model, proving that the CSD-based AI models are the best forecasting tools. The reason for the superiority of CSD-AIs is simple: the pattern of NO_2 concentration is not linear, which means modeling this data cannot be done with traditional models, and AI techniques can better model this nonlinear data.

4.2. Performance of CSD-Based Denoising Methods

Verifying CSD's superiority over other denoising techniques is vital before moving on to accurate forecasting. Figure 9 displays the results of using several denoising methods for evaluation. These approaches include ES, HP, and WD. Comparing the *MAPE*s of ARIMA, CSD-ARIMA, ES-ARIMA, HP-ARIMA, and WD-ARIMA, we have affirmed the minimum error value of *MAPE* in the case of the CSD-based ARIMA. In the case of robust least squares, CSD-RLS has a lower *MAPE* for the training and testing data. However, CSD-RLS outperforms the other denoising methods. Regarding the *MAPE*s of SVR, CSD-SVR, ES-SVR, HP-SVR, and WD-SVR, HP-SVR has a lower *MAPE* for the training and testing data. However, CSD-SVR outperforms the other denoising methods. The third model is GAN, and it can be noted that, based on the *MAPE*s of GAN, CSD-GAN, ES-GAN, HP-GAN, and WD-GAN, CSD-GAN has a lower *MAPE* for the training and testing data. However, CSD-GAN outperforms the other denoising methods. Hence, it is not wrong to say that among the CSD-RLS, CSD-SVR, and CSD-GAN, the CSD-GAN has the lower *MAPE*; however, according to the *MAPE*, the CSD-GAN outperforms the other hybrid models, proving the superiority of this model.

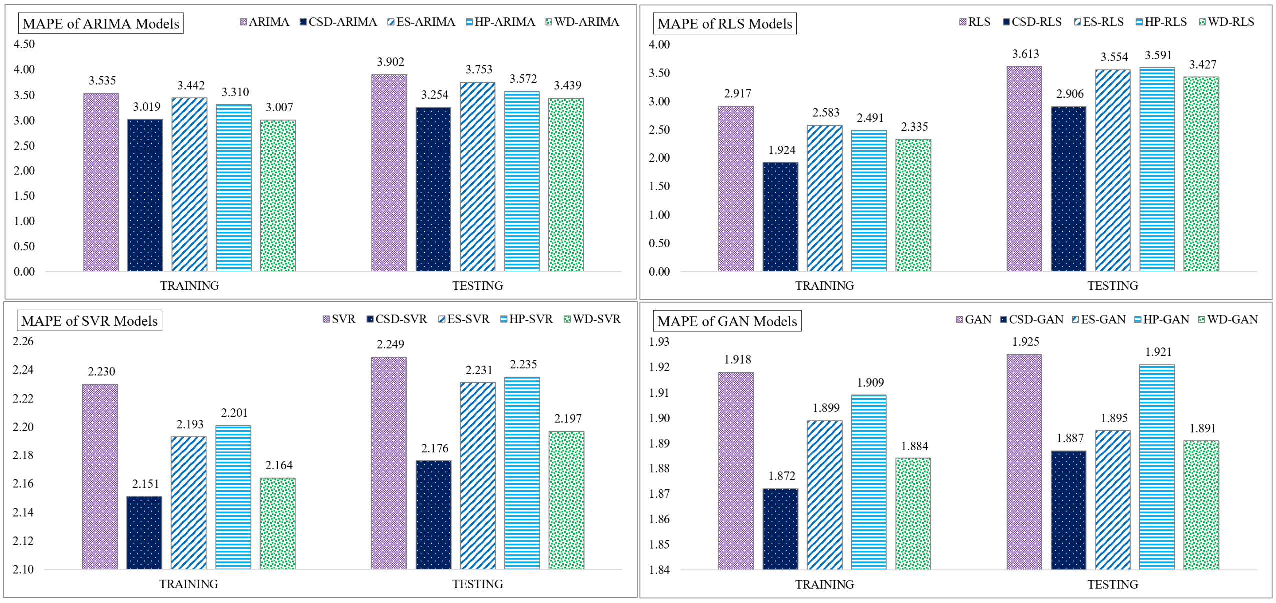

Figure 9. Comparing the *MAPE* of single and denoised models.

Aside from the *MAPE*, *MSE* is also used to compare the performance of various denoising techniques, and the results are presented in Figure 10. In a comparison of the *MSE* of ARIMA and denoised-based ARIMA models, we have reported the lowest *MSE* for CSD-ARIMA. Meanwhile, RLS, CSD-RLS, ES-RLS, HP-RLS, and WD-RLS show that CSD-RLS has a lower *MSE* for the training and testing data. However, CSD-RLS outperforms the other denoising methods. The comparison regarding the MSEs of SVR, CSD-SVR, ES-SVR, HP-SVR, and WD-SVR clearly shows that the CSD-SVR has a lower *MSE* for the training and testing data. However, CSD-SVR outperforms the other denoising methods. The same is true for the *MSE* because the comparison of MSEs among GAN, CSD-GAN, ES-GAN, HP-GAN, and WD-GAN indicates that CSD-GAN has a lower *MSE* for the training and testing data. However, CSD-GAN outperforms the other denoising methods. The *MSE* comparison of all denoising techniques confirms that the CSD-RLS, CSD-SVR, and CSD-GAN have the lower MSEs; however, according to the *MSE*, CSD-GAN outperforms because the least *MSE* value is related to this model.

RMSE is also used to compare the performance of denoising techniques, and the results of the *RMSE* are shown in Figure 11. Here, with the comparison of the *RMSEs* of ARIMA, CSD-ARIMA, ES-ARIMA, HP-ARIMA and WD-ARIMA, we have concluded the lowest error for CSD-ARIMA. In the case of the RLS and denoised RLS models, CSD-RLS has a lower *RMSE* for the training and testing data. However, CSD-RLS outperforms the other denoising methods. In the same way, a comparison of the *RMSE*s of SVR, CSD-SVR, ES-SVR, HP-SVR, and WD-SVR suggests that CSD-SVR has a lower *RMSE* for the training and testing data. However, CSD-SVR outperforms the other denoising methods. Additionally, comparing the *RMSE*s of GAN, CSD-GAN, ES-GAN, HP-GAN, and WD-GAN shows that CSD-GAN has a lower *RMSE* for the training and testing data. However, CSD-GAN outperforms the other denoising methods. Hence, among the CSD-ARIMA, CSD-RLS, CSD-SVR, and CSD-GAN, the CSD-GAN has the lower *RMSE*; however, according to the *RMSE*, the CSD-GAN outperforms the other denoising techniques.

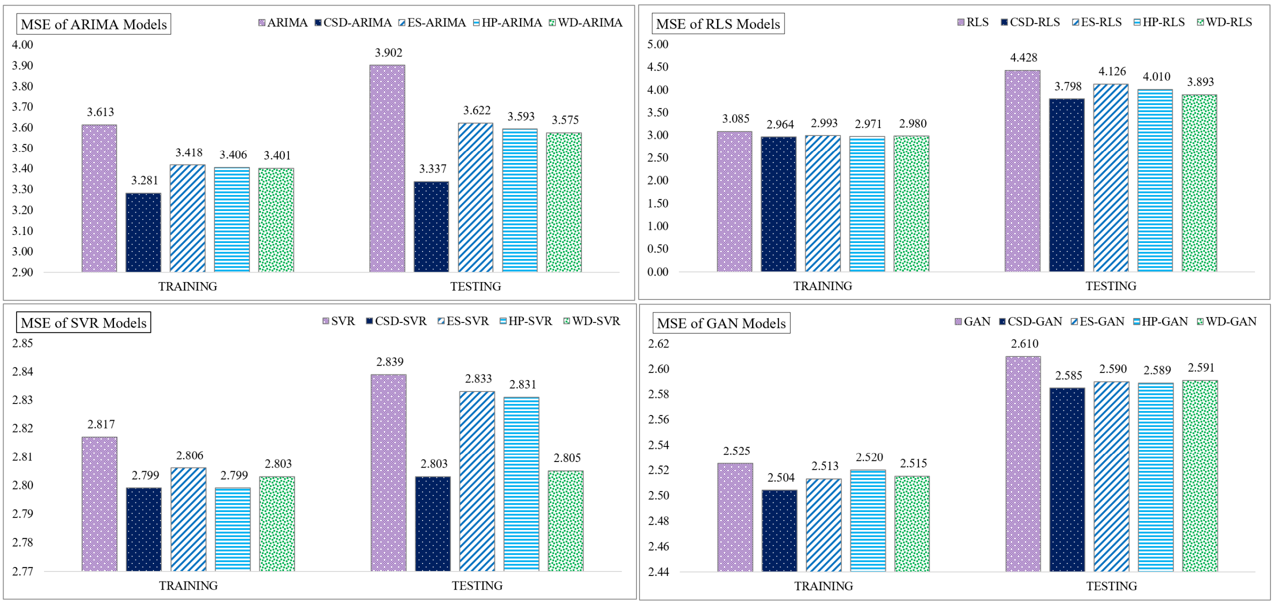

Figure 10. Comparing the *MSE* of single and denoised models.

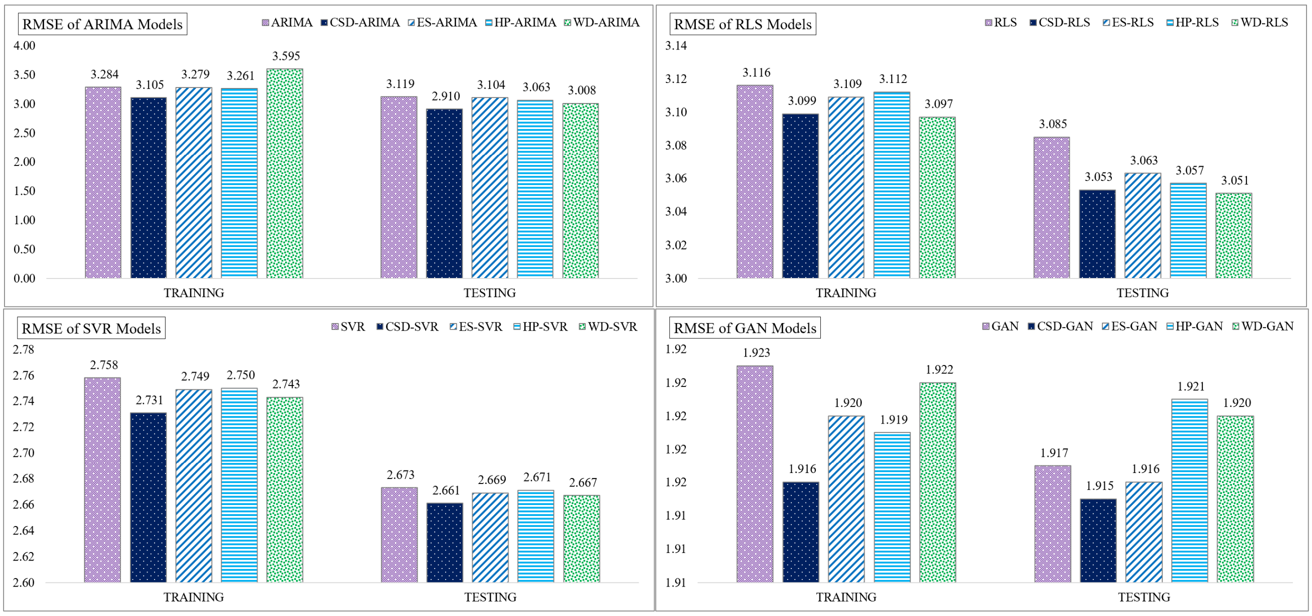

Figure 11. Comparing the *RMSE* of single and denoised models.

The last criterion is *MAD*, which is used to compare the denoising techniques, and the results are presented in Figure 12. In the comparison of the *MAD* values of traditional regression and denoised-based models, we presented that CSD-ARIMA and CSD-RLS have lower *MAD*s for the training and testing data. However, the CSD-based traditional models outperform other denoising methods. When the *MAD* of the SVR model is consulted, it is noted that (from the *MAD* of SVR, CSD-SVR, ES-SVR, HP-SVR, and WD-SVR), CSD-SVR has a lower *MAD* for the training and testing data. However, CSD-GAN outperforms the other denoising methods. Additionally, a comparison of the *MAD*s of GAN, CSD-GAN, ES-GAN, HP-GAN, and WD-GAN validates that CSD-GAN has a lower *MAD* for the training and testing data. However, CSD-GAN outperforms the other denoising methods. From these results, it is clear that among the CSD-ARIMA, CSD-RLS, CSD-SVR, and CSD-GAN, the CSD-GAN has the lower *MAD*; however, according to the *MAD*, the CSD-SVR outperforms.

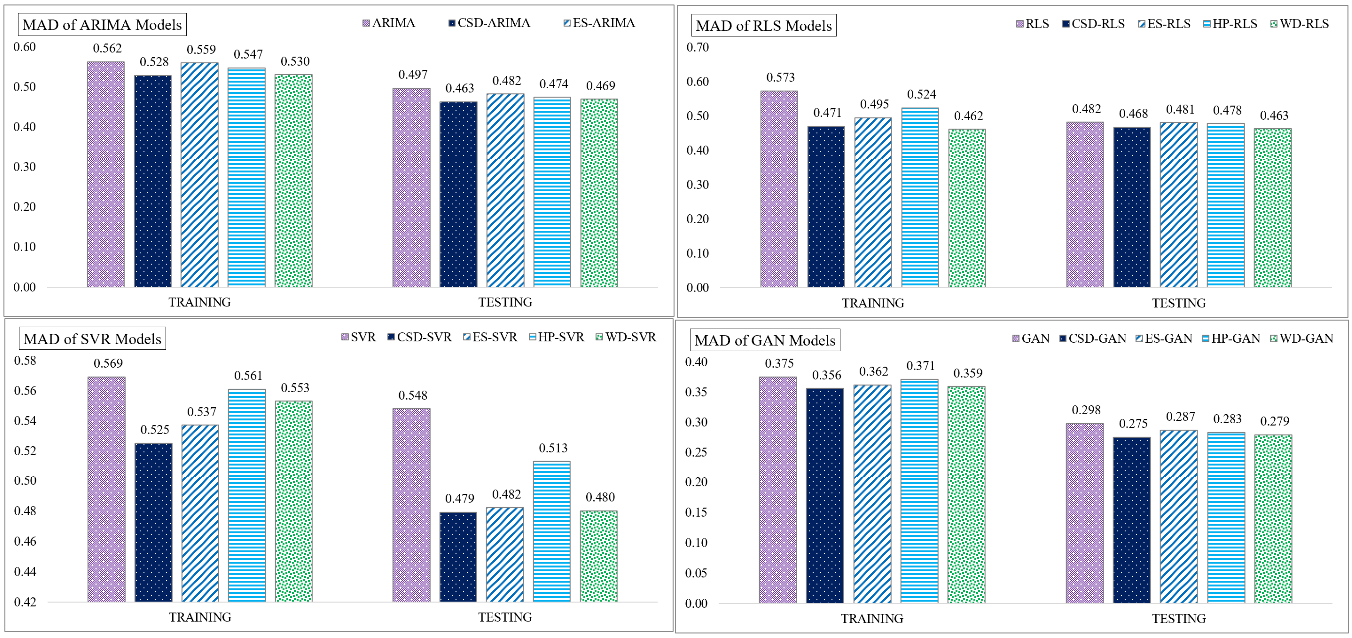

Figure 12. Comparing the *MAD* of single and denoised models.

The results regarding comparing different denoising techniques confirm a few concluding points. It is confirmed that of CSD-GAN, ES-GAN, HP-GAN, and WD-GAN, CSD-GAN proved to be the best model in all cases. CSD-GAN has the lowest *MAPE*, *MSE*, *RMSE*, and *MAD* in one-step- and six-step-ahead predictions. This means CSD-GAN is the best hybrid model to forecast NO_2. This result can also be confirmed through Table 2, where the *MAPE*, *MSE*, *RMSE*, and *MAD* of CSD-GAN are the lowest compared to all other AI and traditional models. Hence, in the current study, the forecasting of NO_2 is achieved based on CSD-GAN.

Table 2. *MAPE*, *MSE*, *RMSE* and *MAD* for single and hybrid models.

		ARIMA	**CSD-ARIMA**	**ES-ARIMA**	**HP-ARIMA**	**WD-ARIMA**
MAPE	*Training data*	3.535	3.019	3.442	3.310	3.007
	Testing data	3.902	3.254	3.753	3.572	3.439
MSE	*Training data*	3.613	3.281	3.418	3.406	3.401
	Testing data	3.902	3.337	3.622	3.593	3.575
RMSE	*Training data*	3.284	3.105	3.279	3.261	3.595
	Testing data	3.119	2.910	3.104	3.063	3.008
MAD	*Training data*	0.562	0.528	0.559	0.547	0.530
	Testing data	0.497	0.463	0.482	0.474	0.469
		RLS	CSD-RLS	ES-RLS	HP-RLS	WD-RLS
MAPE	*Training data*	2.917	1.924	2.583	2.491	2.335
	Testing data	3.613	2.906	3.554	3.591	3.427
MSE	*Training data*	3.085	2.964	2.993	2.971	2.980
	Testing data	4.428	3.798	4.126	4.010	3.893
RMSE	*Training data*	3.116	3.099	3.109	3.112	3.097
	Testing data	3.085	3.053	3.063	3.057	3.051
MAD	*Training data*	0.573	0.471	0.495	0.524	0.462
	Testing data	0.482	0.468	0.481	0.478	0.463

Table 2. *Cont.*

		SVR	CSD-SVR	ES-SVR	HP-SVR	WD-SVR
MAPE	*Training data*	2.230	2.151	2.193	2.201	2.164
	Testing data	2.249	2.176	2.231	2.235	2.197
MSE	*Training data*	2.817	2.799	2.806	2.799	2.803
	Testing data	2.839	2.803	2.833	2.831	2.805
RMSE	*Training data*	2.758	2.731	2.749	2.750	2.743
	Testing data	2.673	2.661	2.669	2.671	2.667
MAD	*Training data*	0.569	0.525	0.537	0.561	0.553
	Testing data	0.548	0.479	0.482	0.513	0.480
		GAN	CSD-GAN	ES-GAN	HP-GAN	WD-GAN
MAPE	*Training data*	1.918	1.872	1.899	1.909	1.884
	Testing data	1.925	1.887	1.895	1.921	1.891
MSE	*Training data*	2.525	2.504	2.513	2.520	2.515
	Testing data	2.610	2.585	2.590	2.589	2.591
RMSE	*Training data*	1.923	1.916	1.920	1.919	1.922
	Testing data	1.917	1.915	1.916	1.921	1.920
MAD	*Training data*	0.375	0.356	0.362	0.371	0.359
	Testing data	0.298	0.275	0.287	0.283	0.279

4.3. Diebold–Mariano (DM) Forecast Accuracy Test

The findings of different error methods lead to confirming the higher predictive performance of the CSD-GAN model. To reconfirm the findings, we have used another forecasting accuracy test, which is proposed by [65]. This technique accounts for the non-Gaussian and nonzero-mean, serially correlated for the errors [65]. As the error comparison presented the outperformance of CSD-GAN, we used this as a base and compared it with the ARIMA, RLS, SVR, GAN, CSD-ARIMA, CSD-RLS and CSD-SVR. Where the null hypothesis H_0 indicates that there is no difference between CSD-GAN and the comparative models, H_1 proposes that the output power of CSD-GAN is better. H_2 is about the higher output power of the comparative models, as compared with CSD-GAN. According to [66,67], we have to adopt the *MSE* for the model comparison estimations, whereas S_0 represents the statistics of the Diebold–Mariano test and p_0 is the p-value. We have used the 95% confidence level, which indicates that the $p > 0.05$ leads towards the non-rejection of the null hypothesis. In the case of a $p < 0.05$, we have to choose H_1 or H_2. If the S_0 statistics are negative, we have to accept the H_1; otherwise, H_2 will be accepted.

The results of the Diebold–Mariano test are reported in Table 3, which confirms that the p-value is less than 0.05 for all the comparative models (ARIMA, RLS, SVR, GAN, CSD-ARIMA, CSD-RLS, and CSD-SVR). However, the null hypothesis is rejected. In such a scenario, we have to focus on the Diebold–Mariano statistics (S_0). The statistics are negative for all the comparative models, which affirm the higher predictive power of CSD-GAN.

Table 3. Diebold–Mariano forecast accuracy test.

MSE	**ARIMA**	**RLS**	**SVR**	**GAN**	**CSD-ARIMA**	**CSD-RLS**	**CSD-SVR**
S_0	−71.102	−63.824	−21.453	−49.822	−57.285	−44.101	−75.086
P_0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Result		Reject H_0; Accept H_1					

Notes: H_0 indicates that there is no difference between CSD-GAN and comparative models. H_1 proposes that the output power of CSD-GAN is better. H_2 is about the higher output power of comparative models, as compared with CSD-GAN.

4.4. CSD-GAN-Based Out-of Sample Forecasting of NO_2 and SO_2

After confirmation regarding the superiority of the CSD-GAN model, the original data for NO_2 and SO_2 are used to confirm the validity of the result; we used the chosen technique to forecast the data. In Figure 13, forecasting for NO_2 was achieved where the observed data is presented in blue, and the CSD-GAN-based prediction data is presented in red. The yellow line shows the forecasted data from 16 July to 30 August. It can be seen that the observed and CSD-GAN-based values are very close, showing that CSD-GAN correctly calculated the NO_2. Also, after checking the superiority of CSD-GAN, the forecasting for NO_2 was conducted from July 2020 to August 2020. In July, NO_2 is lower; however, right from the start of August, the concentration of NO_2 rose sharply. Figure 14 demonstrates the forecasting of SO_2 where the black line is observed, the green is the CSD-GAN-based prediction, and the yellow line highlights the forecasted value.

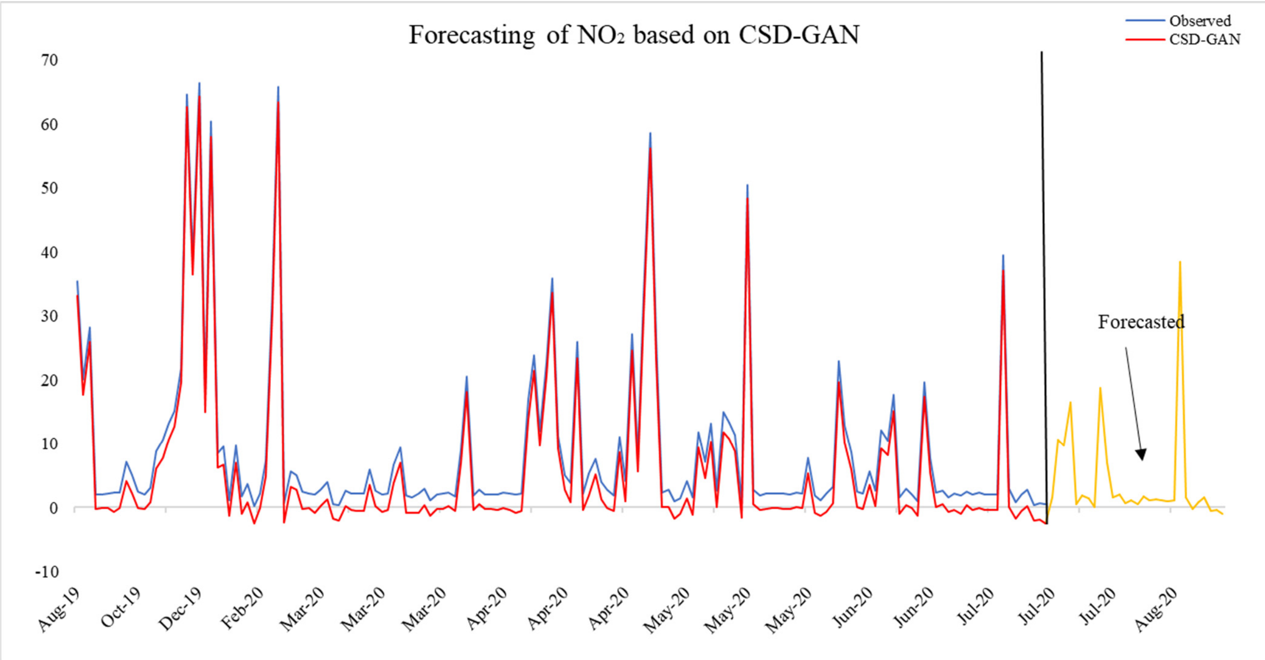

Figure 13. Out-of-sample forecasting of NO_2 by using CSD-GAN.

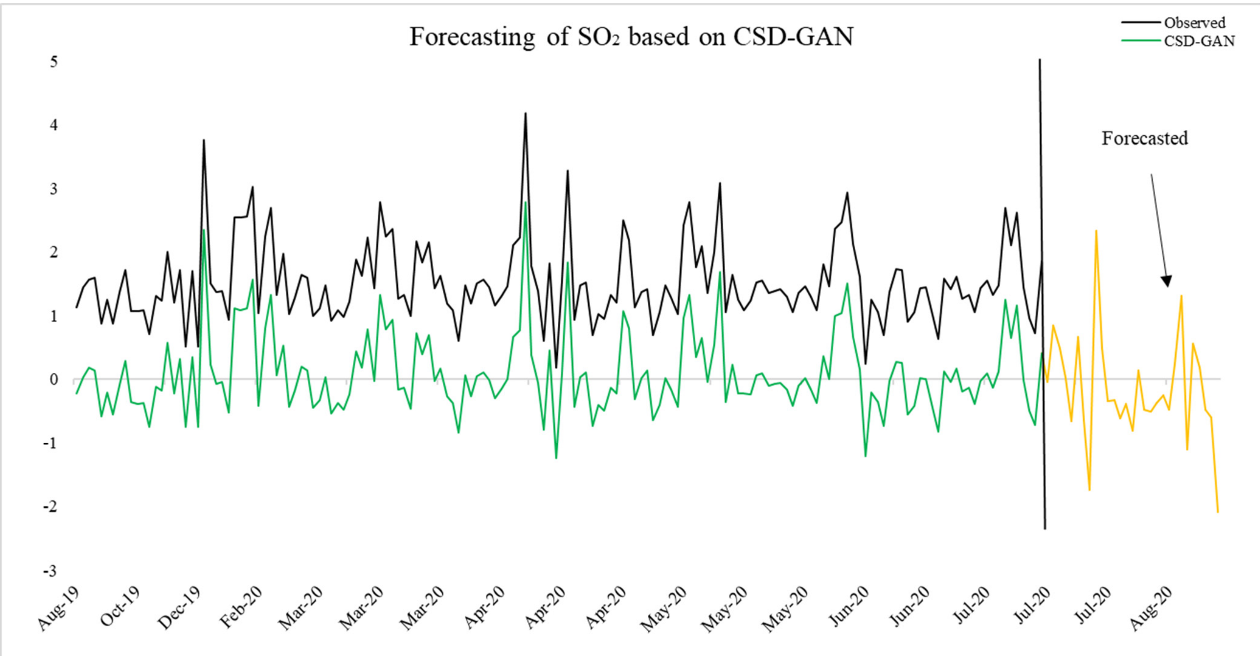

Figure 14. Out-of-sample forecasting of SO_2 by using CSD-GAN.

4.5. Discussion and Summarizations

The analysis regarding the different AI techniques combined with various denoising methods led to various vital points. The first reason that CSD is the superior approach for denoising is that, compared to single models, it has a much-improved capacity to predict future SO_2 and NO_2 concentrations. Second, hybrid models that are based on CSD coupled with AI tools, such as CSD-SVR and CSD-GAN, have a superior performance compared to CSD-RLS, which indicates that AI may successfully describe nonlinear patterns of SO_2 and NO_2. The third and last finding compares several denoising techniques, which demonstrates that CSD is the most effective way for data processing and denoising. These techniques include CSD-GAN, ES-GAN, HP-GAN, and WD-GAN. The fourth and last finding is that CSD-AI models may be the best for making level and directional predictions with different sample sets. This is evidence of their application and consistency. The fifth and last set of results concerns the CSD-based AI learning paradigm's performance in accurately predicting NO_2 and SO_2. When calculating GHG emissions, there is no discrepancy between the observed values and those estimated using CSD-GAN, suggesting that it can adequately anticipate any gas concentration or emission. The findings of higher predictions for denoising-based hybrid models are consistent with [20,68]. Moreover, [43] also confirmed the outperformance of machine learning models in the case of environmental gasses.

5. Conclusions

This study's major purpose is to construct a hybrid model capable of effectively lowering the amount of noise that occurs prior to predictions in order to increase the accuracy of the predictions for SO_2 and NO_2. Combining compressed sensing-based denoising (CSD) with a specialized artificial intelligence (AI) forecasting tool may provide a unique hybrid learning paradigm. CSD is utilized as a preprocessor in this model to obtain clean data from the original NO_2 data by data denoising. To model the clean data and provide the final prediction result, a specialized and potent AI model is applied. SVR and GAN are two examples of these kinds of models. Using NO_2 emission data as sample data, the empirical study reveals that the CSD technique may considerably enhance the forecasting performance of single AI models when utilizing CSD. CSD-AI models surpassed their solo benchmarks in terms of both level and directional predictions, indicating that the CSD technique has the potential to considerably enhance the forecasting performance of single AI models. In terms of both level and directional accuracy, the proposed CSD-GAN outperforms past hybrid models that used more conventional forecasting techniques or other denoising techniques. These models were used to provide forecasts. In addition, the proposed CSD-AI models perform well for SO_2, proving the resilience and generalizability of the novel learning paradigm. This study also reveals that the proposed hybrid CSD-AI model performs very well in predicting NO_2 emissions, a challenging and noisy time series.

This study's results have some practical implications for policymakers concerning atmospheric pollution. First of all, it helps in making an exact prediction of NO_2 and SO_2, which assists in revising the existing policies towards developing proper methods to capture the increased gas emissions. Also, effective measures can be introduced to reduce emissions if the forecast shows that they will increase. In this regard, efficient vehicles can be introduced because gasoline burning is the primary source of NO_2. Additionally, policymakers can forecast other GHG emissions, including CO_2, through this hybrid model. The "hybrid model" is a legitimate choice for policymakers looking to establish air quality and initial warning systems due to its higher performance and prediction capabilities. Also, policymakers can revise the policies regarding control over the usage of fossil fuels by introducing renewable energy sources when they know exactly what the pattern of emissions will be in the future. Through these steps, the harmful impact of these gasses on human health can be controlled.

Similar to previous research, our study encountered limitations that limited our scope. The absence of specific environmental data for Saudi Arabia was a notable constraint.

The limited dataset prevented us from employing a wider range of machine learning tests. The data of different regions of Saudi Arabia is unavailable, which restricts us from cross-regional analysis. These cross-regional analyses are useful for confirming the prediction and forecasting of employed analysis. In the future, we suggest exploring alternative environmental variables for prediction and forecasting. In addition, conducting cross-comparisons with other regions, such as the Gulf Council of Countries (GCC) or other oil-exporting nations, may be beneficial. In addition, future research could contemplate incorporating advanced denoising techniques, especially those based on GARCH (generalized autoregressive conditional heteroskedasticity), in conjunction with machine learning methods.

Author Contributions: S.S. wrote the manuscript. G.A. framed the idea, analyzed the results, and reviewed the manuscript. D.B.-L. supervised the review of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant. No. (UJ-23-SHR-33). Therefore, the authors thank the University of Jeddah for its technical and financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are included in the paper.

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

References

1. Singh, R.L.; Singh, P.K. Global Environmental Problems. In *Principles and Applications of Environmental Biotechnology for a Sustainable Future*; Springer: Singapore, 2016; pp. 13–41. [CrossRef]
2. Baklanov, A.; Molina, L.T.; Gauss, M. Megacities, Air Quality and Climate. *Atmos. Environ.* **2016**, *126*, 235–249. [CrossRef]
3. Moore, M.; Gould, P.; Keary, B.S. Global Urbanization and Impact on Health. *Int. J. Hyg. Environ. Health* **2003**, *206*, 269–278. [CrossRef]
4. Pinault, L.; Crouse, D.; Jerrett, M.; Brauer, M.; Tjepkema, M. Spatial Associations between Socioeconomic Groups and NO2 Air Pollution Exposure within Three Large Canadian Cities. *Environ. Res.* **2016**, *147*, 373–382. [CrossRef]
5. Sonibare, J.A.; Akeredolu, F.A. A Theoretical Prediction of Non-Methane Gaseous Emissions from Natural Gas Combustion. *Energy Policy* **2004**, *32*, 1653–1665. [CrossRef]
6. Turias, I.J.; González, F.J.; Martin, M.L.; Galindo, P.L. Prediction Models of CO, SPM and SO_2 Concentrations in the Campo de Gibraltar Region, Spain: A Multiple Comparison Strategy. *Environ. Monit. Assess.* **2008**, *143*, 131–146. [CrossRef]
7. Wang, P.; Zhang, H.; Qin, Z.; Zhang, G. A Novel Hybrid-Garch Model Based on ARIMA and SVM for PM2.5 Concentrations Forecasting. *Atmos. Pollut. Res.* **2017**, *8*, 850–860. [CrossRef]
8. Pandey, J.S.; Kumar, R.; Devotta, S. Health Risks of NO_2, SPM and SO_2 in Delhi (India). *Atmos. Environ.* **2005**, *39*, 6868–6874. [CrossRef]
9. McKendry, I.G. Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM_{10} and $PM_{2.5}$) Forecasting. *J. Air Waste Manag. Assoc.* **2002**, *52*, 1096–1101. [CrossRef]
10. Dutta, A.; Jinsart, W. Air Pollution in Indian Cities and Comparison of MLR, ANN and CART Models for Predicting PM_{10} Concentrations in Guwahati, India. *Asian J. Atmos. Environ.* **2021**, *15*, 1–26. [CrossRef]
11. Shang, Z.; He, J. Predicting Hourly $PM_{2.5}$ Concentrations Based on Random Forest and Ensemble Neural Network. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 2341–2345. [CrossRef]
12. Bozdağ, A.; Dokuz, Y.; Gökçek, Ö.B. Spatial Prediction of PM10 Concentration Using Machine Learning Algorithms in Ankara, Turkey. *Environ. Pollut.* **2020**, *263*, 114635. [CrossRef]
13. Tripathi, A.K.; Sharma, K.; Bala, M. A Novel Clustering Method Using Enhanced Grey Wolf Optimizer and MapReduce. *Big Data Res.* **2018**, *14*, 93–100. [CrossRef]
14. Wang, P.; Liu, Y.; Qin, Z.; Zhang, G. A Novel Hybrid Forecasting Model for PM10 and SO_2 Daily Concentrations. *Sci. Total Environ.* **2015**, *505*, 1202–1212. [CrossRef]
15. Wang, J.; Bai, L.; Wang, S.; Wang, C. Research and Application of the Hybrid Forecasting Model Based on Secondary Denoising and Multi-Objective Optimization for Air Pollution Early Warning System. *J. Clean. Prod.* **2019**, *234*, 54–70. [CrossRef]
16. Sang, Y.F.; Wang, D.; Wu, J.C.; Zhu, Q.P.; Wang, L. Entropy-Based Wavelet de-Noising Method for Time Series Analysis. *Entropy* **2009**, *11*, 1123–1147. [CrossRef]

17. Niu, L.; Shi, Y. A Hybrid Slantlet Denoising Least Squares Support Vector Regression Model for Exchange Rate Prediction. *Procedia Comput. Sci.* **2010**, *1*, 2397–2405. [CrossRef]
18. de Faria, E.L.; Albuquerque, M.P.; Gonzalez, J.L.; Cavalcante, J.T.P.; Albuquerque, M.P. Predicting the Brazilian Stock Market through Neural Networks and Adaptive Exponential Smoothing Methods. *Expert Syst. Appl.* **2009**, *36*, 12506–12509. [CrossRef]
19. Yuan, C. Forecasting Exchange Rates: The Multi-State Markov-Switching Model with Smoothing. *Int. Rev. Econ. Financ.* **2011**, *20*, 342–362. [CrossRef]
20. Nasseri, M.; Moeini, A.; Tabesh, M. Forecasting Monthly Urban Water Demand Using Extended Kalman Filter and Genetic Programming. *Expert Syst. Appl.* **2011**, *38*, 7387–7395. [CrossRef]
21. Chen, B.T.; Chen, M.Y.; Fan, M.H.; Chen, C.C. Forecasting Stock Price Based on Fuzzy Time-Series with Equal-Frequency Partitioning and Fast Fourier Transform Algorithm. In Proceedings of the 2012 Computing, Communications and Applications Conference, Hong Kong, China, 1–13 January 2012; pp. 238–243. [CrossRef]
22. He, K.; Lai, K.K.; Xiang, G. Portfolio Value at Risk Estimate for Crude Oil Markets: A Multivariatewavelet Denoising Approach. *Energies* **2012**, *5*, 1018–1043. [CrossRef]
23. Sang, Y.F. Improved Wavelet Modeling Framework for Hydrologic Time Series Forecasting. *Water Resour. Manag.* **2013**, *27*, 2807–2821. [CrossRef]
24. Gardner, E.S. Exponential Smoothing: The State of the Art. *J. Forecast.* **1985**, *4*, 1–28. [CrossRef]
25. Hodrick, R.J.; Prescott, E.C. *Postwar U.S. Business Cycles: An Empirical Investigation*; Ohio State University Press: Columbus, OH, USA, 1997; Volume 29, pp. 1–16. Available online: http://www.jstor.org/stable/2953682 (accessed on 5 February 2023).
26. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Fluids Eng. Trans. ASME* **1960**, *82*, 35–45. [CrossRef]
27. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete Cosine Transform. *IEEE Trans. Comput.* **1974**, *100*, 90–93. [CrossRef]
28. Mallat, S.G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 79–85. [CrossRef]
29. Zhu, L.; Zhu, Y.; Mao, H.; Gu, M. A New Method for Sparse Signal Denoising Based on Compressed Sensing. In Proceedings of the 2009 Second International Symposium on Knowledge Acquisition and Modeling, Wuhan, China, 30 November–1 December 2009; pp. 35–38. [CrossRef]
30. Han, B.; Xiong, J.; Li, L.; Yang, J.; Wang, Z. Research on Millimeter-Wave Image Denoising Method Based on Contourlet and Compressed Sensing. In Proceedings of the 2010 2nd International Conference on Signal Processing Systems, Dalian, China, 5–7 July 2010.
31. Sharma, A.; Massey, D.D.; Taneja, A. A Study of Horizontal Distribution Pattern of Particulate and Gaseous Pollutants Based on Ambient Monitoring near a Busy Highway. *Urban Clim.* **2018**, *24*, 643–656. [CrossRef]
32. Li, R.; Cui, L.; Liang, J.; Zhao, Y.; Zhang, Z.; Fu, H. Estimating Historical SO_2 Level across the Whole China during 1973–2014 Using Random Forest Model. *Chemosphere* **2020**, *247*, 125839. [CrossRef]
33. Sheng, T.; Pan, J.; Duan, Y.; Liu, Q.; Fu, Q. Study on Characteristics of Typical Traffic Environment Air Pollution in Shanghai. *China Environ. Sci.* **2019**, *39*, 3193–3200.
34. Wu, L.; Noels, L. Recurrent Neural Networks (RNNs) with Dimensionality Reduction and Break down in Computational Mechanics; Application to Multi-Scale Localization Step. *Comput. Methods Appl. Mech. Eng.* **2022**, *390*, 114476. [CrossRef]
35. Wu, C.-L.; He, H.-D.; Song, R.-F.; Peng, Z.-R. Prediction of Air Pollutants on Roadside of the Elevated Roads with Combination of Pollutants Periodicity and Deep Learning Method. *Build. Environ.* **2022**, *207*, 108436. [CrossRef]
36. Du, W.; Chen, L.; Wang, H.; Shan, Z.; Zhou, Z.; Li, W.; Wang, Y. Deciphering Urban Traffic Impacts on Air Quality by Deep Learning and Emission Inventory. *J. Environ. Sci.* **2023**, *124*, 745–757. [CrossRef]
37. Kurnaz, G.; Demir, A.S. Prediction of SO_2 and PM10 Air Pollutants Using a Deep Learning-Based Recurrent Neural Network: Case of Industrial City Sakarya. *Urban Clim.* **2022**, *41*, 101051. [CrossRef]
38. Aceves-Fernández, M.A.; Domínguez-Guevara, R.; Pedraza-Ortega, J.C.; Vargas-Soto, J.E. Evaluation of Key Parameters Using Deep Convolutional Neural Networks for Airborne Pollution (PM10) Prediction. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 2792481. [CrossRef]
39. Atamaleki, A.; Motesaddi Zarandi, S.; Fakhri, Y.; Abouee Mehrizi, E.; Hesam, G.; Faramarzi, M.; Darbandi, M. Estimation of Air Pollutants Emission (PM_{10}, CO, SO_2 and NO_x) during Development of the Industry Using AUSTAL 2000 Model: A New Method for Sustainable Development. *MethodsX* **2019**, *6*, 1581–1590. [CrossRef] [PubMed]
40. Perez, P.; Menares, C.; Ramírez, C. $PM_{2.5}$ Forecasting in Coyhaique, the Most Polluted City in the Americas. *Urban Clim.* **2020**, *32*, 100608. [CrossRef]
41. Janarthanan, R.; Partheeban, P.; Somasundaram, K.; Navin Elamparithi, P. A Deep Learning Approach for Prediction of Air Quality Index in a Metropolitan City. *Sustain. Cities Soc.* **2021**, *67*, 102720. [CrossRef]
42. Al-Janabi, S.; Mohammad, M.; Al-Sultan, A. A New Method for Prediction of Air Pollution Based on Intelligent Computation. *Soft Comput.* **2020**, *24*, 661–680. [CrossRef]
43. Al Dakheel, J.; Del Pero, C.; Aste, N.; Leonforte, F. Smart Buildings Features and Key Performance Indicators: A Review. *Sustain. Cities Soc.* **2020**, *61*, 102328. [CrossRef]
44. Aggarwal, A.; Toshniwal, D. A Hybrid Deep Learning Framework for Urban Air Quality Forecasting. *J. Clean. Prod.* **2021**, *329*, 129660. [CrossRef]
45. Chiang, P.W.; Horng, S.J. Hybrid Time-Series Framework for Daily-Based $PM_{2.5}$ Forecasting. *IEEE Access* **2021**, *9*, 104162–104176. [CrossRef]

46. Du, S.; Li, T.; Yang, Y.; Horng, S.J. Multivariate Time Series Forecasting via Attention-Based Encoder–Decoder Framework. *Neurocomputing* **2020**, *388*, 269–279. [CrossRef]
47. Du, S.; Li, T.; Gong, X.; Horng, S.J. A Hybrid Method for Traffic Flow Forecasting Using Multimodal Deep Learning. *Int. J. Comput. Intell. Syst.* **2020**, *13*, 85–97. [CrossRef]
48. Du, S.; Li, T.; Yang, Y.; Horng, S.J. Deep Air Quality Forecasting Using Hybrid Deep Learning Framework. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 2412–2424. [CrossRef]
49. Elder, Y.; Kutyniok, G. *Compressed Sensing (Theory and Applications)*; Cambridge University Press: Cambridge, UK, 2012.
50. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2000. [CrossRef]
51. Yin, T.; Wang, Y. Predicting the Price of WTI Crude Oil Futures Using Artificial Intelligence Model with Chaos. *Fuel* **2022**, *316*, 122523. [CrossRef]
52. Broock, W.A.; Scheinkman, J.A.; Dechert, W.D.; LeBaron, B. A Test for Independence Based on the Correlation Dimension. *Econom. Rev.* **1996**, *15*, 197–235. [CrossRef]
53. Zagajewski, B.; Kluczek, M.; Raczko, E.; Njegovec, A.; Dabija, A.; Kycko, M. Comparison of Random Forest, Support Vector Machines, and Neural Networks for Post-Disaster Forest Species Mapping of the Krkonoše/Karkonosze Transboundary Biosphere Reserve. *Remote Sens.* **2021**, *13*, 2581. [CrossRef]
54. Dou, Z.; Sun, Y.; Zhu, J.; Zhou, Z. The Evaluation Prediction System for Urban Advanced Manufacturing Development. *Systems* **2023**, *11*, 392. [CrossRef]
55. Yang, X.; Tan, L.; He, L. A Robust Least Squares Support Vector Machine for Regression and Classification with Noise. *Neurocomputing* **2014**, *140*, 41–52. [CrossRef]
56. Balabin, R.M.; Lomakina, E.I. Support Vector Machine Regression (SVR/LS-SVM)—An Alternative to Neural Networks (ANN) for Analytical Chemistry? Comparison of Nonlinear Methods on near Infrared (NIR) Spectroscopy Data. *Analyst* **2011**, *136*, 1703–1712. [CrossRef]
57. Aggarwal, A.; Mittal, M.; Battineni, G. Generative Adversarial Network: An Overview of Theory and Applications. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100004. [CrossRef]
58. Sahoo, L.; Praharaj, B.B.; Sahoo, M.K. Air Quality Prediction Using Artificial Neural Network. *Adv. Intell. Syst. Comput.* **2021**, *1248*, 31–37. [CrossRef]
59. Shams, S.R.; Jahani, A.; Kalantary, S.; Moeinaddini, M.; Khorasani, N. The Evaluation on Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) Models for Predicting SO_2 Concentration. *Urban Clim.* **2021**, *37*, 100837. [CrossRef]
60. Bowerman, B.L.; O'Connell, R.T.; Koehler, A.B. *Forecasting, Time Series, and Regression: An Applied Approach*; Thomson Brooks/Cole Publishing: Pacific Grove, CA, USA, 2005.
61. Baxter, M.; King, R.G. Approximate Band-Pass Filters for Economic Time Series. *NBER Work. Pap. Ser.* **1995**, *5022*, 1–53.
62. Stoffer, D.S.; Shumway, R.H. An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm. *J. Time Ser. Anal.* **1982**, *3*, 253–264.
63. Struzik, Z.R. Wavelet Methods in (Financial) Time-Series Processing. *Phys. A Stat. Mech. Its Appl.* **2001**, *296*, 307–319. [CrossRef]
64. Donoho, D.L. De-Noising by Modified Soft-Thresholding. *IEEE Asia-Pacific Conf. Circuits Syst.-Proc.* **2000**, *41*, 760–762. [CrossRef]
65. Diebold, F.; Mariano, R. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
66. Hornik, K.; Stinchcombe, M.; White, H. *Presentation on Multilayer Feedforward Networks Are Universal Approximators*; Elsevier: Amsterdam, The Netherlands, 1989.
67. Harvey, D.; Leybourne, S.; Newbold, P. Testing the Equality of Prediction Mean Squared Errors. *Int. J. Forecast.* **1997**, *13*, 281–291. [CrossRef]
68. Yu, L.; Zhao, Y.; Tang, L. A Compressed Sensing Based AI Learning Paradigm for Crude Oil Price Forecasting. *Energy Econ.* **2014**, *46*, 236–245. [CrossRef]