IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 13, 2019, accepted October 9, 2019, date of publication October 30, 2019, date of current version November 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950439

Prediction of Status Particulate Matter 2.5 Using State Markov Chain Stochastic Process and HYBRID VAR-NN-PSO

REZZY EKO CARAKA[1,2,3], RUNG CHING CHEN[1], TONI TOHARUDIN[2], BENS PARDAMEAN[3,4], HASBI YASIN[5], AND SHIH HUNG WU[1]

[1]Department of Information Management, Chaoyang University of Technology, Taichung 41349, Taiwan
[2]Department of Statistics, Padjadjaran University, Sumedang 43563, Indonesia
[3]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia
[4]Computer Science Department, Bina Nusantara University, Jakarta 11480, Indonesia
[5]Department of Statistics, Diponegoro University, Semarang 50251, Indonesia

Corresponding author: Rung Ching Chen (crching@cyut.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST-107-2221-E-324-018-MY2, and part of Padjadjaran University, Indonesia, and by part of Bina Nusantara University, Indonesia.

ABSTRACT Air pollution is the entry or inclusion of living things, energy substances, and other components into the air. Moreover, Air pollution is the presence of one or several contaminants in the outside atmospheric air such as dust, foam, gas, fog, smoke or steam in large quantities with various properties and time intervals of the contaminants in the air resulting in disturbances to the lives of humans, plants or animals. One of the parameters measured in determining air quality is $PM_{2.5}$. However, $PM_{2.5}$ has a higher probability of being able to enter the lower respiratory tract because small particle diameters can potentially pass through the lower respiratory tract. In this paper, we will get two different insight. First, the probability of status change using Markov chain and second, forecasting by using VAR-NN-PSO. More details we classify by three classifications no risk (1-30), medium risk (30-48), and moderate (>49) in Pingtung and Chaozhou. This data is starting from January 2014 to May 2019 and it can be modeled with the Markov chain. At the same time, we perform Hybrid VAR-NN-PSO to forecast $PM_{2.5}$ in Pingtung and Chaozhou. In this optimization, the search for best solutions is carried out by a population consisting of several particles. Based on the results of the discussion, opportunities for the transition from monthly status change are obtained continuous stochastic time with a stationary probability distribution. Regarding the VAR-NN-PSO, we obtained the mean absolute percentage error (MAPE) 3.57% for $PM_{2.5}$ data in Pingtung and 4.87% for $PM_{2.5}$ data in Chaozhou, respectively. This model can be predicted to forecasting 180 days ahead. Besides, the population in PSO has generated randomly with the smallest value and the high value the accuracy.

INDEX TERMS $PM_{2.5}$, Markov chain, stochastic, VAR, PSO, neural network.

I. INTRODUCTION

Air pollution is a change in the composition of air substances so that the air quality of these substances becomes reduced. Polluted air quality generally contains pollutant air with compositions such as COx, NOx, SOx, SPM (suspended particular matter), Ox, and various heavy metals. The level of excessive concentration of pollutants that exceed the tolerance threshold that is permitted. It will have a negative impact that is harmful to the environment, humans, plants, animals, and damage to materials and affect the quality of rainwater (acid rain). Poor air quality or continuous high air pollution can adversely affect the earth and health. The long-term negative impact of poor air quality is ozone depletion, which triggers global warming, while the short-term impact directly on humans is respiratory health problems [1]. Therefore, air quality is an important thing always to monitor [2]. Moreover, it increased exhaust emissions from motorized vehicles, and industrial activities are growing. In urban areas, the level of air pollution results in environmental problems is threatening living creatures because it almost exceeds ambient air quality standards [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

Particle pollution or Particulate matter is a mixture of solid particles and water droplets that can be found in the air. Some types of particles such as dust, dirt, soot or smoke are large enough or dark enough to be seen by the eye, while most others are so tiny that only an electron microscope can detect them [4]. These particles can have various shapes and sizes and can be formed from hundreds of different types of chemicals [5]. The size of particulates in the atmosphere is usually divided into two types — fine particles which have a size smaller than 2.5 micrometers [6] which commonly called $PM_{2.5}$. The coarse particles are more extensive than 2.5 micrometers and smaller than 10 micrometers which often called PM_{10} [7]. Particulates can be divided into primary particles and secondary particles [8]. Primary particles of main particles are pollutant particles emitted directly from sources of emissions, such as development sites, repairs of roads, fields, and forest fires. Secondary particles are particles formed by complicated chemical reactions that transform gases into atmospheric particles such as sulfur dioxide and nitrogen oxides emitted by energy, industrial and motorized vehicles. Particulates emitted into the atmosphere will undergo a process of changing their shape, size, and chemical composition by several mechanisms that will continue to occur until the particles undergo a deposition process. The existence of a transformation mechanism causes the average residence time of particulates in the lower atmosphere to last up to one week. The height influences the particulate deposition process, at low altitudes it usually occurs deposition or dry deposition directly, whereas at the height of more than 100 meters usually occurs wet deposition.

$PM_{2.5}$ types that contain microscopic solid concentrates or tiny liquid droplets can enter the respiratory system to the lungs and can cause serious health problems that can even lead to lung and coronary heart cancer. $PM_{2.5}$ also causes environmental issues such as reduced visibility. One of the factors causing acid rain disturbs the balance of the ecosystem because it can damage the nutrient balance of soil and plants. Moreover, PM_{10} is a solid or liquid particulate that floats in the air with an aerodynamic diameter size of fewer than 10 microns. PM_{10} is more specific is respirable particulate matter and a good predictor of health. PM_{10} has a higher probability of being able to enter the lower respiratory tract because the small particle diameter can potentially pass through the lower respiratory tract. However, particulate matter (PM) is a heterogeneous mixture that varies in physical and chemical properties which depend on meteorological conditions and sources of emissions. Current air quality standards use PM mass concentrations. PMs with an aerodynamic diameter of $\leq$10m (PM_{10}) or $\leq$2.5m ($PM_{2.5}$) as a metric, supported by health studies show a strong association between ambient PM mass concentration and a variety of adverse health effects. Time-series techniques are needed to predict PM concentrations at 2.5.

Due to various factors, $PM_{2.5}$ status will change frequently. These changes are often unexpected. Stochastic analysis is used because of uncertainty factors that exist in hydrological characteristics. For this reason, the researchers propose a variety of stochastic approach methods to determine the pattern of the spread of $PM_{2.5}$. One of them is using the stochastic Markov chain model [9] in which the following state is only affected by the current state and is free of the former state. The transition opportunity matrix is determined by the method of estimating maximum likelihood.

The idea of using mathematical models to explain the behavior of physical phenomena has been done well, as in deterministic models [10]. Nevertheless, not all phenomena are entirely deterministic because unknown factors can occur and affect these physical phenomena [11]. In this case, the time-dependent phenomenon is needed in stochastic models [12]. Neural Network [13] is a nonparametric model that can be used for modelling time series data that does not require various residual assumptions. Several studies have found that this model produces better predictive accuracy than parametric models [14]. Beside of time dimension, data can also have a space dimension known as space-time data [15]. Space-time model is a model that combines dependency between time and location in a multivariate time series data [16]. In the real world, we need methods to solve high dimensional data [17] and non-linear.

The application of Neural Network in time series prediction models [18] is expected to provide more accurate and robust results against data fluctuations [19]. One of the flexibility, of the Neural Network model as a nonparametric model is that there is no need to test model assumptions [20], so the main thing to consider is the formation of a model to get the smallest possible error [21]. Previous authors perform NN with optimization like genetic algorithm [22], gradient descent [23], PSO [24], [25]. Also, [26] uses an artificial neural network (ANN) forecasting the accuracy of daily average concentrations of $PM_{2.5}$ two days in advance is presented. The model was developed from 13 different air pollution monitoring stations in Beijing, Tianjin, and Hebei province (Jing-Jin-Ji area) and obtained high accuracy. Moreover, Chang and Tseng perform NN [27] and the experimental results show that the factory data and stock farming data may be one of the factors influencing $PM_{2.5}$ concentration.

The application of NN in time series forecasting can be the right solution, but the problem is network architecture and the selection of appropriate training methods. Over time the time series method is not only used for univariate cases but also used for multivariate cases [28] one of which is the Vector Autoregressive (VAR) model [29]. Previous studies on VAR modelling have been carried out [30] obtained results that VARIMA modelling can be used on stationary and non-stationary data.

In analyzing, the relationship is between climatological variables in VAR [31]. It is not necessary to distinguish between endogenous and exogenous variables, which means that all variables used in VAR are used as endogenous variables. The problem that appears in neural network modelling for multivariate time series data is how to determine the lag for the input model and to determine the optimal weight for each input model.

Therefore, we need a standard procedure for the optimal input process to obtain an optimal model architecture. The next problem is processing at the neural network layer. Besides many optimal hidden units that must be determined, one crucial thing is the selection of optimisation methods used to estimate the parameters/weight of the model [32]. Some conventional gradient-based methods that are commonly used are often problematic in terms of yield consistency and are trapped in earlier convergence to a local optimum. Therefore a breakthrough is needed to find alternative methods that can overcome this problem.

II. MARKOV CHAIN

Suppose that $S \subset \mathbb{R}$ is a set of values from a random array of variables, then S is called state space. The stochastic process $X = \{X_t, t = 1, 2, \ldots\}$ is a collection of the random variables that map a sample space Ω to state space. So, for every $t \in \{1, 2, \ldots\}$, X_t is a random variable [33]. In this case, assume t as the time and value of the random variable X_t as the state of the process at a time. The Discrete Marko Chain [34] can be explained as a stochastic process $\{X_t, t = 1, 2, \ldots\}$ with state space $\{1, 2, \ldots, N\}$ called a Markov chain with discrete time if for every $t \in \{1, 2, \ldots\}$, applies:

$$P\left(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1\right) = P\left(X_{t+1} = j \mid X_t = i\right) \quad (1)$$

For all possible values of $i_1, i_2, \ldots, i_{t-1}, i, j \in \{1, 2, \ldots, \mathrm{N}\}$. So, for a Markov chain [35], the conditional distribution of any future state X_{t+1} is free of all previous states $X_1, X_2, \ldots, X_{t-1}$, and only depends on the current state X_t. This is called Markov properties. Besides the homogeneous Markov chain can be explained suppose $\{X_t, t = 1, 2, \ldots\}$ is a Markov chain with state space said to be homogeneous if:

$$P\left(X_{t+1} = j \mid X_t = i\right) = P\left(X_2 = j \mid X_1 = i\right) = p_{ij} \quad \text{For} \quad i, j \in \{1, \ldots, N\} \quad (2)$$

The above process can be described as a state Markov chain with probability transition (p_{ij}) with $i, j = 1, \ldots, N$ [36]. The value of the probability transition (p_{ij}) states the probability that if the process is in state i then the next one will switch to state j. However, the probability value is not negative, and because the process must undergo a transition from one state to another, then:

a. $p_{ij} \geq 0$, for all $i, j \in \{1, \ldots, N\}$
b. $\sum N_{j=1} p_{ij}, for\ all\ i \in \{1, \ldots, N\}$

Probability transition can be written in the form of P matrix, which measures ($N \times N$):

$$p = \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix}$$

III. VECTOR AUTOREGRESSIVE

Vector autoregressive (VAR) has several endogenous variables simultaneously [37], but each endogenous variable is explained by the lag of its value and other endogenous variables in the model. The VAR model is built to overcome the relationship between variables so that they can still be estimated without the need to emphasize exogenous issues [38]. In this approach, all variables are considered endogenous, and estimates can be carried out simultaneously or sequentially [39].

VAR only needs to pay attention to two things. First, no need to distinguish between endogenous and exogenous variables. All endogenous and exogenous variables that are believed to be interconnected should be included in the model [40]. Second, to see the relationship between variables in the VAR, several variable lags are needed [41]. The assumption in the VAR model assumes that all variables are interdependent. In general, the VAR model with T variables can be written as follows:

$$Y_{jt} = \beta_j + \sum\nolimits_{i=1}^{p} \gamma_{ji} Y_{1,t-i} + \sum\nolimits_{i=1}^{p} \theta_{ji} Y_{2,t-i} + \ldots + \sum\nolimits_{i=1}^{p} \lambda_{Ti} Y_{T,t-p} + e_{jt} \quad (3)$$

With:

$$\begin{aligned} Y_{jt} &= \text{forecasting number } j \text{ at time} - t \\ t &= \text{forecasting time} \\ T &= \text{number of variables, with} : 1, 2, \ldots, T \\ \beta_j &= \text{constants for variables } j \\ p &= \text{number of lags, with i} : 1, 2, 3, 4, \ldots, p \\ \gamma_{ji} &= \text{parameter in variable 1 lag} - i \\ \theta_{ji} &= \text{parameter in variable 2 lag} - i \\ \lambda_{Ti} &= \text{parameter in variable } T \text{ lag} - i \\ e_{jt} &= \text{residual } j \text{ at time} - t \end{aligned}$$

Based on the VAR model in equation (3), we can write the model with 2 variables ($T = 2$) testing and the number of lag 2 ($p = 2$):

$$Y_{1t} = \beta_1 + \gamma_{11} Y_{1,t-1} + \gamma_{12} Y_{1,t-2} + \theta_{11} Y_{2,t-1} + \theta_{12} Y_{2,t-2} + e_{1t} \quad (4)$$

$$Y_{2t} = \beta_2 + \gamma_{21} Y_{1,t-1} + \gamma_{22} Y_{1,t-2} + \theta_{21} Y_{2,t-1} + \theta_{22} Y_{2,t-2} + e_{2t} \quad (5)$$

To define this model, it is assumed that the two variables Y_1 and Y_2 are stationary and residual in the model is a white noise process. Equation (4) is a model for the first variable, while Equation (5) is a model for the second variable. In general, the VAR model for T variables will consist of T equations where each one equation is an equation with one variable as the dependent variable, and the independent variable is the lag of all other variables. VAR modelling consists of endogenous variables with indices on the left side of the model and a constant component and lagged term component on the right side of the model. Assuming that there is no cross-correlation between the residuals (error term), the VAR model can be estimated using Ordinary Least Square (OLS) sequentially by estimating all equations in turn. The parameters estimated for the VAR model in this research using OLS.

$$\gamma_{11}, \gamma_{12}, \ldots, \gamma_{Ti}, \theta_{11}, \theta_{12}, \ldots, \theta_{Ti}, \lambda_{11}, \lambda_{12}, \ldots, \lambda_{Ti}.$$

Suppose OLS estimation with 2 test variables and the number of lags 1 as follows:

$$\begin{aligned} Y_{1t} &= \beta_1 + \gamma_{11} Y_{1,t-1} + \theta_{11} Y_{2,t-1} + e_{1t} \\ Y_{2t} &= \beta_2 + \gamma_{21} Y_{1,t-1} + \theta_{21} Y_{2,t-1} + e_{1t} \end{aligned} \tag{6}$$

The principle of ordinary least square (OLS) parameter estimation [42] is to minimize the residual sum of square (RSS) written in Equation (6). In this test, parameter estimation will be performed on the Y_{1t} model with the parameters β_1, γ_{11}, and θ_{11}.

$$RSS = \sum_{t=1}^{n} (e_{jt})^2 = \sum_{c=1}^{n} \left(Y_{1c} - \beta_1 - \gamma_{11} Y_{1c,t-1} - \theta_{11} Y_{2c,t-1}\right)^2$$

The RSS equation is derived from all parameters that will be estimated in the model then equated with zero. The RSS equation, after being derived is as follows:

$$\begin{aligned} & n\beta_1 + \gamma_{11} \sum_{c=1}^{n} Y_{1c,t-1} + \theta_{11} \sum_{c=1}^{n} Y_{2c,t-1} \\ &\quad = \sum_{c=1}^{n} Y_{1c}\beta_1 \sum_{c=1}^{n} Y_{1c,t-1} + \gamma_{11} \sum_{c=1}^{n} Y_{c1,t-1}^2 \\ &\quad\quad + \theta_{11} \sum_{c=1}^{n} Y_{2c,t-1} Y_{1c,t-1} = \sum_{c=1}^{n} Y_{1c} Y_{1c,t-1} \\ &\quad\quad \beta_1 \sum_{c=1}^{n} Y_{2c,t-1} + \gamma_{11} \sum_{c=1}^{n} Y_{1c,t-1} Y_{2c,t-1} \\ &\quad\quad + \theta_{11} \sum_{c=1}^{n} Y_{2c,t-1}^2 = \sum_{c=1}^{n} Y_{1c} Y_{2c,t-1} \end{aligned} \tag{7}$$

The form of the matrix equation for Equation (7) is:

$$\left(\boldsymbol{X}'\boldsymbol{X}\right) \widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

With,

$$\boldsymbol{X} = \begin{bmatrix} 1 & Y_{11,t-1} & Y_{21,t-1} \\ 1 & Y_{12,t-1} & Y_{22,t-1} \\ 1 & Y_{13,t-1} & Y_{23,t-1} \\ \vdots & \vdots & \vdots \\ 1 & Y_{1n,t-1} & Y_{2n,t-1} \end{bmatrix}, \quad \widehat{\boldsymbol{B}} = \begin{bmatrix} \beta_1 \\ \gamma_{11} \\ \theta_{11} \end{bmatrix} \text{ and}$$

$$\boldsymbol{y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ \vdots \\ Y_{1n} \end{bmatrix}$$

So that the parameter estimation $\widehat{\boldsymbol{\beta}}$ is obtained as follows:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{y}$$

IV. ANALYSIS

A. PROBABILITY TRANSACTION MARKOV CHAIN

The first step is to calculate the probability transition $PM_{2.5}$ data, which is classified by No Risk (1-30), Medium Risk (30-48), and Moderate (> 49). In this paper, we are using monthly data from January 2014 to May 2019 in 2 Taiwanese locations, Pingtung and Chaozhou. Substitution of $PM_{2.5}$ status with migration has one free random variable, which is a random variable of the actual value $PM_{2.5}$ $\{(t)\}$. Probability for status change at time t can be stated as follows:

$$(t) = \text{Prob}\{I(t) = i\}. \tag{8}$$

In this model, Markov properties apply for the order of real numbers $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1}$.

$$\text{Prob}\{I(t_{n+1}) | I(t_0), I(t_1), \ldots, I(t_n)\} = \text{Prob}\{I(t_{n+1}) | I(t_n)\}. \tag{9}$$

The probability chance of transition at t_{n+1} depends only on time t_n at intervals, $t + \Delta t$ the number of status changes for the month j. The probability of transition from the number of months that change from state i to state j at intervals Δt can be written as follows:

$$(\Delta t) = \text{Prob}\{I(t + \Delta t) = j | I(t) = i\}. \tag{10}$$

(Δt) contains the $\lim\limits_{t \to \infty} = \frac{o(\Delta t)}{\Delta t} = 0$, where $o(\Delta t)$ shows a small probability value and cannot be stated exactly. It is assumed that the selected value of Δt is minimal in the event of a transition so that the probability of changing $PM_{2.5}$ status is a maximum of one month during the time interval Δt. There are three possible transitions namely, the monthly that will be different will transition from state i to state $j = i + 1$ meaning that the month that will change the status of $PM_{2.5}$ will increase by one. This is due to the influence of lag time at a rate of β and the occurrence of incoming migration that occurs in the month of change at a rate of ν. So, the chance of transition in the time interval can be written:

$$\begin{aligned} & Prob\{\Delta l\,(t) = j | l\,(t) = i\} \\ &= \begin{cases} (\nu N + \dfrac{\beta}{N} i\,(N - i)\,\Delta t + o\,(\Delta t)\,, & j = i + 1 \\ (\gamma + \mu)\, i\Delta t + o\,(\Delta t)\,, & j = i - 1 \\ 1 - \left(\nu N + \dfrac{\beta}{N} i\,(N - i)\,\Delta t + (\gamma + \mu)\, i\Delta t\right) + o\,(\Delta t)\,, & \\ & j = i \\ o\,(\Delta t)\,, & j \neq i - 1,\ i + 1,\ i. \end{cases} \end{aligned} \tag{11}$$

Next, calculate of probability changes will be made between No Risk (1-30), Medium Risk (30-48), and Moderate (> 49). Suppose that $\pi = [\pi_i]$ is a probability vector that each component states that the process will be in state i. For time $n \to \infty$ it is called a stationary probability vector or steady-state chance, that is, after the process has been running for several periods, the probability transition value will remain. In this case, π_i can be interpreted as the proportion of long-term time in which the Markov chain is in state i. The vector π

can then be calculated as $= \pi p$, where $[\pi_1, \pi_2, \pi_3]$ is a non-negative solution from equation (12):

$$\begin{aligned} \pi_1 &= \pi_1 P_{11} + \pi_2 P_{21} + \pi_3 P_{31} \\ &\Leftrightarrow \pi_1 (P_{11} - 1) + \pi_2 P_{21} + \pi_3 P_{31} = 0 \\ \pi_2 &= \pi_1 P_{12} + \pi_2 P_{22} + \pi_3 P_{32} \\ &\Leftrightarrow \pi_1 P_{12} + \pi_2 (P_{22} - 1) + \pi_3 P_{32} = 0 \\ \pi_3 &= \pi_1 P_{13} + \pi_2 P_{23} + \pi_3 P_{33} \Leftrightarrow \pi_1 P_{13} + \pi_2 P_{23} \\ &\quad + \pi_2 (P_{33} - 1) = 0 \end{aligned} \tag{12}$$

Based on equation (7) above, it can be formed into the following matrix equation so that it can be formed into the matrix equation as follows:

$$\begin{pmatrix} P_{11} - 1 & P_{21} & P_{31} \\ P_{12} & P_{22} - 1 & P_{32} \\ P_{13} & P_{23} & P_{33} - 1 \end{pmatrix}$$

And can be written

$$\left(P^T - 1\right)\pi = 0 \tag{13}$$

The probability vector of steady-state π is an Eigenvector determined by Eigenvalue λ so that π must fulfil in equation (14).

$$\left(P^T - \lambda I\right)\pi = 0 \tag{14}$$

With Eigen values $\lambda = 1$ and I are identity matrices with dimensions of 3×3. Because π is a probability vector, π must also fulfill:

$$\begin{aligned} \pi_1 + \pi_2 + \pi_3 &= 1 \\ \pi_1 &= \left((1 - r) - \frac{(1 - s)\, p}{q}\right)^{-1} \\ \pi_2 &= \left(-r - \frac{ps}{q}\right)\pi_1, \\ \pi_3 &= \left(\frac{p}{1}\right)\pi_1, \end{aligned} \tag{15}$$

With:

$$p = \left(\frac{P_{13}}{P_{33} - 1} - \frac{P_{23}\,(P_{11} - 1)}{P_{23}(P_{33} - 1)}\right) \tag{16}$$

$$q = 1 - \frac{P_{23} P_{31}}{P_{21}(P_{33} - 1)'} \tag{17}$$

$$r = 1 - \frac{P_{23} P_{33}}{P_{23} P_{33} - 1'} \tag{18}$$

$$s = \frac{P_{32}}{P_{22} - 1} \tag{19}$$

Suppose that T_i is the unit of time spent by the process in state i before switching to another state. T_i the time is considered as the number of repeating random trials independently whose results fail or succeed until the first success with the probability of success $(1 - P_{ii})$. If the next time remains in state i, then the chance for the process to remain in state i is P_{ii} (Probability of failure). T_i Time also has memoryless properties of the Markov chain. The geometric distribution is the distribution of discrete opportunities that have a memoryless. Thus, T_i is a random variable that spreads geometrically with a chance period function. However, $PM_{2.5}$ distribution in Pingtung can be seen in Figure 1 (left) and Chaozhou di (right) Figure 1 and Figure 2 shows the $PM_{2.5}$ frequency and in general, at the beginning of 2014 was the worst $PM_{2.5}$ status in these two locations. However, the 44th month is the lowest $PM_{2.5}$ status for the last five years. Many factors might influence the increase and decrease of air pollution in cities and Taiwan, such as meteorological conditions, including rainfall, wind, and high pressure, which would require further analysis.

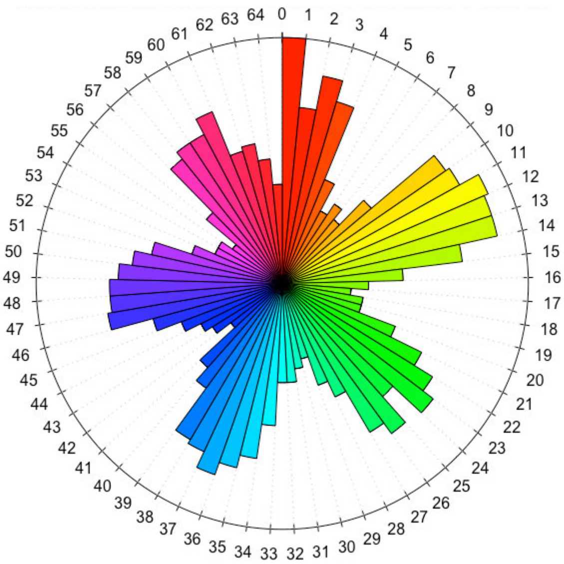

FIGURE 1. Monthly PM 2.5 in Pingtung January 2014 to May 2019.

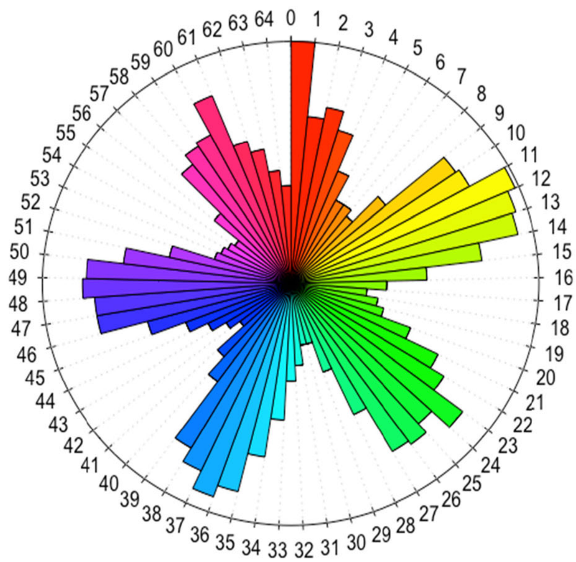

FIGURE 2. Monthly PM 2.5 in Chaozhou January 2014 to May 2019.

In what follows, our Markov models are assumed to be homogeneous which stationary distributed and shall denoted

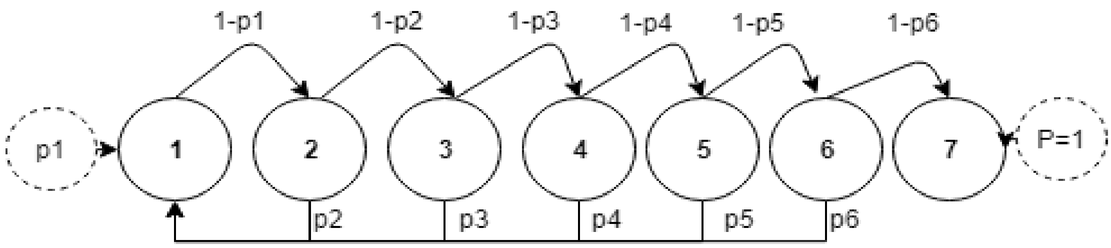

FIGURE 3. **Modeling $PM_{2.5}$ with the Markov chain model.**

by the vector ($\delta = \delta_1, \delta_2, .., \delta_m$). Suppose $\{S_t\}$ is stationary, so δ is for all t distribution of S_t.

Moreover, we can define $v = (1, 2, \ldots, m)$, $V = diag(v)$ which is a diagonal matrix with v on the diagonal principle, also $\gamma_{ij}(k) = (\Gamma^k)_{ij}$ will have the results for mean of S_t and the covariance of S_t and S_{t+k} for all non-negative integers k can be written as $E(S_t)$, $E(S_t, S_{t+k})$, *and cov* $E(S_t, S_{t+k})$.

$$E(S_t) = \sum\nolimits_{i=1}^{m} i\delta_i = \delta v' \quad (20)$$

$$E(S_t, S_{t+k}) = \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{m} ij\delta_i P(S_{t+k} = j | S_t = i) \quad (21)$$

$$E(S_t, S_{t+k}) = \sum\nolimits_{i,j} (i\delta_i)\, \gamma_{ij}(k)\, j = \delta V \Gamma^k v' \quad (22)$$

$$Cov(S_t, S_{t+k}) = \delta V \Gamma^k v' - \left(\delta V'\right)^2 \quad (23)$$

The probability status change in $PM_{2.5}$ is a function of recency that is the number of periods since the last day of status change. Based on Figure 3 if $PM_{2.5}$ status changes at the end of the previous period, it will be at recency 1 for the current period. We assumed that if the status change has reached recency 7, or 7 consecutive periods. So, it would be there is no status change, and then it can be categorised in that week, which tends to be constant. At the same time, the PM2.5 status that has been in recency *1, 2, 3, 4, 5*, and *6* if there is a change in a period, it will return to recency 1 in the next period. When $PM_{2.5}$ status is at recency r, with r = *1, 2, 3, 4, 5, 6, 7*. Whereas *1-pr* is a probability that $PM_{2.5}$ status will not change at the end of the period when it is in the r recency.

Figure 5 and Figure 6 show the condition in Chaozhou. Besides, it can be seen in Figure 4, Figure 5 and Table 1 that the probability of changing the status of $PM_{2.5}$ to no risk in May 2019 to June 2019 is equal to 0.808; the probability medium risk is 0.182, and probability moderate is 0. So that can be said it is estimated that $PM_{2.5}$ status in the Pingtung area is no risk in the future.

Based on the analysis, it can be seen in Figure 4 and Table 2 that the probability to change the status of $PM_{2.5}$ to no risk in May 2019 to June 2019 is 0.828, and the probability of medium risk is 0.231, the probability of moderate at 0. So that can be said at one month in the future it is estimated that $PM_{2.5}$ status in the Pingtung area is also no risk.

Overall, the probability transition from state 1 has the highest value in the no risk, and the probability for the transition from state 2 has the highest value in the no risk.

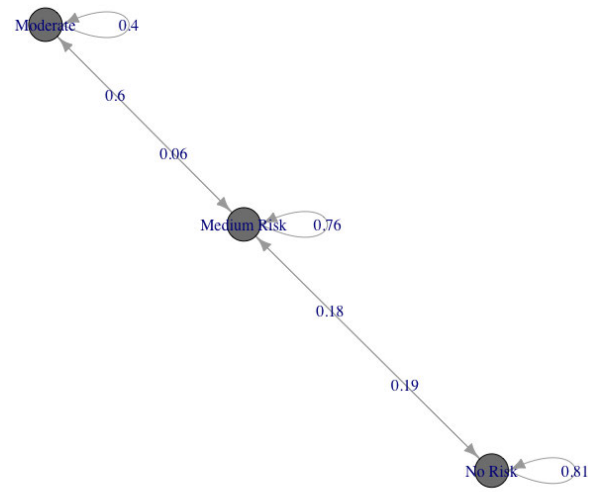

FIGURE 4. **Line probability transition $PM_{2.5}$ in Pingtung.**

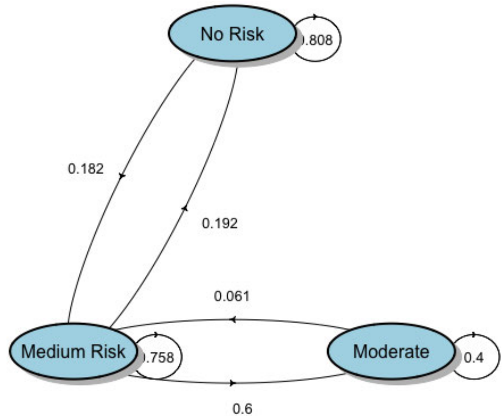

FIGURE 5. **Probability transition $PM_{2.5}$ in Pingtung.**

TABLE 1. **Probability transition in Pingtung.**

	Prob No Risk	Prob Medium Risk	Prob Moderate
If next month no Risk	0.808	0.192	0
If next month Medium Risk	0.182	0.758	0.061
If next month Moderate	0	0.6	0.4

The transition opportunity for the moderate any state to state 3 in the case has a higher chance of transition compared to case 2 (Chaozhou). When compared to the two regions,

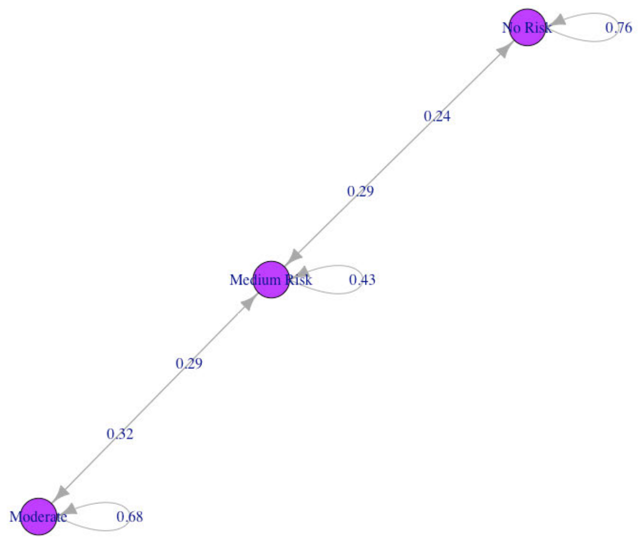

FIGURE 6. Line probability transition $PM_{2.5}$ in Chaozhou.

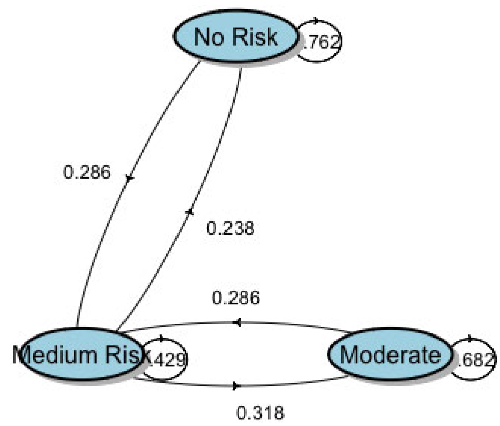

FIGURE 7. Probability transition $PM_{2.5}$ in Chaozhou.

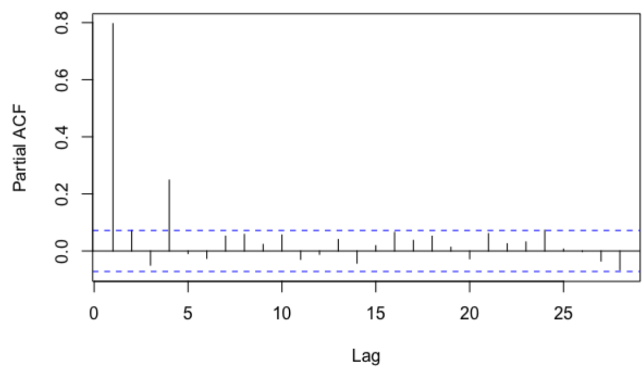

FIGURE 8. Partial autocorrelation function in Pingtung.

TABLE 2. Probability transition in Chaozhou.

	Prob No Risk	Prob Medium Risk	Prob Moderate
If next month no Risk	0.762	0.238	0
If next month Medium Risk	0.286	0.429	0.286
If next month Moderate	0	0.318	0.682

the most significant change was status in Chaozhou. There are more significant than Pingtung in $PM_{2.5}$ in Chaozhou. Based on Cramér's V [43] the p-value = 0.70 is obtained. In both regions, weather anomalies often occur at certain times, especially during the summer season. These conditions can cause an increase in extreme weather in the form of rain in the category of light rain to heavy rain so that the average length of the rainy period is still quite high in the dry season which causes an increase in $PM_{2.5}$.

B. STEP CONSTRUCTION VAR-NN-PSO

This model is based on the FFNN model, which differs when determining input variables. Because the VAR model is used, the input is the lag variable of each predicted variable, in this case, the $PM_{2.5}$ data in Pingtung as Y_1 and Chaozhou as Y_2. Each of these data was taken from January 2019 to May 2019. The lag selection is based on the value and plot of the partial autocorrelation function (PACF) of each variable. Moreover, we perform VAR-NN with training algorithms using Hybrid Particle Swarm Optimization (PSO) and Backpropagation.

Based on the partial autocorrelation function in Figure 7 and Figure 8, it was found that significant lag variables were lags at t_{-1} and lag t_{-2}. Then, after the lag of the input variable is obtained an FFNN model is formed for these two variables. In FFNN, neurons are arranged in layers and signals from the input to the first layer, then to the second layer, and so on [22]. The assumption in the VAR that all variables depend on each other.

VAR (p) model or Vector Autoregressive model with sequence p, which means the independent variable of the model is p-value of the independent lag variable:

$$\mathbf{Y}_t = \boldsymbol{\varphi}_0 + \sum\nolimits_{i=1}^{p} \boldsymbol{\varphi}_i \mathbf{Y}_{t-i} + \varepsilon_t \tag{24}$$

In this paper, multilayer networks are used with the Feed Forward Neural Networks (FFNN) model. In FFNN, neurons are arranged in layers (layers) and signals flow from the input to the first layer, then to the second layer.

$$Y_t = \psi_0 \left\{ v_{b0} + \sum\nolimits_{n=1}^{H} v_{out} \psi_k \left(w_{bi} + \sum\nolimits_{j=1}^{p} w_{in} Y_{t-j} \right) \right\} \tag{25}$$

In Table 3 and Table 4, the $(w_{bn}, w_{in}, v_{out}, v_{bo})$ is the weight parameter in the FFNN model and (ψ_o, ψ_k) is an activation function. Before training artificial neural networks, input, and target scales are often needed so that data enters a certain range. Where is the weight parameter in the FFNN model and is an activation function which in this case uses *tansig* and *purelin* functions which can be seen in Figure 9. To choose the best model is to see the smallest MAPE value. Based on the results of data processing, VAR models for Y_1 and Y_2 are obtained as follows:

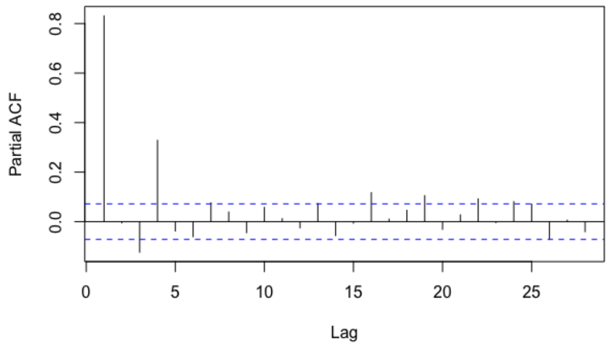

FIGURE 9. Partial autocorrelation function in Chaozhou.

TABLE 3. Weight or parameters of the FFNN model in Pingtung (MAPE =3.15%).

w_{bi}	$w_{1,n}$	$w_{2,n}$	w_{3n}	w_{4n}	v_{bo}	v_{no}
-2.0921	4.7170	2.1080	-0.5804	8.5954	-0.4677	0.16995
-0.9671	-1.1501	0.8792	-1.3633	-0.3329		-0.40412
-3.0586	0.5767	-2.8631	-3.1734	3.8262		-0.42254
-0.0911	-0.6600	-1.3662	0.1320	1.0629		0.30523
-0.8377	0.8439	2.8495	0.1195	-1.0485		0.29357

TABLE 4. Weight or parameters of the FFNN model in Chaozhou (MAPE =4.87%).

w_{bi}	$w_{1,n}$	$w_{2,n}$	w_{3n}	w_{4n}	v_{bo}	v_{no}
-18.7108	22.3541	0.2837	-3.2729	5.6314	-0.2882	-0.22714
-1.2619	-1.3126	5.9300	-1.9960	-5.0606		-0.28212
4.9200	-9.8502	-2.4710	-6.2832	-0.4224		0.28045
-0.7925	-0.4370	-1.0599	1.3845	-1.1663		-0.28309
0.9221	-1.0971	-1.2262	-0.0524	-0.4961		-0.73964

The steps for estimating the VAR-NN model using the Backpropagation algorithm in this study are as follows:

1. Identify the lags of independent variables using values and Partial Autocorrelation Function (PACF).
2. Determine the number of neurons in the hidden layer.
3. Initialize all weights in the hidden layer and output layer.
4. Calculate the output obtained from the neurons in the hidden layer with the sigmoid logistic activation function.
5. Calculate the output obtained from neurons in the output layer.
6. Calculate the error gradient for neurons in the output layer.
7. Calculate the weight correction for the output layer.
8. Update all weights in the output layer.
9. Calculate the error gradient for neurons in the hidden layer.
10. Calculate the weight correction for hidden layers
11. Fix all weights in the hidden layer
12. Calculate the predictions of VAR-NN using mean absolute percentage error (MAPE), to get the best model.
13. Forecasts with the best models.

We perform an FFNN network with neurons in the hidden layer of 5 neurons, the VAR-NN model (2, 2, 5). First perform by Particle Swarm Optimization and Backpropagation (PSO-BP) Hybrid algorithm. PSO starts with a set of particles solutions randomly generated. The following is the PSO calculation formula:

$$v_j^{t+1} = w.v_j^t + c_1.r_1\left(pBest_j^t - x_j^t\right) + c_2.r_2\left(gBest_j^t - x_j^t\right)$$
$$x_j^{t+1} = x_j^t + v_j^{t+1} \tag{26}$$

where:

v_j^t : Velocity
x_j^t : Particle Position
w : Inertia Weight
$c_1\&c_2$: learning rates
$r_1\&r_2$: Random value from 0 to 1
$pBest_j^t$: Best position of the particle
$gBest_j^t$: Global optimum

Learning rates (c_1 and c_2) or often called velocity parameters that show the weight of the memory of a particle on the memory of a swarm. However, the values of c_1 and c_2 are usually 2 so that the multiplication of c_1r_1 and c_2r_2 Ensures that the particles will approach the target by about half the difference. Inertia weight (w) is a weight used to reduce the speed of the speed update formula. An enormous w value is useful for exploration of search space while a small value of w is good for intensification. To further improve PSO performance, this w value is made varied during the solution search process, and more detail can be seen in Table 5.

TABLE 5. Parameter estimation VAR-NN.

Algorithm: **Parameter estimation of the VAR-NN model with the PSO-BP hybrid algorithm**

Input:
Observation data
Lag Input Variable p,
The number of particles (N), the value of c_1 and c_2 and the inertia weight of the PSO
The number of neurons in the hidden layer
The activation function ψ_o, ψ_k on FFNN

Output:

1. Preprocessing Data to standard normal form
2. Initialization of parameters $(w_{bn}, w_{in}, v_{out}, v_{bo})$ corresponds to the number of p inputs and Hidden layers neurons as PSO particles by generating random data (-1,1) as many as N particles.
3. **Repeat**:
   1. Evaluate the fitness value of each particle using MSE
   2. Update the value of *pBest* (best position) and *gBest* (global optimal).
   3. Update the speed and position of each particle.
   4. Until fitness convergence
   5. Use the parameters $(w_{bn}, w_{in}, v_{out}, v_{bo})$ as initial weights on the FFNN model with the Backpropagation training Algorithm
   6. Postprocessing Data
   7. Return $(w_{bn}, w_{in}, v_{out}, v_{bo})$

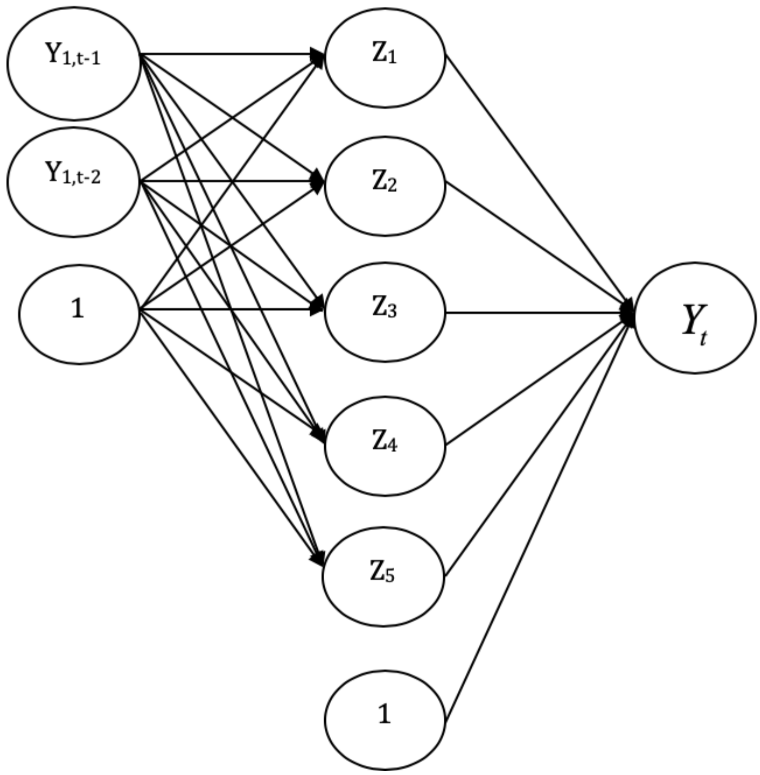

FIGURE 10. Illustration VAR-NN predict $PM_{2.5}$ in Pingtung ($Y_{1,t-1}$) and Chaozhou ($Y_{1,t-2}$).

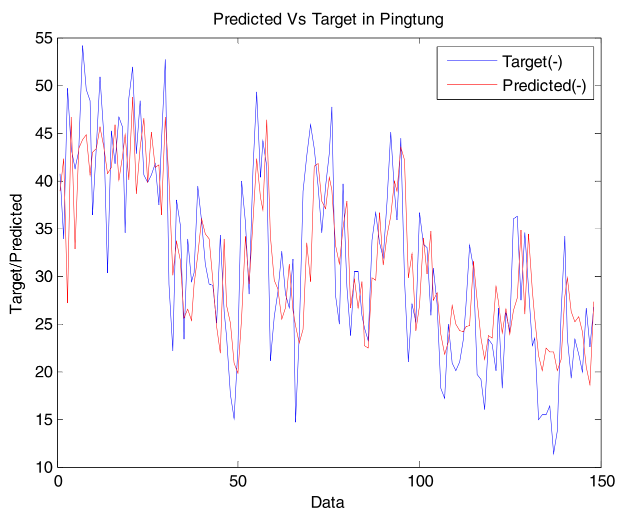

FIGURE 11. Predicted VS target in Pingtung.

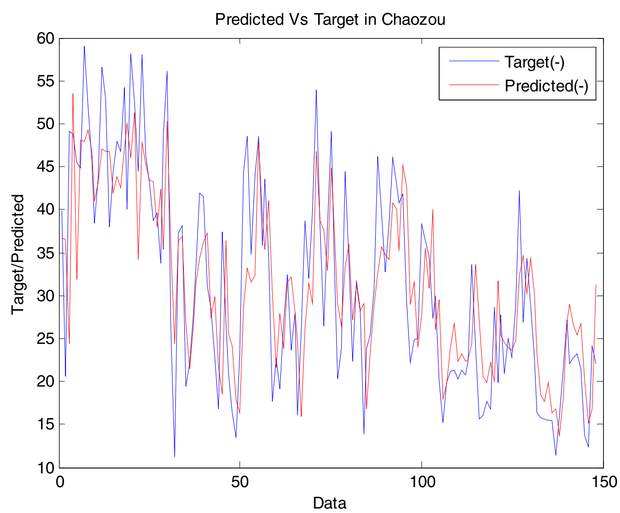

FIGURE 12. Predicted VS Target in Chaozhou.

Based on the simulation, each particle in VAR-PSO-NN presents the position and location of the problem at hand. Each particle searches for the optimal solution with the intelligence of the individual experience by crossing the dimensions of the D search space. This is done by way of each particle adjusting to the best position and the best adjustment of the particle position of the best value of the whole flock (global best) while crossing the search space. At each iteration, each solution represented by the position of the particle evaluates its performance by entering the solution into the fitness function. Each particle is like a point in a certain dimension of space. After the simulation obtained with MAPE 3.57% for $PM_{2.5}$ data in Pingtung and 4.87% for $PM_{2.5}$ data in Chaozhou, the prediction vs actual data plot can be seen in Figure 11 and Figure 12, respectively.

After all the tests are carried out and meet the VAR modelling, forecasting is done as the final result. Data to be forecast is forecasting $PM_{2.5}$ data in June 2019 can be seen

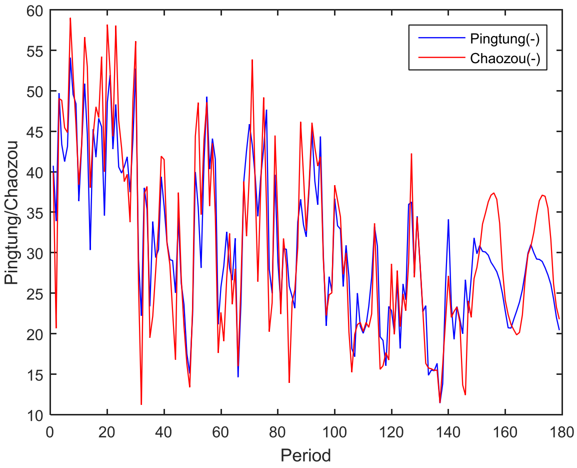

FIGURE 13. Forecasting $PM_{2.5}$ 180 days in Pingtung and Chaozou.

in Figure 13. Based on the forecast it can be seen that $PM_{2.5}$ it tends to increase which can be caused by industrial activities.

V. CONCLUSION

The three-state Markov chain model can be used to determine the change of $PM_{2.5}$ status. The status change fulfils the Markov nature and can determine the opportunity for status change (No risk, risk, moderate) in the next month. We use the most significant lag Y1,t-1, Y1,t-2. From these lags, a combination will be made between lags to determine the existence of nonlinear patterns in the $PM_{2.5}$ data. Regarding the chance of the status change the relationship between $PM_{2.5}$ in Pingtung and Chaozhou is high (0.7) and significant at $\alpha = 5\%$. After performing the VAR-NN-PSO if, without a selection operation, individuals will be trapped in individuals who have low fitness values. Also, the novelty of this research is to examine the VAR-NN-PSO Model and the optimum parameters that have been obtained, used to predict $PM_{2.5}$ levels for the next 180 days with high accuracy. This research can be a recommendation to the government to maintain air pollution besides gets the probability of transition with Markov chain or forecasting with short-term or long-term using VAR-NN-PSO because this model performs high accuracy which justifies by MAPE.

APPENDIX

Proof of Estimating Value $\boldsymbol{P_i}$*:*

Define the likelihood function for first order markov

$$f(p) = P(X_t = 1)\, P(X_2 = x_2 \mid X_1 = 1) \ldots P(X_n = x_n \mid X_{n-1} = x_{n-1})$$

$$f(p) = P(X_t = 1) \prod_{t=2}^{n} P(X_t = x_t \mid X_{t-1} = x_{t-1})$$

$$f(p) = 1 \prod_{t=2}^{n} P_{x_{t-1}x_t}, \text{ we assumed that } P(X_1 = 1) = 1$$

$$f(p) = 1 \prod_{i=1}^{N} \prod_{j=1}^{N} p_{ij}^{f_{ij}} \text{ with } i, j \epsilon \{1, 2, \ldots, N\}$$

where f_{ij} is the number of transitions from state i to state j. p_{ij} is the transition from state i to state j, and N is the number of states. We can define the likelihood function $\mathcal{L}(P)$.

$$\mathcal{L}(P) = \log f(p) = \log 1 + \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} \log p_{ij}$$

With obstacles

$$\sum_{j=1}^{N} P_{ij} = 1, \quad i = 1, 2, \ldots, N$$

Then, maximize the log likelihood function with the Lagrange multiplier method. Suppose the Lagrange multiplier $\lambda_1, \lambda_2, \ldots, \lambda_N$ then the new objective function is

$$\boldsymbol{g}(\boldsymbol{P}, \lambda) = L(P) - \sum_{i=1}^{N} \lambda_i \left(\sum_{j=1}^{N} p_{ij} - 1 \right)$$

$$\boldsymbol{g}(\boldsymbol{P}, \lambda) = \log 1 + \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{N} f_{ij} \log P_{ij} - \sum\nolimits_{i=1}^{N} \lambda_i \left(\sum\nolimits_{j=1}^{N} P_{ij} - 1 \right)$$

The function g is maximized by being derived from $P_{ij}, j = 1, 2, \ldots, N$ and equal to 0.

$$\frac{\vartheta g(P, \lambda)}{\vartheta P_{ij}} = -\frac{f_{ij}}{p_{ij}} - \lambda_i, \quad \forall i, j \epsilon 1, 2, \ldots, N$$

$$\frac{f_{ij}}{p_{ij}} - \lambda_i = 0$$

$$\widehat{P_{ij}} \frac{f_{ij}}{p_{ij}}$$

$$\sum_{j=1}^{N} \widehat{P_{ij}} = \sum_{j=1}^{N} \frac{f_{ij}}{\lambda_i} = 1 \longrightarrow \lambda_i = \sum_{j=1}^{N} f_{ij} \quad \forall i, j \epsilon 1, 2, \ldots, N$$

$$\widehat{P_{ij}} = \frac{f_{ij}}{\sum_{j=1}^{N} f_{ij}}, \quad \forall i, j \epsilon 1, 2, \ldots, N$$

$\therefore$ ***Proof***

DATA AVAILABILITY

The analysis datasets used in this paper are available at Environmental Protection Administration, R.O.C. (Taiwan) https://taqm.epa.gov.tw/taqm/en/ and available from the corresponding author upon reasonable request.

REFERENCES

[1] S. S. Young, "Air pollution and mortality in the medicare population," *JAMA*, pp. 2513–2522, 2018.
[2] J. A. Bernstein, "Health effects of air pollution," *J. Allergy Clin. Immunol.*, vol. 114, no. 5, pp. 1116–1123, 2004.
[3] D. A. Vallero, *Fundamentals of Air Pollution*. New York, NY, USA: Academic, 2014.
[4] X. Querol, "Speciation and origin of PM_{10} and $PM_{2.5}$ in selected European cities," *Atmos. Environ.*, vol. 38, no. 38, pp. 6547–6555, 2004.
[5] M. Zheng, L. G. Salmon, J. J. Schauer, L. Zeng, C. S. Kiang, Y. Zhang, and G. R. Cass, "Seasonal trends in $PM_{2.5}$ source contributions in Beijing, China," *Atmos. Environ.*, vol. 39, no. 22, pp. 3967–3976, 2005.
[6] S. Feng, D. Gao, F. Liao, F. Zhou, and X. Wang, "The health effects of ambient $PM_{2.5}$ and potential mechanisms," *Ecotoxicol. Environ. Saf.*, vol. 128, pp. 67–74, Jun. 2016.

[7] N. M. Rasidi, S. A. Bakar, and F. A. Razak, "Modelling the correlation of PM_{10} concentration and location of air quality monitoring stations in Malaysia using network method," in *Proc. AIP Conf.*, vol. 1750, 2016, Art. no. 030024.
[8] Y.-L. Zhang and F. Cao, "Fine particulate matter ($PM_{2.5}$) in China at a city level," *Sci. Rep.*, vol. 5, Oct. 2015, Art. no. 14845.
[9] D. O. Logofet and E. V. Lesnaya, "The mathematics of Markov models: What Markov chains can really predict in forest successions," *Ecol. Model.*, vol. 126, nos. 2–3, pp. 285–298, 2000.
[10] Jamilatuzzahro, R. E. Caraka, and R. Herliansyah, *Aplikasi Generalized Linear Model dengan R*, 1st ed. Yogyakarta, Indonesia: INNOSAIN Graha Ilmu Yogyakarta, 2018.
[11] P. Hänggi, "Connection between deterministic and stochastic descriptions of nonlinear systems," *Helvetica Phys. Acta*, vol. 53, pp. 491–496, 1980. [Online]. Available: http://www.physik.uni-augsburg.de/theo1/hanggi/Papers/29.pdf
[12] R. E. Caraka, S. A. Bakar, and M. Tahmid, "Rainfall forecasting multi kernel support vector regression seasonal autoregressive integrated moving average (MKSVR-SARIMA)," in *Proc. AIP Conf.*, Jun. 2019, Art. no. 020014.
[13] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*, vol. 2. Boston, MA, USA: PWS, 1995, p. 734.
[14] R. E. Caraka, S. A. Bakar, M. Tahmid, H. Yasin, and I. D. Kurniawan, "Neurocomputing fundamental climate analysis," *Telkomnika*, vol. 17, no. 4, pp. 1818–1827, 2019.
[15] Suhartono, B. Maghfiroh, and S. P. Rahayu, "Hybrid VARX-SVR and GSTARX-SVR for forecasting spatio-temporal data," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, no. 4, pp. 212–218, 2019.
[16] Suhartono, I. M. G. M. Dana, and S. P. Rahayu, "Hybrid model for forecasting space-time data with calendar variation effects," *Telkomnika*, vol. 17, no. 1, pp. 118–130, 2019.
[17] X.-D. Wang, R.-C. Chen, F. Yan, Z.-Q. Zeng, and C.-Q. Hong, "Fast adaptive K-means subspace clustering for high-dimensional data," *IEEE Access*, vol. 7, pp. 42639–42651, 2019.
[18] A. I. McLeod, H. Yu, and E. Mahdi, "Time series analysis with R," in *Handbook of Statistics*, vol. 30. Amsterdam, The Netherlands: Elsevier, 2012, pp. 661–712.
[19] B. Warsito, R. Santoso, H. Yasin, and Suhartono, "Gradient based optimization in cascade forward neural network model for seasonal data," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 21, pp. 7007–7017, 2018.
[20] C. Chatfield, "Time-series forecasting," *Significance*, vol. 2, no. 3, pp. 131–133, 2005.
[21] S. E. M. Artha, H. Yasin, B. Warsito, R. Santoso, and S. Suparti, "Application of wavelet neuro-fuzzy system (WNFS) method for stock forecasting," *J. Phys., Conf. Ser.*, 2018.
[22] T. Toharudin, R. E. Caraka, A. Bachrudin, S. A. Bakar, and D. Ambarsari, "Copula-based feedforward neural network genetic algorithm cargo forecasting," *Int. J. Adv. Soft Comput. Appl.*, vol. 11, no. 1, pp. 22–23, 2019.
[23] Y. Hasbi, W. Budi, and S. Rukun, "Feed forward neural network modeling for rainfall prediction," in *Proc. E3S Web Conf.*, 2018, p. 05017.
[24] S. Saroha and S. K. Aggarwal, "Hybrid WT-PSO based neural networks for single step-ahead wind power prediction for Ontario electricity market," *Int. J. Elect. Eng. Inform.*, vol. 7, no. 2, pp. 237–249, 2015.
[25] B. Warsito, H. Yasin, and A. Prahutama, "Particle swarm optimization versus gradient based methods in optimizing neural network," *J. Phys., Conf. Ser.*, vol. 1217, no. 1, 2019, Art. no. 012101.
[26] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, "Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation," *Atmos. Environ.*, vol. 107, pp. 118–128, Apr. 2015.
[27] J. H. Chang and C. Y. Tseng, "Analysis of correlation between secondary $PM_{2.5}$ and factory pollution sources by using ANN and the correlation coefficient," *IEEE Access*, vol. 5, pp. 22812–22822, 2017.
[28] J. D. Hamilton, *Time Series Analysis*, vol. 39, no. 1. Princeton, NJ, USA: Princeton Univ. Press, 1994.
[29] N. Hashimzade, M. Thornton, and H. Lütkepohl, "Vector autoregressive models," in *Handbook of Research Methods and Applications in Empirical Macroeconomics*. 2013.
[30] M. Apriliadara, S. Suhartono, and D. D. Prastyo, "VARI-X model for currency inflow and outflow forecasting with Eid Fitr effect in Indonesia," in *Proc. AIP Conf.*, 2016.
[31] R. E. Caraka and M. Tahmid, *Statistika Klimatologi*, 1st ed. Yogyakarta, Indonesia: Mobius Graha Ilmu, 2019.
[32] D. Oliva, M. A. Elaziz, and S. Hinojosa, "Optimization," in *Studies in Computational Intelligence*. Springer, 2019.
[33] V. Krishnan, *Probability and Random Processes*. Hoboken, NJ, USA: Wiley, 2005.
[34] F. Biagini and M. Campanino, "Discrete time Markov chains," in *Proc. 3rd UNITEXT Math.*, vol. 2, 2016, pp. 1–34.
[35] J. Blanchet, G. Gallego, and V. Goyal, "A Markov chain approximation to choice modeling," *Oper. Res.*, vol. 64, no. 4, pp. 886–905, 2016.
[36] D. A. Levin, P. Yuval, and E. L. Wilmer, *Markov Chains and Mixing Times*. Choice Rev., 2008.
[37] E. Zivot and J. Wang, "Vector autoregressive models for multivariate time aeries," in *Modeling Financial Time Series With S-Plus*. Springer, 2003.
[38] H. Yasin, B. Warsito, R. Santoso, and Suparti, "Soft computation vector autoregressive neural network (VAR-NN) GUI-based," in *Proc. E3S Web Conf.*, vol. 73, 2018, p. 13008.
[39] B. N. Ruchjana, S. A. Borovkova, and H. P. Lopuhaa, "Least squares estimation of generalized space time autoregressive (GSTAR) model and its properties," in *Proc. AIP Conf.*, vol. 1450, 2012, pp. 61–64.
[40] D. D. Prastyo, F. S. Nabila, Suhartono, M. H. Lee, N. Suhermi, and S. F. Fam, "VAR and GSTAR-based feature selection in support vector regression for multivariate spatio-temporal forecasting," in *Proc. Commun. Comput. Inf. Sci.*, 2019, pp. 46–57.
[41] R. E. Caraka, W. Sugiyarto, G. Erda, and E. Sadewo, "Pengaruh inflasi terhadap impor dan ekspor di provinsi riau dan kepulauan riau menggunakan generalized spatio time series," *J. Badan Pendidikan Pelatihan Keuangan*, vol. 9, no. 2, pp. 181–198, Dec. 2016.
[42] D. J. Olive, *Linear Regression*. Springer, 2017.
[43] W. Bergsma, "A bias-correction for Cramér's V and Tschuprow's T," *J. Korean Stat. Soc.*, vol. 42, no. 3, pp. 323–328, 2013.

REZZY EKO CARAKA received the B.S. degree (S.Si.) from the Department of Statistics, Diponegoro University, the M.Sc. degree research from the School of Mathematical Sciences, National University of Malaysia, and the Ph.D. degree from the College of Informatics, Chaoyang University of Technology, Taiwan. He acts as a Researcher of the Bioinformatics and Data Science Research Center, Bina Nusantara University (BINUS), and the Department of Statistics, Padjadjaran University. He is also a Fellow of the GLM-H Lab, Department of Statistics, Seoul National University, South Korea. At the same time, he was the Co-Founder Statistical Calculator (STATCAL). His research interests include statistical climatology, climate modeling, ecological modeling, statistical machine learning, and large-scale optimization.

RUNG CHING CHEN received the B.S. degree from the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 1990, the M.S. degree from the Institute of Computer Engineering, National Taiwan University of Science and Technology, in 1987, and the Ph.D. degree from the Department of Applied Mathematics and Computer Science, National Chung Hsing University, in 1998. He is currently a Distinguished Professor with the Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan. His research interests include network technology, pattern recognition, and knowledge engineering, the IoT and data analysis, and applications of artificial intelligence.

TONI TOHARUDIN received the M.Sc. degree from the University of Leuven Belgium, in 2005, and the Ph.D. degree in spatial sciences from the University of Groningen, in 2010. He is currently an Associate Professor with the Department of Statistics, Padjadjaran University. Moreover, he acts as the head of the research group in time series and regression.

BENS PARDAMEAN received the bachelor's degree in computer science and the master's degree in computer education from California State University, Los Angeles, and the Ph.D. degree in informative research from the University of Southern California (USC). He has over 30 years of global experience in information technology, bioinformatics, and education. After successfully leading the Bioinformatics Research Interest Group, he currently holds a dual appointment as the Director of Bioinformatics and Data Science Research Center (BDSRC), and an Associate Professor of computer science with Bina Nusantara University (BINUS), Jakarta, Indonesia.

HASBI YASIN received the B.S. (S.Si.) degree from the Department of Mathematics, Diponegoro University, in 2005, and the M.Sc. degree from the Department of Statistics, ITS, Surabaya, in 2009. He is currently an Assistant Professor with the Department of Statistics, Diponegoro University. His research interests include spatial statistics and computational statistics.

SHIH HUNG WU received the B.S. degree in physics from National Taiwan University, Taiwan, in 1991, and the M.C.S. and Ph.D. degrees in computer science from National Tsing Hua University, Taiwan, in 1993 and 1999, respectively. Previously, he was a Postdoctoral Fellow of the Institute of Information Science and Academia Sinica. He is leading a group of natural language understanding research, in 2005, and acts as an Assistant Professor with the Department of Computer Science and Information Engineering, Chaoyang University of Technology. His research interests include ontology, natural language understanding, rational agent negotiation, and machine learning.