Jurnal
Teknologi

Full Paper

APPLYING EXPONENTIAL STATE SPACE SMOOTHING MODEL TO SHORT TERM PREDICTION OF NO_2

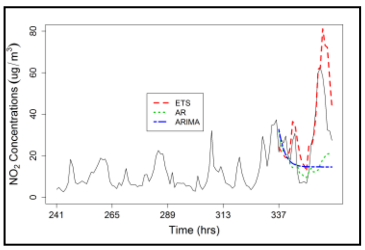
Arie Dipareza Syafei*

Department of Environmental Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 60111

Article history
Received
14 January 2015
Received in revised form
09 April 2015
Accepted
09 June 2015

*Corresponding author
dipareza@enviro.its.ac.id

Graphical abstract



Abstract

Predicting air pollutant level has been important aspect as part of air quality management. A time series model exponential state space smoothing (ESSS) method was employed to short-term predict traffic-related pollutant, nitrogen dioxide (NO_2) during January 2013. Compared with autoregression (AR) and autoregressive integrated moving average (ARIMA) the ESSS model performed better with R^2 0.673 respectively. The performance was also consistent for prediction over days in terms of R^2. For correlation between prediction and observation, the R^2 ranged from 0.4 to 0.6, showing that ESSS model has exceptional performances compared to AR and ARIMA. Hence, ESSS has potential to be applied as part of air quality management for daily air quality warning purposes.

Keywords: Exponential state space smoothing, air pollutant prediction, Surabaya, time series model

1.0 INTRODUCTION

Air pollution has been primary issue to every country due to its potential effect to human health and ecosystem. The concentration above permissible level may cause interference to the health as well as environmental ecosystem. Air pollution management is the key to adapt and prevent the events where the concentrations breach the threshold level. Besides understanding factors that affect pollution level, prediction has become important issue to be implemented. The information from prediction is useful for either end user eg., road users, and city government officers. For instance, when one of the predicted pollutant level will be higher than the threshold level, drivers may consider taking another route or alternatively wearing a mask to reduce getting exposed to the pollutants.

The prediction of air pollutants using time series models has been done by many researchers [1,2]. Common stationary time models such as autoregression (AR), Autoregression Moving Average (ARMA), Autoregression Integrated Moving Average (ARIMA) were popular. However, the pattern of air pollutants is complex where they contain seasonality and trends. Many models were developed to deal with complexity of pollutants, including the exponential state space smoothing (ESSS) approach [3]. ESSS consists of fifteen methods as shown in Table 1 [3,4].

ESSS has been employed to model U.S. Gasolide data, Turkey electricity demand, call center data [5], and solar irradiance [4] with good performance, but none as far as author concerns this model were employed to predict air pollutant levels. In this paper we use NO_2 as the object because high concentrations of NO_2 may cause lung irritation and damage [1]. NO_2 is also important pollutant as indicator to the vehicle flow. The number of vehicle has been increasing due to the growth of economy and industrial sector and this has been primary issue especially in the developing countries such as Indonesia. The aim of this study is to obtain performance of forecasted values of NO_2 during 15th January 2013 using first 14 days data in the same month and to check performance consistency over days by extending the forecasted day into 21 days. We compare its performance with AR and ARIMA

Table 1 Fifteen exponential smoothing methods

		Seasonal		
		N (None)	**A (Additive)**	**M (Multiplicative)**
Trend	N (None)	NN	NA	NM
	A (Additive)	AN	AA	AM
	A_d (Additive damped)	A_dN	A_dA	A_dM
	M (Multiplicative)	MN	mA	MM
	M_d (Multiplicative damped)	M_dN	M_dA	M_dM

2.0 EXPERIMENTAL

2.1 Data and Study Location.

Data were obtained from Air Quality Laboratory, Environmental Agency of Surabaya City Government from Station 6 located in Wonorejo where it represents mixed land use. We focused on this station as it was able to provide adequate data. The data were available as 30-mins interval and we averaged it into hourly data. The data used for training was NO_2 pollutant concentrations (ug/mg) taken from 1st January to 14th January 2013, the 24 hours data on 15th January were taken as validation set. To check its consistency of forecasting we also forecast another 21 days based on prior 14 days data (training data).

2.2 Exponential State Space Smoothing Model

ESSS was emerged to solve data problem subject to complex seasonal patterns [5]. There are fifteen forecast equations which include seasonal and trend components where the observed time series is denoted by y_1, y_2, ... y_n. The seasonal component consists of None (None), Additive (A), and Multiplicative (M) approaches whereas the trend consists of None (N), Additive (A), Additive Damped (A_d), Multiplicative (M), and Multiplicative Damped (M_d). The forecast for h steps ahead up to time t is denoted by y_{t+h}. The forecast equation is shown below:

$$\hat{y}_{t+h} = f\left(l_t, g_t, s_{t-m+h_m}\right) \quad (1)$$

$$l_t = \alpha A_t + (1-\alpha) B_t \quad (2)$$

$$g_t = \beta C_t + (1-\beta) D_t \quad (3)$$

$$S_t = \gamma E_t + (1-\gamma) S_{t-m} \quad (4)$$

Where l_t is the level component at time t, g_t is the growth at time t, s_t is the seasonal component at time t and h_m = [(h-1)mod m]+1 where m represents the number of seasons within the data. The formula details are explained elsewhere [3,4]. Initial states of those components ($l_0, g_0, S_{1-m}, ..., S_0$) and smoothing parameters (α, β, γ) are estimated training data. The values of A_t, B_t, C_t, D_t, E_t are estimated based on from which the method belongs to [3].

The inclusion of state space framework into the smoothing model allows the production of prediction interval and other properties. The smoothing models themselves can only yield a point forecast. From these 15 models there are two models inside one with additive errors and one with multiplicative errors. Since there are 30 models presented, best model from ESSS is automatically selected by the framework based on Akaike's Information Criterior (AIC), Corrected Akaike's Information Criterion (AICc) and Bayesian Information Criterion (BIC).

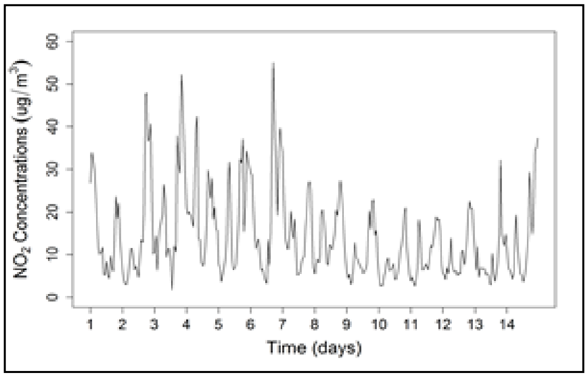
2.3 Performance Measurements.

We compared the performance of ESSS with the common approaches ARIMA and AR. We employed the mean error (ME), the mean absolute error (MAE), root mean square error (RMSE), R^2 and correlation coefficient to judge the performance index.

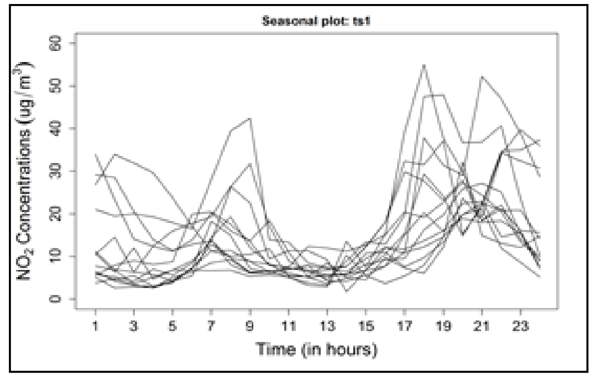
3.0 RESULTS AND DISCUSSION

Figure 1a below shows the daily concentration of NO_2 on Wonorejo site (Station 6). The figure shows seasonal pattern daily as reflected by two-peaks in the morning and night. Furthermore, no apparent trend appeared. The second week of January 2013 showed lower average concentrations than the first week. Figure 2b shows increase of concentrations between 7am to 9am in the morning and after 5pm in the evening suggesting an increase of vehicle volume in this region.

Before applying the model, we checked the stationarity of data because the foundation of time series model is stationarity. Figure 2a and Figure 2b displays ACF and PACF showing that the data was stationary. Box-Ljung test and Augmented Dickey-Fuller (ADF) test also suggested similar conclusion. Small p-value indicated that there is no significant evidence for non-zero correlation at least after lag 6. The p-value of Box-Ljung Test was < 2.2 x 10-6 and the p-value of ADF test was 0.01. Moreover, the ACF and PACF shows weak autocorrelation between data as the values degrade quickly (Figure 2).



(a)

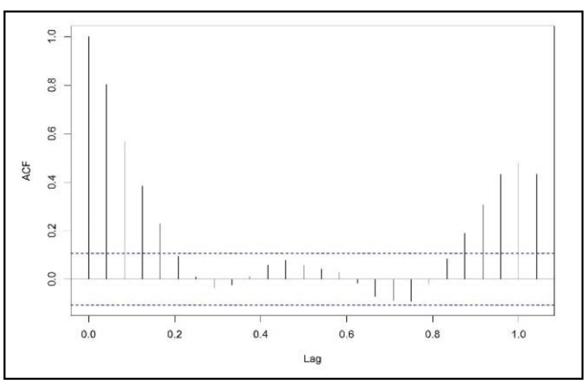


(b)

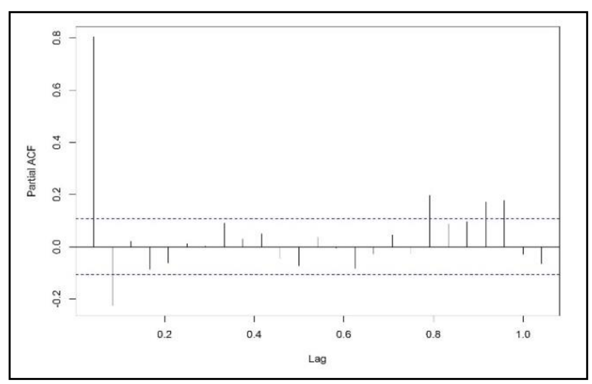
Figure 1 NO_2 concentrations pattern (a) over 14 days from 1st January 2013 (Thursday) to 14th January 2013 (Monday) and time plot (b) pattern.

Internal validation was done to check the performance one-step ahead of model performance and it showed that the performance of AR was better than ESSS and ARIMA. The MAE, and RMSE of ARIMA model were lower than both ESSS and AR, and the R^2 and R of AR was higher than other two models suggesting that the autoregression approach still works better. On the other hand, the forecasted values indicated by external validation of NO_2 were better for

ESSS model as measured by all performance criteria. The R^2 of ESSS was higher than AR and ARIMA which shows higher correlation between forecast with observed values used as validation set. Figure 3a shows the visual figure of forecasting result from the three models and the forecasting line from ESSS appeared to be following the pattern of NO_2 concentration well, Figure 3b shows the confidence interval of the predicted values taken from ESSS model. The residuals of forecasted values show that the pattern follows white noise and have weak correlation, as indicated by ACF and PACF on Figure 4 below. The p-value from Box-Ljung Test (0.000115) and ADF Test (0.07) supports the claim suggesting the ESSS model is appropriate with the assigned data.



(a)

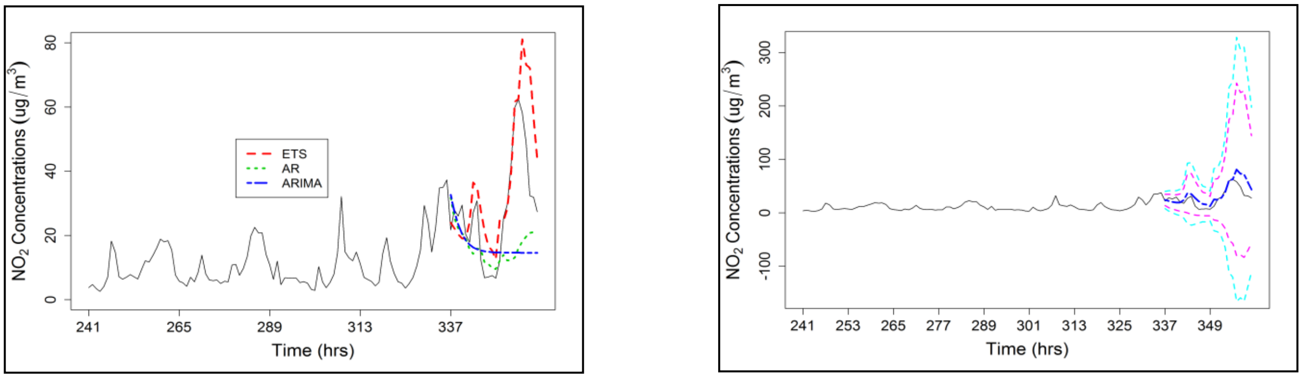


(b)

Figure 2 Autocorrelation plot of NO_2 hourly data: a) ACF, b) PACF

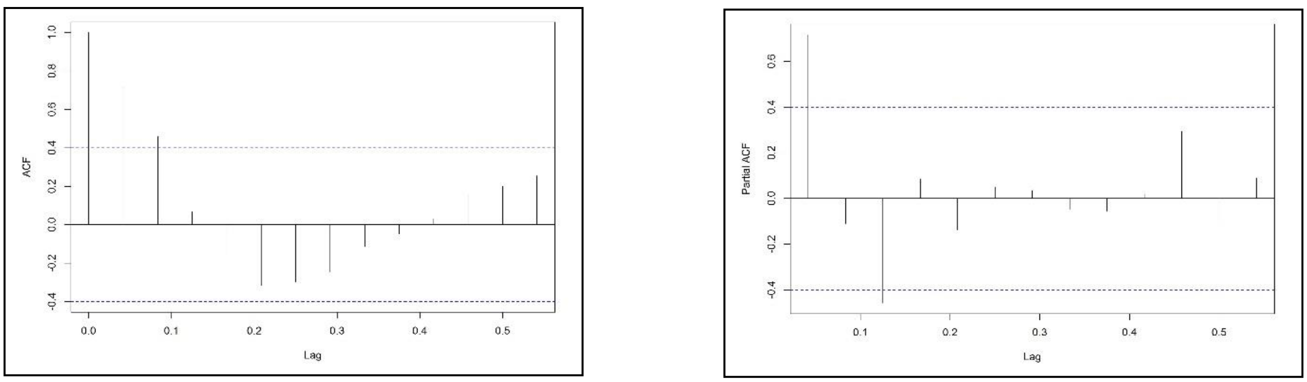
Table 2 Internal validation of each model for NO_2

Validation	Models	ME	MAE	RMSE	R^2	R
Internal	ESSS	-0.673	4.088	6.316	0.66	0.812
	AR(23)	-0.028	3.849	5.476	0.71	0.843
	ARIMA(2,0,0)	-0.026	3.985	5.797	0.675	0.822
External	ESSS	-7.879	10.261	14.019	0.673	0.82
	AR(23)	11.206	13.49	19.438	0.046	0.213
	ARIMA(2,0,0)	11.01	15.012	20.497	0.014	-0.119



(a) (b)

Figure 3 Forecast result from three models, only the last 4 days were shown in the figure (a) for clear picture of forecast result, (b) the 90% and 95% interval of predicted values taken from ESSS model



(a) (b)

Figure 4 Autocorrelation plot of residual from forecast yielded by ESSS model: a) ACF, b) PACF

In order to check consistency over days, we extend the forecasting. We forecast next 24-hours ahead concentrations using prior 14 days NO_2 concentrations for two weeks. We obtain the performance indexes of each day. We average the index and compile them into days. Table 3 shows that the performance of ESSS was not better than AR and ARIMA in terms of RMSE, however, the ESSS significantly improved the R^2 and correlation between prediction and observation

4.0 CONCLUSION

In this paper we explore the new approach of short term prediction of NO^2 using exponential state space smoothing method (ESSS). We find out that the forecasting result to be good thus proofing our hypothesis that ESSS model has good forecasting performance although the R^2 is less than 0.8 as we expected. Furthermore ESSS has better prediction result than more popular models used as comparison AR and ARIMA. The use of ESSS also produced consistent performance over days in terms of correlation between predicted values and observation values. It is therefore the present study has successfully applied ESSS model to one of air pollutants with satisfactory results. This promising result can further be developed and integrated with spatial analysis for a spatio-temporal model which is very useful for regions with limited air quality monitoring stations.

Table 3 Forecasting performance average each day of ESSS, AR, and ARIMA model

Model	Days	ME	MAE	RMSE	R^2	Cor
ESSS	Tue	-11.703	16.854	22.142	0.519	0.719
	Wed	-3.116	8.793	10.87	0.576	0.738
	Thu	-9.942	18.195	22.585	0.405	0.537
	Fri	-0.014	9.99	13.254	0.353	0.476
	Sat	-25.477	29.482	40.892	0.349	0.532
	Sun	-27.854	29.341	39.115	0.599	0.727
	Mon	-14.361	16.428	22.431	0.414	0.614
AR	Tue	2.603	9.425	12.085	0.412	0.64
	Wed	4.927	10.341	13.155	0.465	0.661
	Thu	2.968	13.635	16.206	0.36	0.475
	Fri	-1.982	8.85	10.483	0.248	0.466
	Sat	-2.667	9.498	10.931	0.353	0.432
	Sun	2.279	11.653	14.443	0.349	0.503
	Mon	0.615	7.151	8.803	0.446	0.657
ARIMA	Tue	4.738	12.579	15.327	0.034	-0.043
	Wed	6.879	14	17.422	0.077	0.105
	Thu	6.615	15.74	19.898	0.012	-0.053
	Fri	-0.311	9.706	11.712	0.037	0.117
	Sat	-3.203	11.253	13.047	0.064	0.157
	Sun	1.884	13.973	17.126	0.055	0
	Mon	3.079	10.732	12.777	0.008	0.08

References

[1] Ibrahim, M. Z., Zailan, R., Ismail, M. and Lola, M. S. 2009. Forecasting and Time Series Analysis of Air Pollutants in Several Area of Malaysia. *American Journal of Environmental Science*. 5(5): 625–632.

[2] Kumar, U. and Jain, V. K. 2009. ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). *Stochastic Environmental Research Risk Assessment*. 24(5): 751–760.

[3] Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S. A. 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*. 18: 439-454

[4] Dong, Z., Yang, D., Reindl, T. And Walsh, W. M. 2013. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy*. 55: 1104-1113

[5] Livera, A. M. D., Hyndman, R. J. and Snyder. R. D. 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*. 106: 1513-1527.