

Hybrid Particle Swarm and Conjugate Gradient Optimization in Neural Network for Prediction of Suspended Particulate Matter

Budi Warsito^{1,2*}, Alan Prahutama¹, and Hasbi Yasin¹, and Sri Sumiyati³

¹Department of Statistics, Faculty of Sciences and Mathematics, Diponegoro University, Semarang - Indonesia

²Master Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang – Indonesia

³Department of Environmental Engineering, Faculty of Engineering, Diponegoro University, Semarang - Indonesia

Abstract. The scope of this research is the use of artificial neural network models and meta-heuristic optimization of Particle Swarm Optimization (PSO) for the prediction of ambient air pollution parameter data at air quality monitoring stations in the city of Semarang, Central Java. The observed parameter is an indicator of ambient air quality, Suspended Particulate Matter (SPM). Based on air quality parameter data in previous times which is a time series data, modeling is done using Neural Networks (NN). Estimation of weights from NN is done using a hybrid method between meta-heuristic and gradient optimization. The meta-heuristic optimization method used is Particle Swarm Optimization (PSO) while the gradient based method is the Conjugate Gradient. Optimization with PSO is done first, then proceed with optimization using the Conjugate Gradient. Four scenarios of iteration selection at the PSO stage are 10, 25, 50 and 100. At the Conjugate Gradient, stage iteration is carried out up to 1000 epochs. The predicted results were compared with the PSOs and Conjugate Gradient respectively. The results show that the hybrid method provides better predictions. The number of iterations needed at the PSO stage is not too much so it is efficient in combining the two methods.

Keywords: air pollution; SPM; PSO; Neural Network; Conjugate Gradient; Prediction.

1 Introduction

Air pollution is one of the important problems in the industrial era. Air pollution becomes a major problem worldwide raising several issues for wellbeing and survival of humans as well as environment [1]. It is a major concern that has a serious toxicological effect on human health and the environment [2]. The level of air pollution is measured by the air quality index. Air Quality Index (AQI) is a standardized summary measure of ambient air quality used to express the level of health risk related to particulate and gaseous air pollution [3]. The index, first introduced by US EPA in 1998 classified ambient air quality according to concentrations of such principal air pollutants. Air monitoring assessment is an important task for the stakeholders. It is an imperative investigation undertaken for determining and understanding the degree, nature and status of the ambient air quality of an area. Some of the major air pollutants are Sulphur dioxide, Nitrogen oxides, Suspended particulate matter and Respirable particulate matter. Particulate matter (PM) is actually a mixture of several substances which may differ in various sizes. The particulate matter mainly consists of suspended and the respirable particulate matter [4].

Suspended particulate matter (SPM) consists of smoke, dust, fumes and droplets of viscous liquid. SPM are particulate having aerodynamic diameter less than

100 microns which tends to remain suspended in the atmosphere for a long period of time [5]. SPM is one of the major air pollutants responsible for degradation of the ambient air quality. These particulate are small enough to be inhaled deep into respiratory tract and pulmonary system of human beings and pose health related problems to person inhaling it. Once the magnitude of the impact caused by SPM, it becomes very important to be able to predict the pattern of pollution. Several studies have been carried out to model air pollution patterns, specifically SPM. Some of them are Goyal et al [6] which has implemented regression, ARIMA model and combination of the two and Aarnio et al [7] which use deterministic model. Comparison of the SARIMA model from the R statistical package and Holt Winter of the atmospheric Particulate Matter (PM) concentrations in the city center of SA Carlos also has been done by Pozza et al [8]. Similar research in [9] also discussed the respirable suspended particulate matter (RSPM) modeling by using ARIMA and SARIMA for the Pune city India. Nonparametric modeling also has been used as in [10] which has developed an artificial neural network (ANN) model for forecasting respirable suspended particulate matter (RSPM) concentration in a major urban area Pune Maharashtra, whereas hybrid NN-ARIMA has been conducted [11].

* Corresponding author: budiwrst2@gmail.com

A review of the 21-St century studies about particulate matter has been conducted in [12]. The review demonstrates that Support Vector Machines, Neural Network, and hybrid techniques show promise for suitable temporal particulate matter prediction. This is consistent with other studies that research that uses machine learning such as Neural Network provides better results. Neural network is particularly useful in learning a training dataset without prior knowledge [13]. But, one of the complicated problems is about weight adjusting. The using of a gradient based methods may result in the local minimum problem. One of the popular methods to avoid this problem is repeated training with random starting weights, but it requires extensive computational time. The traditional algorithms are local search algorithms that exploit current solution to generate a new solution. However, they lack in exploration ability to finds local minima of an optimization problem [14]. This problem underlies the use of meta-heuristic optimization to estimate the weight of a neural network. Results show that meta-heuristic training is capable to obtain higher accuracy than stochastic gradient based method [15]. Besides, hybrid optimization has also been proposed as in [16, 17]. In this research, a hybrid optimization between the Particle Swarm Optimization (PSO) and Conjugate Gradient is developed for training neural networks (NN) and undertaking SPM data of time series problems. We called the method as PSO-CG optimization. The four scenarios have been proposed and compared with the PSOs and Conjugate Gradient respectively.

2 Material and Methods

2.1. Data

The data used in this study is the monthly Suspended Particulate Matter (SPM) data in Semarang City, Central Java from January 2008 to December 2017. They are obtained from the Meteorology Climatology and Geophysics Agency of Central Java Province. The data length is 120 and is divided into two parts, the first 96 data as training and the remaining 24 as testing. The input model is identified by using the correlation between past and current series. Significant correlation values from past data with the current are used as input candidates.

2.2 Neural Network Architecture

Neural network architecture used in this research is Feed Forward Neural Network. Due to its use for time series modeling, input of the network is past value of the series, in this case, is a series of SPM data. The past values are determined until a certain lag. A hidden layer is between the input layer and the output layer. This layer is used to perform the weighted sum of the inputs with a nonlinear activation function, the commonly used is the sigmoid logistics function. After processing, the network sends the results to the output layer. The sending is also done with a weighted sum of the hidden layers. At the output

layer, processing is carried out with a linear activation function. The network architecture can be described as in Figure 1.

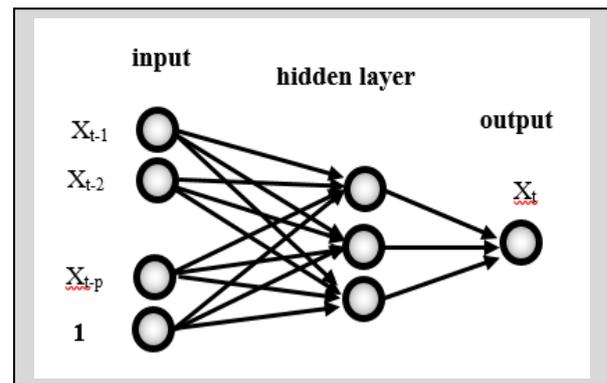


Fig. 1. Neural network architecture for time series prediction with n neuron in hidden layer and the input at lags 1, ..., p.

Mathematically, architecture as in Fig. 1 can be written as follows:

$$x_t = w^b + \sum_{j=1}^k w_j^o f^h \left(w_j^b + \sum_{i=1}^p w_{ji}^h x_{t-i} \right) \quad (1)$$

Symbol f^h represents the activation function at the hidden layer and $w = (w^b, w_j^o, w_j^b, w_{ji}^h)$ is the weights vector.

Logistic sigmoid, $f(x) = 1/(1 + \exp(-x))$, is the activation function commonly used. The number of hidden units can be taken as many as the input unit. The task of optimization is to obtain weights of the vector w so that it gets the smallest possible error or the resulting output is as close as possible to the target. In this research, the measure of accuracy used is the mean square error.

2.3 Hybrid PSO and Conjugate Gradient

The weights of FFNN described in eq. (1) should be estimated. In the hybrid PSO and Conjugate Gradient, the first stage is obtaining the weights of a network by using PSO. The initial weights were generated first by randomly generating of the initial position in each particle. The initial velocity also has been generated randomly. If the initial condition was motionless, the initial velocity is considered as null. In each experiment, the population size (swarm size) is 10 and the PSO parameters are determined as follows: Inertia Weight = 1, Inertia Weight Damping Ratio = 0.99, Personal Learning Coefficient = 1.5, and Global Learning Coefficient = 2.0. In the PSO stage, the maximum number of iterations is determined with four scenarios, i.e. 10, 25, 50 and 100 iterations, respectively. The weights obtained from this stage are the initial weights of the Conjugate Gradient stage. In this last stage, the maximum epoch is 1000. The results are compared with the PSOs and Conjugate Gradient with 1000 iterations, respectively.

3 Results and Discussion

The first stage of neural network modeling for time series prediction is determining lags of input. From the results of identification, the values of lag 1 and 2 were chosen as input, i.e. x_{t-1} and x_{t-2} . According to this result, the number of the hidden units can be determined as two. Through the addition of bias as input, and considering the architecture in Fig. 1 and eq. (1), the number of weights should be estimated is 9. By using logistic sigmoid as activation function, eq. (1) became:

$$x_t = w^b + \sum_{j=1}^2 w_j^o \left(\frac{1}{1 + \exp\left(w_j^h + \sum_{i=1}^2 w_{ji}^h x_{t-i}\right)} \right) \quad (2)$$

Obtaining the weights vector w of the model in eq. (2) been the main task in this stage. After processing the network in eq. (2) using PSO-CG optimization with four scenarios for several times, the results can be summarized as in table 1. In each experiment, the CG stage needs no finish until 1000 epoch to reach the convergence. The comparison with PSO and Conjugate Gradient also has been conducted.

Table 1. Error prediction of SPM by using a neural networks with various optimization.

Method	Scenario	MSE-train ($\times 10^3$)	MSE-test ($\times 10^3$)
PSO-CG	10 iteration – 1000 epoch	3.0464	1.8735
	25 iteration – 1000 epoch	3.1243	1.4918
	50 iteration – 1000 epoch	3.0887	1.8392
	100 iteration – 1000 epoch	3.0049	1.9422
PSO	1000 iteration	2.9217	1.8910
CG	1000 epoch	3.0715	1.6372

An interesting result could be discussed from table 1. We found that the best in-sample prediction was PSO only. It got the least MSE-train, better than the others. Some of PSO-CG made an approach to the PSOs and the others have so far results. The CG also reached a near the PSOs but still higher. However, we could also found that the best in-sample prediction does not guarantee the best out-sample prediction. We found that the best out-sample prediction, the least MSE-test, was given from PSO-CG with the scenario of 25 iteration – 1000 epoch, although it was not so fine at in-sample prediction. The result of out-sample prediction from CG also better than the PSOs. It has a lower MSE-test. From the results, we can say that the PSO has given the best for in-sample prediction whereas the PSO-CG given the best result for

out-sample prediction. The PSO-CG does not need to require a lot of iterations in the PSO stage to obtain the initial weights of the CG state. It could be the strength of the proposed procedure. Plot of in-sample and out-sample predictions are shown in Figure 2 and Figure 3, respectively.

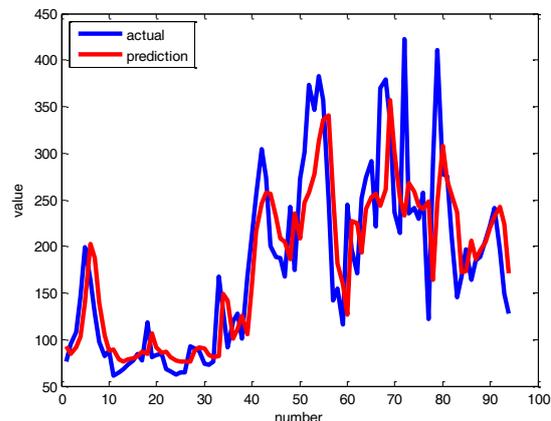


Fig. 2. Plot of the actual and in-sample prediction of SPM by using training data.

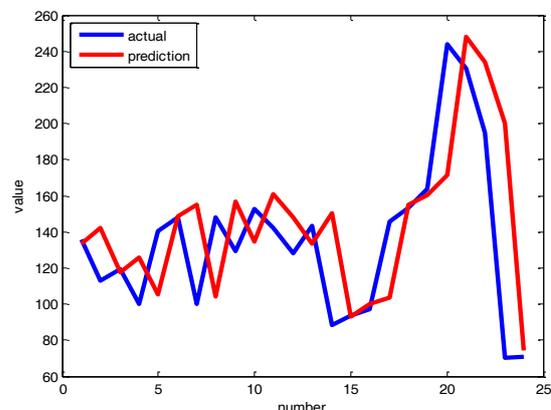


Fig. 3. Plot of the actual and out-sample prediction of SPM by using testing data.

Plot of Figure 2 showed that the in-sample prediction of SPM data, the red line, can approach the actual which was described by the blue line. The model output versus desired output as seen in Fig. 2 showed that most of the forecasted values of SPM were close to the expected values, which indicate that the procedure of the modeling has been quite successful. However, the error at the high values period is still higher than the other period. It also occurred in the results of the research by Wongsathan and Seedadan [11]. A similar result was also obtained in the out-sample prediction shown in Fig. 3, where the forecast does not deviate far from the actual one. It indicated that the comparison between measured data and the predicted monthly SPM data can be considered fairly good. We can state that the proposed procedure is capable of producing good predictions for both training data and testing data. It in accordance with the results in [12, 18] that non-linear approaches like neural networks have suitable for forecasting the SPM

data. This also answers the contradictory from [19] which was stated the currently available ANN models are not applicable for future particulate matter prediction.

4 Conclusion

Prediction of suspended particulate matter (SPM) by using neural network has been conducted. Procedure for obtaining weights of neural networks was done by using combination of meta-heuristic and gradient based optimization. Hybrid PSO-CG was developed as a way to parse the gradient-based optimization problem which is often stuck at the local optimum. The predictions obtained from the SPM data corroborate the argument for choosing this procedure in a more effective way. Comparison of this procedure with a hybrid between meta-heuristics can be an open problem in the future.

This research was supported by Institute of Research and Community Service (LPPM) Diponegoro University with the scheme of Research of Development and Application (RPP) with the contract number: 474-29/UN7.P4.3/PP/2019.

References

1. P. Hardik, P. Darshana, & D. Nishith, *Assessment of Air Quality by Air Quality Index of an Urban Area of Arid Zone of India*. Int. Journal of Advance Research in Science and Engineering, **5**(09) (2016)
2. A. Ghorani-Azam, B. Riahi-Zanjani, & M. Balali-Mood, *Effects of air pollution on human health and practical measures for prevention in Iran*. Journal of research in medical sciences, the official journal of Isfahan University of Medical Sciences, **21**, (2016)
3. M. Kowalska, L. Ośródk, K. Klejnowski, J.E. Zejda, E. Krajny & M. Wojtylak, *Air quality index and its significance in environmental health risk communication*. Archives of Environmental Protection, **35**(1), 13-21 (2009)
4. S.J.P. Buddharatna, J. Godbole, International Journal for Scientific Research & Development, **4**(02), ISSN (online): 2321-0613 (2016)
5. R.E. Lamare, & S.S. Chaturvedi, *Suspended particulate matter in ambient air of Shillong city, Meghalaya, India*. Ind J Sci Res Technol, **2**, 37-41 (2014)
6. P. Goyal, A.T. Chan, & N. Jaiswal, *Statistical models for the prediction of respirable suspended particulate matter in urban cities*. Atmospheric environment, **40**(11), 2068-2077 (2006)
7. M.A. Aarnio, J. Kukkonen, L. Kangas, M. Kauhaniemi, A. Kousa, C. Hendriks, & T. Elolähde, *Modelling of particulate matter concentrations and source contributions in the Helsinki Metropolitan Area in 2008 and 2010* (2016)
8. S.A. Pozza, E.P. Lima, T.T. Comin, M.L. Gimenes, & J.R. Coury, *Time series analysis of PM_{2.5} and PM_{10-2.5} mass concentration in the city of Sao Carlos, Brazil*. International Journal of Environment and Pollution, **41**(1-2), 90-108 (2010)
9. V. Naveen & N. Anu, *Time Series Analysis to Forecast Air Quality Indices in Thiruvananthapuram District, Kerala, India*. J Eng Res Appl, **7**(6), 66-84 (2017)
10. N.P. Sonaje, S.J. Mane, & A.S. Kote, *Artificial Neural Network for One-Day-Ahead Forecasting of Respirable Suspended Particulate Matter Concentration*. **1**(6). pp. 95-99 (2013)
11. R. Wongsathan, & I. Seedadan, *A hybrid ARIMA and Neural Networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area*. Procedia Computer Science, **86**, 273-276 (2016)
12. H.T. Shahraiyni & S. Sodoudi, *Statistical modeling approaches for PM₁₀ prediction in urban areas; A review of 21st-century studies*. Atmosphere, **7**(2), 15 (2016)
13. A. Atakulreka & D. Sutivong, *Avoiding local minima in feedforward neural networks by simultaneous learning*. Australasian Joint Conference on Artificial Intelligence (pp. 100-109). Springer, Berlin, Heidelberg (2007)
14. V. K. Ojha, A. Abraham & V. Snášel, *Metaheuristic design of feedforward neural networks: A review of two decades of research*. Engineering Applications of Artificial Intelligence, **60**, 97-116 (2017)
15. J.R. Delgado, *Meta-heuristic Optimization Algorithms for Neural Networks Training Thesis for: Master Degree*, University of Information Sciences (2018)
16. H.H. Örkçü, M. Doğan & M. Örkçü. *A Hybrid Applied Optimization Algorithm for Training Multi-Layer Neural Networks in the Data Classification*. Gazi University Journal of Science, **28**(1), 115-132 (2015)
17. V. K. Ojha, A. Abraham & V. Snášel, *Metaheuristic design of feedforward neural networks: A review of two decades of research*. Engineering Applications of Artificial Intelligence, **60**, 97-116 (2017)
18. A. Chaloulakou, G. Grivas, N. Spyrellis, *Neural network and multiple regression models for PM prediction in Athens: A comparative assessment*. J. Air Waste Manag. Assoc. **53**, 1183-1190 (2003)
19. J. Kukkonen, L. Partanen, A. Karppinen, J. Walden, R. Kartastenpää, P. Aarnio, T. Koskentalo, R. Berkowicz, *Evaluation of the OSPM model combined with an urban background model against the data measured in 1997 in Runeberg Street, Helsinki*. Atmos. Environ. **37**, 1101-1112 (2003)