

Comparison of Suspended Particulate Matter Prediction Based on Linear and Non- Linear Models

by Sri Sumiyati

Submission date: 18-Mar-2023 04:56PM (UTC+0700)

Submission ID: 2039954675

File name: PROSID_C.16.pdf (710.58K)

Word count: 3174

Character count: 16243

PAPER · OPEN ACCESS

Comparison of Suspended Particulate Matter Prediction Based on Linear and Non-Linear Models

To cite this article: S Sumiyati and B Warsito 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **448** 012029

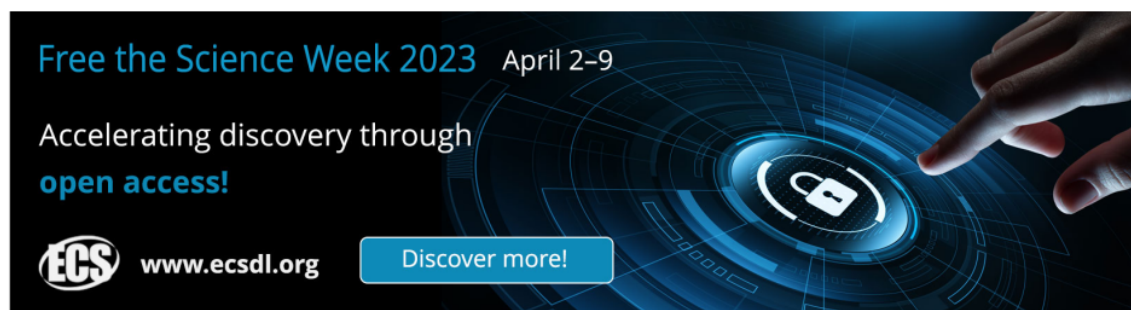
View the [article online](#) for updates and enhancements.

You may also like

- [Impact identification on concrete panels using a surface-bonded smart piezoelectric module system](#)
Ayumi Manawadu and Pizhong Qiao


- [Ferromagnetic and Superparamagnetic Contributions in the Magnetoresistance of Electrodeposited Co-Cu/Cu Multilayers](#)
Qun-Xian Liu, László Péter, József Pádár et al.

- [Study of user influence in routine SPM data processing](#)
D Neas and P Klapetek



Free the Science Week 2023 April 2-9

Accelerating discovery through
open access!

 www.ecsd.org [Discover more!](#)

The banner features a dark blue background with a futuristic, glowing blue interface. A hand is shown interacting with a circular element that has a padlock icon, symbolizing the unlocking of science through open access. The text is white and light blue, providing clear information about the event and a call to action.

Comparison of Suspended Particulate Matter Prediction Based on Linear and Non-Linear Models

S Sumiyati¹ and B Warsito²

²⁵
¹Department of Environmental Engineering, Faculty of Engineering, Diponegoro University Semarang Indonesia

²Department of Statistics, Faculty of Science and Mathematics, Diponegoro University Semarang Indonesia

budiwrst2@gmail.com

¹³
Abstract. Air pollution has been a serious problem in recent years. Air pollutants consist of gaseous pollutants, odours, and suspended particulate matter (SPM) such as dust, fumes, mist, and smoke. SPM has the potential to cause environmental and health problems. With the aim to anticipate the impact, SPM prediction from time to time is needed. In this research, we compared four models for predicting the SPM data. The two linear models selected were ARIMA and wavelet whereas the two non-linear models were neural networks based models, i.e. Feed Forward Neural Network (FFNN) and General Regression Neural Network (GRNN). All four models are built with the same input, which were the past data at the same lagged time based on the best ARIMA model. By using lagged time data as input, the goal is to predict the current of SPM data. Model accuracy is measured based on RMSE values, both in training and testing data. Data processing has provided interesting results that show the superiority of non-linear models over linear models, especially in the training data.

1. Introduction

Air pollution is defined as the anthropogenic emission of harmful chemicals that alter the chemical composition of the natural atmosphere and have an adverse effect on the health of living things, an adverse effect on anthropogenic or natural non-living structures or reduce the air's visibility [1]. The characterization of particulate matter is influenced by meteorological conditions, including temperature, humidity, rainfall and wind speed [2]. Suspended particulate matter (SPM) is an important component of the air pollution mixture. It is the most complex pollutant of the six pollutants regulated by the US EPA as criteria of air pollutants [3]. SPM is one of the most important indoor air pollutants [4]. It also led to serious health hazards in human beings [5]. It has been connected to various detrimental health outcomes as well as having general environmental effects [6]. The source of the particles plays a role in determining the health effects of particles. As further effect, it also responsible for changing global climate [7]. By definition, particulate matter are small solid particles which are able to move freely in the atmosphere [8]. In the simplest terms, particulate air pollution is defined as anything solid or liquid suspended in the air [9].

Development of statistical modeling for air pollution as SPM was very important for predicting the future values. Many studies have been carried out to model SPM data [10-11]. In time series fields, various modelling also have been developed in both linear [12] and non-linear models [13]. The main problem in time series modelling is how to obtain the best model between the various types available.



¹⁵
Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

Therefore, it is important to develop experiments for comparing the two basic methods i.e linear and non-linear models so that the more efficient one would be obtained. This research discusses the use of ARIMA and wavelet models as parts of linear models and Feed Forward Neural Network (FFNN) and General Regression Neural Network (GRNN) as representative of non-linear models. Several previous studies relating to the use of ARIMA [14,15] and wavelet for air pollution prediction [16]. Some experiments of the using neural network for air quality also have done [17,18]. In this paper, the use of the two types of time series model were applied on the monthly SPM data in Central Java province, Indonesia. The comparative studies were analyzed by using RMSE criteria in both in-sample and out-sample predictions.

2. Methods

The monthly data of SPM in Semarang City, Central Java province from January 2008 to December 2017 taken from the Meteorology Climatology and Geophysics Agency of Central Java Province was analyzed. The data is divided into two parts for goodness of fit purposes. The first 108 data is used as training and the remaining as testing. The training data set is needed for obtaining the best model based on the in-sample prediction whereas the testing data set is required for reaching the best accuracy based on the out-sample prediction. We have investigated the data using four linear and non-linear models i.e ARIMA, Wavelet, FFNN and GRNN. A brief explanation of the four models is given in the following section.

2.1. ARIMA model

ARIMA model is a practical importance in forecasting which inclusive Autoregressive, Integrated and Moving Average. The Autoregressive (AR) model is suitable for stationary time series data patterns and the Interated (I) is needed for nonstationary data through differencing. In AR(p) model, the current value of the series x_t is a linear combination of the p most recent values of itself. Combination of AR(p) and MA(q) forms ARMA(p,q) model.

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \tag{1}$$

By considering d = 1, we can write ARIMA (p,1,q) process as:

$$x_t = (1 + \phi_1)x_{t-1} + (\phi_2 - \phi_1)x_{t-2} + \dots + (\phi_p - \phi_{p-1})x_{t-p} - \phi_p x_{t-p-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \tag{2}$$

In the case where the seasonal components are included, the model is called as seasonal ARIMA (SARIMA). The generalized form of SARIMA model can be written as:

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D x_t = \theta_q(B)\Theta_Q(B)^S \varepsilon_t \tag{3}$$

where

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ \Phi_p(B) &= 1 - \Phi_1 B^s - \dots - \Phi_p B^{ps} \\ \theta_q(B) &= 1 - \theta_1 B - \dots - \theta_q B^q \\ \Theta_Q(B) &= 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs} \end{aligned}$$

2.2. Wavelet model

Let $f(x)$ be a function in $L^2(\mathbb{R})$. We can write $f(x)$ as:

$$f(x) = \sum_k c_{j,k} \phi_{j,k}(x) + \sum_{j_0}^j \sum_k d_{j,k} \psi_{j,k}(x), \quad \forall k \in \mathbb{Z} \tag{4}$$

where

$$c_{j,k} = \langle f, \phi_{j,k} \rangle = \int f(x) \phi_{j,k}(x) dx$$

and

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int f(x) \psi_{j,k}(x) dx$$

are collectively known as an approximation and detailed coefficients. A signal f can be decomposed as in simplest form; $f = A + D$, where A and D are called Approximation and Details of the given signal f . Approximation is average of the signal and hence represents low frequency components, while Detail is the difference of the signal and hence represents high frequency components [19].

2.3. Feed forward neural network

FFNN is the main class of neural network model. There are three layers in this architecture namely input layer, hidden layer and output layer. In time series modelling, the input consists of lagged variables of the series whereas the output is current values. The hidden layer contains a non-linear activation function which is transfer signal from input layer to output layer. The architecture of FFNN can be seen in figure 1.

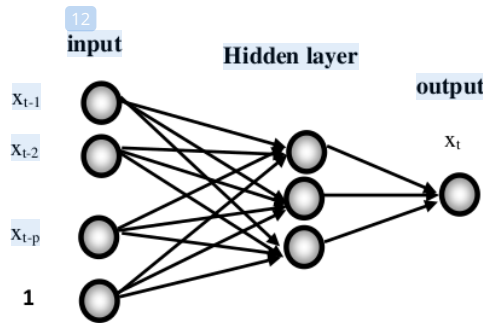


Figure 1. FFNN architecture for time series data

2.4. General regression neural network

GRNN is the other class of neural network. Different with FFNN, the GRNN consists of four layers processing, namely input layer, pattern layer, summation layer and output layer. The architecture can be seen in figure 2.

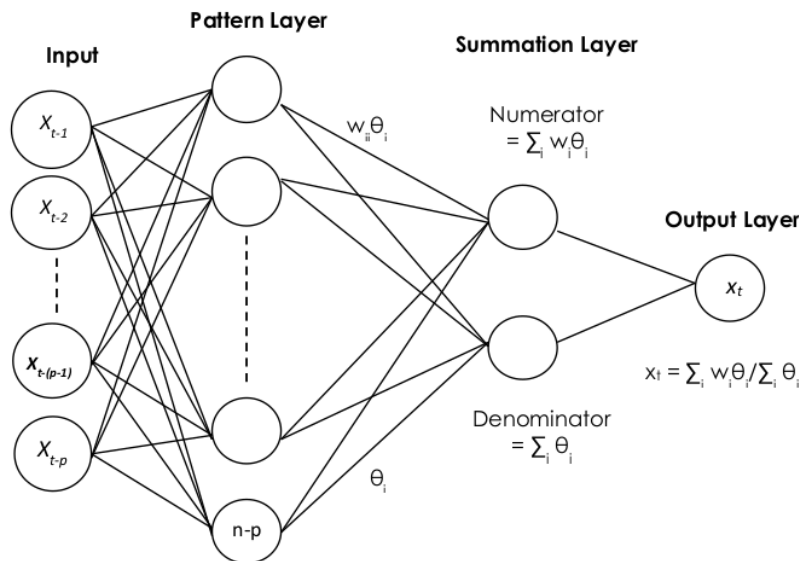


Figure 2. GRNN architecture for time series data

3. Results and discussion

The results of the investigation by using ARIMA model was developed first. The best model for the intended data was SARIMA (0, 1, 0) (0, 1, 1)¹². The result of ARIMA model determine the input of the three other models, i.e lagged variables of 1, 12 and 13. In wavelet model, we used Maximal Overlap Discrete Wavelet Transform (MODWT) as decomposition with haar wavelet. The level of decomposition and the number of coefficients in each level are determined by using Murtagh et al. [20]. In FFNN model, the number of units in hidden layer is determined by some experiments of 1-6 units. The activation function in hidden layer is logistic sigmoid whereas linear function is used as activation function in output layer. All other components in FFNN is determined as in MATLAB routine. In GRNN model, the number of units in pattern layer is the same with the input. The experimental results of the four chosen models are presented in table 1.

Table 1. Results of various scenarios.

No	Model	RMSE	
		in-sample	out-sample
1	ARIMA	7.4171	6.7453
2	Wavelet	6.0384	6.8436
3	FFNN	4.9058	6.5729
4	GRNN	3.6503	7.2678

The main results as shown in table 1 can be described as follows. GRNN has obtained the best in-sample prediction but is not able to get a good out-sample prediction. Its whereas FFNN has given the best out-sample prediction. In fact, that value is the worst, worse than the linear models. On the other hand, FFNN has given the best out-sample prediction. It also has provided a good result for in-sample prediction. Both predictions were better than the linear models. In this case, what is seen as consideration for determining the best model is not only the result of in-sample prediction, but also the out-sample prediction. With these considerations in mind, we have chosen FFNN for the selected model. Based on the analysis of table 1, overall, we also noticed that the non-linear models, or specifically neural network models, showed the superiority over the non-linear models. The goodness of the neural network model for predicting air pollution data or more specifically particulate matter can also be found in the previous studies [21,22]. The non-linear techniques that have been tried for predicting SPM data outperform the linear techniques. The similar results also stated by Shahraiyini and Sodoudi [23], where neural network was one of the recommended techniques. Masih [23] concluded that neural network and Support Vector Machine (SVM) based approaches are preferred for air pollution forecasting. Contrary but not extremely, Ibarra-Berastegi [24] stated that simple linear models work as well as more complicated neural networks or slightly worse. The statement was contradiction with this research that resulting the extreme different between linear and non-linear models. In the comparison of different machine learning approaches, Karimian et al. [25] concluded that the LSTM model -a new hybrid model based on long short-term memory- achieved the best results, better than multiple additive regression trees (MART) and deep feedforward neural network (DFNN). On this side, air pollution modelling or specifically particulate matter becomes more interesting if it involves various machine learning-based methods.

Now, we display the goodness of the chosen model visually. A good prediction is the one that approaches the actual. Figures 3 and 4 show plots of the in-sample and out-sample predictions against the actual, respectively. In the plots, the actual is plotted in the blue line whereas the prediction is the red one. Figure 3 shows that the results of the in-sample prediction are closest to the actual for all of training data. Prediction results at most observation points are very close to the actual. Fluctuation patterns from the data can also be followed by the model very well. This proves that the model has succeeded in making in-sample predictions. Figure 4 shows plot of the comparison of out-sample

prediction with actual values. As in modelling, the out-sample prediction can also approach the pattern of the actual value for the next few stages. Therefore, the model chosen can be said to be robust for modeling and forecasting.

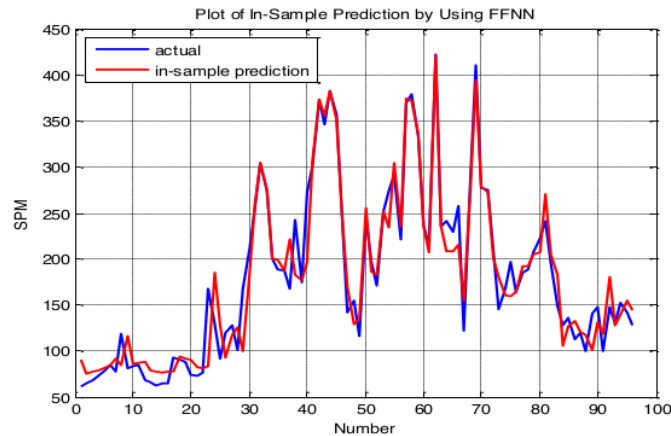


Figure 3. The in-sample prediction of the SPM data by using FFNN

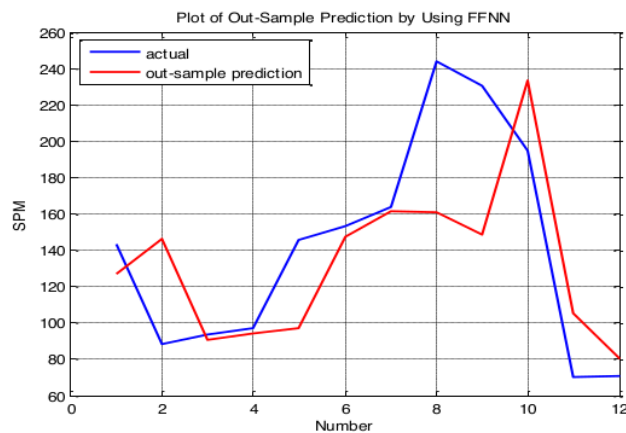


Figure 4. The out-sample prediction of the SPM data by using FFNN

4. Conclusion

17

Comparison of some linear and non-linear time series models for SPM data have been developed. The non-linear models are more suitable for predicting the SPM data. GRNN given the best in-sample prediction whereas FFNN given the best out-sample prediction. It can be done by varying the gradient based optimization methods and make a comparison between them. As an alternative for the future works, hybrid of linear and non-linear models and hybrid of neural network with other non-linear models are recommended. The using of heuristic optimization for various neural network models like RBFNN, CFNN, RNN and the hybrid with other technique can be used as consideration. The use of external variables can also be considered as one way to improve the performance of the model.

References

- [1] Daly A & Zannetti P 2007 An introduction to air pollution—definitions, classifications, and history *Ambient air pollution. P. Zannetti, D. Al-Ajmi and S. Al-Rashied, The Arab School for Science and Technology and The EnviroComp Institute*, 1-14.
- [2] Araújo I, Costa D & De Moraes R 2014 Identification and characterization of particulate matter concentrations at construction jobsites *Sustainability* **6(11)** 7666-7688
- [3] Koenig J Q 2000 Health effects of particulate matter *In: Health Effects of Ambient Air Pollution* 115-137 Springer Boston MA
- [4] Ešťoková A, Številová N & Kubincová L 2010 Particulate matter investigation in indoor environment *Global Nest Journal* **12(1)** 20-26
- [5] Nagar J K, Akolkar A B & Kumar R 2014 A review on airborne particulate matter and its sources, chemical composition and impact on human respiratory system *Int J Environ Sci* **5(2)** 447-463
- [6] Vallius M 2005 Characteristics and sources of fine particulate matter in urban air *Publications of the National Public Health Institute Mannerheimintie 166 FI-00300 Helsinki Finland*
- [7] Khan M F, Hirano K & Masunaga S 2012 Assessment of the sources of suspended particulate matter aerosol using US EPA PMF 3.0 *Environmental monitoring and assessment* **184(2)** 1063-1083.
- [8] Kraus M & Šenitková I J 2017 Particulate Matter Mass Concentration in Residential Prefabricated Buildings Related to Temperature and Moisture *IOP Conference Series: Materials Science and Engineering* **245(4)** p 042068 IOP Publishing
- [9] Dockery D W 2009 Health effects of particulate air pollution *Annals of Epidemiology* **19(4)** 257-263
- [10] Aarnio M A, Kukkonen J, Kangas L, Kauhaniemi M, Kousa A, Hendriks C, ... & Elolähde T 2016 Modelling of particulate matter concentrations and source contributions in the Helsinki Metropolitan Area in 2008 and 2010 *Boreal Environment Research* **21** 445-460
- [11] Ronquillo-Lomeli G, Herrera-Ruiz G, Ríos-Moreno J, Ramirez-Maya I & Trejo-Perea M 2018 Total Suspended Particle Emissions Modelling in an Industrial Boiler *Energies* **11(11)** 3097
- [12] Rahman M & Hossain M A 2012 Time series analysis model for particulate matter of air pollution data in Dhaka City *Asian Journal of Water, Environment and Pollution* **9(4)** 63-69
- [13] Warsito B, Prahutama A, Yasin H & Sumiyati S 2019 Hybrid Particle Swarm and Conjugate Gradient Optimization in Neural Network for Prediction of Suspended Particulate Matter *E3S Web of Conferences* **125** 25007 EDP Sciences
- [14] Guarnaccia C, Breton J G C, Breton R M C, Tepedino C, Quartieri J & Mastorakis N E 2018 ARIMA models application to air pollution data in Monterrey, Mexico *AIP Conference Proceedings* **1982(1)** p. 020041 AIP Publishing
- [15] Lee M H, Rahman N H A, Latif M T, Nor M E & Kamisan N A B 2012 Seasonal ARIMA for forecasting air pollution index: A case study *American Journal of Applied Sciences* **9(4)** 570-578
- [16] Su Y, Sha Y, Zhai G, Zong S & Jia J 2019 Comparison of air pollution in Shanghai and Lanzhou based on wavelet transform *Environmental Science and Pollution Research* **26** 16825-16834
- [17] Valencia R, Sanchez G & Diaz I 2016 A General Regression Neural Network for modeling the behavior of PM10 concentration level in Santa Marta, Columbia *ARP Journal of Engineering and Applied Sciences* **11(11)**
- [18] Russo A, Lind P G, Raischel F, Trigo R & Mendes M 2015 Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales *Atmospheric Pollution Research* **6(3)** 540-549
- [19] Kumar A 2017 Wavelet Analysis of Air Pollution due to Carbon Monoxide *International Journal of Advance Research, Ideas and Innovations in Technology* **3(4)** 131-136
- [20] Murtagh F, Starck J L & Renaud O 2004 On neuro-wavelet modeling *Decision Support Systems* **37(4)** 475-484

- [21] Ronquillo-Lomeli G, Herrera-Ruiz G, Ríos-Moreno J, Ramirez-Maya I & Trejo-Perea M 2018 Total Suspended Particle Emissions Modelling in an Industrial Boiler *Energies* **11(11)** 3097
- [22] Saeed S, Hussain L, Awan I A & Idris A 2017 Comparative analysis of different statistical methods for prediction of PM_{2.5} and PM₁₀ concentrations in advance for several hours *International Journal of Computer Science and Network Security* **17(11)** 45-52
- [23] Masih A 2019 Machine learning algorithms in air quality modeling *Global Journal of Environmental Science and Management* **5(4)** 515-534
- [24] Ibarra-Berastegi G 2006 Short-term prediction of air pollution levels using neural networks *WIT Transactions on Ecology and the Environment* **86** 23-31
- [25] Kaimian H, Li Q, Wu C, Qi Y, Mo Y, Chen G, Zhang X & Sachdeva S 2019 Evaluation of different machine learning approaches to forecasting PM_{2.5} mass concentrations *Aerosol and Air Quality Research* **19(6)** 1400-1410

Comparison of Suspended Particulate Matter Prediction Based on Linear and Non-Linear Models

ORIGINALITY REPORT

12%

SIMILARITY INDEX

10%

INTERNET SOURCES

7%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1 He Qun, Shan Wei. "Research on Optimization Algorithm of Data Flow Forecasting Analysis Based on ARIMA Model", 2010 International Conference on Challenges in Environmental Science and Computer Engineering, 2010
Publication <1%

2 Yona Falinie A. Gaus, Hongying Meng, Asim Jan, Fan Zhang, Saeed Turabzadeh. "Automatic affective dimension recognition from naturalistic facial expressions based on wavelet filtering and PLS regression", 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015
Publication <1%

3 api.openalex.org
Internet Source <1%

4 ejurnal.seminar-id.com
Internet Source <1%

paper.ijcsns.org

5	Internet Source	<1 %
6	pt.scribd.com Internet Source	<1 %
7	tel.archives-ouvertes.fr Internet Source	<1 %
8	www.ias.ac.in Internet Source	<1 %
9	www.tib.eu Internet Source	<1 %
10	Submitted to University of Leicester Student Paper	<1 %
11	Submitted to University of Surrey Student Paper	<1 %
12	ndl.ethernet.edu.et Internet Source	<1 %
13	rdhsofficegampaha.org Internet Source	<1 %
14	N. I. Molina-Gómez, J. L. Díaz-Arévalo, P. A. López-Jiménez. "Air quality and urban sustainable development: the application of machine learning tools", International Journal of Environmental Science and Technology, 2020 Publication	<1 %

15	mts.intechopen.com Internet Source	<1 %
16	sportdocbox.com Internet Source	<1 %
17	Christian Francq. "Arma models with bilinear innovations", Communications in Statistics. Stochastic Models, 1999 Publication	<1 %
18	Computational Intelligence in Economics and Finance, 2007. Publication	<1 %
19	assets.researchsquare.com Internet Source	<1 %
20	dspace.knust.edu.gh Internet Source	<1 %
21	dspace.stir.ac.uk Internet Source	<1 %
22	journal.gnest.org Internet Source	<1 %
23	thalis.math.upatras.gr Internet Source	<1 %
24	www.frontiersin.org Internet Source	<1 %
25	A. C. Kumoro, M. Hasan. " MODELING OF SUPERCRITICAL CARBON DIOXIDE	<1 %

EXTRACTION OF ANDROGRAPHOLIDE FROM
LEAVES ", Chemical Engineering
Communications, 2007

Publication

26

Guillermo Ronquillo-Lomeli, Noé Amir
Rodríguez-Olivares, Leonardo Barriga-
Rodríguez, Antonio Ramírez-Martínez et al.
"Nonlinear modeling of industrial boiler NOx
emissions", Journal of the Air & Waste
Management Association, 2022

Publication

<1 %

27

Xu, H.K.. "Viscosity approximation methods
for nonexpansive mappings", Journal of
Mathematical Analysis and Applications,
20041001

Publication

<1 %

28

ipfs.io
Internet Source

<1 %

29

papasearch.net
Internet Source

<1 %

30

Studies in Computational Intelligence, 2008.
Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On

Comparison of Suspended Particulate Matter Prediction Based on Linear and Non-Linear Models

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8
