Indonesian Journal of Science & Technology

Journal homepage: http://ejournal.upi.edu/index.php/ijost/

Hybrid Vector Autoregression Feedforward Neural Network with Genetic Algorithm Model for Forecasting Space-Time Pollution Data

Rezzy Eko Caraka[1,2,3,7], Rung Ching Chen[2**], Hasbi Yasin[4], Suhartono[5], Youngjo Lee[1,6], Bens Pardamean[7,8]*

[1] Lab Hierarchical Likelihood, College of Natural Science, Seoul National University, Republic of Korea.
[2] Department of Information Management, Chaoyang University of Technology, Taiwan (ROC).
[3] Medical Research Center, Seoul National University Hospital (SNU-H), Seoul, Republic of Korea.
[4] Department of Statistics, Diponegoro University, Indonesia
[5] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya,Indonesia.
[6] Department of Statistics, College of Natural Science, Seoul National University, Republic of Korea.
[7] Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia
[8] Computer Science Department, Bina Nusantara University, Jakarta, Indonesia.
Correspondence: *E-mail: crching@cyut.edu.tw; **E-mail: rezzyekocaraka@gmail.com

ABSTRACTS

The exposure rate to air pollution in most urban cities is really a major concern because it results to a life-threatening consequence for human health and wellbeing. Furthermore, the accurate estimation and continuous forecasting of pollution levels is a very complicated task. In this paper, one of the space-temporal models, a vector autoregressive (VAR) with neural network (NN) and genetic algorithm (GA) was proposed and enhanced. The VAR could tackle the issue of multivariate time series, NN for nonlinearity, and GA for parameter estimation determination. Therefore, the model could be used to make predictions, such as the information of series and location data. The applied methods were on the pollution data, including NO_X, $PM_{2.5}$, PM_{10}, and SO_2 in Taipei, Hsinchu, Taichung, and Kaohsiung. The metaheuristics genetic algorithm was used to enhance the proposed methods during the experiments. In conclusion, the VAR-NN-GA gives a good accuracy when metric evaluation is used. Furthermore, the methods can be used to determine the phenomena of 10 years air pollution in Taiwan.

ARTICLE INFO

Article History:
Received 21 October 2020
Revised 09 March 2021
Accepted 12 March 2021
Available online 15 March 2021

Keywords:
Pollution,
VAR,
FFNN,
Genetic algorithm,
Hybrid forecasting

1. INTRODUCTION

To improve the accuracy of forecasting methods, the M-Competition was put into practice by Spyros Makridakis. The M-1, being the first was held in 1982 by using 1001 subsamples from 111 methods, with 9 variations. The result showed that the statistical complex methods sometimes do not give accurate results as the simple ones. Therefore, to improve the accuracy, a combination of methods derived from the horizon was carried out. And, in the second competition a larger scale was used.

Also, the M-2 was given to the forecaster to perform a combination of methods and personal judgement. Consequently, the M-2 competition aimed to evaluate the model based on the feedback from every participant. In early 2000, M-3 competitions (Makridakis & Hibon, 2000) were held by doing time series on daily, monthly, quarterly, and also annual basis. This competition therefore helps to improve the accuracy of the model, adjusted to the threshold of each observation. In 2018, the M-4 competition was performed by implementing the Artificial Intelligence (AI) (Wang *et al*., 2009; Yu & Schwartz, 2006) and machine learning (Makridakis *et al*., 2020b; Makridakis *et al*., 2020a).

Furthermore, the use of a hybrid method which is a combination of traditional statistics with machine learning, gives very impressive results. However, the latest competition was held on the 2nd of March to 30th of June, 2020. The most popular method for generating forecast is the De-seasonalization and Adaptive normalization, using Recurrent Neural Network (NN) and Ensemble. Apart from the aspects of M1-M5, these methods are very useful for all domains and applications in forecasting.

The scientists grasped past decades to observe the impact of climatic changes throughout the survival of all living, whether such an uncertainty was reflected in the context of disasters (Nasution *et al*., 2020). In addition to all the technological innovations, as well as advanced statistical (Hybrid methods) which have contributed to higher sensitivity of global warming (Searle & Gow, 2010), there are many remarkable accomplishments in the estimation, measurement, forecasting, and an early warning detection of climatic changes.

Consequently, climate forecasting and prediction could be divided into two main areas, namely empiric (or statistical), and numerical weather (or climate) forecasting of dynamic models. The empirical methods could be straightforward such as the persistence, where previous weather is predicted to occur for a certain long period, or perhaps more advanced and powerful, like the regression techniques.

However, the numerical and environmental analysis, uses the dynamic mathematical models (Timbal *et al*., 2008) of the climate structure (Caraka *et al*., 2019). Also, the simulations is mostly used for short term forecasting when evaluating accuracy. Decision making was highlighted as the end result of climate forecasting, as well as precipitation events to annual and seasonal predictions. Therefore, this weather knowledge would form another aspect of decision-making (Supari *et al*., 2017).

Furthermore, numerous people are likely to engage in the processes where weather forecasting are appropriate, rather than those directly affected by climatic change (Siagian *et al*., 2014; Kaban *et al*., 2019) and disaster reduction (Kurniawan *et al*., 2018). For example, resulting in fatalities, damage of homes and agricultural industry, unrest of

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

food production and distribution systems, and health negative consequences were the challenges experienced in urban cities (Ng *et al.*, 2009). Therefore, the summary emergency preparedness processes, such as the early warning systems, may be used to prevent future damages.

Also, in the improvement of time series forecasting, certain important and necessary phenomena often address nonlinear anomalies such as the interaction between previous events (Kravtsov *et al.*, 2005; Zuhairoh & Rosadi, 2020). Therefore, it would be inappropriate to use the linear time series in this scenario. However, in the recent years, the emphasis has been on the nonlinear time series forecasting, which is a deep challenge of the researchers and industry practitioners (De Gooijer & Hyndman, 2006; AL-Dhurafi *et al.*, 2018).

Therefore, the aim of this study is to demonstrate the concept of a data-driven strategy to reconstruct vector autoregressive, based on neural network and genetic algorithm. This paper was organized as follows, in section 2, the study area and materials were previewed. Also, the way to employ vector autoregressive (VAR) with neural network (NN) and genetic algorithm (GA) was discussed, while VAR with NN and particle swarm optimization (PSO) was constructed in previous study (Caraka *et al.*, 2019). Furthermore, the key result of the investigations was explained in section 3 while the conclusion and future studies was described in section 4.

2. METHODS

2.1. Previous Studies

From the year 1996, the method of machine learning have been used, because it has numerous advantages, especially in time efficiency, accuracy, and does not need to fulfil the classical assumptions of traditional methods. Nevertheless, the model is mostly black-box (Chen *et al.*, 2020), that does not provide much information like the traditional time series methods, Autoregressive integrated moving average (ARIMA), and its derivatives (Suhartono, 2005). For example, the ARIMA can be used to explain the meaning of a coefficient obtained in the model (Suhartono, 2011).

In the cases of small-dimensional data, machine learning could provide predictable results, and at the training stage, it does not seem inclusive thereby giving large errors = (Lee *et al.*, 2017). Therefore, to overcome this, the combination of the models and traditional time series is needed (Suhartono, *et al.*, 2019). For example, performing feature selection, using Vector Autoregressive (VAR) General Space-Time Autoregressive (GSTAR) on Support Vector Regression (SVR), employing VAR on neural networks and particle swarm optimization (Caraka *et al.*, 2019), combination of ARIMA and SVR (Yang & Lin, 2016). Consequently, besides being able to increase the accuracy, this combination is proven to be used for a more complete explanation, especially for decision making.

The problem associated with real-time is often an inadequate or incomplete information (Singh & Huang, 2019). However, Feedforward neural network is robust, and could accurately model the nonlinear connections for both input and output (Tang & Fishwick, 1993). The method imitates the operations of human nerves which tried to solve real issues based on past experience. Therefore, in developing a neural network model, the common principles that was evaluated includes, activation of input and hidden layers, activation of hidden and output layers, the amount of hidden layers and weights between all the three units (Yasin *et al.*, 2018). Also, long short term

memory, nowadays, becomes popular in facing big data including unexpected fluctuation on the data. (Toharudin *et al.*, 2021; Fischer & Krauss, 2018).

Besides, the Multilayer Perceptron (MLP) network optimized with three different training algorithms, such as variable learning rate (MLP-GDX), resilient back-propagation (MLP-RP), and Levenberg-Marquardt (MLP-LM) were studied in terms of ability to estimate sediment transport in a clean pipe (Ebtehaj & Bonakdari, 2016). Also, Radial Basis Function (RBF) including particle swarm optimization (PSO) (RBFN-PSO) and the back-propagation algorithms could be used to predict the densimetric Froude number (Fr) with high accuracy (Qasem *et al.*, 2017). Furthermore, the decision tree (DT) was employed to predict sediment transport in sewer pipes at the limit of deposition (Ebtehaj *et al.*, 2016).

In 2012, atmospheric 2.5-µm particulate matter ($PM_{2.5}$) was introduced as an air pollutant in Taiwan. Besides, its spatial-temporal distribution and characteristics involves a variety of natural and anthropogenic tools (Chang *et al.*, 2020). There are also several anthropogenic activities in major metropolitan areas, such as transportation and industry that further contributes to pollution (Paoletti *et al.*, 2017). Air substances, either solid or liquid, when present in high concentrations, could pose health risks to wildlife and also affect the ecosystems. In addition, the increase in the contaminants present in the atmosphere have a negative consequence on human respiratory system and could also result to cardiovascular disease (Brook *et al.*, 2010). Several pollutions data were data namely Nitrogen Oxide (NOx), $PM_{2.5}$, atmospheric 10-µm particulate matter (PM_{10}), and sulfur dioxide (SO_2). $PM_{2.5}$ (Feng *et al.*, 2016), PM_{10} (Masseran & Safari, 2020), SO_2 (Caraka *et al.*, 2020), NOx (as nitrogen dioxide (NO_2) and nitrogen trioxide (NO_3)) are the major sources of pollution in urban regions due to high traffic and the massive industrial developments (De Vito *et al.*, 2009).

2.2. Study area and materials

The study areas included Taipei, Hsinchu, Taichung, and Kaohsiung city, which consists of pollution data namely NOx, $PM_{2.5}$, PM_{10}, and SO_2. Furthermore, the locations of these areas, as established by the Taiwan Environmental Protection Administration Executive Yuan, is shown in **Figure 1**. Also, preprocessing was done to the raw dataset obtained on the 1st untill 5th of May, 2020, before being used for analysis and interpretation. This was carried out in order to transform raw data, which is usually incomplete, inconsistent, redundant, and noisy to a more appropriate information.

In this study, the data were standardized and divided into two, namely training and testing, which are used for construction and model testing respectively. Also, because the data were obtained daily for 10 years, a long series was formed. In the training stage, 1-step-ahead also known as k-step-ahead was applied, however, this does not produce an accurate forecast because it is a long period. Therefore, the ratio of 80:20, 70:30, and 60:40 was set, and the one that generated the small error was chosen by inserting it into the FFNN-GA model.

The statistical summaries for the concentrations of the studied air pollutants in 4 location are shown in **Table 1**. The results generally showed that the concentrations, PM_{10}, $PM_{2.5}$, NO_X are higher in Taichung. However, the highest pollutants in Kaohsiung are SO_2.

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

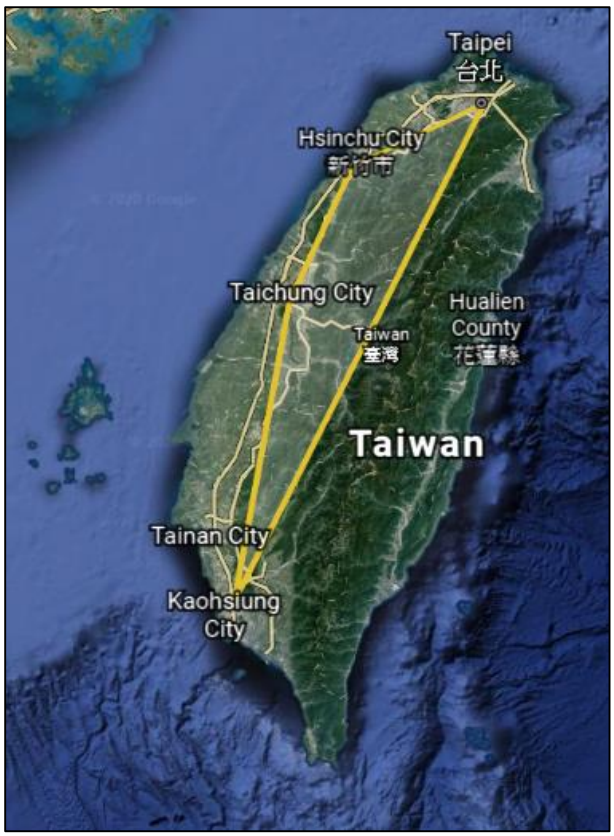

Figure 1. Study area established by the Taiwan environmental protection administration.

Table 1. Basic information and statistics of pollution at 4 location stations (May 2010 May 2020).

Pollu-tion	Location	N	Mean	SE Mean	StDev	Variance	Mini-mum	Q1	Me-dian	Q3	Maxi mum	Range
$PM_{2.5}$	TAICHUNG	3632	26.623	0.264	15.66	245.244	1	15	23	35	106	105
	TAIPEI	3632	23.635	0.203	12.161	147.892	3.97	15.107	20.84	29.023	109.83	105.86
	HSINCHU	3632	18.179	0.13	7.763	60.265	0.63	13.04	16.34	21.102	76.72	76.09
	KAOHSIUNG	3632	23.854	0.163	9.747	95.014	5.27	16.49	21.495	29.697	68.96	63.69
NO_X	TAICHUNG	3632	22.944	0.171	10.26	105.269	4.35	15.23	20.57	28.63	81.43	77.08
	TAIPEI	3632	6.196	0.106	6.315	39.874	0.09	1.85	4.15	8.28	65.14	65.05
	HSINCHU	3632	3.2786	0.0546	3.2643	10.6559	0	1.52	2.3	3.79	45.65	45.65
	KAOHSIUNG	3632	4.0401	0.0544	3.2501	10.5631	0.38	2.01	3.07	4.88	37.49	37.11
PM_{10}	TAICHUNG	3632	50.642	0.419	24.949	622.476	5	32	45.5	65	173	168
	TAIPEI	3632	21.244	0.208	12.412	154.052	1	12	19	27	100	99
	HSINCHU	3632	22.46	0.219	13.135	172.521	1	13	19	29	103	102
	KAOHSIUNG	3632	31.719	0.31	18.477	341.414	1	17	29	44	123	122
SO_2	TAICHUNG	3632	2.8706	0.017	1.0208	1.042	0	2.2	2.7	3.4	9.3	9.3
	TAIPEI	3632	2.9835	0.0263	1.5742	2.4781	0.4	1.9	2.6	3.7	16.2	15.8
	HSINCHU	3632	2.6778	0.0186	1.1125	1.2377	0.1	1.9	2.5	3.2	13.2	13.1
	KAOHSIUNG	3632	5.3129	0.0515	3.0833	9.5066	0	3.3	4.5	6.4	33.8	33.8

2.3. *Vector Autoregressive (VAR)*

Zhang (2003) introduced a hybrid model to improve the accuracy, also, space-time forecasting could be used to solve the case of seasonal (Suhartono *et al*., 2016) and non-seasonal time series (Suhartono, Maghfiroh, et al., 2019). The VAR method is a sophisticated statistical approach which is commonly applied to a wide range of variable time series. As a result, this model describes the relationship between the observations of the variables itself and the original period, as well as the one with other possible factors at the previous time.

VAR is also widely used in empiric macro-economics (Caraka *et al*., 2016; Ang & Piazzesi, 2003; Lanne & Luoto, 2020; Bernanke *et al*., 2005) which might not put constraints on parameters as opposed to the VAR (Suhartono & Subanar, 2006). However, the model needs to access a complex structure that present a risk over parametrization when using the sets of data with small dimensions. Also, the observed value from one location at a periodic interval is determined by the measurements gotten from the previous delay at other areas. Therefore, the model would be in form of VAR 1 as represented in the equation 1.

$$z_t = \phi_1 z_{t-1} + \varepsilon_t \quad (1)$$

When the observed value in a location during time t is determined by actual values at the recent "*p*" delay, together with the ones of other areas, the model would therefore be in form of VAR (p), as written in Equation 2.

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + \varepsilon_t \quad (2)$$

According to this, $\boldsymbol{z}_t$ is the vector of observations at *t*-time and *n*-th is the location of size $(n \times 1)$. Matrix parameters ϕ_p VAR *p*-order of size$(n \times n)$. ε_t explains the white noise vector $\varepsilon_t \sim MN(0, \Sigma)$ with size $(n \times 1)$. The identification step for determining the order of VAR model can be be seen from the Multivariate Partial Autocorrelation Function (MPACF) plot (Suhartono *et al.,* 2018). The VAR (1) model has a Partial Autocorrelation Function cut-off pattern, after the 1st lag, while the stationarity could be seen from the characteristic root value of the matrix parameter. Furthermore, the selection of the best VAR (1) model uses the Akaike Information Corrected Criterion (AICC), which is the one with the least value (Suhartono & Subanar, 2007).

Vector Autoregressive Integrated Moving Average (VARIMA) is a generalization of the ARIMA univariate model. As in ARMA, this analysis also takes into account the stationarity of the data and this can be seen from the Matrix Cross Correlation Function (MCCF) and Partial Cross Correlation Function (MPCCF) plots and Box-Cox plots. The treatment of multivariate data which is not stationary is the same as univariate data which is not stationary. Multivariate data that are not stationary in variance are transformed so that the data can be stationary, and if it is not stationary in the mean, differencing is used to stationary it. In general, the VARIMA model (*p, d, q*) can be written in the (Eq.3).

$$\phi_p(B)D(B)\hat{Y}_t = \Theta_q(B)\varepsilon_t \quad (3)$$

When $\hat{Y}_t = \left(y_1, y_2, \ldots, y_{m,t}\right)'$ is a response vector that corrects average, $\phi_p(B)$ and $\Theta_q(B)$ is a coefficient matrix AR(*p),* MA(*q*). Besides, $D(B)$ is a differentiation process operator represented by $diag\left((1-B)^{d_1}, (1-B)^{d_2}, \ldots, (1-B)^{d_m}\right)$ and $\varepsilon_t \sim iidn\,(0, \Omega)$. The partial autocorrelation function (PACF) is required in time series

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

univariate to determine the order in the AR model. The partial autoregression matrix on lag s with the notation (s) , as the coefficient of the last matrix when data is applied toin a Vector Autoregressive (VAR) process of order s. $P(s)$ equals $\phi_{s,s}$ in multivariate linear regression. Partial autoregression matrix equation given in (Eq.4).

$$P(s) = \begin{cases} \Gamma'(1)[\Gamma(0)]' and \\ \{\Gamma'(s) - c'(s)[A(s)]^{-1}b(s)\}\{\Gamma(0) - \mathrm{b}'(s)[A(s)]^{-1}b(s)\}^{-1} \end{cases} \quad (4)$$

Then we can expand the definition of partial univariate autocorrelation be a time series vector and obtain the correlation matrix between Y and Y_{y+s} in (Eq.5) and (Eq.6). The correlation matrix which is affirmed as a correlation between residual vectors has the following equation.

$$\mathrm{u}_{s-1,t+s} = \mathrm{Y}_{t+s-\alpha_{s-1,1}}\mathrm{Y}_{t+s-1} - \cdots - \alpha_{s-1,1}Y_{t+1}$$

$$= \begin{cases} \mathrm{Y}_{t+s} - \sum_{k=1}^{s-1} \alpha_{s-1,k}\mathrm{Y}_{t+s-k} & s \geq 2 \\ \mathrm{Y}_{t+1}, & s = 1 \end{cases} \quad (5)$$

And

$$\mathrm{v}_{s-1,t} = \mathrm{Y}_t - \beta_{s-1,1}\mathrm{Y}_{t+1} - \cdots - \beta_{s-1,s-1}Y_{t+s-1}$$

$$= \begin{cases} \mathrm{Y}_{t+s} - \sum_{k=1}^{s-1} \beta_{s-1,k}\mathrm{Y}_{t+k} & s \geq 2 \\ \mathrm{Y}_t, & s = 1 \end{cases} \quad (6)$$

Multivariate linear regression coefficient matrix$\alpha_{s-1,k}$ and $\beta_{s-1,k}$ be minimized

$$E\left[\left(\mathrm{u}_{s-1,t+s}\right)^2\right] \text{ and } E\left[\left(\mathrm{v}_{s-1,t}\right)^2\right]$$

Where x_i,y_h, and z_o are used to represent the node value in input layer, hidden, and output respectively. Besides, n_i and n_h represent the number of units in input and hidden layer. Also, w_{ih} and w_{ho} represent the weight connecting input node as well as hidden node. According to this, b_o and b_h represent bias from the models. In addition, the Genetic algorithms are metaheuristic frameworks which are appropriate for a wide scope of optimization (Whitley, 1994).

Furthermore, its usefulness makes them to act in practice for many solution spaces while the natural selection seems to be the base of genetic algorithms. Also, the recent variety and growth of life is a probable reason to believe in the thrust of natural selection. Meanwhile, **Algorithm 1** demonstrates the pseudo-code of the Main Genetic, which serve as the starting point for a variety of similar features. Consequently, at an early stage, this solution could be explained through population selection. Also, this initialization is implied to represent the full quick fix area at random, as well as to model and integrate knowledge.

Algorithm 1 Genetic Algorithm
1: Initialize population
2: **repeat**
3: crossover
4: mutation and phenotype mapping
5: fitness
5: **until** reach the population
6: parental selection
7: **until** convergence

2.4. *Metric evaluation*

It is necessary to review the performance of prediction using real forecast, however the scale of residues is not a good indicator of how strong certain real forecast errors are

expected to have been. Therefore, the accuracy could only be determined by considering how well the method fits on new data that have not been used when fitting the model. For example, when we have the trained and tested data, $\{y_1, y_2, \dots, y_T\}$ and $\{y_{T+1}, y_{T+2}, \dots\}$ respectively, the error is defined as the difference of forecast and the data. The T index describes each portion of the testing as explained in (Eq.9).

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T} \tag{9}$$

The error forecast could be distinguished from residuals explicitly in two ways. First, the errors are measured on the trained dataset, but the predicted values are determined on the tested values. Also, the residuals are based on one-step forecasting, although errors could demand multistage prediction models. Lastly, to determine the accuracy of the forecast, errors in different perspectives need to be highlighted (Hyndman & Koehler, 2006), namely the Mean Absolute (MAE), Root Mean Square (RMSE), and Symmetric Mean Absolute percentage (SMAPE) expressed in (Eq.10), (Eq.11), and (Eq.12) respectively.

$$Mean\ Absolute\ Error = \frac{1}{n} \sum |e_t| \tag{10}$$

$$Root\ Mean\ Square\ Error = \sqrt{\frac{1}{n} \sum e_t^2} \tag{11}$$

$$sMAPE = \frac{100\%}{n} \left(\frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2} \right) \tag{12}$$

3. RESULTS AND DISCUSSION

Once the training framework for neural networks has been established, the next issue that emerges is the preprocessing. Furthermore, at the early stages, the processes required to build the system is fairly challenging. Besides, initializing certain weights to zero prohibits any input from being transmitted to the output. Also, the first and last layers were often configured separately from the data, resulting to a pointless exercise. Rather, all layers need to be placed in such a way that the universal network performance would not blast, thereby reflecting on the input. Considering the neuron in (Eq.13), the layer *L* performs the operation.

$$z_{l+1} = \emptyset(y_l) = \emptyset(\mathbf{w}_l \cdot \mathbf{x}_l) \tag{13}$$

where $\mathbf{x}_l$ describes the input vector, z_{l+1} is the output of the neuron, $\mathbf{w}_l$ is the weight, and $\emptyset$ is the activation function (Suhartono *et al.*, 2016). In this study, $\mathbf{w}_l$ is assumed to be independent and identically distributed as expressed in (Eq.14).

$$\begin{aligned} Var\ (y_l) &= Var\ (\mathbf{w}_l \cdot \mathbf{x}_l) \\ &= n_{l-1}\ (w_l x_l) \end{aligned} \tag{14}$$

According to this, the random variables w_l and x_l is described in (Eq.15) which has the same distribution with the dimension n_{l-1}.

$$Var\ (y_l) = n_{l-1}\, Var\ (w_l) \mathbb{E} x_l^2 \tag{15}$$

In Taichung and Kaohsiung, the quality of air in the winter is really bad, especially in the days when there are little wind and lots of sunshine. Throughout the city, there are many manufacturing companies, and since the place is a basin region with mountains on three sides, ambient emission will be present when there is no wind. In addition, Kaohsiung, which is the hub of Taiwan has heavy duty industries, such as petrochemicals, construction, and shipbuilding as well as a large coal-fired power generation. As a result, the majority of the pollutants found in the air are from these sources, especially the greenhouse gases. For example, SO_2 is a recognized health risk and aspect of acid rain. Also, NOx, which might result to a decreased lung function, is

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

generated into nitrate aerosols (the major source of fine particulate matter).

Figure 2 showed that 5 hidden layers are used to create the model, and the ratio used is calculated by glancing at the error value of the testing results using the ratio shown in **Table 1**. Furthermore, during the construct stage of VAR-NN, the pollution dataset is from the cities of Taichung (Y_1), Taipei (Y_2), Hsinchu (Y_3), Kaohsiung (Y_4) in Taiwan. The best parameter of neural network such as x_i, y_h, and z_o, obtained by using the genetic algorithm, were described in (Eq.4) and (Eq.5), respectively. Also, it was compared with the VAR-OLS, VAR-FFNN (BP), as well as VAR-FFNN (BP) and GA. The model used to measure the ratio, has one unit as well as the number of hidden layers (setup 1), using sigmoid activation function. Therefore, the parameters used for the selection setting includes, tournament = 10, Cross over fraction = 0.8, Cross over function with cross over single points, Population size = 50, and Generation = 1000, explained with the following example:

Binary Gene String

1	0	1
...		
1	0	1

Multi-Valued Gene String

2	0.1	0.2

Consequently, the values on these parameters are established as the default, before being modified for the next simulation model. Based on this, the **Algorithm 2** describes the pseudo-code of constructing VAR-NN-GA.

In its simplest form, this algorithm picks two chromosomes randomly and then selects the one with the highest fitness value to be the first parent. Furthermore, same process is repeated in order to choose a second parent. However, taking chromosomes randomly is a complicated tournament selection method.

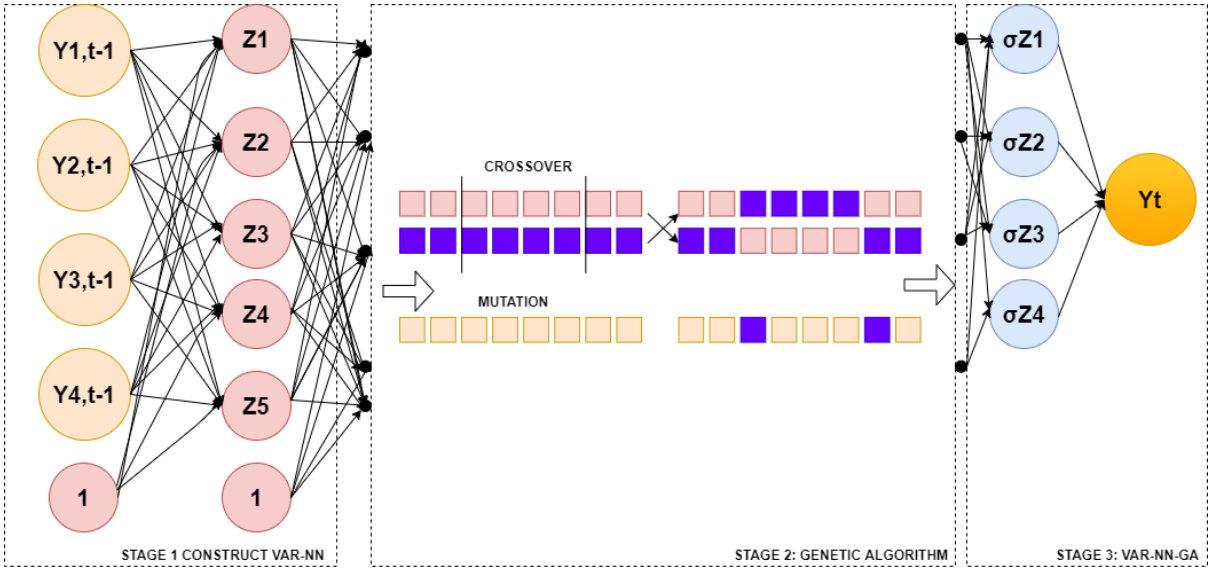

Figure 2. Proposed Hybrid Methods VAR-NN-GA

Algorithm 2 VAR-NN- GA
1: Input
Require training data $\{y_1, y_2, \dots, y_T\}$
Require Lag input *p* variable
Require GA Criteria (minimum MSE)
Require Number of Neuron in the hidden layer (n_i and n_h)
Require Activation function
2**: Generate** initial population
3: **Evaluate** population
4: **While {**stopping **GA criteria** not satisfied}
5: **Repeat**
6: {**for 1 to** (number of tournaments)
7: **Select** chromosome for tournament
8: **Find** lowest fitness
9: **remove** chromosome with lowest fitness
10: **crossover** (create new chromosome)}
11: **Mutation**
12: **Evaluation** (mutated chromosomes) }
13: use the parameters of FFNN (x_i, y_h, and z_o)
14: post processing data
15: **return** parameters of FFNN (x_i, y_h, and z_o)
16: Forecasting

The models developed were evaluated and conducted using the data test and the compute function respectively, which produces all the expected values. In addition, the tested and expected values were modified to real pollution data prior to normalization. Therefore, good results were produced by calculating the values of RMSE, MAE, and SMAPE. Also, by paying attention to the details, a good model is described as when the value of the three-consistency metrics used are close to 0, and the VAR parameter that was employed is the number of lags. Consequently, in some simulations, the values of RMSE and MAE are often low, but SMAPE yields different values. Conceptually, the SMAPE was achieved by subtracting the actual from the absolute prediction for each measured period, then square the result and find the square root of the previous estimate.

However, this procedure was imprecise for data that has zero value, for example, in climatology, 0 might mean that there is no phenomenon that occurred on that day or time. The genetic algorithm functions developed in this paper could be accessed in the appendix. In addition, **Table 2** shows that VAR-FFNN(BP)-GA gives an accurate forecast for PM_{10}, NO_X, SO_2, and $PM_{2.5}$.

The comparison for testing dataset was illustrated in the Figures above. In addition, the actual and predicted data of No_x, PM_{10}, $PM_{2.5}$, and SO_2 were given in (**Figures 3, 4, 5, and 6**) respectively. Even though, the expected value was distant from the real, they both follows the actual line. Besides, there are likely to be high error values due to many datasets, for example, the ranges of $PM_{2.5}$, PM_{10}, NO_X, and SO_2 are large, though there are some with zero values that are real, yet, making estimation is difficult. In addition, the zero value of the pollution variable is also far from the first quartile. Consequently, the data that could be included in the prediction results are missing. The general equation for forecasting pollution data in the four locations is represented in (Eq16), as a result, (Eq.17), (Eq.18), (Eq.19), and (Eq.20) were used for prediction purpose in this study areas.

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

Table 2. Models Comparison based on Pollution.

Pollution	Methods	Portion	Training			Testing			Average			Elapsed time
			RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	RMSE	MAE	SMAPE	
PM_{10}	VAR-OLS	80:20	13.1073	9.1243	0.0015	9.2268	6.6171	0.0075	11.16705	7.8707	0.0045	0.606828
		70:30	13.3261	9.3	0.0018	10.0568	7.1275	0.0062	11.69145	8.21375	0.004	0.465683
		60:40	13.5078	9.4491	0.0021	10.7137	7.6503	0.0048	12.11075	8.5497	0.00345	0.328773
	VAR-FFNN (BP)	80:20	12.9849	8.9858	3.6302	9.2998	6.5971	4.1837	11.14235	7.79145	3.90695	2.810811
	VAR-FFNN (BP)-GA*	80:20	13.0214	9.0262	3.6086	9.2151	6.5517	4.2210	11.1183	7.7890	3.9148	21.683287
$PM_{2.5}$	VAR-OLS	80:20	9.3542	6.574	0.002	6.8294	5.0888	0.0128	8.0918	5.8314	0.0074	0.610011
		70:30	9.6188	6.7839	0.0024	7.0345	5.2173	0.0071	8.32665	6.0006	0.00475	0.455977
		60:40	9.8556	6.9442	0.0028	7.3669	5.466	0.0048	8.61125	6.2051	0.0038	0.338347
	VAR-FFNN (BP)	80:20	9.2725	6.5062	3.8288	6.9241	5.1339	5.1339	8.0983	5.82005	4.48135	1.49469
	VAR-FFNN (BP)-GA*	80:20	9.19	6.3959	3.8103	6.8276	5.0316	3.5308	8.0088	5.71375	3.67055	24.210176
NOx	VAR-OLS	80:20	5.2721	3.3133	0.0024	3.7211	2.4615	0.0137	4.4966	2.8874	0.00805	0.731557
		70:30	5.3744	3.4148	0.0029	4.0092	2.5716	0.0069	4.6918	2.9932	0.0049	0.499683
		60:40	5.5415	3.5316	0.0035	4.0788	2.6498	0.0054	4.81015	3.0907	0.00445	0.447416
	VAR-FFNN (BP)	80:20	5.2084	3.2521	6.864	3.7083	2.4149	5.9453	4.45835	2.8335	6.40465	1.904269
	VAR-FFNN (BP)-GA*	80:20	5.1972	3.2383	6.7979	3.6839	2.3625	5.7141	4.44055	2.8004	6.256	20.657192
SO_2	VAR-OLS	80:20	1.5419	1.0526	0.0028	0.9347	0.714	0.0127	1.2383	0.8833	0.00775	0.628418
		70:30	1.6006	1.0934	0.0033	0.9922	0.7602	0.0065	1.2964	0.9268	0.0049	0.464132
		60:40	1.6582	1.1377	0.004	1.0688	0.8053	0.0054	1.3635	0.9715	0.0047	0.342162
	VAR-FFNN (BP)	80:20	1.5301	1.0466	3.4086	0.9227	0.7033	3.5282	1.2264	0.87495	3.4684	2.541875
	VAR-FFNN (BP)-GA*	80:20	1.5221	1.0421	3.3835	0.9207	0.6996	3.5535	1.2214	0.8709	3.4685	21.485214

General Equation for our models is

$$\widehat{\boldsymbol{Y}_t} = \psi_2\left\{v_0 + \sum_{k=1}^{5} v_0\,\psi_2\left\{w_0 + \sum_{j=1}^{4} w_{kj}\boldsymbol{Y}_{t-1}\right\}\right\};\ \widehat{\boldsymbol{Y}_t} = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \\ Y_{4,t} \end{bmatrix}, \boldsymbol{Y}_{t-1} = \begin{bmatrix} Y_{1.t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \\ Y_{4,t-1} \end{bmatrix} \qquad (16)$$

Forecasting NO_X using VAR-FFNN (BP)-GA is

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix} = \psi_2\left\{\begin{bmatrix}1.9175\\ -6.0002\\ -1.6315\\ -4.3720\end{bmatrix} + \begin{bmatrix}-0.0970 & 0.0143 & -2.4493 & 0.7231 & 0.2263\\ -0.8919 & -0.4898 & 5.1375 & 1.5031 & -0.3546\\ -0.3478 & -0.2637 & 0.7840 & 0.5540 & 0.0122\\ 0.2142 & 0.0332 & 3.9513 & -1.4269 & 1.3991\end{bmatrix}\left(\psi_1\left\{\begin{bmatrix}1.7907\\ -2.6447\\ 9.9445\\ 0.7890\\ 0.7318\end{bmatrix} + \begin{bmatrix}1.3704 & -1.2065 & 2.1402 & -0.0168\\ 0.1754 & 1.0877 & -3.8872 & -0.3911\\ -7.1599 & -3.2717 & 0.2860 & -2.1287\\ 0.9826 & -0.1159 & 0.2590 & 0.0177\\ 0.8529 & 0.0371 & 0.0936 & 0.4458\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right)\right\}$$

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix} = \begin{bmatrix}1.9175\\ -6.0002\\ -1.6315\\ -4.3720\end{bmatrix} + \begin{bmatrix}-0.0970 & 0.0143 & -2.4493 & 0.7231 & 0.2263\\ -0.8919 & -0.4898 & 5.1375 & 1.5031 & -0.3546\\ -0.3478 & -0.2637 & 0.7840 & 0.5540 & 0.0122\\ 0.2142 & 0.0332 & 3.9513 & -1.4269 & 1.3991\end{bmatrix}\left(tansig\left\{\begin{bmatrix}1.7907\\ -2.6447\\ 9.9445\\ 0.7890\\ 0.7318\end{bmatrix} + \begin{bmatrix}1.3704 & -1.2065 & 2.1402 & -0.0168\\ 0.1754 & 1.0877 & -3.8872 & -0.3911\\ -7.1599 & -3.2717 & 0.2860 & -2.1287\\ 0.9826 & -0.1159 & 0.2590 & 0.0177\\ 0.8529 & 0.0371 & 0.0936 & 0.4458\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right) \quad (17)$$

Forecasting $PM_{2.5}$ using VAR-FFNN (BP)-GA

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix} = \psi_2\left\{\begin{bmatrix}6.8828\\ 7.1349\\ 8.1110\\ 0.4537\end{bmatrix} + \begin{bmatrix}0.5423 & -0.0970 & -0.1403 & 0.1027 & -7.3576\\ 0.3842 & 0.6266 & -0.0419 & -0.5936 & -7.3618\\ 0.3571 & 0.4513 & -0.0487 & -0.02283 & -8.4794\\ 0.0628 & 0.0490 & -0.3886 & 0.0795 & -0.7081\end{bmatrix}\left(\psi_1\left\{\begin{bmatrix}0.5834\\ 0.5483\\ -0.4701\\ 1.3593\\ 10.1354\end{bmatrix} + \begin{bmatrix}1.2811 & 0.0434 & 0.1169 & -0.2718\\ -0.4915 & -0.5563 & 1.8954 & 0.3924\\ -0.0993 & -0.4873 & 0.6237 & -2.4012\\ 0.5437 & -1.4507 & 2.0842 & 0.5177\\ -1.8870 & 5.2310 & -3.5550 & 5.2919\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right)\right\}$$

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix}=\begin{bmatrix}6.8828\\ 7.1349\\ 8.1110\\ 0.4537\end{bmatrix}+$$

$$\begin{bmatrix}0.5423 & -0.0970 & -0.1403 & 0.1027 & -7.3576\\ 0.3842 & 0.6266 & -0.0419 & -0.5936 & -7.3618\\ 0.3571 & 0.4513 & -0.0487 & -0.02283 & -8.4794\\ 0.0628 & 0.0490 & -0.3886 & 0.0795 & -0.7081\end{bmatrix}\left(tansig\left\{\begin{bmatrix}0.5834\\ 0.5483\\ -0.4701\\ 1.3593\\ 10.1354\end{bmatrix}+\right.\right.$$

$$\left.\left.\begin{bmatrix}1.2811 & 0.0434 & 0.1169 & -0.2718\\ -0.4915 & -0.5563 & 1.8954 & 0.3924\\ -0.0993 & -0.4873 & 0.6237 & -2.4012\\ 0.5437 & -1.4507 & 2.0842 & 0.5177\\ -1.8870 & 5.2310 & -3.5550 & 5.2919\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right) \quad (18)$$

Forecasting PM_{10} using VAR-FFNN (BP)-GA is

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix}=\psi_2\left\{\begin{bmatrix}-7.0571\\ -2.2792\\ 5.6476\\ -4.7282\end{bmatrix}+\right.$$

$$\begin{bmatrix}-0.0300 & -0.2660 & -7.1537 & -0.1195 & -1.6996\\ 0.9963 & 0.3288 & -2.9428 & 0.6457 & -2.0250\\ 1.0267 & 0.1429 & 6.8891 & -0.8146 & -1.6718\\ 0.6962 & -0.3004 & -10.9293 & 6.0641 & -1.5019\end{bmatrix}\left(\psi_1\left\{\begin{bmatrix}0.3618\\ -0.6761\\ -2.0482\\ -11.7361\\ 0.1141\end{bmatrix}+\right.\right.$$

$$\left.\left.\left.\begin{bmatrix}-0.5184 & -0.1385 & 0.3463 & 0.7161\\ -0.4403 & 1.3597 & -0.7319 & -1.2717\\ 0.2805 & -0.4054 & 0.5048 & -0.3257\\ 2.8764 & -0.2176 & 0.8796 & -9.2759\\ -0.4193 & -0.0748 & -0.0723 & 0.2072\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right)\right\}$$

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix}=\begin{bmatrix}-7.0571\\ -2.2792\\ 5.6476\\ -4.7282\end{bmatrix}+$$

$$\begin{bmatrix}-0.0300 & -0.2660 & -7.1537 & -0.1195 & -1.6996\\ 0.9963 & 0.3288 & -2.9428 & 0.6457 & -2.0250\\ 1.0267 & 0.1429 & 6.8891 & -0.8146 & -1.6718\\ 0.6962 & -0.3004 & -10.9293 & 6.0641 & -1.5019\end{bmatrix}\left(tansig\left\{\begin{bmatrix}0.3618\\ -0.6761\\ -2.0482\\ -11.7361\\ 0.1141\end{bmatrix}+\right.\right.$$

$$\left.\left.\begin{bmatrix}-0.5184 & -0.1385 & 0.3463 & 0.7161\\ -0.4403 & 1.3597 & -0.7319 & -1.2717\\ 0.2805 & -0.4054 & 0.5048 & -0.3257\\ 2.8764 & -0.2176 & 0.8796 & -9.2759\\ -0.4193 & -0.0748 & -0.0723 & 0.2072\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right) \quad (19)$$

Forecasting SO_2 using VAR-FFNN (BP)-GA is

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix} = \psi_2\left\{\begin{bmatrix}0.7886\\ -6.6331\\ 4.9191\\ 0.4431\end{bmatrix} + \begin{bmatrix}-0.0261 & 0.3170 & 3.6011 & 2.4089 & -0.3611\\ 0.1399 & 0.1987 & 8.8856 & 15.2573 & -0.1413\\ 0.0603 & 0.0172 & 3.3978 & -2.1660 & -0.2668\\ -0.0043 & -0.2524 & 4.0224 & 3.2053 & -0.2737\end{bmatrix}\left(\psi_1\left\{\begin{bmatrix}2.1996\\ -0.0193\\ -2.3356\\ 2.0016\\ -1.0644\end{bmatrix} + \begin{bmatrix}-1.3017 & 1.4578 & 4.1478 & -1.0995\\ 0.9824 & 0.1957 & 0.0667 & -1.9023\\ 0.0541 & 0.7071 & 0.5233 & -1.0439\\ -0.0151 & -0.0513 & -0.3781 & 0.4320\\ -0.7927 & 0.3619 & -0.4259 & -1.7062\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right)\right\}$$

$$\begin{bmatrix}\widehat{Y_{1,t}}\\ \widehat{Y_{2,t}}\\ \widehat{Y_{3,t}}\\ \widehat{Y_{4,t}}\end{bmatrix} = \begin{bmatrix}0.7886\\ -6.6331\\ 4.9191\\ 0.4431\end{bmatrix} + \begin{bmatrix}-0.0261 & 0.3170 & 3.6011 & 2.4089 & -0.3611\\ 0.1399 & 0.1987 & 8.8856 & 15.2573 & -0.1413\\ 0.0603 & 0.0172 & 3.3978 & -2.1660 & -0.2668\\ -0.0043 & -0.2524 & 4.0224 & 3.2053 & -0.2737\end{bmatrix}\left(tansig\left\{\begin{bmatrix}2.1996\\ -0.0193\\ -2.3356\\ 2.0016\\ -1.0644\end{bmatrix} + \begin{bmatrix}-1.3017 & 1.4578 & 4.1478 & -1.0995\\ 0.9824 & 0.1957 & 0.0667 & -1.9023\\ 0.0541 & 0.7071 & 0.5233 & -1.0439\\ -0.0151 & -0.0513 & -0.3781 & 0.4320\\ -0.7927 & 0.3619 & -0.4259 & -1.7062\end{bmatrix}\begin{bmatrix}Y_{1.t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\\ Y_{4,t-1}\end{bmatrix}\right\}\right) \quad (20)$$

Climatologically, $PM_{2.5}$ and PM_{10} concentrations determines the solar radiation intensity, because the higher the edge of the radiation, the more the surface temperature. Furthermore, the average temperature of the earth during the day is higher than that of the air, implying that temperature is passed from the earth's surface to the air. In addition, climatic conditions are becoming unpredictable due to the impact of pollutants dispersed. According to the data obtained from the four study areas, the concentration of PM_{10} at night till the morning tends to be constant, moreover, it rises during the normal working hours. Therefore, the constant concentration of PM_{10} was affected by the stable atmosphere at night. Based on the characteristics of pollution in Taipei, Hsinchu, Taichung, and Kaohsiung, the concentration of $PM_{2.5}$ and PM_{10} will significantly reduce unless the air temperature is increased. Also, the high air temperatures throughout the day depends on the solar surface heating, which could cause gusts of wind that mostly have an impact on meteorological conditions. Furthermore, the increase in the surface temperature could make the atmosphere unpredictable, as a result, the PM_{10} could be contaminated, and its concentration will be massively reduced.

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

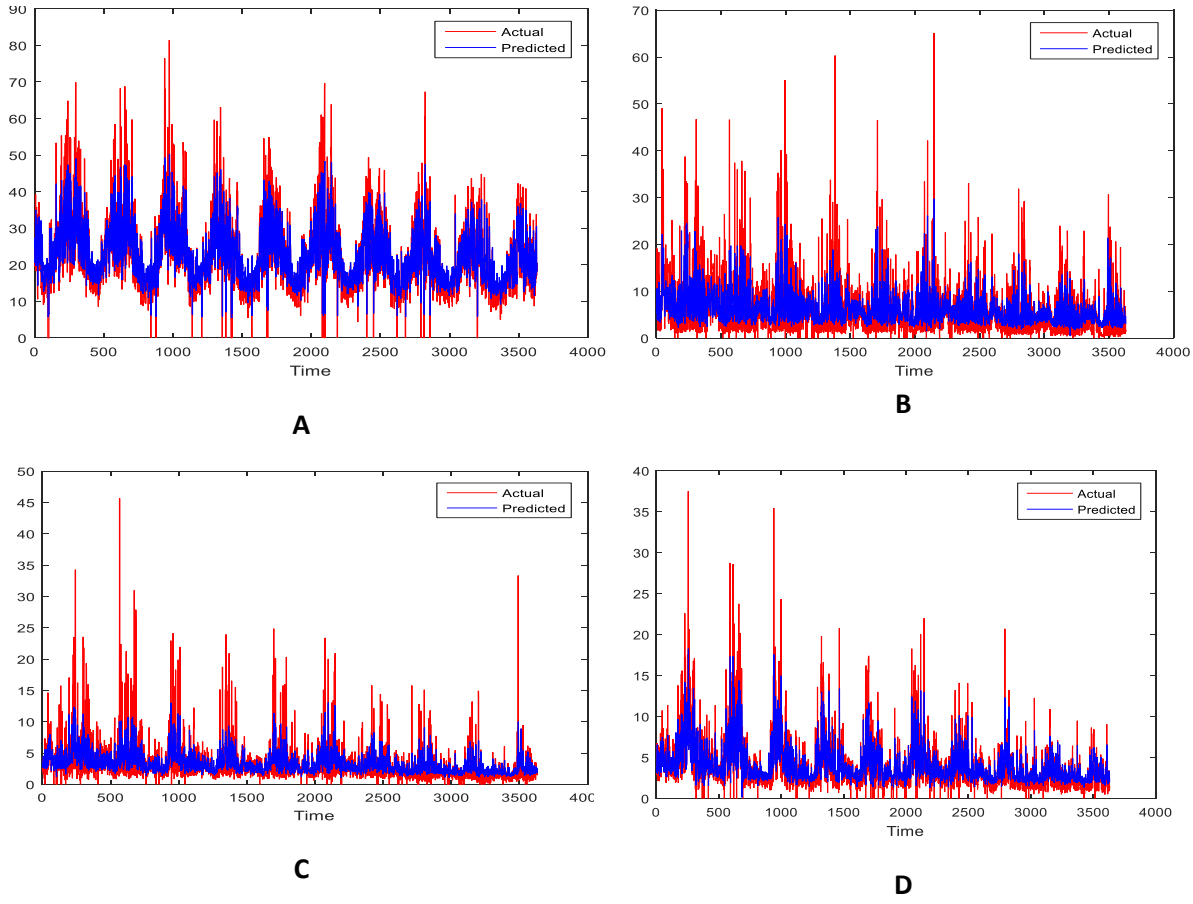

Figure 3. NO_X VAR Train FFNN using Genetic Algorithm and Backpropagation in Taichung (A), Taipei (B), Hsinchu (C), Kaohsiung (D).

Figure 7 shows all the findings measured from the 4 locations chosen for the forecast in Taiwan, namely, Taichung, Taipei, Hsinchu and Kaohsiung. However, it was discovered that Taichung is having the highest rate of emissions in PM_{10}, $PM_{2.5}$, SO_2, and NO_x. Consequently, the VAR-FFNN-GA is not completely biased in estimating the max and min values, nevertheless, the model could predict the data pattern very well. Long story short, VAR-FFNN-GA provided good accuracy is shown in **Figure 8.**

Generally, when there are fewer raining days during winter in the central and southern regions, air pollution is more problematic. Furthermore, when the quality of air is bad, the concentration of PM10, PM2.5, SO2, and NOx is increased, and the level of pollution would be very high. Also, news reports had revealed that the northeast Monsoon often brings haze from China to Taiwan, which has made Taiwan's air quality worse. The reason for this is because, pollution is only serious in few areas of the country, however, when the index burst across the country at the same time, the source is usually related to the overseas, such as China. To analyze this problem, Kaohsiung and Taipei City was taken as an example, because they are often affected by the overseas air pollution.

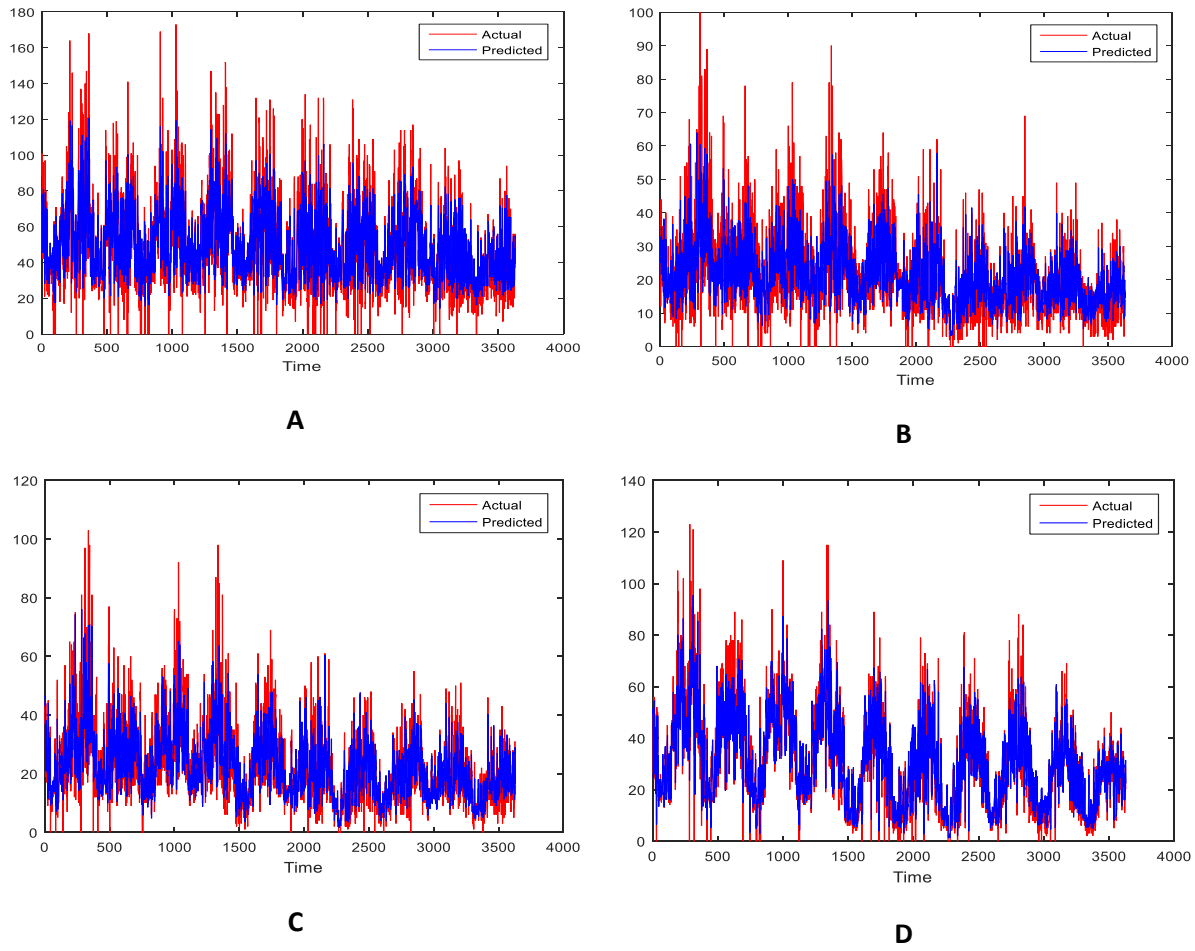

Figure 4. PM_{10} VAR Train FFNN using Backpropagation in Taichung (A), Taipei (B), Hsinchu (C), Kaohsiung (D).

Taiwan has developed so far, both economically and socially, nevertheless, deep thought should be taken on how to further advance in these areas. In doing this, the economy and environmental protection may conflict because there are many factors to be considered. For instance, as the role of natural science is very important, that of humanity should not be ignored.

Therefore, it is the role of government to develop and implement relevant policies on environmental protection and public engagement to mitigate pollution. Considering the government of Taipei as an example, this year's carbon control initiatives of the Environmental Protection Authority include, replacing old diesel cars with modern ones, creating a friendly environment for hybrid vehicles, going to subsidize the replacements of two-stroke locomotives for vulnerable ones, increasing the use of green transport, and including anti-pollution facilities in the food service industry.

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

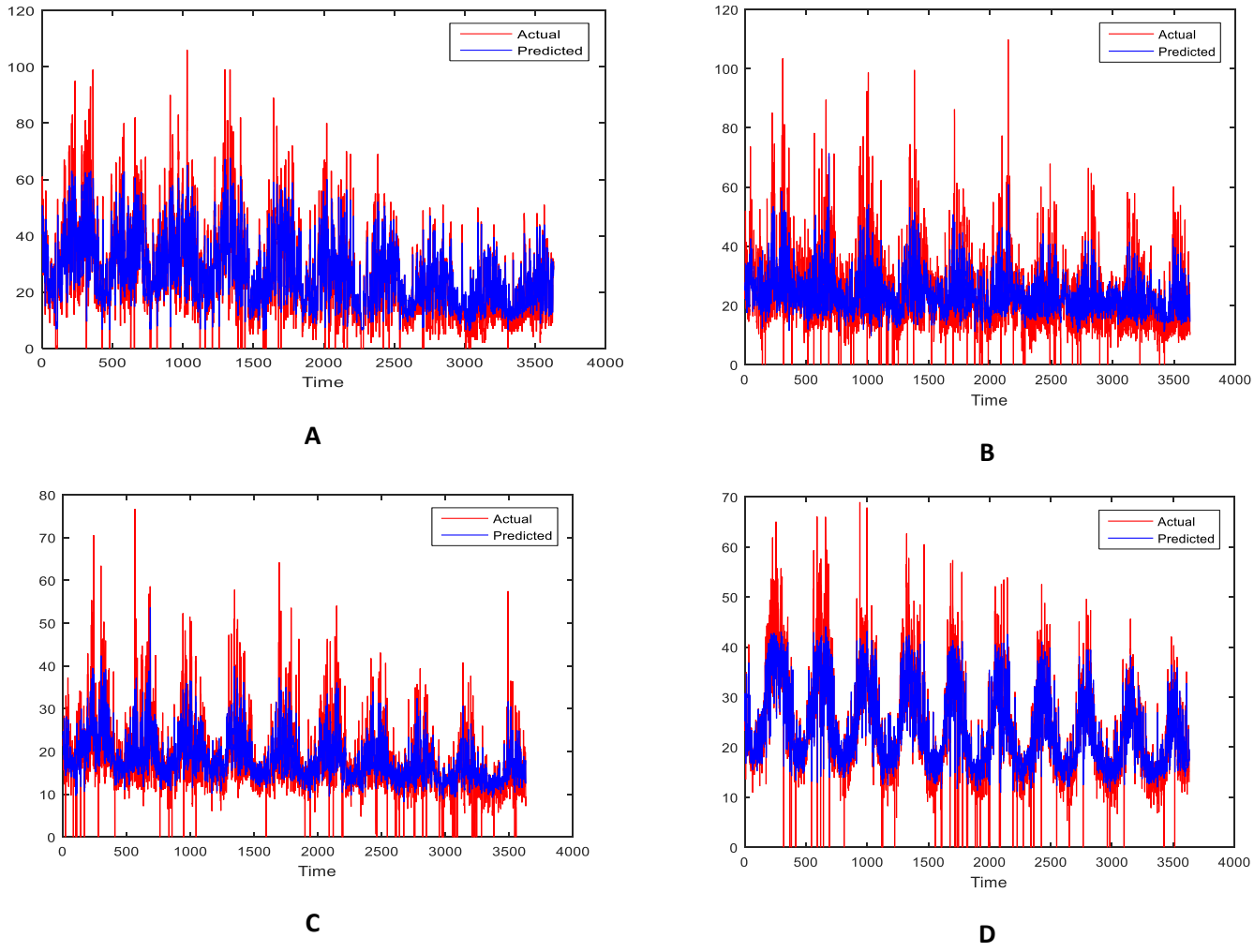

Figure 5. $PM_{2.5}$ VAR Train FFNN using Genetic Algorithm and Backpropagation in Taichung (A), Taipei (B), Hsinchu (C), Kaohsiung (D).

Although the government has taken the necessary actions, however, in order to cooperate with the emission reduction policies, the citizen need to play the role of supervising the relevant decisions of the central and local governments. Taiwan is a democratic country; however, this should not only be practiced only on the day of voting, rather, it should be implemented every day of life. Therefore, by continuing to understand related issues, paying attention to related policy developments, and even participating in various actions actually means voting for a common future.

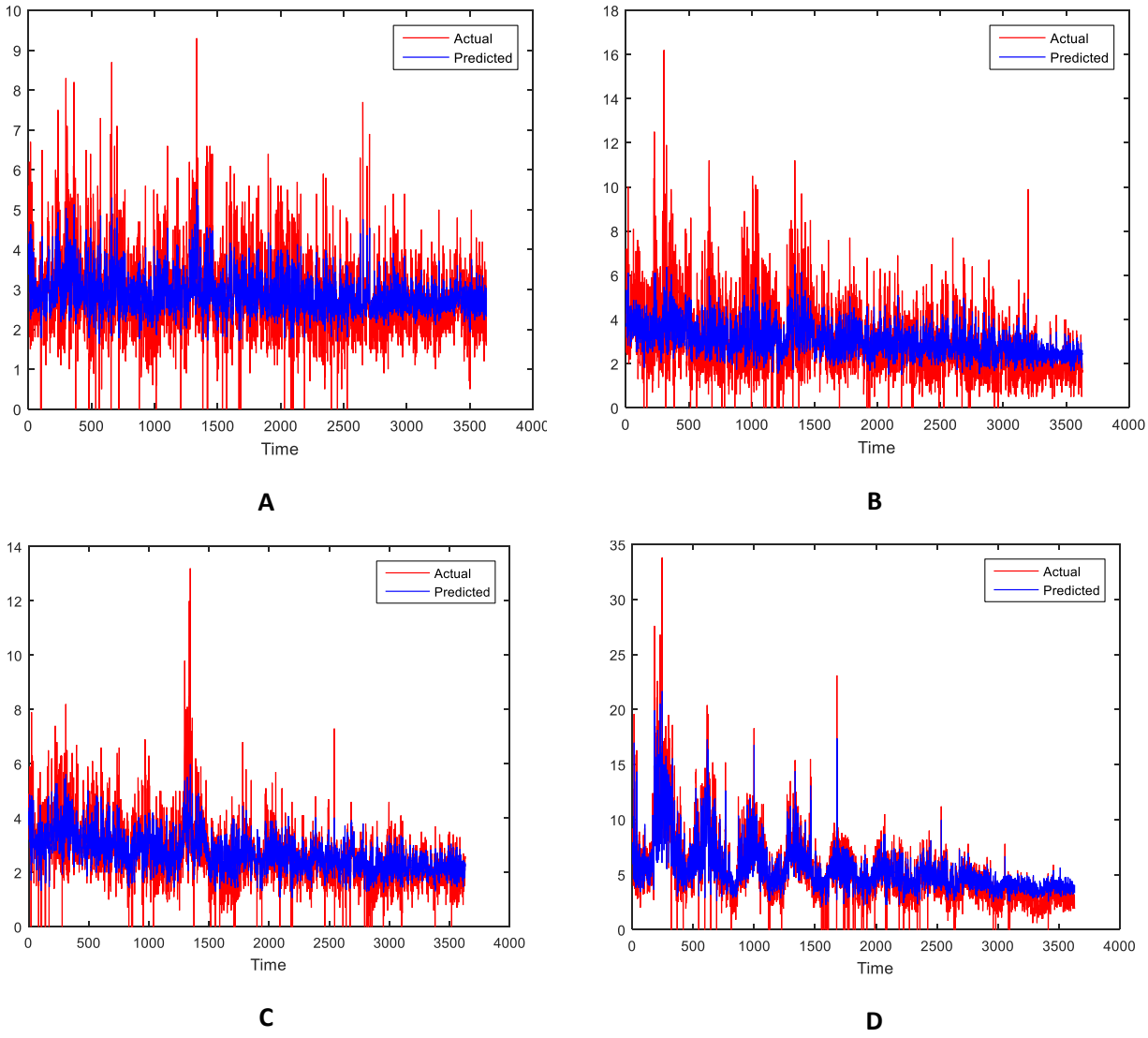

Figure 6. SO_2 VAR Train FFNN using Backpropagation in Taichung (A), Taipei (B), Hsinchu (C), Kaohsiung (D).

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

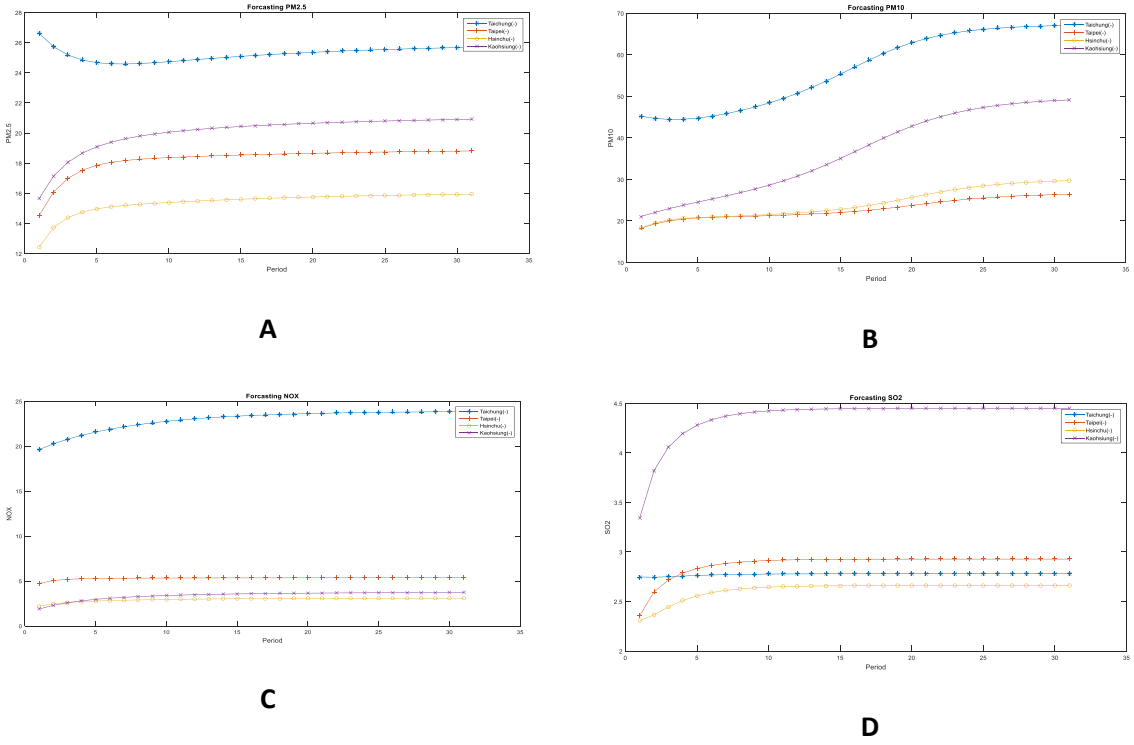

Figure 7. Forecasting 30 days ahead in Taichung (A), Taipei (B), Hsinchu (C), Kaohsiung (D).

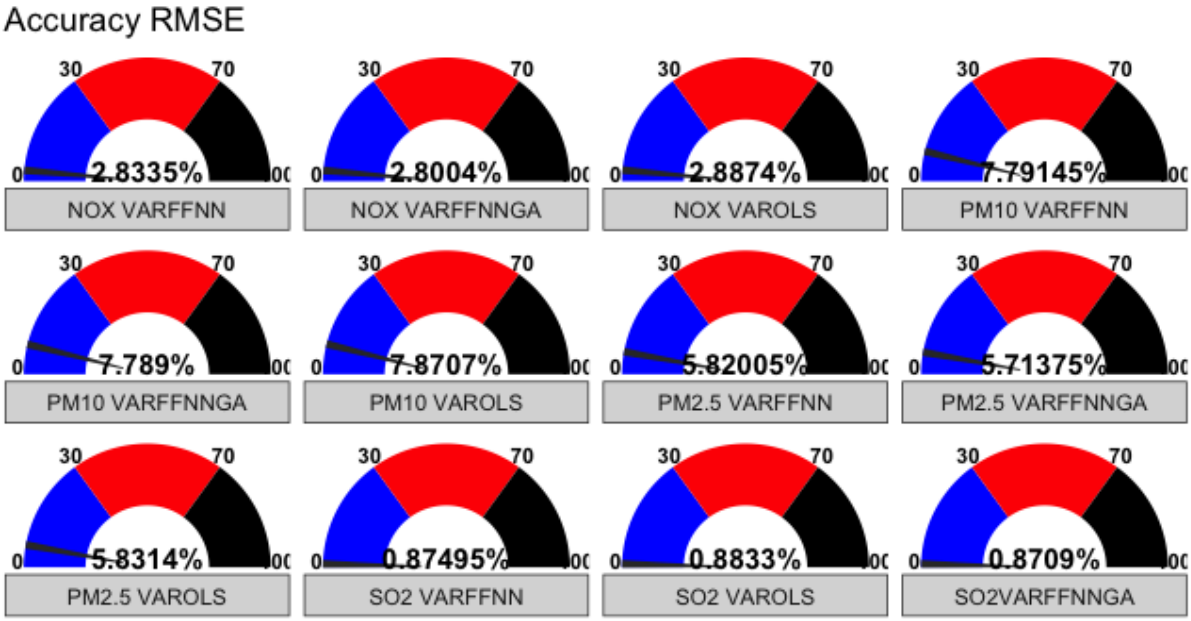

Figure 8. Accuracy RMSE from All Dataset.

4. CONCLUSION

In this study, the relationship pattern employed is the non-linear using the FFNN model. Furthermore, each output variable is associated with the input and weight parameter which is optimized by GA, in order to obtain the smallest possible error. In addition, this study combined vector autoregressive with neural network, and also employ metaheuristics genetic algorithm. Based on this construction, the model aimed

at providing good accuracy rather than making long-term prediction, resulting in one point. Taichung and Kaohsiung have the highest level of pollutant, compared to Taipei and Hsinchu. Also, after running the analysis, the ratio of bad air quality days to the total number monitored in these four cities, showed the highest in Taiwan. Meanwhile, some of the heaviest population densities in these four areas are located in the various night markets. Based on the current air pollution control of the Environmental Protection Agency, the proportion of bad air quality events to the number of days counted in Taichung and Kaohsiung are perhaps the highest in Taiwan. Long story short, there are limitations in this study due to the seasonal patterns in these locations. As a result, the data are not stationer in mean, and need to be differentiated yearly, for example D=1, S=364 or 365. Furthermore, the time series plotted in fig 3, 4, and 5 showed clearly that the data have seasonal pattern due to the areas. Therefore, it means that, VAR (1) or VAR (1)-NN-GA would not cover this issue and tend to yield flat forecast in long term. Based on this, the future study should perform Seasonal VAR and perform the advanced technology and innovations aimed at reducing pollution.

5. ACKNOWLEDGEMENTS

This research fully supported by Ministry of Science and Technology, Taiwan [MOST-109-2622-E-324-004]. This research in part of Chaoyang University of Technology and the Higher Education Sprout Project, Ministry of Education (MOE), Taiwan, under the project name: "The RandD and the cultivation of talent for health-enhancement products". This research supported by National Research Foundation of Korea grants [NRF-2019R1A2C1002408]. The datasets provided by the Environmental Protection Administration of Taiwan are acknowledged.

6. AUTHOR'S NOTE

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The analysis codes and datasets used in this paper available from the corresponding author upon reasonable request. The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

7. REFERENCES

AL-Dhurafi, N. A., Masseran, N., and Zamzuri, Z. H. (2018). Compositional time series analysis for air pollution index data. *Stochastic Environmental Research and Risk Assessment*, *32*(10), 2903-2911.

Ang, A., and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, *50*(4), 745-787.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, *120*(1), 387-422.

Brook, R. D., Rajagopalan, S., Pope III, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., and Kaufman, J. D. (2010). Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation*, *121*(21), 2331-2378.

Caraka, R. E., Bakar, S. A., Tahmid, M., Yasin, H., and Kurniawan, I. D. (2019). Neurocomputing fundamental climate analysis. *Telkomnika*, *17*(4), 1818-1827.

Caraka, R. E., Chen, R. C., Toharudin, T., Pardamean, B., Yasin, H., and Wu, S. H. (2019). Prediction of status particulate matter 2.5 using state Markov chain stochastic process and HYBRID VAR-NN-PSO. *IEEE Access*, *7*, 161654-161665.

Caraka, R. E., Lee, Y., Kurniawan, R., Herliansyah, R., Kaban, P. A., Nasution, B. I., and Pardamean, B. (2020). Impact of COVID-19 large scale restriction on environment and economy in Indonesia. *Global Journal of Environmental Science and Management*, *6*(Special Issue (Covid-19)), 65-84.

Caraka, R. E., Sugiyarto, W., Erda, G., and Sadewo, E. (2016). Pengaruh inflasi terhadap impor dan ekspor di provinsi riau dan kepulauan riau menggunakan generalized spatio time series. *Jurnal BPPK: Badan Pendidikan dan Pelatihan Keuangan*, *9*(2), 180-198.

Chang, F. J., Chang, L. C., Kang, C. C., Wang, Y. S., and Huang, A. (2020). Explore spatio-temporal PM2.5 features in northern Taiwan using machine learning techniques. *Science of the Total Environment*, *736*(2020), 1–14.

Chen, R. C., Dewi, C., Huang, S. W., and Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 1-26.

De Vito, S., Piga, M., Martinotto, L., and Di Francia, G. (2009). CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic Bayesian regularization. *Sensors and Actuators B: Chemical*, *143*(1), 182-191.

Ebtehaj, I., and Bonakdari, H. (2016). Bed load sediment transport estimation in a clean pipe using multilayer perceptron with different training algorithms. *KSCE Journal of Civil Engineering*, *20*(2), 581-589.

Ebtehaj, I., Bonakdari, H., and Zaji, A. H. (2016). An expert system with radial basis function neural network based on decision trees for predicting sediment transport in sewers. *Water Science and Technology*, *74*(1), 176-183.

El Brahmi, A., Abderafi, S., and Ellaia, R. Artificial neural network analysis of sulfide production in a Moroccan sewerage network. *Indonesian Journal of Science and Technology*, *6*(1), 193-204.

Feng, S., Gao, D., Liao, F., Zhou, F., and Wang, X. (2016). The health effects of ambient PM2. 5 and potential mechanisms. *Ecotoxicology and environmental safety*, *128*, 67-74.

Fischer, T., and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654-669.

Hyndman, R. J., and Koehler, A. B. (2006). Effect of Question Formats on Item Endorsement Rates in Web Surveys. *International Journal of Forecasting*, *22*(4), 679-688.

Jan, G. (2006). De Gooijer and Rob J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, *22*(3), 443-473.

Kaban, P. A., Kurniawan, R., Caraka, R. E., Pardamean, B., and Yuniarto, B. (2019). Biclustering method to capture the spatial pattern and to identify the causes of social vulnerability in Indonesia: a new recommendation for disaster mitigation policy. *Procedia Computer Science*, *157*, 31-37.

Kravtsov, S., Kondrashov, D., and Ghil, M. (2005). Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *Journal of Climate*, *18*(21), 4404-4424.

Kurniawan, R., Siagian, T. H., Yuniarto, B., Nasution, B. I., and Caraka, R. E. (2018). Construction of social vulnerability index in Indonesia using partial least squares structural equation modeling. *International Journal of Engineerng and Technology*, *7*(4), 6131-6136.

Lanne, M., and Luoto, J. (2020). Identification of economic shocks by inequality constraints in Bayesian structural vector autoregression. *Oxford Bulletin of Economics and Statistics*, *82*(2), 425-452.

Lee, Y., Ronnegard, L., and Noh, M. (2017). *Data analysis using hierarchical generalized linear models with R*. CRC Press.

Makridakis, S., and Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451-476.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2020a). Predicting/hypothesizing the findings of the M4 competition. *International Journal of Forecasting*, *36*(1), 29-36.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2020b). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54-74.

Masseran, N., and Safari, M. A. M. (2020). Risk assessment of extreme air pollution based on partial duration series: IDF approach. *Stochastic Environmental Research and Risk Assessment*, *34*(3), 545-559.

Nasution, B. I., Kurniawan, R., Siagian, T. H., and Fudholi, A. (2020). Revisiting social vulnerability analysis in Indonesia: an optimized spatial fuzzy clustering approach. *International Journal of Disaster Risk Reduction*, *51*, 101801.

Ng, S. M., Angelier, J., and Chang, C. P. (2009). Earthquake cycle in Western Taiwan: insights from historical seismicity. *Geophysical Journal International*, *178*(2), 753-774.

Paoletti, E., De Marco, A., Anav, A., Gasparini, P., and Pompei, E. (2018). Five-year volume growth of European beech does not respond to ozone pollution in Italy. *Environmental Science and Pollution Research*, *25*(9), 8233-8239.

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |

Qasem, S. N., Ebtehaj, I., and Bonakdari, H. (2017). Potential of radial basis function network with particle swarm optimization for prediction of sediment transport at the limit of deposition in a clean pipe. *Sustainable Water Resources Management*, *3*(4), 391-401.

Searle, K., and Gow, K. (2010). Do concerns about climate change lead to distress? *International Journal of Climate Change Strategies and Management*, *2*(4), 362-379.

Siagian, T. H., Purhadi, P., Suhartono, S., and Ritonga, H. (2014). Social vulnerability to natural hazards in Indonesia: driving factors and policy implications. *Natural hazards*, *70*(2), 1603-1617.

Singh, P., and Huang, Y. P. (2019). A high-order neutrosophic-neuro-gradient descent algorithm-based expert system for time series forecasting. *International Journal of Fuzzy Systems*, *21*(7), 2245-2257.

Suhartono, D. E. A., Prastyo, D. D., Kuswanto, H., and Lee, M. H. (2019). Deep Neural Network for Forecasting Inflow and Outflow in Indonesia. *Sains Malaysiana*, *48*(8), 1787-1798.

Suhartono, Maghfiroh, B., and Rahayu, S. P. (2019). Hybrid VARX-SVR and GSTARX-SVR for forecasting spatio-temporal data. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(4), 212–218.

Suhartono, S. (2005). A comparative study of forecasting models for trend and seasonal time series does complex model always yield better forecast than simple models. *Jurnal Teknik Industri*, *7*(1), 22-30.

Suhartono, S., and Subanar, S. (2006). The optimal determination of space weight in GSTAR model by using cross-correlation inference. *Quantitative Methods*, *2*(2), 45-53.

Suhartono, S., and Subanar, S. (2007). Some comments on the theorem providing stationarity condition for gstar models in the paper by Borovkova Et Al. *Jurnal of The Indonesian Mathematical Society*, *13*(1), 115-121.

Suhartono, S., Prastyo, D. D., Kuswanto, H., and Lee, M. H. (2018). Comparison between VAR, GSTAR, FFNN-VAR and FFNN-GSTAR models for forecasting oil production. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, *34*(1), 103-111.

Suhartono, Wahyuningrum, S. R., Setiawan, and Akbar, M. S. (2016). GSTARX-GLS model for spatio-temporal data forecasting. *Malaysian Journal of Mathematical Sciences*, *10*, 91–103.

Suhartono. (2011). Time series forecasting by using seasonal autoregressive integrated moving average: subset, multiplicative or additive model. *Journal of Mathematics and Statistics, 7*, 20-27.

Supari, Tangang, F., Juneng, L., and Aldrian, E. (2017). Observed changes in extreme temperature and precipitation over Indonesia. *International Journal of Climatology*, *37*(4), 1979–1997.

Tang, Z., and Fishwick, P. A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA journal on computing*, *5*(4), 374-385.

Timbal, B., Hope, P., and Charles, S. (2008). Evaluating the consistency between statistically downscaled and global dynamical model climate change projections. *Journal of Climate*, *21*(22), 6052-6059.

Toharudin, T., Pontoh, R. S., Caraka, R. E., Zahroh, S., Lee, Y., and Chen, R. C. (2020). Employing long short-term memory and Facebook prophet model in air temperature forecasting. *Communications in Statistics-Simulation and Computation*, 1-24.

Wang, W. C., Chau, K. W., Cheng, C. T., and Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, *374*(3–4), 294–306.

Warsito, B., Yasin, H., and Prahutama, A. (2019). Particle swarm optimization to obtain weights in neural network. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, *35*(3).

Whitley, D., and Tutorial, A. G. A. (1994). Statistics and computing, 4. *Kluwer Academic Publishers*, *4*(2), 65-85.

Yang, H. L., and Lin, H. C. (2016). An integrated model combined ARIMA, EMD with SVR for stock indices forecasting. *International Journal on Artificial Intelligence Tools*, *25*(02), 1650005.

Yasin, H., Warsito, B., and Santoso, R. (2018). Feed forward neural network modeling for rainfall prediction. *E3S Web of Conferences*, 1–5.

Yu, G., and Schwartz, Z. (2006). Forecasting short time-series tourism demand with artificial intelligence models. *Journal of travel Research*, *45*(2), 194-203.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.

Zuhairoh, F., and Rosadi, D. (2020). Real-time Forecasting of the COVID-19 Epidemic using the Richards Model in South Sulawesi, Indonesia. *Indonesian Journal of Science and Technology*, *5*(3), 456-462.

DOI: https://doi.org/10.17509/ijost.v6i1.32732 |
p- ISSN 2528-1410 e- ISSN 2527-8045 |