Prediction of Carbon Monoxide Concentration with Variation of Support Vector Regression Kernel Parameter Value

Erniwati Halawa[1,2], Yazid Bindar[3], Acep Purqon[1], and Wahyu Srigutomo[1*]

[1]Physics of Earth and Complex System, Physics Department,
Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung,
Jl. Ganesa 10 Bandung 40132, Indonesia
[2]Physics Department, Universitas Negeri Medan, Medan Estate 20221, Indonesia
[3]Department of Chemical Engineering, Faculty of Industrial Technology,
Institut Teknologi Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia
*E-mail: wahyu@fi.itb.ac.id

Abstract. Human and industrial activities produce air pollutants that can cause a decline in air quality. In urban areas, transportation activities are the main source of air pollution. One of the emitted air pollutants produced by transportation is carbon monoxide (CO). The understanding of CO concentration is crucial since its overabundance beyond a certain limit will have a negative impact on human health and the environment. In this study, the support vector regression (SVR) method was used to predict CO concentration. The purpose of this study was to predict the hourly CO concentration in the Ujung Berung district, Bandung City, West Java, Indonesia with optimal prediction accuracy. An experiment was carried out by modeling the CO concentration with varying kernel parameter values to obtain accurate prediction results. The suitability of the values between error (ε), a trade-off constant (C), and variation mismatch (γ) is vital to obtain optimal prediction results. The results showed that the best prediction accuracy value was 97.68% with kernel parameter values $\varepsilon = 0.02$, $\gamma = 30$, and $C = 0.006$. These results may lead to proper decision making on environmental issues and can improve air pollution control strategies.

Keywords: *air pollution; carbon monoxide; kernel; prediction; support vector regression.*

1 Introduction

Human activities in the form of transportation, industry, and households contribute to air pollution in urban areas. A study of air pollution conducted in urban areas in 51 countries [1] concluded that 25% of urban air pollution comes from transportation or traffic activities, 15% comes from industrial activities, 20% comes from domestic fuel burning activities, and the rest comes from dust and natural salt. Road vehicles emit multiple air pollutants and greenhouse gases,

Received September 30th, 2020, Revised November 26th, 2021, Accepted for publication January 27th, 2021

DOI: 10.5614/j.math.fund.sci.2022.54.1.3

such as particulate matter (PM), hydrocarbons (HCs), carbon dioxide (CO_2), nitrogen oxides (NO_X), and carbon monoxide (CO) [2]. CO emissions result from imperfect fuel combustion, where the carbon in the fuel is incompletely oxidized, producing CO instead of CO_2. CO is extremely toxic and has no color or odor. In humans, inhaling air contaminated by CO decreases the flow of oxygen in the bloodstream, thus affecting the quality of health. Together with HCs, CO also plays a role in the formation of smog and ground-level ozone.

Accurate information about the concentration of air pollutants is significantly important for the general public. Accuracy of predicting the concentration of air pollutants ensures the accuracy of efforts in preventing and controlling air pollution in an area. The uncertainty model used is an important aspect of air pollution modeling because it can improve air pollution control strategies. In trying to control and prevent air pollution, governments, private companies and researchers directly measure the concentration of air pollutants and conduct modeling using various methods.

Many researchers have studied air pollution modeling using different approaches, which can be grouped into regression techniques [3-5], deterministic models [6-7], and artificial intelligence based models (AI) [8-11]. These modeling approaches provide satisfactory results in air pollution forecasting.

Regression and AI methods have two main advantages over deterministic models. Firstly, both models do not require real-time data about emissions, and secondly, their structure is often better known than with a deterministic model [12]. However, according to Moazami *et al.* [13], both approaches have weaknesses, especially regarding the high uncertainty of the input data, the calculations, and the inherent properties of the processes taking place in the atmosphere. Therefore, uncertainty analysis is an important aspect when modeling air pollution. The support vector machine (SVM) paradigm developed by Boser, Guyon, and Vapnik in 1992 [14] is an important method to solve forecasting problems related to air pollution [15-19]. The SVM method has a branch called support vector regression (SVR), which is primarily designed for regression problems. SVR considers not only the estimated error of the data, but also the generalizability of the model, that is, SVR's is able to improve the model's predictions when evaluating new data.

The use of SVR to predict the concentration of several air pollutants is discussed in Ortiz *et al.* [20] and Castelli *et al.* [21]. Moazami *et al.* [13] used SVR to analyze the uncertainty of air pollution and predict CO levels in Tehran, Iran. The results were compared to those from other artificial intelligence models, showing that the prediction of CO with SVR was more accurate than with the other models. In addition, Akbarzadeh *et al.* [22] predicted the daily CO concentration in the

Tehran atmosphere using SVR to minimize two error functions that reduced the output error and improved the performance of the model. They concluded that the SVR model is better than the adaptive fuzzy neuro inference system (ANFIS) and the artificial neural network (ANN) methods in predicting short-term CO concentrations. According to Brereton *et al.* [23], SVR is one of the supervised learning algorithms that exhibit high accuracy and optimal performance, and is superior to other artificial intelligence methods such as ANFIS and ANN.

Although these results suggest that the performance and accuracy of the SVR method are better than those of the AI methods, the optimization of SVR's performance remains a topic of interest for researchers. This is because the parameter values used in those studies were still very limited. In the present study, the effect of kernel parameter variation in SVR on the predicted CO concentration values was analyzed to evaluate the performance of the prediction approach. The scheme was applied to CO concentration data measured in Ujung Berung district, Bandung City, West Java, Indonesia. The Environmental Management Agency (BPLH) of Bandung City [24] has reported that the distribution of CO emissions in Bandung City is mainly contributed by the transportation sector (97.4%), followed by settlements, waste, and industry, who contributed 0.1%, 2.4%, and 0.1%, respectively.

2 Materials and Methods

This section describes the formulation, error calculation, and accuracy of SVR. The formulation was applied to CO concentration data obtained in the Ujung Berung District, Bandung City, West Java, Indonesia as the case study area of this research.

2.1 SVR Formulation

SVR is an SVM that is designed to perform a regression. SVR was introduced in 1997 by Vapnik, Golowith & Smola [25]. The basic principle is the same as for SVM, namely to find the best hyperplane. The SVR maps the input vector to a higher dimension. The goal of SVR is to find a function $f(x)$ that is a hyperplane in the form of a regression function that fits all input data with the smallest possible error [14].

The error ε is the deviation of the data sample from the hyperplane. Data deviations less than or equal to ε are acceptable. Data deviations slightly larger than ε are given a tolerance, called a soft margin, the amplitude of which is known as ξ (slack variable). The slack variable is used to solve the sampling mismatch in the optimization problem, so that the equation for obtaining the optimal hyperplane becomes:

$$\min \frac{1}{2}\|\bar{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (1)$$

with the following equation of boundary conditions:

$$\begin{aligned} & y_i - \bar{w}.\bar{x}_i - b \le \varepsilon + \xi_i \\ & \bar{w}.\bar{x}_i + b - y_i \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0 \end{aligned} \quad (2)$$

C is the capacity constant, where $C > 0$ sets the trade-off between the flatness of the function f and its sum in order to tolerate deviations higher than ε [15].

Different kernel functions can be selected to create different types of SVR models. The radial basis function (RBF) is a kernel function that has fewer tuning parameters than polynomial and sigmoid kernel functions [26] and tends to perform well. RBF, defined by parameter γ, is a better choice than other kernel functions [27]. In addition, RBF is a kernel function that can solve data with nonlinear cases. The core function parameters γ, ε, and C need to be optimized, since the correct adjustment of these parameters greatly influences the results of the prediction.

To generate different types of SVR models, different kernel functions can be chosen. The radial basis function (RBF) is a well-performing kernel function with fewer tuning parameters than polynomial and sigmoid kernel functions [26]. RBF is specified as a superior choice to other kernel functions by parameter [27]. Furthermore, RBF is a kernel function that can solve nonlinear data. The fundamental function parameters γ, ε, and C must be tuned because the correct modification of these parameters has a significant impact on the prediction results.

In this study, the first step was filtering, which involved deleting unnecessary data and modifying the data through normalization and display to make it easier to process. After that, the data were divided into three groups: training data, test data, and forecast data. The data, which included variations in kernel parameter values γ, ε, and C, were simulated using SVR. The experiment was repeated several times in order to obtain the best modeling results. Error (R^2) calculation was used to determine the best prediction outcomes.

2.2 Error Calculation and Accuracy

The fundamental purpose of utilizing modeling to find knowledge about statistical data is to create a representative model. A procedure of measuring errors and the accuracy of the modeling results with actual data is required to determine a model's representativeness. Small inaccuracies show how well the

model matches the real-world situation, and vice versa. In this study, R^2 was used to calculate the error.

In a regression model, *R*-squared, or R^2 (also known as the coefficient of determination), is a statistical metric that determines how much of the variance in the dependent variable can be explained by the independent variable. R-squared, indicates how well the data fits the regression model (the goodness of fit). R^2 is expressed as a decimal number between -∞ to 1. The basic technique of R^2 error calculation is mean squared error (MSE).

The MSE of a model is calculated by squaring the difference between the predicted and the actual data. After that, all the squares are added together and divided by the total number of samples. This is expressed in the following equation:

$$MSE(model) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

where n is the total number of data, i is the index number of the data, y_i is the i-th actual data, and $\hat{y}_i$ is the i-th predictive data.

Meanwhile, by squaring the difference between the actual data and the average of the actual data, the baseline MSE is computed. After that, all the squares are added together and divided by the total number of data. This is expressed in the following equation:

$$MSE(baseline) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{12}$$

where $\bar{y}$ is the average value of the actual data. The value of R^2 can be determined by:

$$R^2 = 1 - \frac{MSE(model)}{MSE(baseline)}. \tag{13}$$

The following equation is used to determine the level of precision:

$$Accuracy(\%) = \frac{100}{n}\sum_{i=1}^{n}\left(1 - \left|\frac{y_i - \hat{y}_i}{y_i}\right|\right). \tag{14}$$

2.3 Performance Comparison of SVR Utilizing RBF Kernel Function

To validate the SVR method utilizing the RBF kernel function, a comparison was made between SVR using various kernel functions, particularly linear and polynomial kernels. Time series data on the closing price of Bank Central Asia (BCA)'s stock market, one of Indonesia's commercial banking institutions, was used to compare the three kernel functions [28]. This type of time series data were

used because it has two distinctive behaviors: high frequency variation and low frequency variation, which are both needed for testing SVR performance. When compared to the other two kernel functions on the daily closing price data (Figure 1), the RBF kernel function had the best R^2 modeling error of 0.978, whereas the best R^2 modeling errors for the linear and polynomial kernel functions were 0.908 and 0.970, respectively, ensuring that the RBF kernel was the finest kernel function, at least within the scope of the above comparison. Nevertheless, the above results show that the SVR with RBF kernel function provides a high degree of performance.

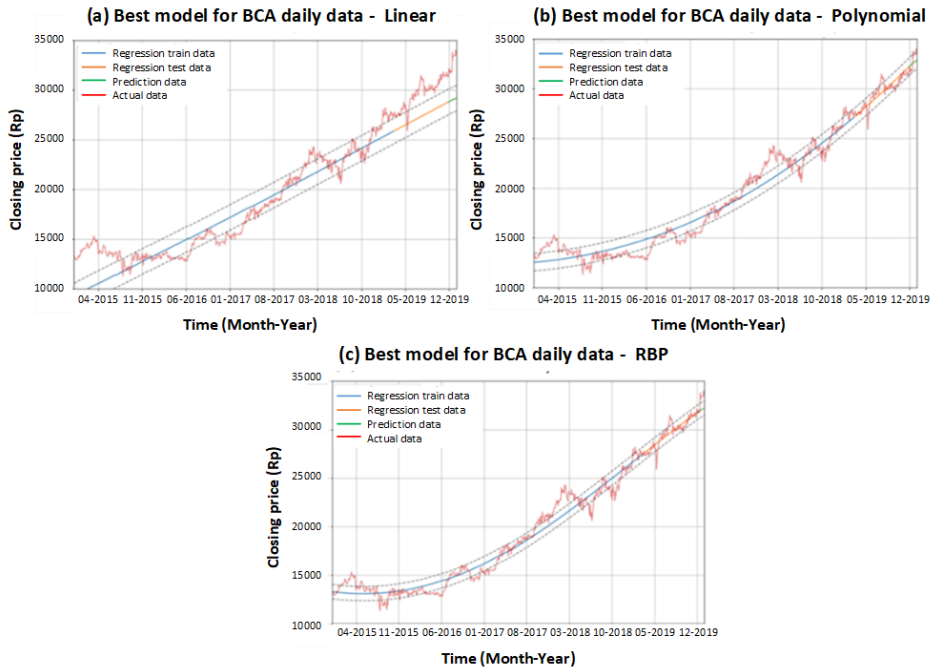

Figure 1 Comparison of SVR prediction performance of (a) linear kernel function, (b) polynomial kernel function, and (c) RBF kernel function.

2.4 Case Study Area

Ujung Berung is a district in the eastern part of Bandung City, West Java, Indonesia. Ujung Berung district is located at 107°43’8.11” to 107°42’20.50” east longitude and 6°55’4.81” to 6°53’13.61" south latitude, at an altitude of ±700 meters above sea level. The area of Ujung Berung district is shown in Figure 2.

Ujung Berung, with a land size of 661.2 hectares, is located east of the city of Bandung. The Cillengkrang district and Bandung district border on the north of the Ujung Berung area, the Panyileukan, Cinambo, and Arcamanik districts

border on the south, the Mandalajati district borders on the east, and the Cibiru district borders on the west.

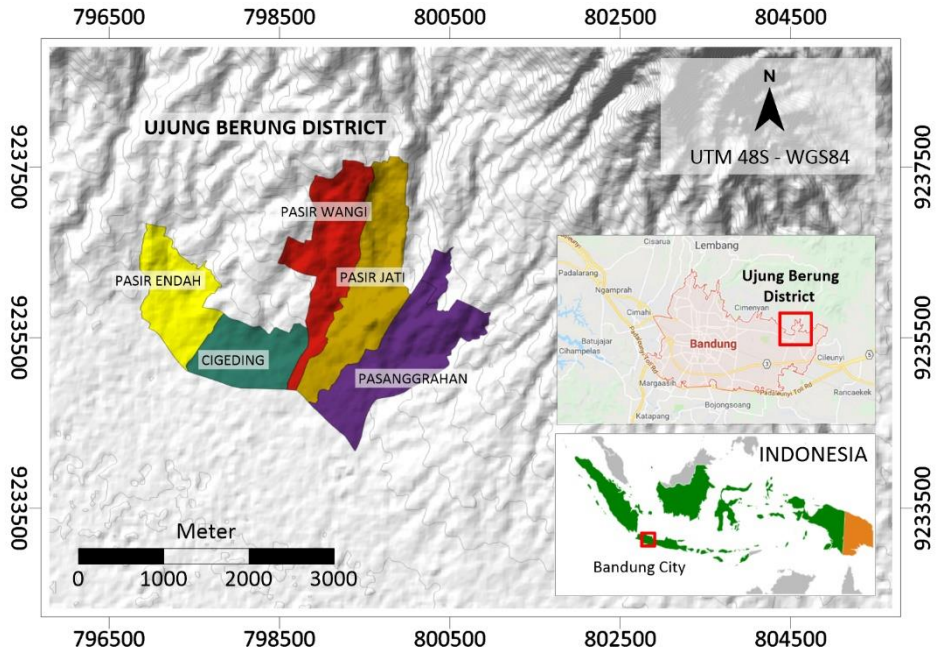

Figure 2 Map of Ujung Berung District, Bandung City, West Java, Indonesia.

Ujung Berung can be reached via three different routes. The first is the route from Bandung's downtown area, specifically the Bandung city plaza or Asia-Africa Street. The second route is the Soekarno-Hatta Road. The third route is from the east, across the Cileunyi-Cibiru region. The three routes to Ujung Berung result in a high level of transportation traffic in the Ujung Berung area, particularly on the AH Nasution Route, which is the district's main road. One reason to monitor air quality in this area is the high level of traffic activity.

Year after year, the rate of population growth in the Ujung Berung area increases. The population of Ujung Berung district rose from 2015 to 2019, according to the Bandung City Central Statistics Agency (BPS) [29]. With a population growth rate of 1.55 percent, the population of Ujung Berung was 85,887 people in 2019. There were 1,513 births and 238 deaths in 2019. Increased population growth has an impact on the number of people who drive and the kind of vehicles they drive. It also has an impact on transportation operations.

3 Results and Discussion

The CO concentration was predicted using the SVR technique and an RBF kernel function with a 95:5 training to test data ratio. The CO concentration data for Ujung Berung, Bandung came from the city of Bandung's Department of Environment and Hygiene (DLHK) database. As input data, hourly CO concentration measurements were used. The modeled CO concentration data was separated into 3,024 training data and 168 test data from April 26, 2018 to September 5, 2018, while the projected CO concentration data was divided into 120 data from September 6, 2018 to September 10, 2018. Figure 3 depicts the actual CO concentration data.

CO concentrations increased at 02.00, 05.00 to 07.00, 11.00 to 13.00, and 17.00 to 20.00 in Figure 4 throughout the course of 24 hours. This is linked to the high transportation activity in the Ujung Berung area, which forms a bottleneck in Bandung City when moving outside of the city, particularly in the eastern section of the city.

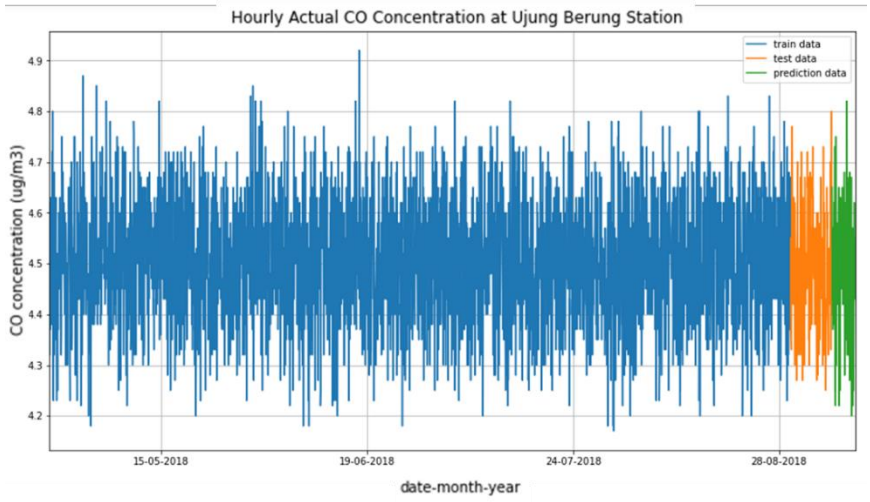

Figure 3 Hourly actual CO concentration at Ujung Berung Station.

For each kernel parameter, 15 different values were tested in this study. Table 1 displays the experimental outcomes in terms of parameter values, model accuracy, and prediction accuracy.

Based on the results of varying the kernel parameters on carbon monoxide concentration data, it was discovered that a smaller ε value causes an overfitting pattern, in which the model accuracy increases but the prediction accuracy drops, as was shown in the second and third experiments. Trials with bigger fluctuations

in the value of ε reveal that the model's and predictions' accuracy declines since the resulting model does not match well enough due to the large data deviation tolerance, as was seen in the fifth, sixth, seventh, twelve, and fifteenth experiments.

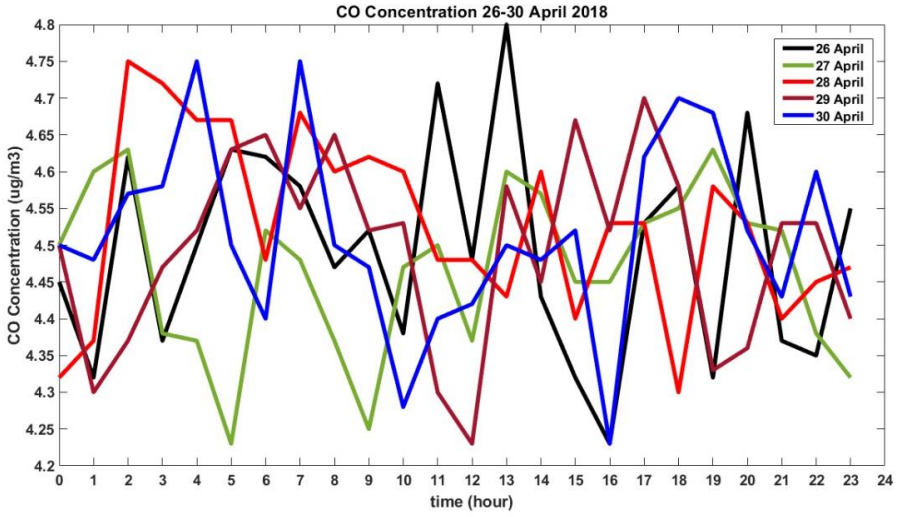

Figure 4 CO concentration for 24 hours from 26/04/2018 to 30/04/2018.

The accuracy of the model increases or decreases as the γ value changes. Because the model cannot forecast the training data or does not develop a model from the extracted data, a large increase in γ results in a poorly fitted model. Changing the value of γ to a very small number, on the other hand, produces overfitting, which causes the model to only store predictable training data, become unstable, and lose generalization, reducing its predictive power.

The trade-off between maximizing margin and training error is controlled by parameter *C*. Parameter *C* is critical for determining the model's predictive ability. When there is overfitting that follows the curvature of the data, a value of *C* that is too large tends to yield a more curvilinear model, resulting in erroneous predictions using a curved model. On the other hand, if the change in *C* is too small, the model will be underfitted since it will tolerate far larger deviations than lowering the model's accuracy and predictions, as was seen in the first, second, third, seventh, and tenth experiments.

The values of ε, γ, and *C* are all related to each other in some way. As a result, the suitability of the three parameters has a significant impact on the accuracy of a model and prediction. Thus, it is essential to attempt many different

variations/combinations of parameter values to get the best results. Because the results of the calculation of errors and their correctness can experience considerable changes at specific value limits, it is also necessary to utilize a smaller value scale, i.e. two to three digits after the decimal point.

Table 1 Experimental results of varying the kernel parameters of the CO concentration data.

Observation(s)	Parameters			Level of Accuracy (%)		R^2
	ε	γ	C	Model	Prediction	
1	0.1	30	0.006	97.85	97.66	0.001479
2	0.045	30	0.006	97.85	97.62	0.000692
3	0.053	29	0.006	97.81	97.63	0.0178
4	0.0035	30	1	97.84	97.46	0.15616
5	1	20	1	97.7	97.42	-0.15616
6	1	30	1	97.7	97.42	-0.15616
7	1	10	0.006	97.7	97.42	-0.15616
8	0.034	29	1	97.84	97.44	0.17861
9	0.079	30	0.14	97.85	97.67	0.17861
10	0.079	30	0.006	97.85	97.64	0.17861
11	0.02	30	0.006	97.84	97.68	0.17861
12	1	0.1	1	97.7	97.42	-0.15985
13	0.04	0.1	1.5	97.85	97.67	0.0042
14	0.04	0.1	6	97.85	97.67	0.0042
15	8	30	1	97.7	97.42	-0.15985

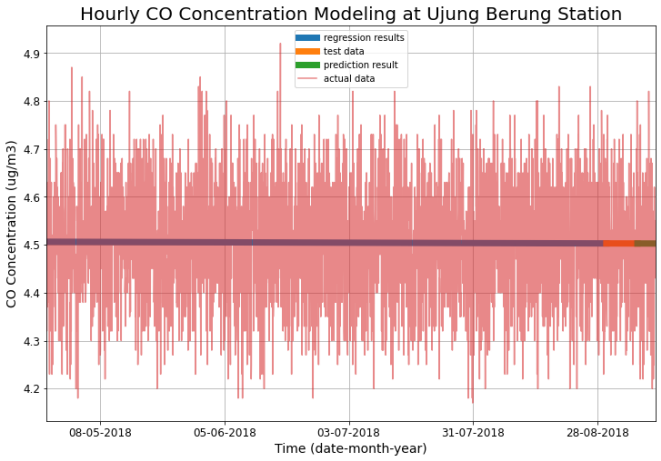

Figure 5 CO concentration model in Ujung Berung with an almost straight curve for $\varepsilon = 8$, $\gamma = 30$, and $C = 1$.

According to the experimental results, the regression curve, which is a data hyperplane, changes when the kernel parameter values are changed. The result of the 15th trial, which was a straight line, is shown in Figure 5. Due to the large

value of ε, the resulting hyperplane is still not appropriate. Because it tolerates large deviations, a high value of ε creates an underfitting model, diminishing its predictive power.

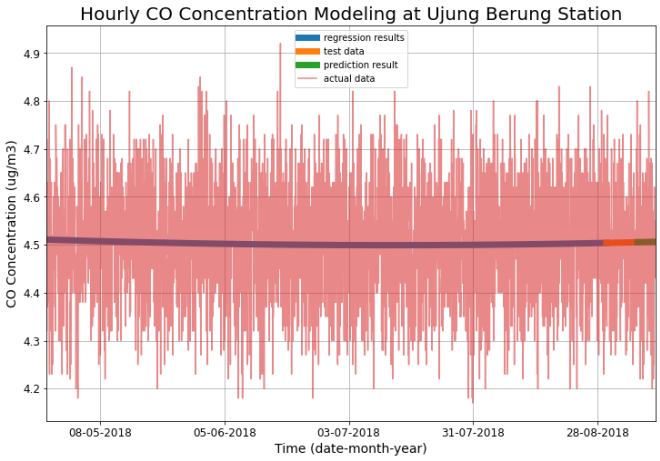

Figure 6 CO concentration in Ujung Berung model with an almost flat hyperplane for $\varepsilon = 0.04$, $\gamma = 0.1$, and $C = 6$.

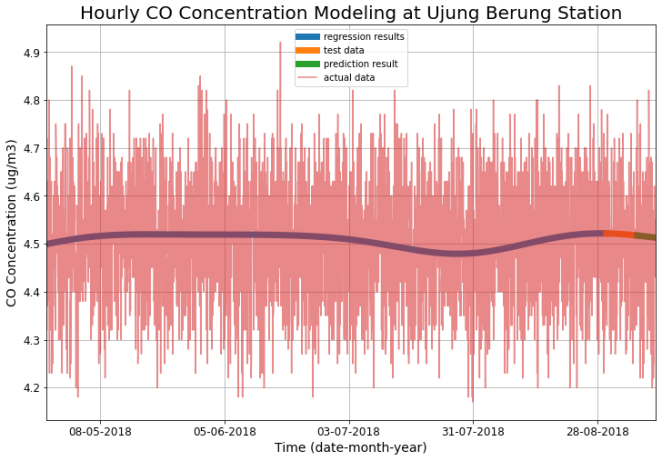

Figure 7 CO concentration model in Ujung Berung with the best prediction accuracy for $\varepsilon = 0.02$, $\gamma = 30$, dan $C = 0.006$.

The hyperplane is depicted in Figure 6 as a virtually flat line with simple curves at both ends. C generated a large shift in the value of the regression result in the 14th experiment, whereas ε and γ only saw a minor change. When the γ value is

low, it will neglect most of the supporting vectors, resulting in underfitting of trained points and a reduction in predictive power.

Figure 7 shows the graph for predicting CO concentration in Ujung Berung with the best forecast accuracy. The findings of the 11th trial are shown by a line with R^2 = 0.17861 and a highest prediction value of 97.68% that follows the graph of the actual data. To produce the finest regression results, a small ε value minimizes the deviation and is supported by the proper γ and C values.

Figure 8(a) depicts the actual CO concentration, with an average value of 4.502 $\mu g/m^3$. Meanwhile, Figure 8(b) shows the predicted CO concentration results with an average of 4.504 $\mu g/m^3$. The prediction findings demonstrate that they are in line with the actual CO concentration data, indicating that the prediction results are highly accurate.

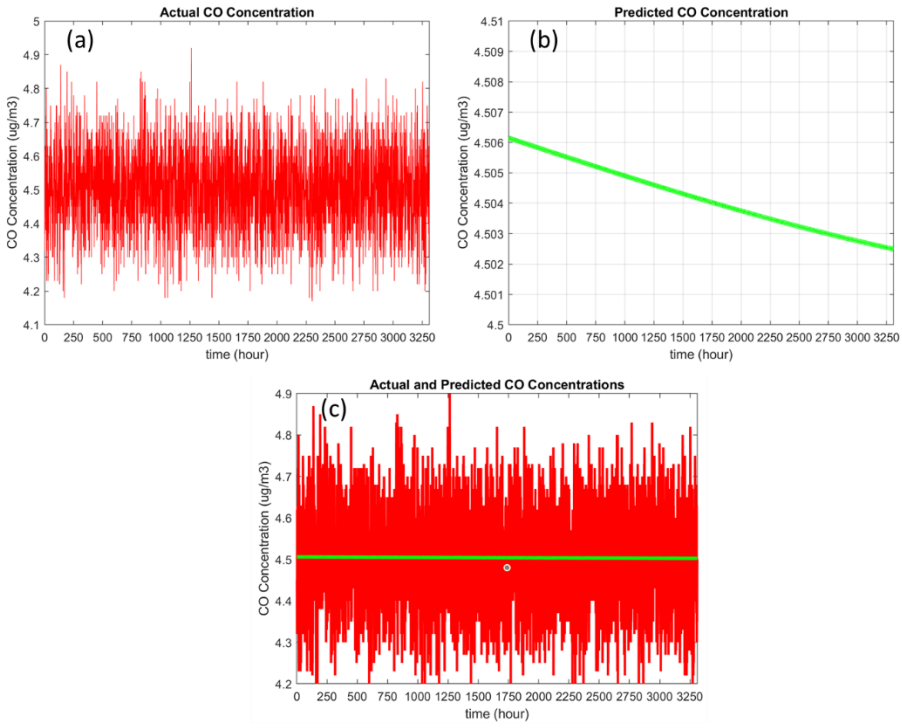

Figure 8 CO concentration: (a) actual data with an average value of 4.502 $\mu g/m^3$, (b) predicted results with an average value of 4.504 $\mu g/m^3$, (c) actual data and predicted data.

4 Conclusion

The best hyperplane is determined by the adequacy of the kernel parameter values. The best prediction accuracy value was 97.68%, for $\varepsilon = 0.02$, $\gamma = 30$, and $C = 0.006$, according to the experimental data, which reveal $R^2 = 0.17861$. The average value of the actual CO concentration data and the predicted values of 4.502 μg/m^3 and 4.504 μg/m^3, respectively, support the conclusion that the CO concentration prediction accuracy was very high. The kernel parameter values have a large impact on the model and prediction results since they can produce overfitting and underfitting. Overfitting will occur if parameter ε is set to a small value, while underfitting will occur if the value is increased. Parameters γ and C, on the other hand, have a reverse effect.

References

[1] Karagulian, F., Belis, C.A., Dora, C.F.C., Prüss-Ustün, A.M., Bonjour, S., Adair-Rohani, H. & Amann, M., *Contributions to Cities Ambient Particulate Matter (PM): A Systematic Review of Local Source Contributions at Global Level*, Atmospheric Environment, **120**, pp. 475-483, 2015.

[2] European Environment Agency, *Explaining Road Transport Emissions Non-Technical Guide*, 2016.

[3] Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M. & Pereira, M.C., *Multiple Linear Regression and Artificial Neural Networks Based on Principal Components to Predict Ozone Concentrations*, Environ. Modelling & Software, **22**, pp. 97-103, 2007.

[4] Sousa, S.I.V., Pires, J.C.M., Martins, F.G., Pereira, M.C. & Alvim-Ferraz, M.C.M., *Potentialities of Quantile Regression to Predict Ozone Concentrations*, Environmetrics, **20**(2), pp. 147-158, 2009.

[5] Załuska M. & Gładyszewska-Fiedoruk K., *Regression Model of PM2.5 Concentration in a Single-Family House*, Sustainability, **12**, p. 5952, 2020.

[6] Markakis, K., Valari, M., Perrussel, O., Sanchez, O. & Honore, C., *Climate Forced Air-Quality Modeling at Urban Scale: Sensitivity to Model Resolution, Emissions and Meteorology*, Atmos. Chem. Phys., **15**, pp. 4767-4821, 2015.

[7] Mehrani, M.J., Filvantorkaman, M. & Markade, F.A., *Using Box Model to Estimate the Impact of Air Pollution from an Industrial Plant*, MOJ Civil Engineering, **2**(4), pp. 124–127, 2017.

[8] Rafati, L., Ehrampousha, M.H., Talebib, A., Mokhtaria, M., Pisheha, Z. K. & Dehghana, H. R., *Modelling the Formation of Ozone in the Air by Using Adaptive Neuro-Fuzzy Inference System (ANFIS) (Case Study: City of Yazd, Iran)*, Desert, **19**(2), pp. 131-135, 2014.

[9] Elangasinghe, M.A., Singhal, N., Dirks, K.N. & Salmond, J.A., Development *of an ANN-Based Air Pollution Forecasting System with Explicit Knowledge Through Sensitivity Analysis*, Atmospheric Pollution Research, **5**, pp. 696-708, 2014.

[10] Halawa, E., Zaharo, A., Fitrasari, D. & Purqon, A., *Prediction of Carbon Monoxide Concentration Using Artificial Neural Network Method* (*ANN*), Seminar Kontribusi Fisika ITB, pp. 140-144, 2015.

[11] Shahrainy, H.T., Shahsavani, D., Sargazi, S. & Habibi-Nokhanda, M., *Evaluation of MARS for the Spatial Distribution Modeling of Carbon Monoxide in An Urban Area*, Atmospheric Pollution Research, **6**, pp. 581-588, 2015.

[12] Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R. & Chatterton, T., *Modelling SO_2 Concentration at a Point with Statistical Approaches*, Environmental Modelling & Software, **19**(10), pp. 887-905, 2004.

[13] Moazami, S., Noori, R., Amiri, B. J., Yeganeh, B., Partani, S., & Safavi, S., *Reliable Prediction of Carbon Monoxide Using Developed Support Vector Machine*, Atmospheric Pollution Research, **7**, pp. 412-418, 2016.

[14] Smola, A. J., & Scholkopf, B., *A Tutorial on Support Vector Regression*, Statistics and Computing, **14**, pp. 199-222, 2004.

[15] Lu, W.-Z. & Wang, W-J., *Potential Assessment of the Support Vector Machine Method in Forecasting Ambient Air Pollutant Trends*, Chemosphere, **59**(5), pp. 693-701, 2005.

[16] Osowski, S. & Garanty, K., *Forecasting of the Daily Meteorological Pollution Using Wavelets and Support Vector Machine*, Engineering Applications of Artificial Intelligence, **20**(6), pp. 745-755, 2007.

[17] Wang, W., Men, C. & Lu, W., *Online Prediction Model Based on Support Vector Machine*, Neurocomputing, **71**(4-6), pp. 550-558, 2008.

[18] Luna, A.S., Paredes, M.L.L., de Oliveira, G.C.G. & Corrêa, S.M., *Prediction of Ozone Concentration in Tropospheric Levels Using Artificial Neural Networks and Support Vector Machine at Rio De Janeiro, Brazil*, Atmospheric Environment, **98**, pp. 98-104, 2014.

[19] Leong, W.C., Kelani, R.O. & Ahmad, Z., *Prediction of Air Pollution Index (API) Using Support Vector Machine (SVM)*, Journal of Environmental Chemical Engineering, 2019.

[20] Ortiz-García, E.G., Salcedo-Sanz, S., Perez-Bellido, A.M., Portilla-Figueras, J.A., & Prieto, L., *Prediction of Hourly O_3 Concentrations Using Support Vector Regression Algorithms*, Atmospheric Environment, **44**, pp. 4481-4488, 2010.

[21] Castelli, M., Clemente, F.M., Popovič, A. & Silva, S., *A Machine Learning Approach to Predict Air Quality in California*, Complexity, 2020.

[22] Akbarzadeh, A., Vesali Naseh, M. R. & Nodefarahani, M., *Carbon Monoxide Prediction in the Atmosphere of Tehran Using Developed Support Vector Machine*, Pollution, **6**(1), pp. 43-57, 2020.

[23] Brereton, R.G. & Lloyd, G.R., *Support Vector Machines for Classification and Regression*, The Analyst: Tutorial Review, 2009.
[24] Environmental Management Agency of Bandung City, *Bandung City Environmental Management Agency Strategic Plan 2013-2018*, 2018.
[25] Basak, D., Pal, S., & Patranabis, D.C., *Support Vector Regression*, Neural Information Processing – Letter and Reviews. **11**(10), pp. 203-224, Sept. 2019.
[26] Li, X., Lord, D., Zhang, Y. & Xie, Y., *Predicting Motor Vehicle Crashes Using Support Vector Machine Models*, Accident Analysis and Prevention, **40**(4), pp. 1611-1618, 2008.
[27] Noori, R., Abdoli, M.A., Ghasrodashti, A.A. & Jalili Ghazizade, M. *Prediction of Municipal Solid Waste Generation with Combination of Support Vector Machine and Principal Component Analysis: A Case Study of Mashhad*, Environmental Progress & Sustainable Energy, **28**(2), pp. 249-258, 2009.
[28] Bank Central Asia Tbk Stock Price Today – BBCA Live Ticker, Investing.com, https://id.investing.com/equities/bnk-central-as, (November 26th, 2021).
[29] Central Statistics Agency of Bandung City, *Ujung Berung District in Numbers 2020*, 2020.