

A Study on Ultrafine Dust Prediction Model Estimation Using ARIMA Model and Multiplicative SARIMA Model

Chang-Ho An^{1*}, Jae-Hyun Kim², In-Kyu Song³.

^{1,3}Department of Financial Information Engineering, Seokyeong University, Seoul 02713; E-mail: choan@skuniv.ac.kr

²Department of Computer Engineering, Seokyeong University, Seoul 02713

Abstracts: Ultrafine dust data has seasonality. Therefore, in this study, the ARIMA model and the multiplicative SARIMA model, which are time series modeling methods, were estimated in consideration of seasonality, and the accuracy of the estimated prediction model was proposed using the MAPE measure to propose an ultrafine dust prediction model. For the ultrafine dust data used for the estimation of the ARIMA model, a seasonally adjusted estimate using the decomposition method was used assuming a multiplicative model, and the original ultrafine dust data was used for the multiplicative SARIMA model. The estimated prediction models were the ARIMA(0,1,4) model and the multiplicative SARMA(0,1,1)(0,1,1)₁₂ model. Residual analysis to validate the estimated ARIMA(0,1,4) model showed that the histogram of the Portmanteau statistic p-value was found to be significant. The predicted result increased in January and March, and no increase was observed from April to December. For the multiplicative SARMA(0,1,1)(0,1,1)₁₂ model, the significance probability of the chi-square statistic was significant at all lags. The prediction result showed that it increased in January and February, decreased continuously from March, and increased again in November. The prediction accuracy of the ARIMA (0,1,4) model was about 82.1%, and the multiplicative SARMA(0,1,1)(0,1,1)₁₂ model about 89.5%. The multiplicative SARMA(0,1,1)(0,1,1)₁₂ model was found to be about 7.4% better than ARIMA (0,1,4) in terms of the accuracy of the prediction model.

Keywords: ARIMA Model, Decomposition Method, Multiplicative SARIMA Model, Seasonal Adjustment, MAPE, Residual Analysis.

1. INTRODUCTION

The main causes and response strategies for fine dust have become important social issues. High concentrations of fine dust caused by climate change and environmental pollution are a great threat to our daily lives, such as health, life expectancy, and economic activities. Fine dust is a very small particle pollutant that floats or scatters in the air and is mainly generated when chemical fuels such as coal and oil are burned or when gas is discharged from factories and automobiles. It is divided into fine dust smaller than 10 μm (PM10) and ultrafine dust smaller than 2.5 μm (PM2.5) in diameter. As for the cause of fine dust in Korea, it is being argued over whether it comes from China or from domestic power plants, heating, vehicles, industrial facilities, etc. And it is also being confused over how to deal with fine dust generated during cooking of grilled meat and fish in indoor environments such as restaurants and houses. As shown in Figure 1, the public's greatest anxiety among environmental issues was found in the fine dust sector, which accounted for 72.9% in 2020 and 64.6% in 2022 [1].

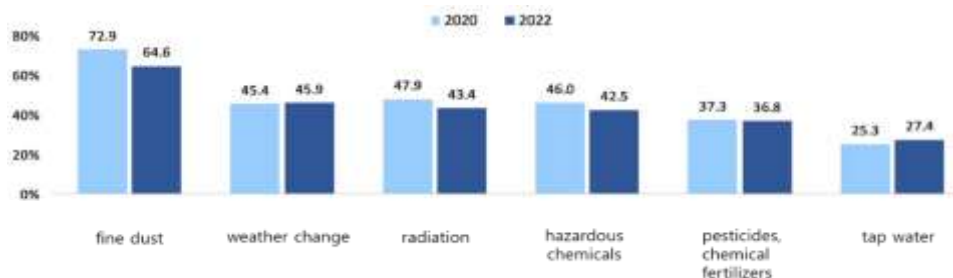


Fig. 1. Anxiety about environmental issues

Since fine dust is so small in size, it stays in the air and can adversely affect health by moving into the body through the respiratory tract. The impact of fine dust on health is direct and has a fatal effect on children, the elderly, and people with respiratory diseases. Ultrafine dust (PM_{2.5}) has a relatively smaller particle diameter than fine dust (PM₁₀), and is deposited in the alveoli and has a greater effect on the respiratory system [2]. According to recent research results, automobile exhaust gas, heating and power generation sectors, and each workplace account for the largest causes of fine dust emissions in Korea. According to the Seoul Research Institute, in the case of Seoul, the heating and power generation sector (39%) has the highest contribution to ultrafine dust (PM_{2.5}) emissions, followed by the automobile sector (25%). Among automobiles, especially diesel vehicles accounted for 60% [3]. The National Fine Dust Information Center explains that the concentration of ultrafine dust (PM_{2.5}) is related not only to the direct emission of air pollutants, but also to meteorological and topographical conditions that affect the spread and accumulation of pollutants. The level of fine dust in Korea is gradually improving thanks to the efforts of the people and the government, but it is still 1.5 to 2 times higher than that of major overseas cities such as Paris, Tokyo, London, and LA. The annual average concentration of fine dust in Seoul showed a clear trend of decreasing until 2012, and then repeatedly increasing and decreasing [4].

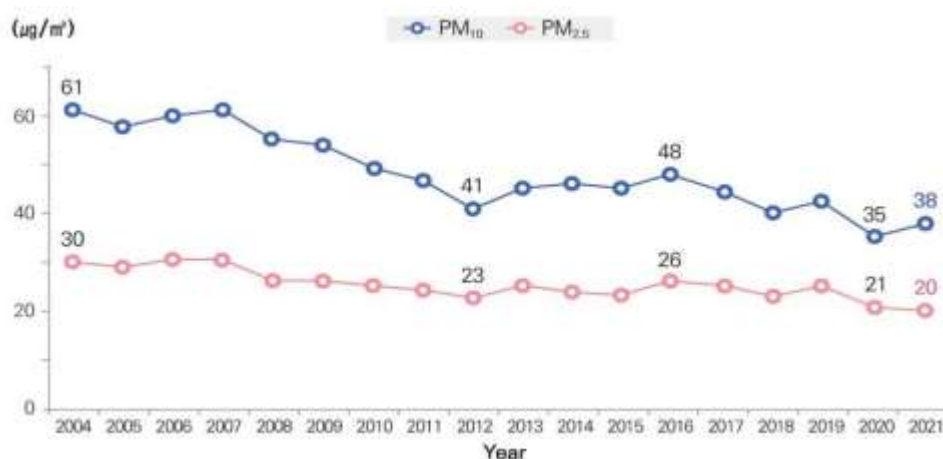


Fig. 2. Changes in fine dust air pollution in Seoul by year

At a time when public anxiety about fine dust continues, it is more important than anything else to minimize public anxiety. It is very hard that individual citizens can solve the fine dust problem, including the air quality problem. Therefore, the government should comprehensively and specifically identify the problems and establish improvement directions and response strategies. In this study, based on ultrafine dust data measured in Dongjak-gu, Seoul, we propose a prediction model for ultrafine dust concentration using a time series model considering seasonality.

2. Review of Previous Studies

Fine dust was classified by the World Health Organization in 2013 as a group 1 carcinogen, most of which is air pollutant. In other words, fine dust has a fatal effect on people's lives and health. Therefore, the need for research in various aspects such as causes and countermeasures for fine dust is increasing. Most of the meteorological data and air quality data are used to predict the concentration of fine dust. In the methodology, various prediction methods such as nonlinear models using machine learning and deep learning, multilinear regression models that are statistical models, and time series models are used. The following is a detailed look at previous studies using various methods.

K. P. Ra et al. predicted the concentration of fine dust using meteorological data and air pollutant data in an RNN/LSTM model [5]. Kang and Kang combined meteorological data and traffic data to predict fine dust based on machine learning using the Domain Adaptation method [6]. A. Chaloulakou et al. evaluated the fine dust prediction performance by comparing the PM₁₀ concentration as the response variable and meteorological data as the

explanatory variable through an neural network model and a regression model [7]. Kim and Moon analyzed machine learning-based seasonal forecasting ability using PM10 as the dependent variable and 7 variables (PM2.5, NO2, SO2, diurnal temperature range, wind speed, humidity, visibility) as independent variables of meteorological and air pollution factors [8]. Lim studied a fine dust concentration prediction model using machine learning with PM10 as the response variable and 11 atmospheric factors as explanatory variables in consideration of correlation and multicollinearity [9]. Lee and Oh analyzed the correlation between humidity and fine dust concentration through a study on the effect of humidity on light scattering fine dust measurement [10]. Choo et al. applied a multilinear model and found that the cause influencing the concentration of PM2.5 in Seoul was the meteorological factor [11]. In addition, the concentration of fine dust has a seasonal characteristic in that the concentration increases in spring and winter and decreases in summer and autumn. Previous studies on this are as follows. The concentration of ultrafine dust in Korea, including China, Japan, and Mongolia, increases in spring and winter in average monthly concentration and high concentration occurrence days [12,13,14]. And among the techniques for predicting the concentration of fine dust, the application of the Stochastic model (time series) is known to produce high prediction performance when data are accumulated [15].

3. Research Method

Ultrafine dust data has seasonality. Firstly, therefore, the autoregressive integrated moving average (ARIMA) model was applied using the seasonally adjusted values by removing the seasonality from the ultrafine dust data. The ARIMA(p,d,q) model is defined as follows.

$$\phi(B)(1 - B)^d Z_t = \delta + \theta(B)\varepsilon_t \tag{Equation 1}$$

where, B is backward shift operator, and $(1 - B)^d Z_t$ is ultrafine dust time-series data with d-th order difference (non-seasonal differencing) to remove the trend. $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$.

And the seasonally adjusted estimates are calculated as follows.

$$\widehat{Z}_t^{sa} = \frac{Z_t}{\widehat{S}_t} \tag{Equation 2}$$

where, Z_t is the original value of ultrafine dust data, and \widehat{S}_t is an estimated seasonal component.

Secondly, since the ultrafine dust data includes a seasonal factor, the multiplicative seasonal ARIMA model was applied. This model requires differencing (seasonal differencing) to remove seasonality, apart from differencing (nonseasonal differencing) to eliminate trend factors. The multiplicative SARIMA(p, d, q)(P, D, Q)_s model is defined as follows.

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D Z_t = \delta + \theta(B)\Theta(B^s)\varepsilon_t \tag{Equation 3}$$

where, B is backward shift operator, and $(1 - B)^d(1 - B^s)^D Z_t$ is the D-th order seasonally differenced ultrafine dust time series data with d-th order nonseasonal difference and seasonal period s. $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$, $\Theta(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs}$.

And in (Equation 1) and (Equation 3), $\varepsilon_t \sim (0, \varepsilon_\varepsilon^2)$ follows the white noise process.

The measure used to evaluate the predictive power of the ultrafine dust prediction model was the MAPE (mean absolute percentage prediction error) measure, which is defined as a function of prediction error, and the equation is as follows.

$$MAPE = \frac{100}{L} \sum_{l=1}^L \left| \frac{e_{n-1+l}(1)}{z_{n+l}} \right| \tag{Equation 4}$$

MAPE is mainly used when it affects outliers, and the lower the value, the higher the accuracy of the prediction model.

4. RESEARCH RESULTS

4.1 Constant term and ADF test

(Table 1.) shows the results of the ADF test performed by log transformation and first order difference for variance stabilization and trend removal for ultrafine dust data with seasonality removed by decomposition assuming a multiplicative model. In the result, the p-value of the Tau statistic is less than $\alpha = 0.05$, so $H_0: \phi = 1$ is rejected and the ultrafine dust data is stationary time series data. And since the significance probability of the t-statistic in the constant term t-test is 0.8429, which is greater than $\alpha = 0.05$, $H_0: \delta = 0$ is adopted and the constant term is not included.

Table 1. Ultrafine dust ADF test

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
ZM	0	-100.592	0.0001	-20.54	<.0001
	1	-290.999	0.0001	-11.84	<.0001
	2	101743.1	0.9999	-6.83	<.0001
	3	117.6012	0.9999	-6.43	<.0001
	4	49.0232	0.9999	-6.97	<.0001
	5	42.7959	0.9999	-5.15	<.0001
SM	0	-100.592	0.0001	-20.36	0.0001
	1	-291.67	0.0001	-11.74	0.0001
	2	152657.6	0.9999	-6.76	0.0001
	3	117.7542	0.9999	-6.37	0.0001
	4	49.0445	0.9999	-6.89	0.0001
	5	42.7688	0.9999	-5.09	0.0002

4.2 ARIMA(0,1,4) model estimation

The results of model identification using the default values of MINIC (minimum information criterion) using BIC (Bayesian information criterion) are shown in (Table 2.), and the parameter estimation results using ML (maximum likelihood) estimation method are shown in (Table 3.).

In (Table 2.), the minimum value is MINIC=-3.24474, so the identified model is the MA (4) model.

Table 2. MA (4) Model Identification

Model Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	-1.92739	-2.60686	-2.78833	-3.18523	-3.24474	-3.19052
AR 1	-2.78686	-3.11044	-3.04615	-3.15371	-3.20518	-3.19505
AR 2	-2.98471	-3.05391	-3.00847	-3.20697	-3.21153	-3.18537
AR 3	-2.98264	-3.04544	-3.02377	-3.14775	-3.14747	-3.1186
AR 4	-2.99968	-3.11487	-3.04763	-3.14307	-3.0807	-3.05279
AR 5	-3.22188	-3.20984	-3.17289	-3.14973	-3.08899	-3.13259

As a result of estimating the parameters by applying the ARIMA (0,1,4) model to the ultrafine dust data with seasonality removed in (Table 3.), the p-value of parameter $\theta_1, \dots, \theta_4$ is less than $\alpha = 0.05$, so $H_0: \theta_1, \dots, \theta_4 = 0$ was rejected and all parameters are statistically significant.

Table 3. ARIMA (0,1,4) parameter estimation

Parameter	Estimate	S.E	t -Value	Pr > t
MA1,1	1.76347	0.17625	10.01	<.0001
MA1,2	-1.0954	0.24855	-4.41	<.0001
MA1,3	0.82778	0.23858	3.47	0.0005
MA1,4	-0.5057	0.15665	-3.23	0.0012

Therefore, the estimated ARIMA (0,1,4) prediction model is as follows.

$$(1 - B)\ln Z_t = (1 - 1.7634B + 1.09954B^2 - 0.8277B^3 + 0.5057B^4) \quad (\text{Equation 5})$$

4.3 ARIMA (0,1,4) model test and prediction

As a result of the white noise test of the residuals for the ARIMA (0,1,4) model, the histogram of the p-value was located below the significance level $\alpha = 0.05$ in the Portmanteau test. And the ultrafine dust predicted by the ARIMA (0,1,4) prediction model is as shown in (Figure. 3.), where it increased in January and March of 2023, and did not increase from April to December and maintained a certain level.

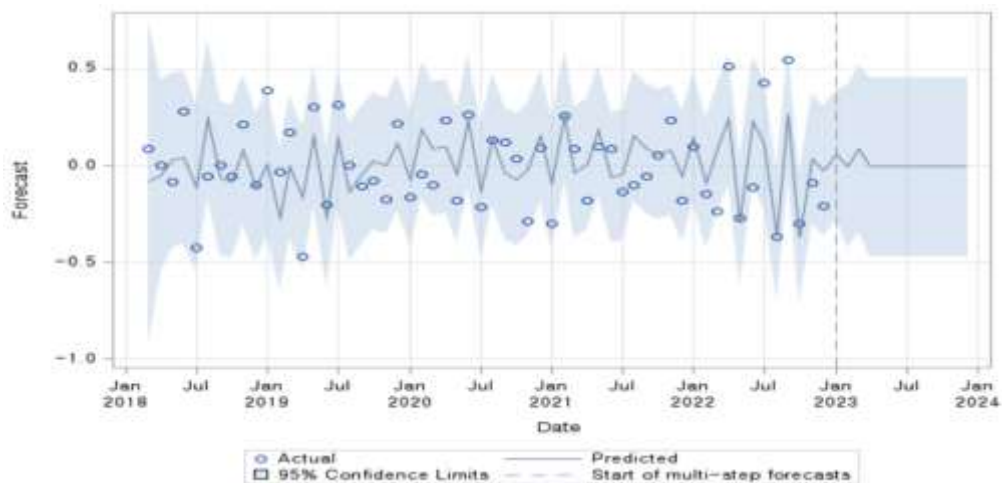


Fig. 3. Prediction by ARIMA (0,1,4) model (constant term not included)

4.4 Multiplicative SARIMA(p, 1, d)(P, 1, Q)₁₂ model identification

As a result of performing autocorrelation analysis by logarithmic transformation and first-order non-seasonal differencing and seasonal differencing on the ultrafine dust raw data, it appeared as a stationary time series in the sample autocorrelation function (SACF) and partial autocorrelation function (SAPCF) figures. As a result of the constant term test, the p-value of the t-statistic was 0.9017, indicating that the constant term was not included at the significance level $\alpha = 0.05$. Therefore, among several candidate models with the model identification criterion set to $0 \leq p, q, P, Q \leq 1$, the models that passed the Portmanteau test criteria for the residuals were the multiplicative SARIMA(1,1,0)(0,1,1)₁₂ model and the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model. Then, between the two models, the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model was selected on the criterion for the Bayesian information criterion (BIC) with the minimum value and the minimum variance ($\widehat{\sigma}_\epsilon^2$).

Table 4. Model identification of SARIMA(0,1,1)(0,1,1)₁₂

SARIMA(p, 1, q)(P, 1, Q) ₁₂ Candidate models	portmanteau test p – value > 0.05	BIC	$\widehat{\sigma}_\epsilon^2$
SARIMA(1,1,0)(1,1,0) ₁₂	reject		
SARIMA(0,1,1)(0,1,1) ₁₂	accept	7.174967	0.060472
SARIMA(1,1,0)(0,1,1) ₁₂	accept	9.588324	0.063659
SARIMA(1,1,1)(0,1,1) ₁₂	reject		
SARIMA(1,1,1)(1,1,0) ₁₂	reject		
SARIMA(1,1,0)(1,1,1) ₁₂	reject		
SARIMA(1,1,1)(1,1,1) ₁₂	reject		
SARIMA(0,1,1)(1,1,0) ₁₂	reject		
SARIMA(0,1,1)(1,1,1) ₁₂	reject		

4.5 Multiplicative SARIMA(0, 1, 1)(0, 1, 1)₁₂ model estimation

(Table 5.) shows the parameter estimation results of the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model. Since the t-statistic p-value of the estimated parameter is less than the significance level $\alpha = 0.05$, the estimated parameter is statistically significant.

Table 5. Parameter estimation of SARIMA(0,1,1)(0,1,1)₁₂

Parameters by Maximum Likelihood Estimation				
Parameter	Estimate	S.E	t -Value	Pr > t
MA1,1	0.81574	0.08675	9.4	<.0001
MA2,1	0.55592	0.15741	3.53	0.001

Therefore, the estimated multiplicative SARIMA(0,1,1)(0,1,1)₁₂ prediction model equation is as follows.

$$(1 - \phi B)(1 - B^{12})\ln Z_t = (1 - 0.81574B)(1 - 0.55592B^{12}) \quad \text{(Equation 6)}$$

4.6 Multiplicative SARIMA(0, 1, 1)(0, 1, 1)₁₂ model test and prediction

(Table 6.) is the portmanteau test result of the residuals for the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model. At all lags, since the p-value of the chi-square (χ^2) statistic is less than $\alpha = 0.05$, the residuals follow white noise.

Table 6. Portmanteau test of residuals

Autocorrelation Test of Residuals								
Lags	Chi-Square	Pr > ChiSq	Autocorrelation Coefficient					
6	5.59	0.2321	0.043	0.207	-0.129	-0.173	-0.007	-0.122
12	8.47	0.5827	-0.148	0.015	0.113	0.038	0.094	-0.053
18	14.13	0.5892	0.126	0.204	0.01	-0.052	-0.054	-0.126
24	18.79	0.6586	0.185	-0.119	0.015	-0.006	0.049	0.063

(Figure. 4.) is the predicted result of ultrafine dust by multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model. Looking at the forecast results, it was found that it increased in January and February 2023, decreased continuously from March, and increased again in November.

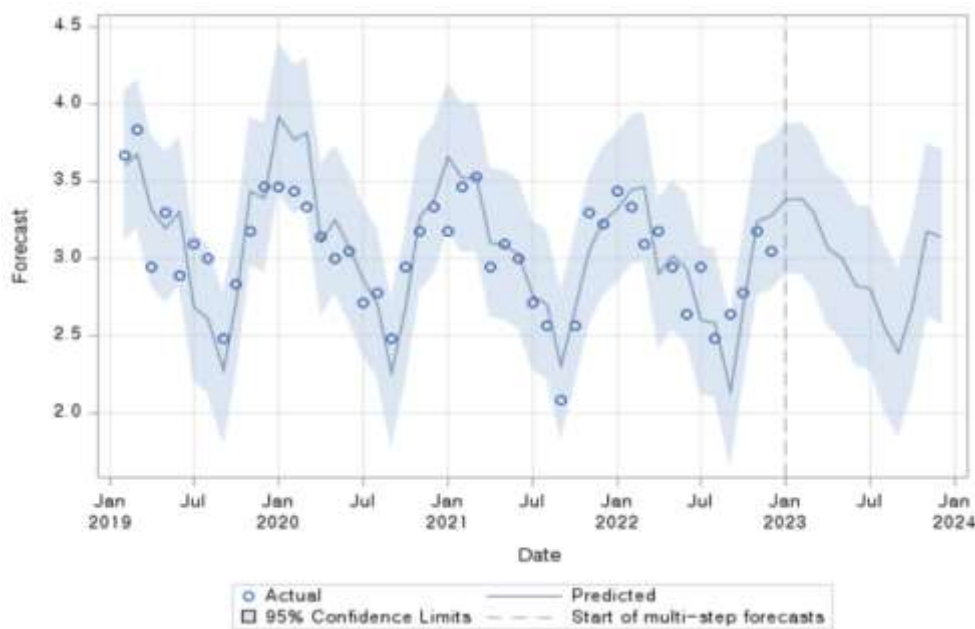


Fig. 4. Prediction by SARIMA(0,1,1)(0,1,1)₁₂ model

4.7 Prediction Model Accuracy Evaluation

The ARIMA (0,1,4) prediction model and the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ prediction model were found to follow white noise as a result of residual analysis, so the two prediction models are models that can be used for prediction. As a result of using the MAPE (mean absolute percentage prediction error) measure to evaluate the prediction accuracy of the two models, it was found that the ARIMA (0,1,4) model had a prediction error of about 17.9% on average with MAPE = 0.179241, and that the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model's MAPE=0.105281 showed an average prediction error of about 10.5%. That is, the prediction accuracy of the ARIMA (0,1,4) model was about 82.1%, and the prediction accuracy of the multiplicative model was about 89.5%. Therefore, the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model was estimated to have about 7.4% better predictive power than ARIMA (0,1,4).

CONCLUSION

In a situation where people's anxiety and fear about ultrafine dust are amplifying, fine dust prediction research is very important. In this study, a prediction model for ultrafine dust was proposed using a time-series modeling method. The ultrafine dust data used in the study was the final definitive measurement data provided by Air Korea

(<http://www.airkorea.or.kr>) [16]. The data was measured from January 2018 to December 2022 in Dongjak-gu, Seoul, and was used after converting it to a monthly average. The Stochastic model of time series for prediction were the ARIMA model and the multiplicative SARIMA model. Both the ARIMA (0,1,4) model, which was estimated after removing seasonality through decomposition, and the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model, which was estimated after performing first-order nonseasonal differences and seasonal differences on the raw ultrafine dust data, turned out to be valid models. However, the prediction results and accuracy were different. The prediction results of the ARIMA (0,1,4) model showed an increase in January and March of 2023 and no increase from April to December. The prediction result of the multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model showed an increase in January and February, a decrease from March, and an increase again in November, and showed higher prediction accuracy in terms of prediction accuracy. Therefore, we propose a multiplicative SARIMA(0,1,1)(0,1,1)₁₂ model with less prediction error as a prediction model for ultrafine dust. As seen in this study, it is difficult to build a prediction model for ultrafine dust with high prediction accuracy. There are several predictive modeling methods, and even if the validity of the prediction models is proven, the prediction results bring different results. Therefore, continuous research on prediction models to improve the prediction accuracy, such as machine learning and deep learning methods including time series modeling methods, is required.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT, MOE) and (No. 2019M3E7A1113102).

※ MSIT: Ministry of Science and ICT, MOE: Ministry of Education

REFERENCES

- [1] Statistics Korea, (2022). 2022 Social Survey Results.
- [2] Atkinson, R. W., Mills, I. C., Walton, H. A. & Anderson, H. R., Fine particle components and health – asystematic review and meta-analysis of epidemiological time series studies of daily mortality and hospital admissions. *Journal of Exposure Science and Environmental Epidemiology*, 25 (2015), 208-214. <https://doi.org/10.1038/jes.2014.63>
- [3] <http://www.daenews.co.kr/13201>
- [4] Seoul, (2021). 2021 Seoul Air Quality Report.
- [5] Ra, K. P., Kim, M. C., Kim, M. J., Lim, S. T. & Sim. Y. G., A Study on The Prediction of The Fine-Dust Concentration Using RNN/LSTM. *The Institute of Electronics and Information Engineers*, (2019), 1400-1405.
- [6] Jam, F. A., Mehmood, S., & Ahmad, Z. (2013). Time series model to forecast area of mangoes from Pakistan: An application of univariate ARIMA model. *Acad. Contemp. Res*, 2, 10-15.
- [7] Kang, T. C. and Kang, H. B., Machine Learning based Estimation of the Concentration of Fine Particulate Matter Using Domain Adaptation Method. *Journal of Korea Multimedia Society*, 20 (2017), 1208-1215. <https://doi.org/10.9717/kmms.2017.20.8.1208>
- [8] Chaloulakou, A., Grivas, G. and Spyrellis, N., Neural Network and multiple regression models for PM10 prediction in Athens: A comparative assessment. *Journal of the Air & Waste Management Association*, 53 (2003), 1183-1190. <https://doi.org/10.1080/10473289.2003.10466276>
- [9] Kim, H. R. and Moon, T. H., Machine learning-based fine dust prediction model using meteorological data and fine dust data. *Journal of the Korean Association Geographic Information Studies*, 24 (2021), 92-111. doi:10.11108/kagis.2021.24.1.092
- [10] Lim, J. M., An estimation model of fine dust concentration using meteorological environment data and machine learning. *Journal of Information Technology Services*, 18 (2019), 173-186. doi:<https://doi.org/10.9716/KITS.2019.18.1.173>
- [11] Lee, C. H. and Oh, S. N., Effect of atmospheric humidity on fine dust measurement using the light scattering method. *Journal of the Korean Society of Hazard Mitigation*, 20 (2020), 391-399. <https://doi.org/10.9798/kosham.2020.20.1.391>
- [12] Choo, G. H., Lee, K. T. and Jeong, M. J., Analysis of empirical multiple linear regression models for the production of PM2.5 concentrations. *Journal of Korean Earth Sci*, 38 (2017), 283-292. doi:10.5467/JKESS.2017.38.4.283
- [13] Bae, M., Kim, B.-U., Kim, H. C., Kim, S., A Multiscale Tiered Approach to Quantify Contributions: A Case Study of PM2.5 in South Korea During 2010-2017. *Atmosphere*, 11 (2020a), 141. <https://doi.org/10.3390/atmos11020141>
- [14] Bae, C., Kim, B.-U., Kim, H. C., Yoo, C., Kim, S., Long-Range Transport Influence on Key Chemical Components of PM2.5 in the Seoul Metropolitan Area, South Korea, during the Years 2012-2016. *Atmosphere*, 11 (2019), 48. <https://doi.org/10.3390/atmos11010048>
- [15] Kim, Ju. H., Kim, H. C., Kim, B.-U., Shin, H. J. and Kim, S., Long-term Trend Analysis of Key Criteria Air Pollutants over Air Quality Control Regions in South Korea using Observation Data and Air Quality Simulation. *Journal of Korean Society for Atmospheric Environment*, 34 (2018), 101-119. <https://doi.org/10.5572/kosae.2018.34.1.101>
- [16] Oh, J. M., Shin, H. S., Shin, Y. S. and Jeong, H. C., Forecasting the Particulate Matter in Seoul using a Univariate Time Series Approach.

Journal of the Korean Data Analysis Society., 19 (2017), 2457-2468. <https://doi.org/10.37727/jkdas.2017.19.5.2457>
https://www.airkorea.or.kr/web/last_amb_hour_data?pMENU_NO=123

DOI: <https://doi.org/10.15379/ijmst.v10i4.1884>

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.