atmosphere

MDPI

Article

Evolving Hybrid Generalized Space-Time Autoregressive Forecasting with Cascade Neural Network Particle Swarm Optimization

Toni Toharudin [1,*], Rezzy Eko Caraka [1,2,*], Hasbi Yasin [3] and Bens Pardamean [4,5]

1 Department of Statistics, Padjadjaran University, Bandung 45363, West Java, Indonesia
2 Research Center for Data and Information Sciences, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Bandung 40135, West Java, Indonesia
3 Department of Statistics, Diponegoro University, Semarang 50275, Central Java, Indonesia; hasbiyasin@live.undip.ac.id
4 Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Palmerah, Indonesia; bpardamean@binus.edu
5 Computer Science Department, BINUS Graduate Program—Master of Computer Science Program, Bina Nusantara University, Jakarta 11480, Palmerah, Indonesia
* Correspondence: toni.toharudin@unpad.ac.id (T.T.); rezzy.eko.caraka@brin.go.id or r.eko.caraka@unpad.ac.id (R.E.C.)

Abstract: Background: The generalized space-time autoregressive (GSTAR) model is one of the most widely used models for modeling and forecasting time series and location data. Methods: In the GSTAR model, there is an assumption that the research locations are heterogeneous. In addition, the differences between these locations are shown in the form of a weighting matrix. The novelty of this paper is that we propose the hybrid time-series model of GSTAR uses the cascade neural network and obtains the best parameters from particle swarm optimization. Results and conclusion: This hybrid model provides a high accuracy value for forecasting $PM_{2.5}$, PM_{10}, NO_x, and SO_2 with high accuracy forecasting, which is justified by a mean absolute percentage error (MAPE) accuracy of around 0.01%.

Keywords: hybrid; time series; cascade; neural network; space-time; pollution

Citation: Toharudin, T.; Caraka, R.E.; Yasin, H.; Pardamean, B. Evolving Hybrid Generalized Space-Time Autoregressive Forecasting with Cascade Neural Network Particle Swarm Optimization. *Atmosphere* **2022**, *13*, 875. https://doi.org/10.3390/atmos13060875

Academic Editor: Eduardo García-Ortega

Received: 24 April 2022
Accepted: 24 May 2022
Published: 27 May 2022

1. Introduction

Many countries experienced an increased rate of urbanization and economic growth in the past few decades, leading to an increase in population, resource consumption, and environmental pollution. Air pollutants are a mixture of particulate matter, including $PM_{2.5}$ and PM_{10}, and gas-phase pollutants, including NO_2, CO, O_3, and SO_2, which have been shown to start reducing visibility and cause respiratory, cardiovascular, and immune system disorders, as well as mental disorders [1,2]. The World Health Organization (WHO) classifies some particulate matter as a class 1 cancer-causing agent, a designation given when there is sufficient evidence of carcinogenic effects [3]. The importance of data science in the environment can be beneficial in supporting national health risk management and environmental planning [4].

Precise forecasting is needed for decision-making. In the big data era paradigm, data availability is very abundant, diverse, and complex, and a high volume of data is accompanied by high data speed [5–7]. The motivation to provide high forecasting results has also been described in M-Competitions [6–8]. These M-Competitions are available studies that have compared the efficiency of a huge series of prominent forecasting models using acknowledged scientists who provide estimations of their knowledge and experience, using methods that include the application of recent statistics and parametric, semi-parametric, machine learning, deep learning, and expert systems [9–14].

Several studies that used hybrid time series by combining neural networks with traditional time series proved more beneficial for reasoning, such as the roll-motion Arima-neural networks [15], hybrid VAR-SVR [16], SARIMA-ANN [17], SARIMA-CNN-LSTM [18], and fuzzy neural network [19]. Meanwhile, the challenges in forecasting neural networks include determining the right pre-processing method and the appropriate neural network architecture, and finding the optimal parameter values. Thus, the most reasonable method during the existence of the time series model is to conduct pre-processing or inputting using the partial autocorrelation function (PACF) by justifying the correlation of a stationary time series with its own lagged values [20–22].

The neural network architecture is made of individual units called neurons that mimic the biological behavior of the brain. Regarding the neural network architecture, it can be used to develop a feed-forward neural network [22–28], radial basis neural network [29,30], deep feed-forward neural network [26], recurrent neural network [31], long-short term memory [32,33], gated recurrent unit (GRU) neural network [34], and much more. Optimization techniques also have an essential role in obtaining optimal machine learning parameters. Many types of optimization are used, such as a genetic algorithm (GA) nearest neighbor search (NNS), or simulated annealing (SA). In addition, metaheuristics, such as a tabu search (TS), can reduce the risk of falling into the local optimal according to the calculation used or by adjusting the parameter value. Using a metaheuristic makes it possible to increase the diversity of the population, thereby reducing the risk of falling into the local optimum. In the case of particle swarm optimization (PSO), the uniform arrangement of particles initially located in the solution space is of considerable value to the inertia constant and to the randomness added to the acceleration stability. The risk of falling into the local optimum can be reduced through the size of the swarm. Since PSO swarms solutions with slightly different directions of improvement (particle), it cannot be called a complete greedy hill-climbing approach with individual particles. Several studies have also conducted neural networks with PSO that provided excellent accuracy, such as the PSO-enabled LSTM neural network [35], artificial neural network, and PSO algorithm [36–38].

In this paper, we enhance the cascade neural network with particle swarm optimization (PSO). PSO improves the solutions in pbest and gbest, but it is not known if they add randomness through random numbers. It can be said that the risk is inherent in the algorithm itself. In this study, to solve the early convergence of the basic PSO, the swarm is divided into positive and negative swarms, and the positive swarm improves the solution in the direction of the basic PSO. The particle swarm moves the solution position in the opposite direction of the fitness value (pbest) and best value of any particle in its neighborhood (gbest). A bidirectional repositioning strategy was used. Particles belonging to the swarm played a role in improving the solution by following pbest and gbest over generations, similar to players in the existing PSO. Particles belonging to the particle swarm moving in the opposite direction were, until then, moving in the same direction and speed as the swarm. At a certain point in time, they start to move in the opposite direction to the pbest and gbest to increase the diversity of the entire swarm. To spread the particles evenly and radially in the multidimensional space centered on the negative swarm gbest, it is appropriate to use the time when it is judged that the local optimum has been reached to some extent, such as when the improvement of the solution is not made for several generations, as the detonation time. This article is organized as follows; in Section 2, we review the multivariate time series VAR and GSTAR and training using cascade neural network, and employ the model using PSO. Their performance is examined in Section 3 via simulation studies and four accurate air pollution data benchmark datasets. Finally, Section 4 presents the summary and conclusion.

2. Methodology

Multivariate Time Series VAR and GSTAR

Time series analysis is used when the data used are time-bound. Therefore, there is a correlation between the current event data and data from one previous period [5]. It means that events also influence current events in an earlier period [6,7]. Time series analysis that uses only one variable is called a univariate time series. In everyday life, a variable is often related to several other variables and when conducting research, more than one variable is involved [39–41]. Analysis using many variables is called a multivariate time series. Multivariate time series analysis is generally used to model and explain the interactions and movements between several time series variables [42].

A multivariate time series is a series of data consisting of several variables taken from time to time and recorded sequentially according to the time of occurrence with fixed time intervals. Time intervals include seconds, minutes, hours, days, months, years, and other periods. Similar to univariate time series, the stationarity of multivariate time series data can also be justified from the matrix autocorrelation function (MACF), and matrix partial autocorrelation function (MPACF) plots [43,44]. The MACF plot that decreases slowly indicates that the data do not have a stationary average; thus, it is necessary to perform differencing to stabilize the data. Likewise, when the information is not stationary in the variance, it must transform to obtain stationary data. If there is a time series vector with n observations, that is, $Z_1, Z_2, \ldots, Z_n$, then the sample correlation matrix equation is as follows:

$$\hat{\rho}(k) = \left[\hat{\rho}_{ij}(k)\right] \tag{1}$$

where $\hat{\rho}_{ij}(k)$ is the sample cross-correlation for the $i-th$ and $j-th$ series components, which are expressed in the following equation:

$$\hat{\rho_{ij}}(k) = \frac{\sum_{t=1}^{n-k}\left(Z_{i,t} - \overline{Z_i}\right)\left(Z_{j,t+k} - Z_j\right)}{\left[\sum_{t=1}^{n}\left(Z_{i,t} - \overline{Z_i}\right)^2 \sum_{t=1}^{n}\left(Z_{j,t} - \overline{Z_j}\right)^2\right]^{\frac{1}{2}}} \tag{2}$$

Meanwhile, $\overline{Z}_i$ and $\overline{Z}_j$ are the sample means of the corresponding series components. In line with this, the matrices and graphs become more complex as the dimensions and vectors become more extensive, making identification difficult. To overcome this, [45] introduced a suitable method for summarizing the explanation of sample correlation, namely, by using symbols denoted by (+), (−), and (.) in the sample correlation matrix for $(i)and(j)$. The standard error (SE) of a statistic is the standard deviation of its sampling distribution or an estimate of that standard deviation.

- The symbol (+) is defined as $\hat{\rho}_{ij}(k)$ greater than two times the standard error and indicates the relationship has a positive correlation.
- The symbol $(-)$ represents a value of $\hat{\rho}_{ij}(k)$ less than -2 times the standard error. It indicates that the relationship has a negative correlation.
- The symbol (.) denotes $\hat{\rho}_{ij}(k)$, which is between ± 2 times the standard error and indicates no correlation.

The field carried out the generalization of the PACF concept into a time series vector [45] and defined a partial autocorrelation matrix at lag s with the notation $\mathcal{P}(s)$. The equation for the partial autocorrelation matrix is as follows:

$$\mathcal{P}(s) = \begin{cases} \Gamma'(1)[\Gamma(0)]^{-1}, s = 1 \\ \left(\Gamma'(s) - c'(s)[A(s)]^{-1}b(s)\right)\left(\Gamma(0) - b'(s)[A(s)]^{-1}b(s)\right)^{-1}, s > 1 \end{cases} \tag{3}$$

For $s \geq 2$, then the value of $A(s), b(s), c(s)$ can be expressed as follows:

$$A(s) = \begin{bmatrix} \Gamma(0) & \Gamma'(1) & \dots & \Gamma'(s-2) \\ \Gamma(1) & \Gamma(0) & \cdots & \Gamma'(s-3) \\ \vdots & \vdots & & \vdots \\ \Gamma(s-2) & \Gamma(s-3) & \cdots & \Gamma(0) \end{bmatrix}, b(s) = \begin{bmatrix} \Gamma'(s-1) \\ \Gamma'(s-2) \\ \vdots \\ \Gamma'(1) \end{bmatrix}, c(s) = \begin{matrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(s-1) \end{matrix} \quad (4)$$

where $\Gamma(s)$ is the s-lag covariance matrix. The sample estimate of $\mathcal{P}(s)$ can be calculated by $\Gamma(s)$, replacing the unknown with the sample covariance matrix $\Gamma(s)$:

$$\hat{\Gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}\left(Z_t - \overline{Z}\right)\left(\left(Z_{t+s} - \overline{Z}\right)'\right), s = 1, 2, \dots \quad (5)$$

where $\overline{Z} = \left(\overline{Z_1}, \overline{Z_2}, \overline{Z_3}, \overline{Z_4}\right)$ is the sample mean vector. The data identification based on MPACF values is also denoted in the form of (+), (−), and (.) as in the MACF. Like the partial autocorrelation equation in the univariate case, the partial autocorrelation matrix equation $\mathcal{P}(s)$ also has a cut-off property for the AR process.

3. The Step Construction Cascade Neural Network with Particle Swarm Optimization

The basic idea is to imitate nature-inspired optimization and social behavior patterns such as birds and fish acting in colonies for self-defense against natural enemies or bees and ants acting in groups to find food. When a colony is formed, individuals can effectively achieve their goals through cooperation. The particle swarm optimization (PSO) was first introduced by [46]. PSO is a population-based optimization technique similar to the genetic algorithm developed by Holland in 1975. The genetic algorithm imitates the evolutionary mechanism of natural selection, whereas PSO reflects the social behavior patterns of biological communities such as flocks of birds and fish [47–50].

The particle swarm optimization (PSO) is a tool to solve the optimization problem and imitates the social behavior of these organisms. However, PSO is one of the most widely used optimization techniques to solve several issues looking for the optimal point and solution of the function. In PSO, individual latent solutions are represented by particles, which is explained in Figure 1. If the number of decision variables is m, the search space in PSO becomes the real space of m and the $j-$th particle at time t. In addition, $x_j(t)$ is expressed as a real vector of parent $x_j(t) = x_{j1}(t), x_{j2}(t), \dots, x_{jm}(t)$. Meanwhile, a 'swarm' is a set of these particles. In addition, t can be expressed as $S(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$. Each particle frame has the properties of the current position and 'velocity'. In the iterative process, particles move through the multi-dimensional search space and have the fitness function value evaluated by the fitness function. Based on this, they move to a location with a higher fitness function value. At this time, the particles lock the two optimal positions and move to the next position. The optimal solution (pbest) is found through the iteration of each particle. The overall optimal solution (gbest) found through iteration of all leaflets is the $j-$th particle at time t. Equation (6) explains the velocity of PSO.

$$v_j(t+1) = w(t)\cdot v_j(t) + c_1\cdot r_1\cdot\left(p_{best_j}(t) - x_j(t)\right) + c_2\cdot r_2\cdot\left(g_{best}(t) - x_j(t)\right) \quad (6)$$

Here, $w(t)$ is the inertial load for the previous speed. Meanwhile, C_1, C_2 is the acceleration constant, and r_1, r_2 is a random number of $(0, 1)$. Additionally, it is required to select an appropriate value. The GSTAR model is a specific vector autoregressive (VAR) model. This model shows a linear dependence of location and time, where the main difference with VAR is in the spatial support; in the GSTAR model, it is shown in a weighted matrix [22,51,52]. The GSTAR model also generalizes the space-time autoregressive (STAR) model [53]. The fundamental difference between the GSTAR and STAR models lies in assuming the parameters [9,54]. The STAR model assumes the locations used in the study are the same, thus this model can only be applied to uniform locations. Meanwhile, in the GSTAR model, there is

an assumption that the research locations are heterogeneous [43,51,52]. In line with this, Equation (7) expresses the GSTAR model with p autoregressive orders and $\lambda_1, \lambda_2, \ldots, \lambda_p$ spatial order, as well as GSTAR $(p\ \lambda_1, \lambda_2, \ldots, \lambda_p)$.

$$Z(t) = \sum_{s=1}^{p} \left[\phi_{s0} + \sum_{k=1}^{\lambda s} \phi_{sk} W^{(k)} \right] Z(t-s) + e_t \tag{7}$$

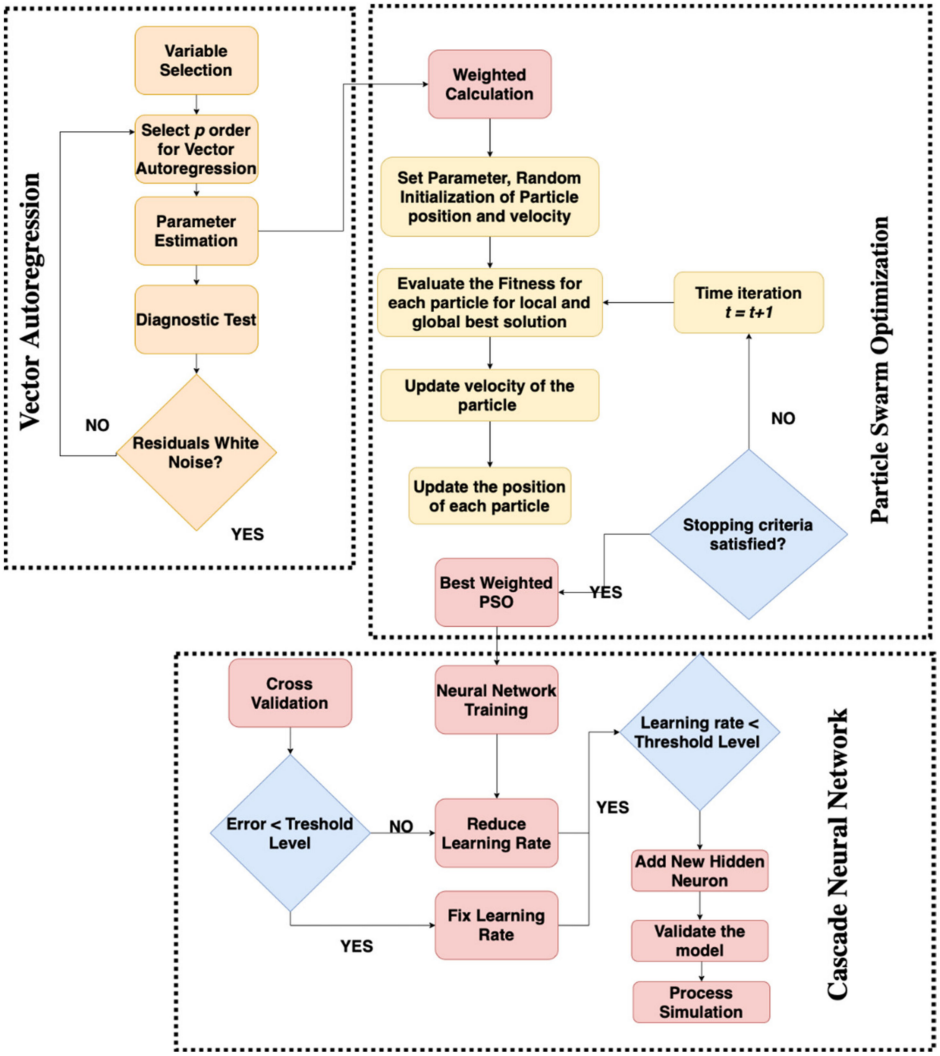

Figure 1. Flowchart VAR-CASCADE NN PSO.

However, ϕ_{s0} is the autoregressive parameter $s = 1, 2, \ldots, p$; ϕ_{sk} is the regression spatial parameter $k = 1, 2, \ldots, \lambda_s$; $\phi_{s0} = \text{diag}\left(\phi'_{s0}, \ldots, \phi^{N}_{s0}\right)$; and $\phi_{sk} = \text{diag}\left(\phi'_{sk}, \ldots, \phi^{N}_{sk}\right)$. Meanwhile, the weighted matrix $N \times N$ can be expressed as W^k with the weighted value selected to meet $w^{(k)}_{ii} = 0$ and $\sum_{i \neq j} w^{(k)}_{ii} = 1$, $i = 1, 2, \ldots, N$. If the spatial order is $k = 1$, then the GSTAR model formed is GSTAR(p1) with the following formula:

$$Z(t) = \sum_{s=1}^{p} [\phi_{s0} + \phi_{s1} W] Z_{t-s} + e_t \tag{8}$$

GSTAR(1,1)

$$Z(t) = [\phi_{s0} + \phi_{s1} W] Z_{t-1} + e_t \tag{9}$$

GSTAR(2,1)

$$Z(t) = \sum_{s=1}^{2} [\phi_{s0} + \phi_{s1} W] Z_{t-s} + e_t$$
$$Z(t) = [\phi_{10} + \phi_{11} W] Z_{t-1} + [\phi_{20} + \phi_{21} W] Z_{t-s} + e_t \quad (10)$$

The study areas included the cities of Taipei, Hsinchu, Taichung, and Kaohsiung, which consist of pollution data, namely, nitrogen oxide (NO_x), atmospheric particulate matter ($PM_{2,5}$), atmospheric particulate matter (PM_{10}), and sulfur dioxide (SO_2), respectively.

4. Results and Discussion

This paper is an extended version of previous research that uses a hybrid VAR neural network genetic algorithm [9] and cascade neural network genetic algorithm [53]. In Table 1, we made a comparison of accuracy using RMSE, MAE, and SMAPE. We decided to use SMAPE as the basis for selecting the best model because this technique can be used to evaluate the diversity between training and testing in more detail. If the cascade neural network method is used for VAR, it provides a speedy computing time compared to GSTAR. This could be due to the complexity of the weights obtained between the different VAR and GSTAR methods. However, we can use modeling for PM_{10} (GSTAR(1)-CFNN-PSO), $PM_{2.5}$ (GSTAR(2)-CFNN-PSO), NO_x (GSTAR(2)-CFNN-PSO), and SO_2 (GSTAR(2)-CFNN-PSO). VAR computing time is faster than GSTAR because the parameters in VAR are fewer than in GSTAR. Nevertheless, in GSTAR we used weighted information for modeling. In line with this, the time-lapse computing of GSTAR should be more complicated than of VAR. The evaluation function of PSO was justified by mean square error. In Figure 2, the variables used as inputs are NOx, $PM_{2.5}$, PM_{10}, and SO_2. Each of these variables already included spatial location. Then, before the initiation using PSO, an analysis using the sigmoid activation function was carried out; after obtaining the best weight, it was retrained with a linear activation function so that forecasting could be performed for each space-time variable.

Table 1. Model Comparison based on Pollution.

GSTAR CFNN	Dataset	FVAL	OBJ	Training			Testing			Average			Elapsed Time (Seconds)
				RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	
GSTAR (1)	NO_X	0.7470	0.7470	5.2660	3.2622	0.0049	4.7143	2.4282	0.0268	4.9902	2.8452	0.0159	229.0739
	$PM_{2.5}$	1.1580	1.1580	9.2831	6.5054	0.0041	6.7955	4.9921	0.0205	8.0393	5.7488	0.0123	91.236504
	PM_{10}	0.735571	0.7356	13.0906	9.1146	0.0038	9.2994	6.5805	0.016	11.1950	7.8476	0.0099 *	90.51314
	SO_2	2.643	2.643	1.5409	1.0592	0.0057	0.9035	0.6893	0.0265	1.2222	0.8743	0.0161	95.544592
GSTAR (2)	NO_X	0.715855	0.7159	5.203	3.2747	0.005	3.7358	2.4901	0.0239	4.4694	2.8824	0.0145 *	106.778741
	$PM_{2.5}$	2.02728	2.0273	9.2237	6.4361	0.0042	6.7847	4.9562	0.0199	8.0042	5.6962	0.0121 *	120.178632
	PM_{10}	2.4187	2.4187	12.9489	8.9896	0.0036	9.1351	6.4255	0.0167	11.0420	7.7076	0.0102	196.033143
	SO_2	7.79396	7.794	1.5172	1.0465	0.0056	0.8858	0.6726	0.0257	1.2015	0.8596	0.0157 *	206.967525
VAR (1)	NO_X	3.20116	3.2012	5.1847	3.2426	13.5834	3.7378	2.4028	12.4158	4.4613	2.8227	12.9996	83.651723
	$PM_{2.5}$	5.99995	6	9.2164	6.4473	7.5802	6.8311	5.0247	6.8707	8.0238	5.7360	7.2255	83.842196
	PM_{10}	3.31837	3.3184	12.9178	8.9778	7.23	9.2731	6.5761	8.4924	11.0955	7.7770	7.8612	84.947817
	SO_2	2.65504	2.655	1.5349	1.0506	6.8211	0.9185	0.7007	6.4913	1.2267	0.8757	6.6562	75.589595
VAR (2)	NO_X	7.50572	7.5057	5.0466	3.1183	12.0234	3.6744	2.3254	11.8713	4.3605	2.7219	11.9474	82.196178
	$PM_{2.5}$	5.31512	5.3151	9.1431	6.3854	7.8689	6.777	5.0306	7.2665	7.9601	5.7080	7.5677	269.645339
	PM_{10}	3.72273	3.7227	12.7213	8.7579	6.9237	9.0049	6.3091	7.9569	10.8631	7.5335	7.4403	105.356041
	SO_2	5.47567	5.4757	1.493	1.027	6.7956	0.9068	0.6823	7.1553	1.1999	0.8547	6.9755	161.550041

* Best model.

The cascade neural network is a pyramid-like cascade of basic FFNN units that is modifiable. Each neural unit has to have a simple network architecture to handle only one task at hand, including data transformation, data integration, or prediction output. A cascade neural network model, in concept, could host vast volumes of data concurrently. Though a single input has poor prediction accuracy, a considered reliable prediction is assured if oriented data is given. The set of input hyperparameters is determined by the ability to inspect for the best model.

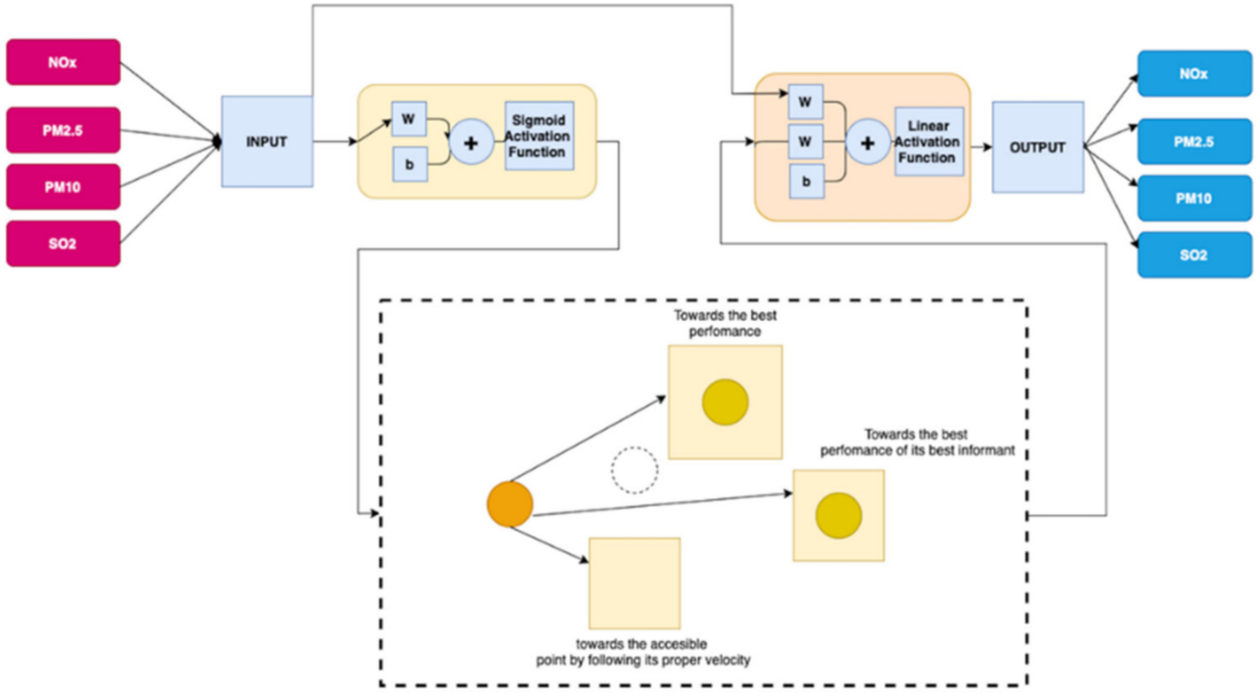

Figure 2. Space-time cascade neural network with particle swarm optimization.

In addition to the time dimension, time series can also have space and time or space-time dimensions. The space-time model is a model that combines elements of time and location dependencies in multivariate time series data, for example, the hourly level of air pollution at several observation points in one area. The generalized space-time autoregressive (GSTAR) model is a model that can overcome the assumption that research locations are heterogeneous; thus, the parameters for each location are different, which is represented in Figures 3–6. Based on the best model we forecasted for the next 30 days, it can be seen in Figure 7 that in Taiwan, Taichung city has the most significant values for $PM_{2.5}$, PM_{10}, NO_x, and SO_2 values compared to Taipei, Hsinchu, and Kaohsiung. It can be said that the city of Taichung is significantly impacting air pollution; thus, special handling is needed to solve this air pollution problem, especially with penta-helix contributors. Excessive industrial activities can also cause air pollution. In addition, the number of fossil-fueled vehicles contributes to an increase in the amount of carbon dioxide and other toxic gases in the air. Forecasting using GSTAR-CFNN-PSO can be used as a reference for long-term forecasting because it provides a high accuracy value [22,35,40,54–56]. In this study, we compared the performance of neural network methods to air pollution. We proposed using two hybrid models that combine VAR and GSTAR with neural network models to forecast pollution. In our proposed model, a time series is considered a nonlinear function of several past observations and random errors. Therefore, VAR and GSTAR models are used to generate the residuals. We first examined the accuracy of these three models using simulated data and then the pollution data. All of our methods yielded almost the same values regardless of leading time. The cascade neural network with PSO was better for the pollution forecast if we consider the training time.

Pollution issues in several countries are specifically related to traffic jams, air and noise pollution, traffic accidents, and delays (increased travel time) [57–66]. In the late 1980s, developed countries reached a much more advanced stage in transportation planning than they had 20 years earlier. The rapid advancement of knowledge in electronics and computer equipment has enabled the development of several previously unimagined transportation infrastructure concepts. On the other hand, many developing countries are dealing with transportation issues, some of which are already critical. Problems caused by limited existing transportation infrastructure have been added to other issues such as

low per capita income, rapid urbanization, limited resources (particularly funds), and the quality and quantity of data related to transportation [4,67–69]. It is simple to determine the cause of traffic jams in Taiwan with our bare eyes. Congestion is caused by an increasing uncontrolled number of vehicles, including cars, motorcycles, public transit, and other vehicles, which then spill into traffic, causing this general problem. Furthermore, the number of private cars is increasing. It can be seen that the number of cars on major roads outnumbers other vehicles such as motorcycles, public transportation, and so on.

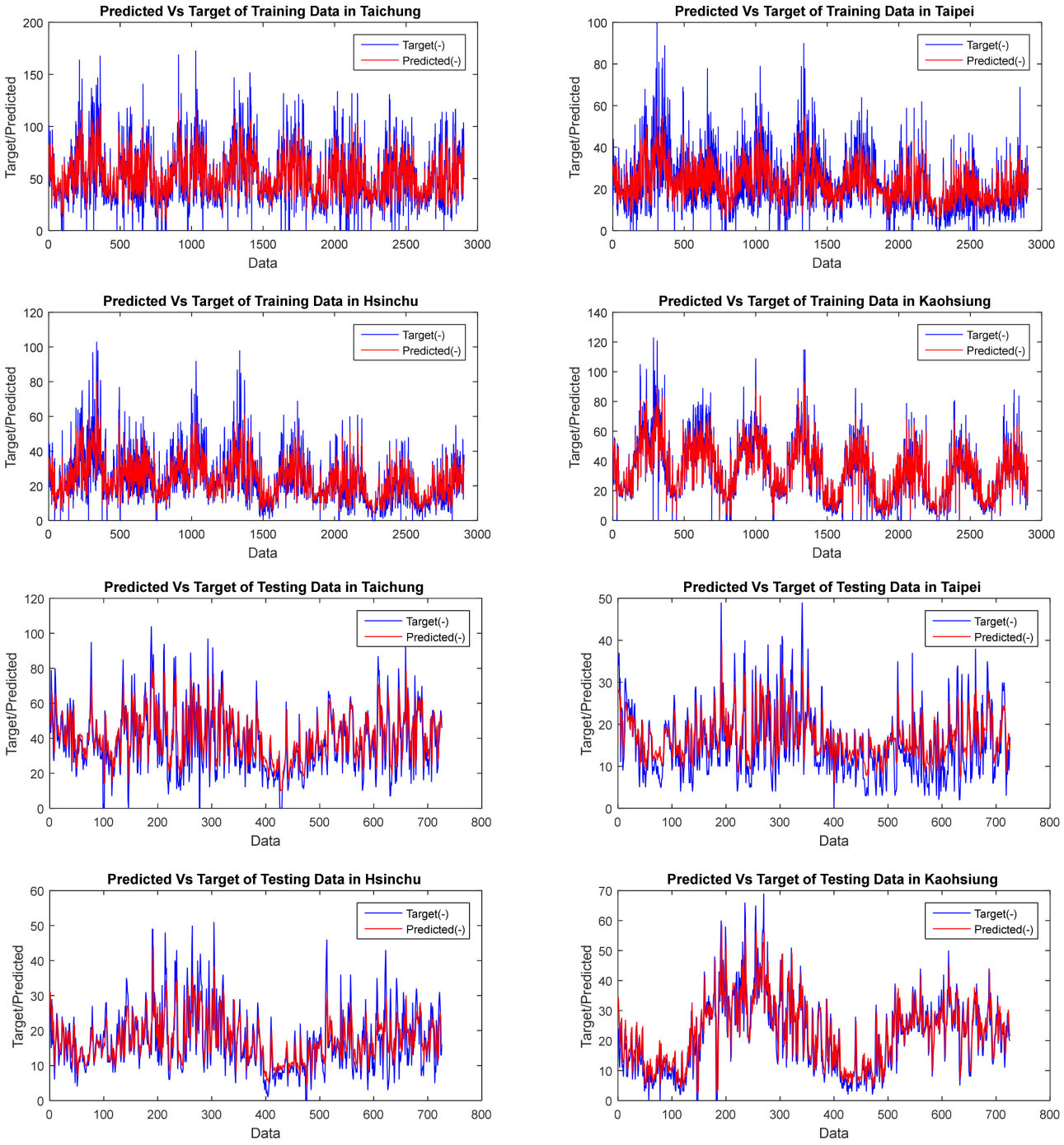

Figure 3. PM_{10} GSTAR(1) CFNN PSO.

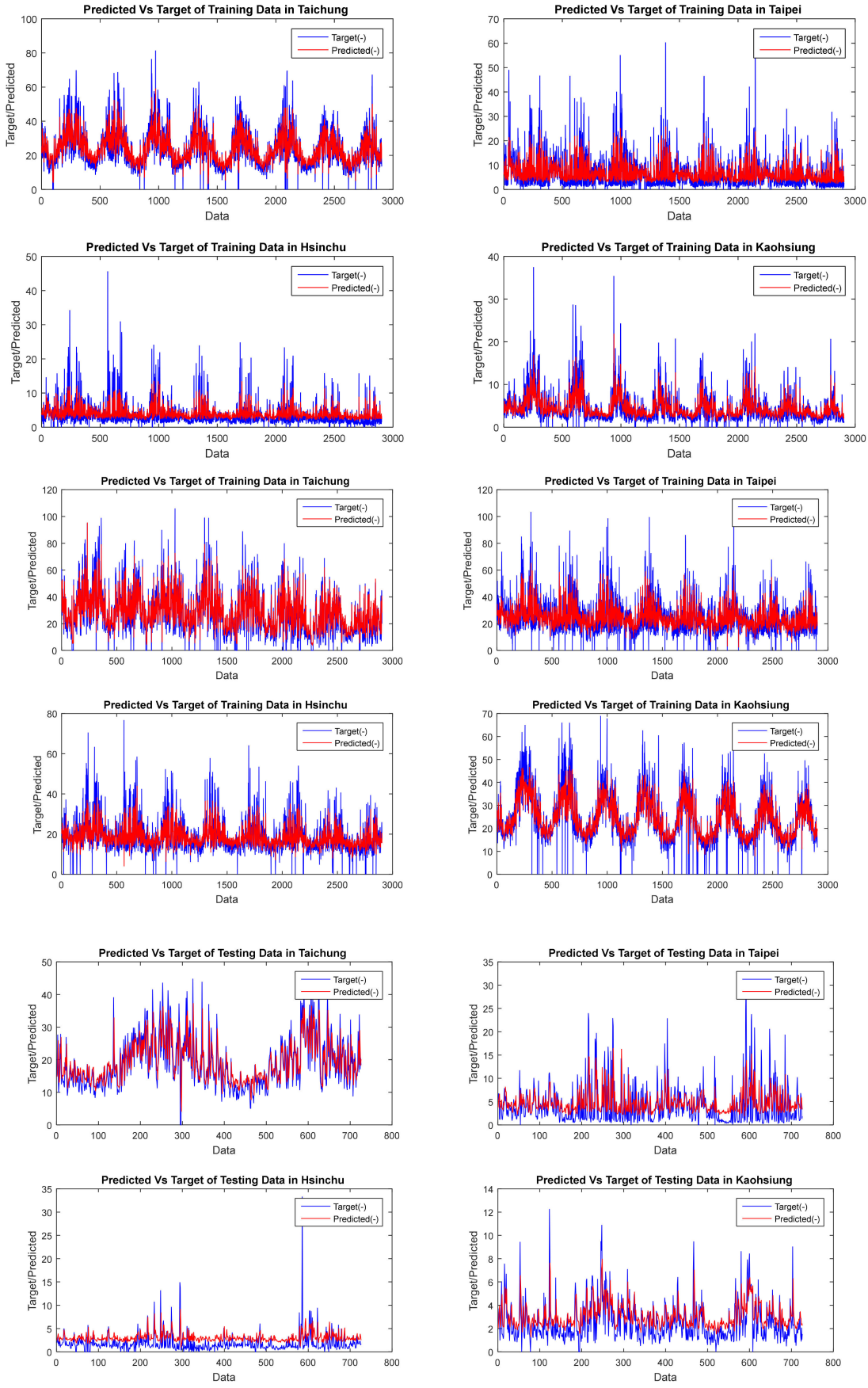

Figure 4. NOx GSTAR(2) CFNN PSO.

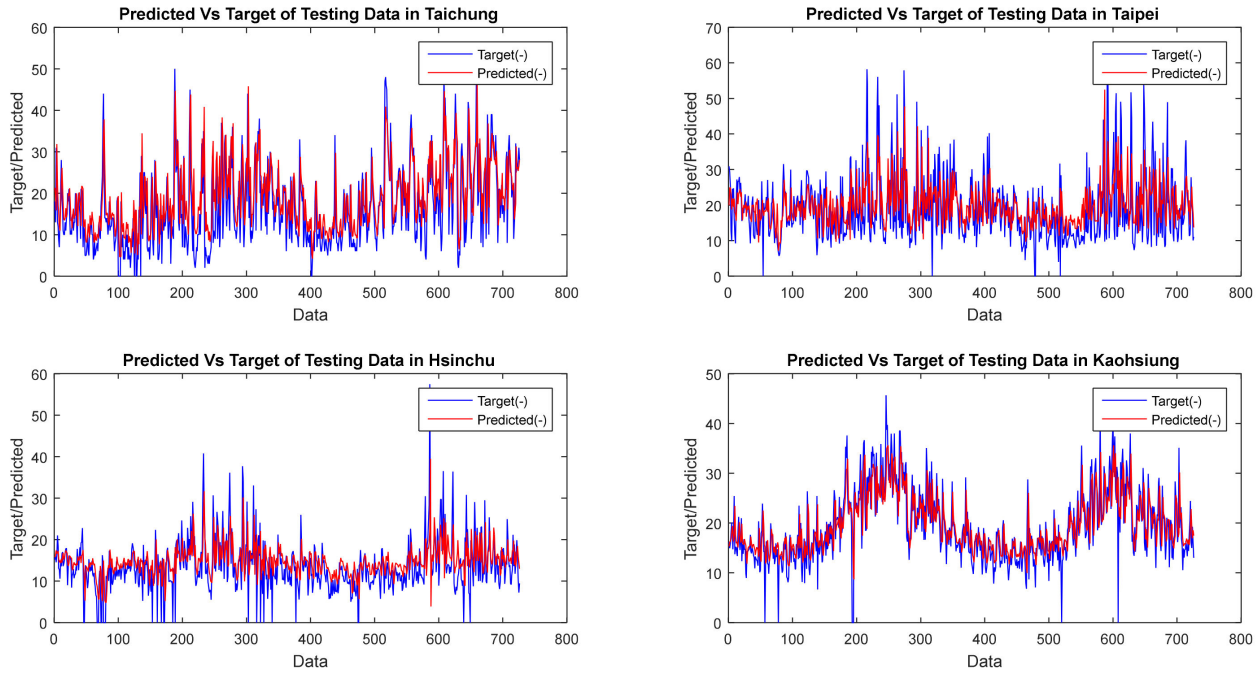

Figure 5. $PM_{2.5}$ GSTAR(2) CFNN PSO.

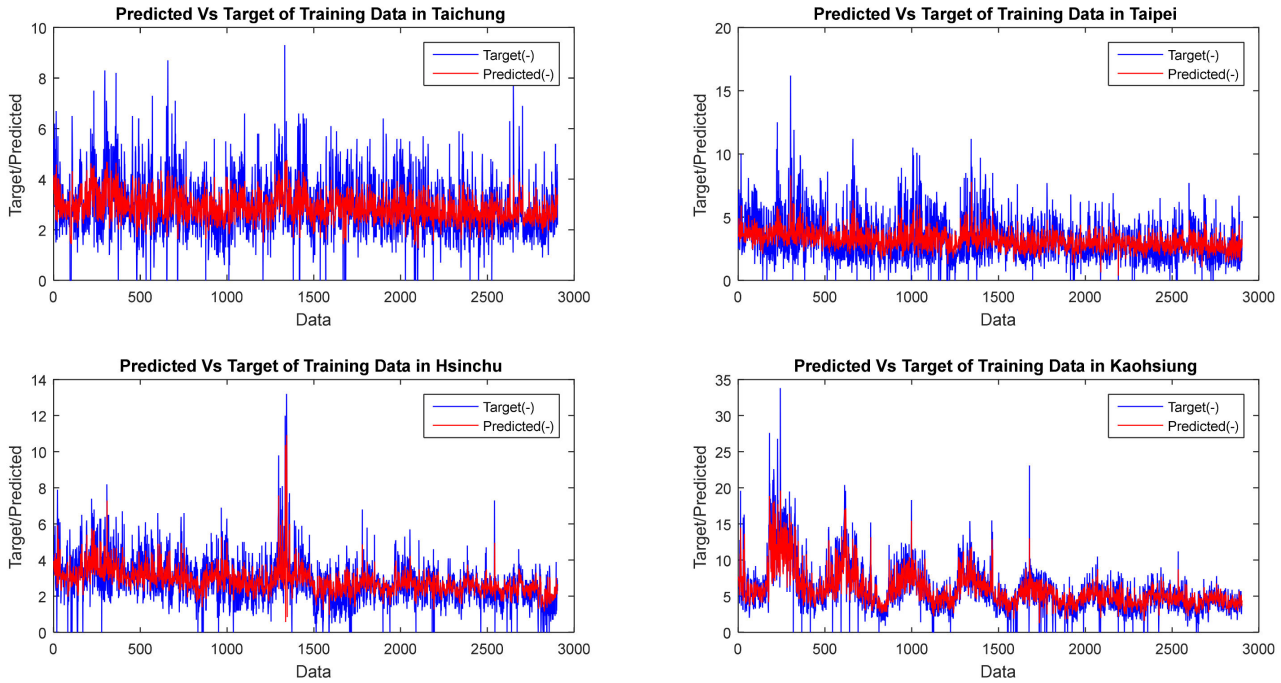

Figure 6. *Cont.*

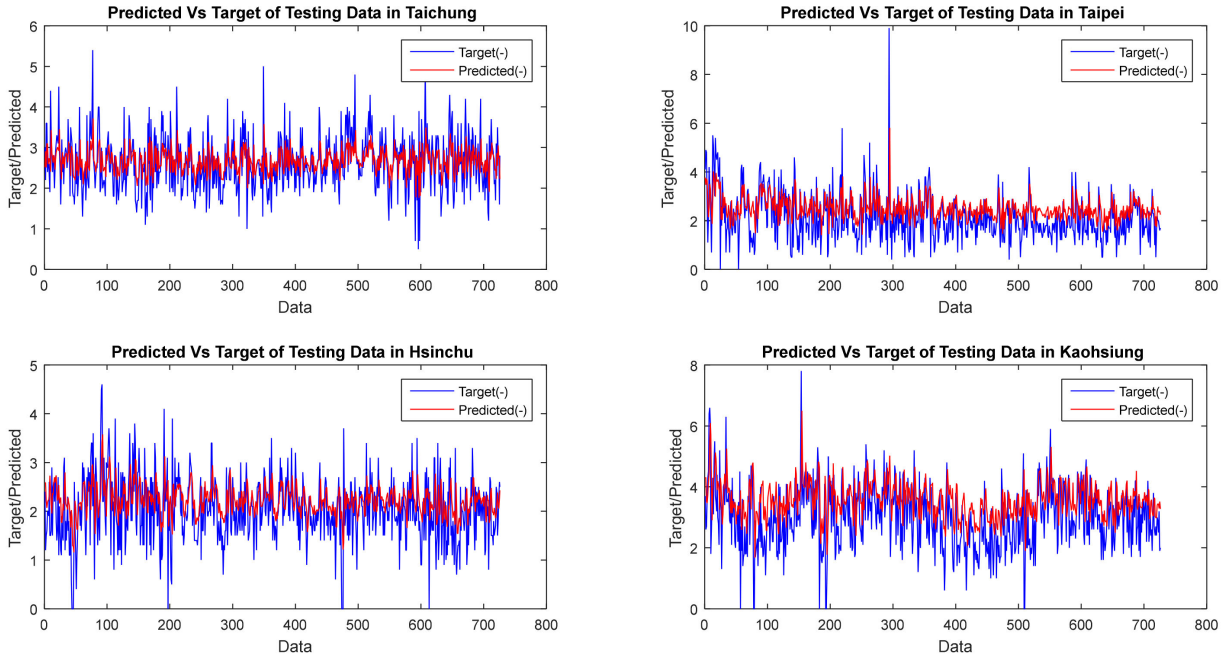

Figure 6. SO_2 GSTAR(2) CFNN PSO.

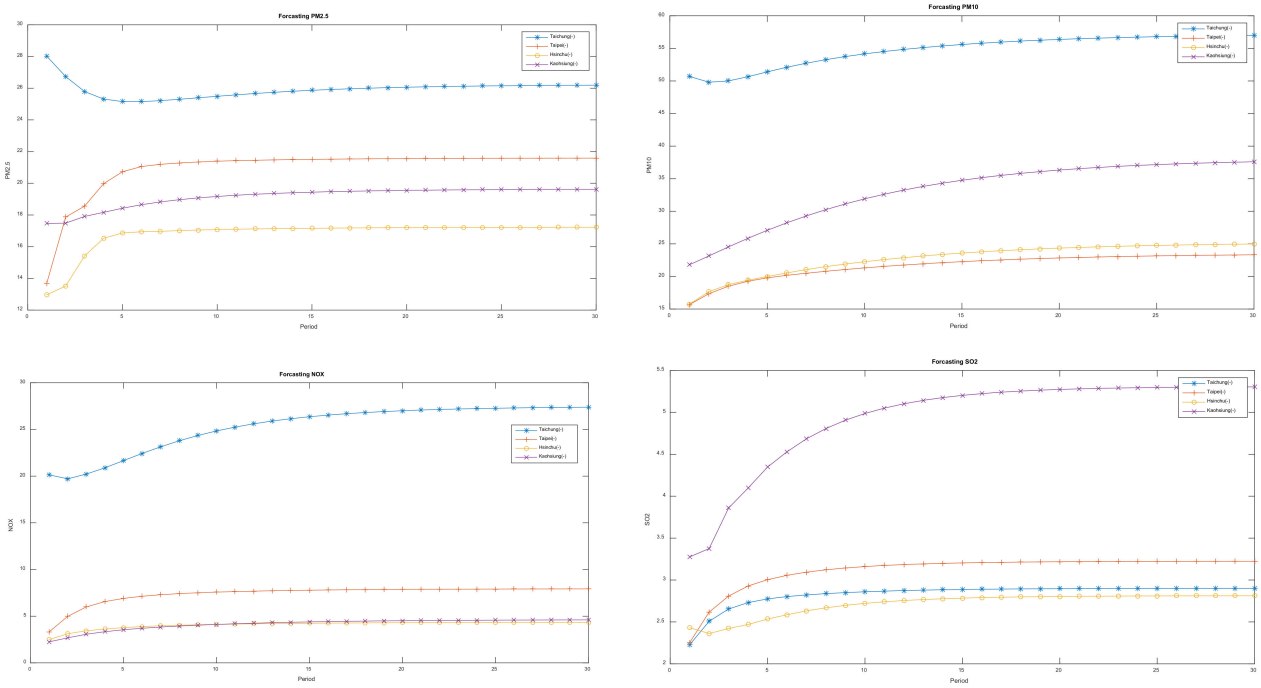

Figure 7. GSTAR(2) CFNN PSO forecasting.

5. Conclusions

The GSTAR model has limitations, as it cannot model nonlinear time series. This can be overcome by applying a hybrid model to the GSTAR. The hybrid model works by combining linear and nonlinear models to handle linear and nonlinear correlation structures. Each model produces different error values, and it has been discussed that, overall, the GSTAR-cascade-neural-network model provides better accuracy when compared to the VAR model. Therefore, the pollution conditions, including NO_x, $PM_{2.5}$, PM_{10}, and SO_2 in the four locations of Taichung, Taipei, Kaohsiung, and Hsinchu, can be generalized

for spatial forecasting using the GSTAR-cascade-neural-network with a reasonably high accuracy up to 0.01% (close to zero). Further research also needs to involve subset elements in the GSTAR model to improve forecast accuracy. In addition, it is possible to develop a GSTARX-ANN 1-level hybrid model for space-time data forecasting as a comparison of the GSTARX-ANN 2-level hybrid model for forecasting data containing trend, seasonal, calendar variation, and linear and nonlinear noise series components. In addition, we recommend using other types of architecture on neural networks and deep neural networks. Further research can be conducted with simulation studies for seasonal data only and seasonal and non-seasonal using the GSTAR-GLS and GSTAR-SUR models. Due to rapid urbanization and industrialization over the last decades, Taiwan has faced severe environmental problems, including air pollution. To address the issue of air pollution, the government has undertaken a number of defensive measures. In line with this, spatial-temporal empirical studies of air quality using monitoring data from the last ten years revealed that air quality changed with urban development. The evaluations are used to make future specific recommendations and policy advice [66,70–74]. In many countries, traffic is the leading cause of serious air pollution. Traffic is a serious threat to the environment not only because of the number of cars on the road, but also because of the fuel efficiency. The per-vehicle mileage in Taiwan is excessively high when compared to other metropolitan cities because Taiwan's transportation policy has emphasized expanding the supply of traffic utilities, such as through road construction, and expanding intersections and parking lots due to low fuel and parking fees.

Many researchers make observations about pollutant gases, such as ozone (O_3), nitrogen oxide (NO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and particulate matter, to determine the air quality in an area and control the stability of its air quality. The research process is carried out by forming an air quality model and then creating a control device to prevent or reduce pollutant substances in the air. Due to the high cost of these controllers and the high maintenance costs, it is not possible to place and use as many tools as possible to measure pollutant concentrations. Therefore, the estimation of pollutant concentration is essential in predicting pollutant concentrations in the future and estimating pollutant concentrations in areas that are not included in the radius of a measuring instrument. The next research can use the air pollution model with chemical transport. This model is a linear dynamic model that is then formed into a state-space model and discrete model, which can be performed using VAR. After that, it is processed at the prediction, correction, and simulation stages.

$$\frac{\partial C}{\partial t} \sim \frac{\partial^2 C}{x^2}\frac{\partial C}{\partial t} \sim K_x \frac{\partial^2 C}{\partial x^2}$$

C is the concentration of the substance in the air and t is time in seconds. However, K_x is the proportional constant in units of $\frac{m^2}{second}$.

We have explained how to perform vector autoregression with a cascade neural network and particle swarm optimization. The disadvantage of VAR models is that they only can capture the relationship between multiple quantities as they change over time. Every region has different uniqueness of climate phenomena and pollution, making the accuracy for forecasting not really robust.

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

Funding: This research is supported by Padjadjaran University (Jatinangor, Jawa Barat, Indonesia). This research is also supported by the Multidisciplinary Digital Publishing Institute under a research grant and by Rezzy Eko Caraka as a guest editor.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: A supplementary code to this article can be found online at https://github.com/Rezzy94/cascadenn (accessed on 24 April 2022).

Conflicts of Interest: The authors declare no conflict of interest. The results are appropriately placed in the context of prior and existing research. This paper has been reviewed in three rounds by independent reviewers. It does not affect the position of the corresponding author as a member of the editorial board in the Special Issue "Recent Topics of Climate Vulnerability: Statistics, Machine Learning, and Data Science, from Theory to Application".

References

1. Liao, L.; Du, M.; Chen, Z. Air Pollution, Health Care Use and Medical Costs: Evidence from China. *Energy Econ.* **2021**, *95*, 105132. [CrossRef]
2. Bakolis, I.; Hammoud, R.; Stewart, R.; Beevers, S.; Dajnak, D.; MacCrimmon, S.; Broadbent, M.; Pritchard, M.; Shiode, N.; Fecht, D.; et al. Mental Health Consequences of Urban Air Pollution: Prospective Population-Based Longitudinal Survey. *Soc. Psychiatry Psychiatr. Epidemiol.* **2021**, *56*, 1587–1599. [CrossRef] [PubMed]
3. WHO. *Review of Evidence on Health Aspects of Air Pollution—REVIHAAP*; WHO: Geneva, Switzerland, 2021.
4. Shi, Y.; Bilal, M.; Ho, H.C.; Omar, A. Urbanization and Regional Air Pollution across South Asian Developing Countries—A Nationwide Land Use Regression for Ambient PM2.5 Assessment in Pakistan. *Environ. Pollut.* **2020**, *266*, 115145. [CrossRef] [PubMed]
5. De Gooijer, J.G.; Hyndman, R.J. 25 Years of Time Series Forecasting. *Int. J. Forecast.* **2006**, *22*, 443–473. [CrossRef]
6. Makridakis, S.; Hibon, M. The M3-Competition: Results, Conclusions and Implications. *Int. J. Forecast.* **2000**, *16*, 451–476. [CrossRef]
7. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 Time Series and 61 Forecasting Methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [CrossRef]
8. Makridakis, S. A Survey of Time Series. *Int. Stat. Rev.* **1976**, *44*, 29. [CrossRef]
9. Caraka, R.E.; Chen, R.C.; Yasin, H.; Lee, Y.; Pardamean, B. Hybrid Vector Autoregression Feedforward Neural Network with Genetic Algorithm Model for Forecasting Space-Time Pollution Data. *Indones. J. Sci. Technol.* **2021**, *6*, 243–266. [CrossRef]
10. Suhartono, S. Time Series Forecasting by Using Seasonal Autoregressive Integrated Moving Average: Subset, Multiplicative or Additive Model. *J. Math. Stat.* **2011**, *7*, 20–27. [CrossRef]
11. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. M5 Accuracy Competition: Results, Findings, and Conclusions. *Int. J. Forecast.* **2022**, *38*. [CrossRef]
12. Theodorou, E.; Wang, S.; Kang, Y.; Spiliotis, E.; Makridakis, S.; Assimakopoulos, V. Exploring the Representativeness of the M5 Competition Data. *Int. J. Forecast.* **2021**. [CrossRef]
13. Hewamalage, H.; Bergmeir, C.; Bandara, K. Global Models for Time Series Forecasting: A Simulation Study. *Pattern Recognit.* **2022**, *124*, 108441. [CrossRef]
14. Li, Y.; Ye, Q.; Liu, A.; Meng, F.; Zhang, W.; Xiong, W.; Wang, P.; Wang, C. Seeking Urbanization Security and Sustainability: Multi-Objective Optimization of Rainwater Harvesting Systems in China. *J. Hydrol.* **2017**, *550*, 42–53. [CrossRef]
15. Suhermi, N.; Suhartono; Prastyo, D.D.; Ali, B. Roll Motion Prediction Using a Hybrid Deep Learning and ARIMA Model. *Procedia Comput. Sci.* **2018**, *144*, 251–258. [CrossRef]
16. Suhartono; Maghfiroh, B.; Rahayu, S.P. Hybrid VARX-SVR and GSTARX-SVR for Forecasting Spatio-Temporal Data. *Int. J. Innov. Technol. Explor. Eng. IJITEE* **2019**, *8*, 212–218.
17. Ruiz-Aguilar, J.J.; Turias, I.J.; Jiménez-Come, M.J. Hybrid Approaches Based on SARIMA and Artificial Neural Networks for Inspection Time Series Forecasting. *Transp. Res. Part E Logist. Transp. Rev.* **2014**, *67*, 37–49. [CrossRef]
18. He, K.; Ji, L.; Wu, C.W.D.; Tso, K.F.G. Using SARIMA–CNN–LSTM Approach to Forecast Daily Tourism Demand. *J. Hosp. Tour. Manag.* **2021**, *49*, 25–33. [CrossRef]
19. Zhang, J. Investment Risk Model Based on Intelligent Fuzzy Neural Network and VaR. *J. Comput. Appl. Math.* **2020**, *371*, 112707. [CrossRef]
20. Soltani, S.; Modarres, R.; Eslamian, S.S. The Use of Time Series Modeling for the Determination of Rainfall Climates of Iran. *Int. J. Climatol.* **2007**, *27*, 819–829. [CrossRef]
21. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting & Control*; John Wiley & Sons: Hoboken, NJ, USA, 1994; ISBN 8131716333.
22. Suhartono; Prastyo, D.D.; Kuswanto, H.; Lee, M.H. Comparison between VAR, GSTAR, FFNN-VAR and FFNN-GSTAR Models for Forecasting Oil Production Methods. *Mat. Malays. J. Ind. Appl. Math.* **2018**, *34*, 103–111.
23. Bebis, G.; Georgiopoulos, M. Feed-Forward Neural Networks. *IEEE Potentials* **1994**, *13*, 27–31. [CrossRef]
24. Huang, S.N.; Tan, K.K.; Lee, T.H.; Giam, T.S. Decentralized Adaptive Control of Nonlinear Systems Using Neural Networks. *Proc. IEEE Conf. Decis. Control* **2002**, *2*, 1757–1761. [CrossRef]
25. Bodyanskiy, Y.V.; Tyshchenko, A.K.; Deineko, A.A. An Evolving Radial Basis Neural Network with Adaptive Learning of Its Parameters and Architecture. *Autom. Control Comput. Sci.* **2015**, *49*, 255–260. [CrossRef]

26. Glorot, X.; Bengio, Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Setti Ballas, Italy, 13–15 May 2010; Volume 9, pp. 249–256.
27. Recknagel, F. Applications of Machine Learning to Ecological Modelling. *Ecol. Model.* **2001**, *146*, 303–310. [CrossRef]
28. Toharudin, T.; Pontoh, R.S.; Caraka, R.E.; Zahroh, S.; Lee, Y.; Chen, R.C. Employing Long Short-Term Memory and Facebook Prophet Model in Air Temperature Forecasting. *Commun. Stat. Simul. Comput.* **2021**, 1–12. [CrossRef]
29. Helmini, S.; Jihan, N.; Jayasinghe, M.; Perera, S. Sales Forecasting Using Multivariate Long Short Term Memory Networks. *PeerJ Prepr.* **2019**, *7*, e27712v1. [CrossRef]
30. Yamak, P.T.; Yujian, L.; Gadosey, P.K. A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. In Proceedings of the 2nd International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China, 20–22 December 2019; pp. 49–55.
31. Li, D. Human Skeleton Detection and Extraction in Dance Video Based on PSO-Enabled LSTM Neural Network. *Comput. Intell. Neurosci.* **2021**, *2021*, 2545151. [CrossRef]
32. AgaAzizi, S.; Rasekh, M.; Abbaspour-Gilandeh, Y.; Kianmehr, M.H. Identification of Impurity in Wheat Mass Based on Video Processing Using Artificial Neural Network and PSO Algorithm. *J. Food Processing Preserv.* **2021**, *45*, e15067. [CrossRef]
33. Shafiei Chafi, Z.; Afrakhte, H. Short-Term Load Forecasting Using Neural Network and Particle Swarm Optimization (PSO) Algorithm. *Math. Probl. Eng.* **2021**, *2021*, 5598267. [CrossRef]
34. Quarto, M.; D'urso, G.; Giardini, C.; Maccarini, G.; Carminati, M. A Comparison between Finite Element Model (Fem) Simulation and an Integrated Artificial Neural Network (Ann)-Particle Swarm Optimization (Pso) Approach to Forecast Performances of Micro Electro Discharge Machining (Micro-Edm) Drilling. *Micromachines* **2021**, *12*, 667. [CrossRef]
35. Suhartono; Wahyuningrum, S.R.; Setiawan; Akbar, M.S. GSTARX-GLS Model for Spatio-Temporal Data Forecasting. *Malays. J. Math. Sci.* **2016**, *10*, 91–103.
36. Astutik, S.; Iriawan, N.; Nair, G.; Suhartono. Sutikno Bayesian State Space Modeling for Spatio-Temporal Rainfall Disaggregation. *Int. J. Appl. Math. Stat.* **2013**, *37*, 26–37.
37. Suhartono, A. Comparative Study of Forecasting Models for Trend and Seasonal Time Series: Does Complex Model Always Yield Better Forecast Than Simple Model. *J. Tek. Ind.* **2005**, *7*, 22–30. [CrossRef]
38. Delforge, D.; Watlet, A.; Kaufmann, O.; Van Camp, M.; Vanclooster, M. Time-Series Clustering Approaches for Subsurface Zonation and Hydrofacies Detection Using a Real Time-Lapse Electrical Resistivity Dataset. *J. Appl. Geophys.* **2021**, *184*, 104203. [CrossRef]
39. Suhartono, S.; Subanar, S. Some Comments on the Theorem Providing Stationarity Condition for Gstar Models in the Paper by Borovkova Et Al. *J. Indones. Math. Soc.* **2007**, *13*, 115–121. [CrossRef]
40. Ruchjana, B.N.; Borovkova, S.A.; Lopuhaa, H.P. Least Squares Estimation of Generalized Space Time AutoRegressive (GSTAR) Model and Its Properties. In *AIP Conference Proceedings*; American Institute of Physics: College Park, MD, USA, 2012; Volume 1450, pp. 61–64.
41. Wei, W.W.S.; Hamilton, J.D. *Time Series Analysis*; Prentice Hall: Hoboken, NJ, USA, 1994; pp. 837–900.
42. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In *1995 IEEE International Conference on Neural Networks Proceedings*; IEEE: Piscataway, NJ, USA, 1995; Volume 4, pp. 1942–1948. [CrossRef]
43. Caraka, R.E.; Chen, R.C.; Yasin, H.; Pardamean, B.; Toharudin, T.; Wu, S.H. Prediction of Status Particulate Matter 2.5 Using State Markov Chain Stochastic Process and HYBRID VAR-NN-PSO. *IEEE Access* **2019**, *7*, 161654–161665. [CrossRef]
44. Rahman, C.M.; Rashid, T.A. Dragonfly Algorithm and Its Applications in Applied Science Survey. *Comput. Intell. Neurosci.* **2019**, *2019*, 9293617. [CrossRef]
45. Khursheed, M.U.N.; Nadeem, M.F.; Khalil, A.; Sajjad, I.A.; Raza, A.; Iqbal, M.Q.; Bo, R.; Rehman, W.U. Review of Flower Pollination Algorithm: Applications and Variants. In Proceedings of the 2020 International Conference on Engineering and Emerging Technologies, ICEET 2020, Lahore, Pakistan, 22–23 February 2020; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2020.
46. Karaboga, D.; Basturk, B. A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm. *J. Glob. Optim.* **2007**, *39*, 459–471. [CrossRef]
47. Setiawan, S.; Prastuti, M. S-GSTAR-SUR Model for Seasonal Spatio Temporal Data Forecasting. *Malays. J. Math. Sci.* **2016**, *10*, 53–65.
48. Parker, P.; Swindlehurst, A. Space-Time Autoregressive Filtering for Matched Subspace STAP. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 510–520. [CrossRef]
49. Storvik, G.; Frigessi, A.; Hirst, D. Stationary Space-Time Gaussian Fields and Their Time Autoregressive Representation. *Stat. Modeling* **2002**, *2*, 139–161. [CrossRef]
50. Cheng, T.; Wang, J.; Haworth, J.; Heydecker, B.; Chow, A. A Dynamic Spatial Weight Matrix and Localized Space-Time Autoregressive Integrated Moving Average for Network Modeling. *Geogr. Anal.* **2014**, *46*, 75–97. [CrossRef]
51. Prastyo, D.D.; Nabila, F.S.; Suhartono; Lee, M.H.; Suhermi, N.; Fam, S.F. VAR and GSTAR-Based Feature Selection in Support Vector Regression for Multivariate Spatio-Temporal Forecasting. In Proceedings of the Communications in Computer and Information Science, Dresden, Germany, 10–11 October 2019; pp. 46–57.

52. Abdullah, A.S.; Matoha, S.; Lubis, D.A.; Falah, A.N.; Jaya, I.G.N.M.; Hermawan, E.; Ruchjana, B.N. Implementation of Generalized Space Time Autoregressive (GSTAR)-Kriging Model for Predicting Rainfall Data at Unobserved Locations in West Java. *Appl. Math. Inf. Sci.* **2018**, *12*, 607–615. [CrossRef]
53. Caraka, R.E.; Yasin, H.; Chen, R.C.; Goldameir, N.E.; Supatmanto, B.D.; Toharudin, T.; Basyuni, M.; Gio, P.U. Evolving Hybrid Cascade Neural Network Genetic Algorithm Space-Time Forecasting. *Symmetry* **2021**, *13*, 1158. [CrossRef]
54. Bonar, H.; Ruchjana, B.N.; Darmawan, G. Development of Generalized Space Time Autoregressive Integrated with ARCH Error (GSTARI—ARCH) Model Based on Consumer Price Index Phenomenon at Several Cities in North Sumatera Province. In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2017.
55. Borovkova, S.; Lopuhaä, H.P.; Ruchjana, B.N. Consistency and Asymptotic Normality of Least Squares Estimators in Generalized STAR Models. *Stat. Neerl.* **2008**, *62*, 482–508. [CrossRef]
56. Yundari, Y.; Huda, N.M.; Pasaribu, U.S.; Mukhaiyar, U.; Sari, K.N. Stationary process in GSTAR (1;1) through kernel function approach. In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2020; Volume 2268, p. 020010. [CrossRef]
57. Zhang, J.J.; Wei, Y.; Fang, Z. Ozone Pollution: A Major Health Hazard Worldwide. *Front. Immunol.* **2019**, *10*, 2518. [CrossRef]
58. Riojas-Rodríguez, H.; Romieu, I.; Hernández-Ávila, M. Air Pollution. In *Occupational and Environmental Health*; Oxford University Press: Oxford, UK, 2017; ISBN 9780190662677.
59. Deryugina, T.; Heutel, G.; Miller, N.H.; Molitor, D.; Reif, J. The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *Am. Econ. Rev.* **2019**, *109*, 4178–4219. [CrossRef]
60. Akimoto, H. Global Air Quality and Pollution. *Science* **2003**, *302*, 1716–1719. [CrossRef]
61. Wang, H.; Zhang, H.; Cai, G. An Application of Phytoremediation to River Pollution Remediation. *Procedia Environ. Sci.* **2011**, *10*, 1904–1907. [CrossRef]
62. Masseran, N.; Safari, M.A.M. Modeling the Transition Behaviors of PM 10 Pollution Index. *Environ. Monit. Assess.* **2020**, *192*, 1–15. [CrossRef]
63. Masseran, N.; Mohd Safari, M.A. Intensity–Duration–Frequency Approach for Risk Assessment of Air Pollution Events. *J. Environ. Manag.* **2020**, *264*, 110429. [CrossRef] [PubMed]
64. AL-Dhurafi, N.A.; Masseran, N.; Zamzuri, Z.H. Hierarchical-Generalized Pareto Model for Estimation of Unhealthy Air Pollution Index. *Environ. Modeling Assess.* **2020**, *25*, 555–564. [CrossRef]
65. Paoletti, E.; de Marco, A.; Pompei, E. Five-Year Volume Growth of European Beech Does Not Respond to Ozone Pollution in Italy. *Environ. Sci. Pollut. Res.* **2017**, *25*, 8233–8239. [CrossRef] [PubMed]
66. Caraka, R.E.; Yusra, Y.; Toharudin, T.; Chen, R.; Basyuni, M. Did Noise Pollution Really Improve during COVID-19? Evidence from Taiwan. *Sustainability* **2021**, *13*, 5946. [CrossRef]
67. Tsai, F.; Tu, J.Y.; Hsu, S.C.; Chen, W.N. Case Study of the Asian Dust and Pollutant Event in Spring 2006: Source, Transport, and Contribution to Taiwan. *Sci. Total Environ.* **2014**, *478*, 163–174. [CrossRef]
68. De Vos, J. Transportation Research Interdisciplinary Perspectives the Effect of COVID-19 and Subsequent Social Distancing on Travel Behavior. *Transp. Res. Interdiscip. Perspect.* **2020**, *5*, 100121. [CrossRef]
69. Wei, G.; Zhang, Z.; Ouyang, X.; Shen, Y.; Jiang, S.; Liu, B.; He, B.J. Delineating the Spatial-Temporal Variation of Air Pollution with Urbanization in the Belt and Road Initiative Area. *Environ. Impact Assess. Rev.* **2021**, *91*, 106646. [CrossRef]
70. Gayialis, S.P.; Kechagias, E.P.; Konstantakopoulos, G.D. A City Logistics System for Freight Transportation: Integrating Information Technology and Operational Research. *Oper. Res.* **2022**, 1–30. [CrossRef]
71. Peng, J.; Parnell, J.; Kessissoglou, N. Spatially Differentiated Profiles for Road Traffic Noise Pollution across a State Road Network. *Appl. Acoust.* **2021**, *172*, 107641. [CrossRef]
72. Hong, W.C. Traffic Flow Forecasting by Seasonal SVR with Chaotic Simulated Annealing Algorithm. *Neurocomputing* **2011**, *74*, 2096–2107. [CrossRef]
73. Li, M.W.; Hong, W.C.; Kang, H.G. Urban Traffic Flow Forecasting Using Gauss-SVR with Cat Mapping, Cloud Model and PSO Hybrid Algorithm. *Neurocomputing* **2013**, *99*, 230–240. [CrossRef]
74. Kechagias, E.P.; Gayialis, S.P.; Konstantakopoulos, G.D.; Papadopoulos, G.A. Traffic Flow Forecasting for City Logistics: A Literature Review and Evaluation. *Int. J. Decis. Support Syst.* **2019**, *4*, 159–176. [CrossRef]