Temporal prediction of carbon monoxide using the Elman Recurrent Neural Network

H Tantriawan[1], I S Sitanggang[2], L Syaufina[3] and H Harsa[4]

[1]Department of Science, Informatic Engineering, Institut Teknologi Sumatera, Lampung 35365, Indonesia
[2]Department of Computer Science, Bogor Agricultural University, Meranti Street IPB Dramaga, Bogor 16680,West Java, Indonesia
[3]Department of Silviculture, Bogor Agricultural University, Lingkar Akademik Street IPB Dramaga, Bogor 16680, West Java, Indonesia
[4] Research and Development Center, Indonesia Climatology and geophysics Agency, Angkasa Street I No.2 Kemayoran, Jakarta 10720, Indonesia

E-mail: hartanto.tantriawan@if.itera.ac.id

Abstract. Peatland fires in Indonesia are considered as regional disasters which occur periodically. These negative impacts, especially on our health, continue to threaten the society across the region. The objective of this study was to create a temporal model for predicting the pollutant concentration from peatland fires using the Elman Recurrent Neural Network (ERNN) and training by gathering data from fires which have been occurred recently in Sumatera, Indonesia. The data describing the haze from the peatland fires were generated using the HYSPLIT model with the input of hotspot sequences and meteorological data from NOAA. The stages of the model development consisted of data pre-processing, pollutant concentrations generating using HYSPLIT, pollutant concentration analysis, network architecture formation, weight determination, model training, and the prediction of the model evaluation. Experimental results indicated that the calculation of the ISPU (standard air pollution index) using the GDAS data of 20.5 g / m^3 obtained an ISPU value of 221. This value indicated that the air in the South Sumatera Province was very unhealthy. Similar to the calculation of ISPU using the WRF-Chem data of 26 g / m^3 obtained an ISPU value of 253. This value indicated that air in the South Sumatera Province was very unhealthy.

1. Introduction

Peatland fires in Indonesia are considered as natural disasters in the regions which occur periodically. This event is part of a recurring regional and global disaster. Especially throughout the year of 2015, the smog spread in the troposphere by bringing CO and CO_2. Through the measurement of Aqua Atmospheric Infrared Sounder (AAIS), the Ozone Monitoring Instrument (OMI) and the Aura Microwave Limb Sounder (MLS) from forest and land fires occurred in Sumatra and Kalimantan, it was revaled that the worst fires wereoccurred in 1991, 1994, and 1997. These fires happened along with the El-Nino event [1]. Indonesia's forest and land fires released large amounts of carbon such as CO, CO_2, CH_4, which spread to the atmosphere with an average CO_2 emissions rate of 11.3 T_g per day from September to October 2015. This result exceeded the CO_2 release rate from fuel fossils in the EU of 8.9 T_g CO_2 per day [2].

The side effects of this forest and peatland fires are the disasters that come from smokes. The disasters that come from smokes release high carbon emissions and are major cause of global warming. In 2009, in the province of Riau, there were bog fires which caused the increase of air pollution level. The resulting CO, CO_2, NO_2, CH_4, and SO_2 were so harmful to human health even reaching the contribution of 74% to the increase of carbon emissions [3].

Indonesia's geographical position has been highly vulnerable to the impacts of climatic change, so it is necessary to mitigate climatic change. The results of a meeting between the Indonesian governments with the G20 in Pittsburgh led to an agreement to reduce greenhouse gas emissions by 26% by self-effort and reaching 41% of international assistances received by 2020 from a condition without Business As Usual (BAU). Subsequent to this matter, the government made a Regulation of the President of the Republic of Indonesia Number 61 of 2011 on the National Action Plan for Green House Gas Emission Reduction [4].

Simulations using the data of forest fires in Singapore on June 19-22nd, 2013 WRF-Chem output indicated that the concentration of pollutants in the form of CO had a correlation value of 0.61-0.98 [5]. The ERNN (Elman Recurrent Neural Network) algorithm was implemented on the spatial data [6]. The objective of this study was to create a temporal model for the prediction of pollutant concentration from peatland fires using the Elman Recurrent Neural Network (ERNN) and training with data from recent fires in Sumatera, Indonesia.

2. Method

The haze data from the peatland fires was generated using the HYSPLIT model with the input of hotspot sequences and meteorological data from NOAA and the Indonesian Climatology and Geophysics Agency. The simulation using HYPSLIT was performed to obtain data on pollutant concentration, longitude, latitude, CO concentration. Prediction of pollutant concentration was by the Elman Recurrent Neural Network algorithm. Steps to predict the Elman Recurrent Neural Network algorithm were mentioned below: Input unit function firstly obtained its value, thenhidden unit calculated its value (number of Input weight and context units), before the context unit as performed to get its value. Afterwards, the output unit calculated all values (the number of Input weight, Hidden and Context units). The analysis was conducted on a model with interesting information based on the amount of CO concentration due to peat land fires. The stages in this study are illustrated in figure 1.

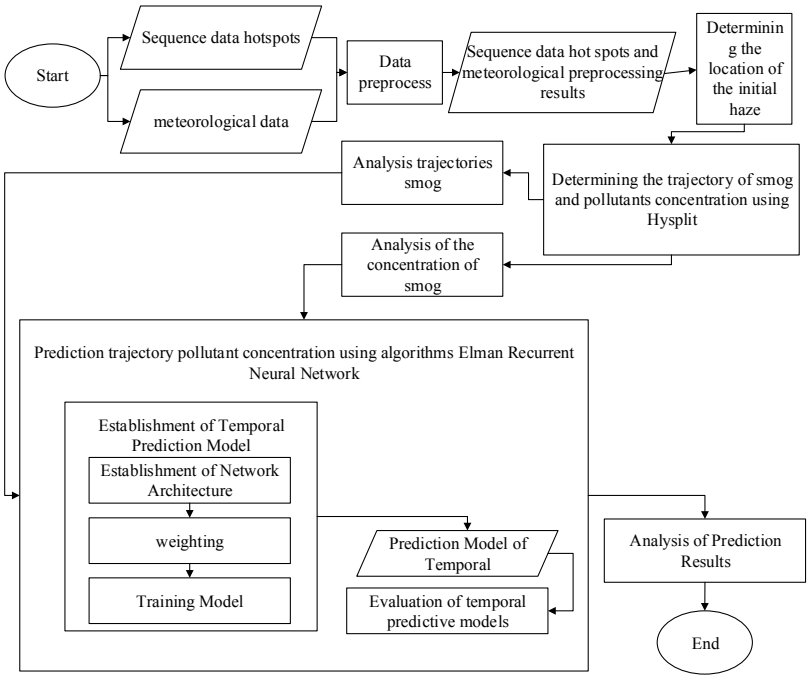

Figure 1. Flowchart of research stages.

One of the instruments to measure the cleanness of the air was the Air Pollution Standard Index (ISPU). There were several basic parameters to measure the ISPU such as Particulate (PM_{10}), Sulfur dioxide (SO_2), Carbon monoxide (CO), Ozone (O_3), Nitrogen dioxide (NO_2). Here was the equation (1) used to calculate the ISPU,

$$I = \frac{Ia-Ib}{Xa-Xb}(Xx-Xb) + Ib \quad (1)$$

where I = ISPU was calculated, Ia = upper limit ISPU, Ib = lower limit ISPU, Xa = upper bound ambient, Xb = lower threshold ambient, Xx = real measured ambient result of measurement [7].

3. Result and analysis

3.1. ERNN Prediction modelling

Before modelling, it was necessary to perform a stationary test in the middle and stable test values in variety, to the result of the simulation data. To understand the stationariy of information in the mean value, the unit roots test using the Dickey-Fuller augmented test on the pollutant data was acquired daily. While, for the stationarity test in variance, the Bartlett and Lavene test was undertaken.

3.1.1. GDAS Data. The P-value of the Dickey-Fuller augmented test resulted in the CO data of 0.01, if the value of α was 0.05, then the value of P-value was smaller than α. Therefore, it could be concluded that the data was considered stationary to the mean. The Bartlett and Levene test results on CO data got P-value less than 2.2 × 10-16. The value was smaller when compared to the value of α which was worth 0.05. Therefore, it could be concluded that the data was considered not stationary to the variance. Therefore, it was necessary to do a box-cox transformation. Figure 2 illustrates the result of the box-cox transformation in the CO data.

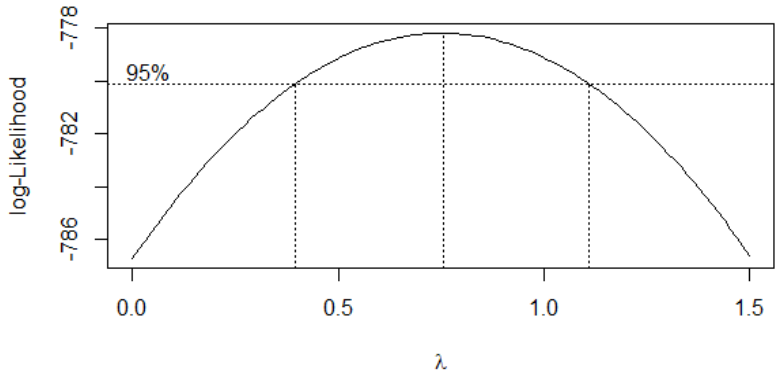

Figure 2. Plot Box-Cox CO with GDAS data.

In figure 3, it can be seen that 95% of the confidence interval was determined as the range value of λ 0.43 to 1.2. Based on table 1, the value of λ was one in the trusted range, and did not need to be transformed. After the data was stationary to the mean and the variance, the PACF plot was performed to determine the significant lag of the data. Figure 3 illustrates the PACF plot of data which were stationary. According to the result of the PACF plot in figure 3, there was a real correlation value on lag-1 and lag-7. Then the number of input nodes in the architecture was as many as two inputs. The first input was the 1st lag; the second input was the 7th lag. Both nodes predicted the output on the 8th lag.

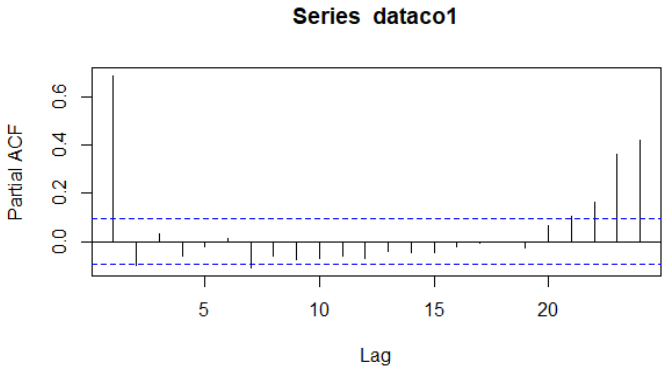

Figure 3. Plot PACF CO with GDAS data.

The success of the first experiment was able to predict the trend according to the pattern in Table 1 then the further experiment was performed by modifying the value of the learning rate of 0.2; 0.3; 0.4; and 0.5. The modified learning rate was calculated to select the best architecture. Figure 4 illustrates the modified learning rate results on the CO data.

Table 1. Box-Cox transformation based on λ value.

Characteristics	Specification
Architecture	1 input layer with 2 *node* 1 hidden layer with 2 *node* 1 output layer
Activation Function	Sigmoid binary
Learn Function Parameters	0.1
Maximum Epoch	1000

Based on figure 4, the learning results of ERNN were obtained by the pattern of carbon monoxide (CO) pollutant data series in the past, both in predicting the data trends. At 2 o'clock (UTC) until the 3rd hour (UTC), the actual value was quite high. This result could happen because during these several hours, it occured during peak fires during the day, thus releasing high levels of carbon dioxide pollutants. Figure 4 illustrates the ERNN both in predicting values 5th, 6th, and 8th hours on LR 0.1. However, this method of learning was less suitable for the last data such as the data on time from 1 to 4 hours. Even at the 9th hour until the 24th hour, the value was relatively high. This relatively large value might be due to seasonal patterns that might occur over a period that has not been analyzed.

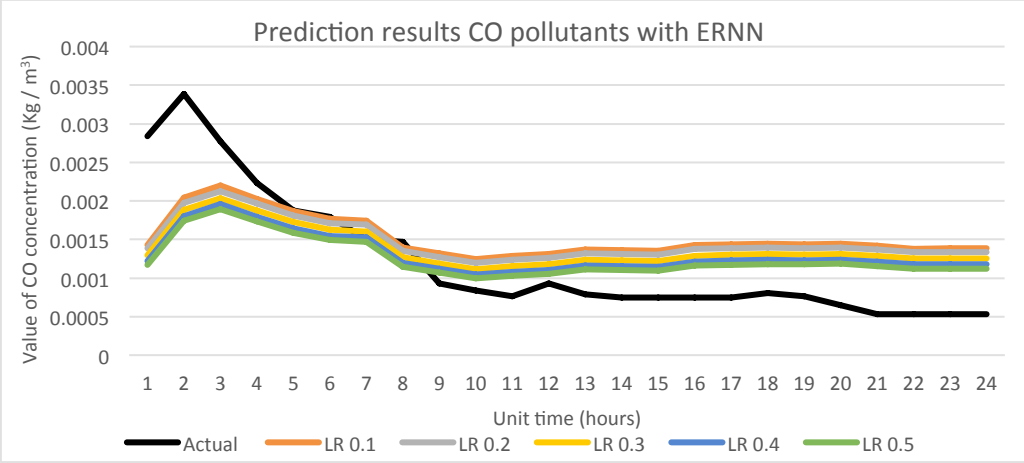

Figure 4. Results of the prediction of CO pollutant with GDAS data.

3.1.2. WRF-Chem Data. The P-value of the Dickey-Fuller augmented test results of a CO data of 0.01, if the value of α was 0.05, then the value of P-value was smaller than α. Therefore, it could be concluded that the data was considered stationary to the mean. The Bartlett and Levene test results on CO data got P-value less than 2.2 × 10-16. The value was smaller when compared to the value of α was worth 0.05. Therefore, it could be concluded that the data was considered not stationary to the variance. Therefore, it was necessary to do a box-cox transformation. Figure 2 illustrates the result of the box-cox transformation in the CO data.

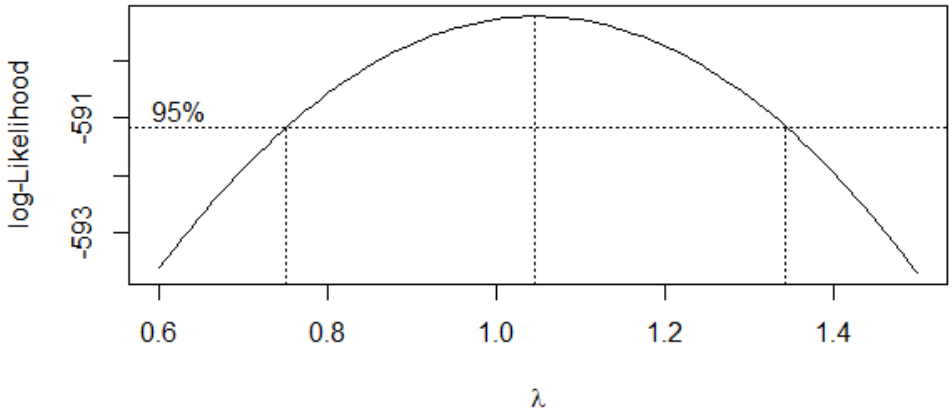

Figure 5. Plot Box-Cox CO with WRF Chem data.

From figure 5, it could be seen that the 95% confidence interval was in the range value of λ 0.78 to 1.35. Based on table 1, because the value of λ was one in the trusted range, it did not need to be transformed. After the data was stationary to the mean and the variance, the PACF plot was performed to determine the significant lag of the data. Figure 6 illustrates the PACF plot of data that were stationary. According to the result of the PACF plot in figure 6, the real correlation values were spotted on lag-1 and lag-18. Then, the number of input nodes in the architecture was as many as two inputs. The first input was the 1st lag; the second input was the 18th lag. Both nodes predicted the output on the 19th lag.

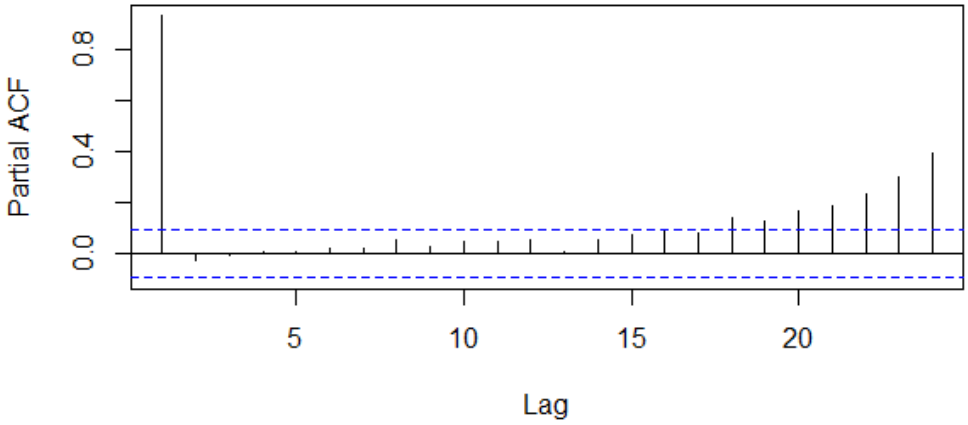

Figure 6. Plot PACF CO with the WRF-Chem data.

Based on figure 7, ERNN learning results were obtained by the pattern of carbon monoxide (CO) pollutant data series in the past, both in the trends of predicting data. At 2 o'clock (UTC) until the 3rd hour (UTC), the actual value was quite high. This result could happen because during these hours it would have occured during peak fires during the day, thus releasing high levels of carbon dioxide pollutants. Figure 7 illustrates the ERNN both in predicting the values 8 to 18 hours on LR 0.4.

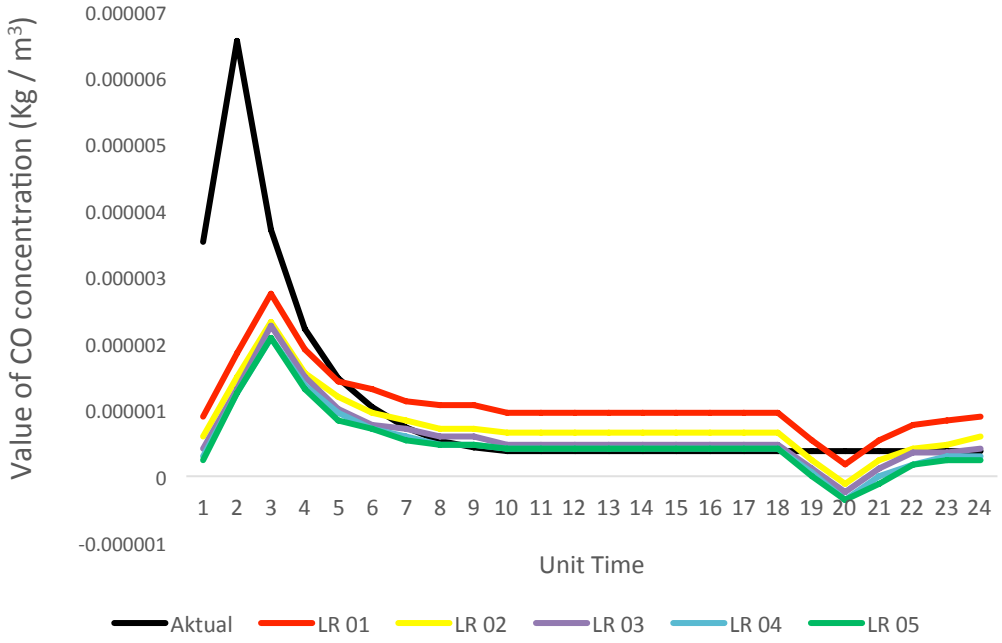

Figure 7. Results of CO pollutant prediction with WRF-Chem data.

3.2. Evaluation

Evaluation was carried out to determine the performance of model forecasting. Evaluations could be rated using some statistical tests. One such statistical test is the MAPE test (Mean Absolute Percentage Error). MAPE calculations can be seen from the following equation (2) [8].

$$MAPE = \frac{\sum_{t=1}^{n} |(y_t - \hat{y}_t)/y_t|}{n} (100); \; y_t \neq 0 \quad (2)$$

Where y_t is the actual value, $\hat{y}_t$ is the predicted value, and n is the amount of predicted data.

The best result of the ERNN modelling of CO with data GDAS was obtained from the 0.1 learning rate. The performance of ERNN is illustrated in table 2. In table 2, it could be seen that the best MAPE results were obtained on average MAPE in the first 8 hours of 17.7% at the 0.1 learning rate. These results were still not good for predicting actual values, but both in the predicted fluctuations the correlation of the CO concentration data. The best result of the ERNN modelling of CO with data WRF-Chem was obtained from learning rate 0.4. The performance of ERNN is illustrated in Table 3. In Table 3, it could be seen that the best MAPE results were obtained on average MAPE in the first 16 hours of 27% at learning rate 0.4. These results were still not good for predicting actual values, but both in the predicted fluctuations correlation of CO concentration data.

According to table 2, there was an anomaly occurring in the first 8-hour prediction on CO with data GDAS. Because the greater the learning rate used, the higher the value of MAPE obtained. While at the 16th and the 24th hour, the higher the learning rate used, the smaller the MAPE value obtained. This result proved that ERNN was good at predicting pollutant concentration values in the first 8 hours but not well in predicting in the next hours.

Table 2. Predicted Pollutant Results with modification of the learning rate.

Unit time (UTC)	Data Testing CO (mg/m^3)	CO with GDAS data (mg/m^3)					Data Testing CO (mg/m^3)	CO with WRF-Chem data (mg/m^3)				
		Lr 0.1	Lr 0.2	Lr 0.3	Lr 0.4	Lr 0.5		Lr 0.1	Lr 0.2	Lr 0.3	Lr 0.4	Lr 0.5
1	2837	1432	1382	1298	1222	1170	3.52	0.89	0.60	0.42	0.30	0.24
2	3388	2046	1973	1882	1802	1744	6.56	1.85	1.49	1.31	1.25	1.25
3	2771	2202	2128	2037	1955	1896	3.71	2.74	2.32	2.26	2.09	2.09
4	2231	2028	1965	1876	1795	1738	2.21	1.91	1.55	1.49	1.37	1.31
5	1878	1868	1810	1723	1644	1588	1.48	1.43	1.19	1.01	0.95	0.83
6	1792	1767	1711	1624	1546	1491	1.04	1.31	0.95	0.77	0.72	0.72
7	1504	1746	1689	1603	1525	1470	0.74	1.13	0.83	0.72	0.60	0.54
8	1468	1396	1350	1266	1193	1144	0.54	1.07	0.72	0.60	0.48	0.48
9	929	1319	1272	1189	1117	1069	0.44	1.07	0.72	0.60	0.48	0.48
10	834	1242	1197	1114	1042	995	0.38	0.95	0.66	0.48	0.42	0.42
11	765	1285	1237	1154	1081	1032	0.38	0.95	0.66	0.48	0.42	0.42
12	931	1310	1262	1178	1104	1055	0.38	0.95	0.66	0.48	0.42	0.42
13	784	1369	1319	1235	1161	1111	0.38	0.95	0.66	0.48	0.42	0.42
14	748	1362	1313	1229	1155	1105	0.38	0.95	0.66	0.48	0.42	0.42
15	748	1356	1307	1223	1149	1099	0.38	0.95	0.66	0.48	0.42	0.42
16	748	1424	1374	1289	1214	1162	0.38	0.95	0.66	0.48	0.42	0.42
17	748	1436	1386	1301	1226	1174	0.38	0.95	0.66	0.48	0.42	0.42
18	804	1444	1394	1310	1234	1182	0.38	0.95	0.66	0.48	0.42	0.42
19	766	1440	1390	1305	1229	1178	0.38	0.54	0.24	0.12	0.06	0.00
20	648	1447	1397	1312	1237	1185	0.38	0.18	0.12	0.24	0.36	0.36
21	527	1417	1368	1284	1208	1157	0.38	0.54	0.24	0.12	0.00	0.12
22	529	1382	1334	1250	1175	1123	0.38	0.77	0.42	0.36	0.18	0.18
23	529	1383	1335	1251	1175	1124	0.38	0.83	0.48	0.36	0.30	0.24
24	529	1383	1335	1251	1175	1124	0.38	0.89	0.60	0.42	0.30	0.24
MAPE (8 hr) %		17.7	19.6	22.4	24.9	27.2		46	38	39	44	46
MAPE (16 hr) %		41.8	39.7	35.9	32.6	30.7		99	55	33	27	28
MAPE (24 hr) %		71	67	59.9	53.6	49.5		99	55	38	39	43

3.2.1. Air Pollution Standar Index (ISPU). Based on the calculation of the ISPU conducted with the value of CO with data GDAS in the test data (October 27th, 2015) of 20.5 g/m^3, the result of the ISPU value was 221. This value indicated that the air content in the South Sumatra Province was very unhealthy. Based on the decision of the Head of the Environmental Impact Control Agency number 107 of 1997, the level of this very harmful ISPU had an impact on the cardiovascular increase in non-smokers with heart disease. In the other side, the calculation of the ISPU generated value of CO with WRF-CHem data in the test data (October 27th, 2015) of 26 g/m^3. The result of the obtained ISPU value was 253. This value indicated that the air content in South Sumatra Province was stated very unhealthy. Based on the decision of the Head of the Environmental Impact Control Agency number 107 of Year of 1997, the level of this very harmful ISPU had an impact on the cardiovascular increase in non-smokers with heart disease.

4. Conclusion

The ERNN algorithm is appropriate for predicting temporal data proved by the results of the following research. The ERNN modeling CO pollutants using GDAS data obtained the best results at a learning rate of 0.1 with the MAPE value of 17.7%. The best-predicted carbon monoxide with GDAS data were obtained at the 5th, 6th, and 8th hour. The ERNN modeling CO pollutants using GDAS data obtained the best results at the learning rate of 0.1 with the MAPE value of 27%. The best-predicted carbon monoxide with GDAS data were obtained at 8th until 18th hour.

To optimize the predicted yield of pollutant concentration, it is recommended to use GDAS data of 0.5 degrees or even 0.25 degrees. When the distance of the pollutant concentration captured by the satellite was smaller, it was expected that the predicted results would be closer to the actual data. While in this study, GDAS data one degree was utilized which meant every data point of pollutant concentration was 100 km. So GDAS data of 0.5 degrees and 0.25 degrees need to be tested.

References

[1] Field R D van der Werf G R Fanin T Fetzer E J Fuller R Jethva H Levy R Livesey N J Luo M Torres O 2016 Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to El Niño-induced drought. *Proceedings of the National Academy of Sciences*. **113**:9204–9209.

[2] Huijnen V Wooster M Kaiser J Gaveau D Flemming J Parrington M Inness A Murdiyarso D Main B Van Weele M 2016 Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. *Scientific reports*. **6:**26886.

[3] Syaufina L 2008 *Forest and land fires in Indonesia: fire behavior, causes, and impacts of fires* (Jakarta: Bayumedia Pub)

[4] [KEMENLHK] Indonesia Ministry of Environment and Forestry. 2011. Regulation of the President of the Republic of Indonesia number 61 of 2011 on the national action plan of reducing greenhouse gas emissions (Jakarta: Indonesia Ministry of Environment and Forestry) (In Indonesian)

[5] Heriyanto E Syaufina L Sobri M 2015 Forecasting simulation of smoke dispersion from forest and land fires in Indonesia *Procedia Environmental Sciences* **24:**111–119.

[6] Robby I S 2017 *The spatio-temporal model for predicting the emergence of hot spots on peatland in Siak District uses SARIMA and Elman Recurrent Neural Network [thesis]* (Bogor : Bogor Agricultural University) (In Indonesian)

[7] [BAPEDAL] Environmental National Control System of Indonesia 1997 *Decree Number 107 Year 1997 on Calculation and Reporting and Information of Air Pollution Standard Index* (Jakarta: Environmental National Control System of Indonesia) (In Indonesian)

[8] Rahman N H A Lee M H Latif M T 2015 Artificial neural networks and fuzzy time series forecasting: an application to air quality *Quality & Quantity* **49:**2633–2647

[9] [KEMENKES] Indonesia Ministry of Health 2011 *Permenkes RI No.**1077**/Menkes/Per/V / 2011 concerning Guideline for air sanitation in the house* (Jakarta: Indonesia Ministry of Health) (In Indonesian)

[10] Wei WW 2006 *Time series analysis: univariate and multivariate methods* (USA: Pearson Addison Wesley)