



OPEN Spatiotemporal distribution, trend, forecast, and influencing factors of transboundary and local air pollutants in Nagasaki Prefecture, Japan

Santos Daniel Chicas^{1✉}, Jair Gaspar Valladarez², Kiyoshi Omine³, Venkataraman Sivasankar⁴ & Sangyeob Kim^{5✉}

The study of PM_{2.5} and NO₂ has been emphasized in recent years due to their adverse effects on public health. To better understand these pollutants, many studies have researched the spatiotemporal distribution, trend, forecast, or influencing factors of these pollutants. However, rarely studies have combined these to generate a more holistic understanding that can be used to assess air pollution and implement more effective strategies. In this study, we analyze the spatiotemporal distribution, trend, forecast, and factors influencing PM_{2.5} and NO₂ in Nagasaki Prefecture by using ordinary kriging, Pearson's correlation, random forest, Mann-Kendall, auto-regressive integrated moving average and error trend and seasonal models. The results indicated that PM_{2.5}, due to its long-range transport properties, has a more substantial spatiotemporal variation and affects larger areas in comparison to NO₂, which is a local pollutant. Despite tri-national efforts, local regulations and legislation have been effective in reducing NO₂ concentration but less effective in reducing PM_{2.5}. This multi-method approach provides a holistic understanding of PM_{2.5} and NO₂ pollution in Nagasaki prefecture, which can aid in implementing more effective pollution management strategies. It can also be implemented in other regions where studies have only focused on one of the aspects of air pollution and where a holistic understanding of air pollution is lacking.

Studies have attributed the increase in air pollution to rapid urban development and modernization¹. Over the years, much emphasis has been placed on the analysis of PM_{2.5} (particulate matter with a diameter of 2.5 µm or less) and NO₂ (Nitrogen Oxide) due to the adverse effects on public health²⁻⁴, global climate^{5,6} and long-range transport, particularly for PM_{2.5}⁷. As a result of public health and environmental implications, countries and international organizations have engaged in regulating and monitoring PM_{2.5} and NO₂ concentrations. For instance, the World Health Organization (WHO) air quality guidelines established in 2005 were revised and released on September 22, 2021. These new guidelines come from decades of research showing that air pollution's health effects result from high exposure and very low concentrations⁸. Therefore, the guidelines of 2005 recommended that the annual average of PM_{2.5} and NO₂ concentrations should not exceed 10 and 40 µg/m³ (21 ppb), respectively. The 2021 guidelines reduce these recommendations to 5 and 10 µg/m³ (5 ppb) for PM_{2.5} and NO₂, respectively.

The analysis and monitoring of PM_{2.5} and NO₂ are essential to assess the effectiveness of mitigation strategies and compliance with standards. Currently, monitoring stations reliably and accurately measure PM_{2.5} and NO₂ concentrations. However, monitoring is often difficult as PM_{2.5} and NO₂ measurements are only done at some locations due to the high costs of installation, maintenance, and management of monitoring stations. As a result, detailed information about the spatiotemporal distribution, trend, and climatic and temporal effect of

¹IRI-THESys and Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. ²Faculty of Science and Technology, University of Belize, Belmopan, Belize. ³Department of Civil Engineering, School of Engineering, Nagasaki University, 1-14 Bunkyo, Nagasaki 852-8521, Japan. ⁴Post Graduate and Research Department of Chemistry, Pachaiyappa's College (affiliated to University of Madras), Chennai, Tamil Nadu 600030, India. ⁵Estuary Research Center, Shimane University, Nishikawatsu-cho 1060, Matsue 690-8504, Japan. ✉email: chicassa@hu-berlin.de; kim@soc.shimane-u.ac.jp

PM_{2.5} and NO₂ is often lacking in locations with few monitoring stations. Therefore, the need exists, in these locations, to implement a multi-method approach that could predict PM_{2.5} and NO₂ where it is not measured and generate fine-grain spatial distribution data, examining the changes of PM_{2.5} and NO₂ concentration over time and understand the climatic and temporal effects on PM_{2.5} and NO₂ concentrations. This information is critical for identifying areas (hotspots) that do not comply with international pollution concentration standards, quantitatively evaluating the air quality policy, and assessing the risks to human health⁹. Furthermore, empirical models are required to describe the general features of the spatial patterns of PM_{2.5} and NO₂, trends, and influencing factors^{10,11}.

In Japan and many regions of the world, studies have only focused on the spatiotemporal distribution, trend, forecast, or influencing factors of PM_{2.5} and NO₂^{12,13}. However, rarely studies have combined these to generate a more holistic understanding that can help identify health risk areas, influencing factors, pollution trends, and the efficacy (or lack thereof) of policy interventions¹¹. For instance, in Nagasaki Prefecture, most of the studies have only focused on either the health impact caused by air pollutants^{14–16}, the long-range transport of air pollutants from the Asian continent^{17–19}, or the effects of climatic variables and the spatial and temporal distribution of PM_{2.5}²⁰. Although these studies provide essential information, a multi-method approach is necessary to understand better the distribution, factors, and current and future trends of PM_{2.5} and NO₂ pollution in Nagasaki Prefecture.

This study used a multi-method approach to analyze PM_{2.5} and NO₂ data in Nagasaki Prefecture from 2013 to 2021. This study aims (1) to estimate PM_{2.5} and NO₂ pollution variability in unmeasured areas using ordinary kriging, (2) to identify and analyze the correlation of the major climatic and temporal factors that influence PM_{2.5} and NO₂ pollution in Nagasaki Prefecture via Pearson's correlation and random forest feature selection and (3) to conduct a trend and forecast analysis of PM_{2.5} and NO₂ based on fitting loess, automated auto-regressive integrated moving average (ARIMA) and error trend and seasonal models (ETS). Using a multi-method approach, we provide a broader analysis and understanding of the spatiotemporal distribution, forecast, trend, and influencing factors of pollutants which is crucial for the improvement, development, and assessment of mitigation strategies and for identifying health risk areas. Furthermore, this proposed multi-method approach can be used in other regions where studies have only focused on one of the aspects of air pollution and where a holistic understanding of air pollution is lacking.

Policy background. As a result of the rapid economic development of the Northeast Asian sub-region and the resulting environmental problems, China, Japan and Korea have held a Tripartite Environment Ministers Meeting (TEMM) annually since 1999²¹. This meeting aims to strengthen environmental cooperation among these countries and address environmental problems at the domestic, regional, and global levels. At the 15th Tripartite Environment Ministers Meeting in 2013, the Tripartite Policy Dialogue on Air Pollution (TPDAP) was established and started in 2014²¹. The objective of the establishment of the TPDAP was to coordinate efforts among the three countries to address the air pollution problem by developing cooperation initiatives and sharing information about air pollution policy implementation and impacts. The 3rd TPDAP in 2016 established two working groups to share air pollution information (Fig. 1).

The establishment of the Tripartite Policy Dialogue on Air Pollution (TPDAP) has resulted in reducing air pollution in the three countries (Table 1).

Results

PM_{2.5} and NO₂ spatiotemporal distribution. The years that had the maximum mean average level of PM_{2.5} concentration were 2014 and 2016, with 16.2 and 14 µg/m³, respectively. The minimum mean average concentrations in 2020 and 2021 were 10.4 and 9.7 µg/m³, respectively (Table 2). There have been dynamic temporal variations of PM_{2.5} concentration with respect to a yearly minimum, maximum, and mean during the study period. These variations are also expressed in the year's seasons, with Winter and Spring being the seasons with the highest PM_{2.5} concentrations. For each year of the study period, 2013–2021, Spring had the highest concentration of PM_{2.5} (Table S2). Amidst these variations, there is an indication of a declining trend of PM_{2.5} concentration from 2013 to 2021.

Regarding the spatial distribution of PM_{2.5} concentration, there is also yearly variation as to the hotspot of PM_{2.5} (Fig. 2). However, to a great degree, the most affected area is the westernmost part of Nagasaki Prefecture, as illustrated by the PM_{2.5} spatial distribution maps of 2015, 2016, 2017, and 2018.

Concerning NO₂, the maximum mean average level of NO₂ concentrations was in 2013 and 2014 at 6.3 ppb. The minimum mean average concentrations were in 2020 and 2021, with 4.7 and 4.4 ppb, respectively (Table 2). The box plots indicated minimal temporal variations of NO₂ concentration during the study period (Fig. 3). Winter and Spring are the seasons with the highest NO₂ concentration, with Winter being the season with the highest NO₂ concentrations for each year, except for 2013 when Spring had the highest concentration (Table S2). The spatial distribution of NO₂ indicated that the hotspots have remained in the same location over the years (Fig. 3). The high concentration of NO₂ is located in Sasebo and Nagasaki, the two largest cities of Nagasaki Prefecture.

Pearson's correlation and Random forest. Pearson's correlation results indicated that meteorological factors influence PM_{2.5} and NO₂ concentrations in Nagasaki Prefecture (Fig. 4). In the case of PM_{2.5}, the factors had weak positive, negative, and mixed correlation results. The factors negatively correlated with PM_{2.5} were the southern oscillation index, average temperature, maximum temperature, minimum temperature, average humidity, and minimum humidity, with the southern oscillation index having the most substantial influence among these factors. On the other hand, minimum wind speed and sunlight time had a positive correlation. The other factors had mixed results in some stations having positive, negative, and no correlation. For NO₂, some

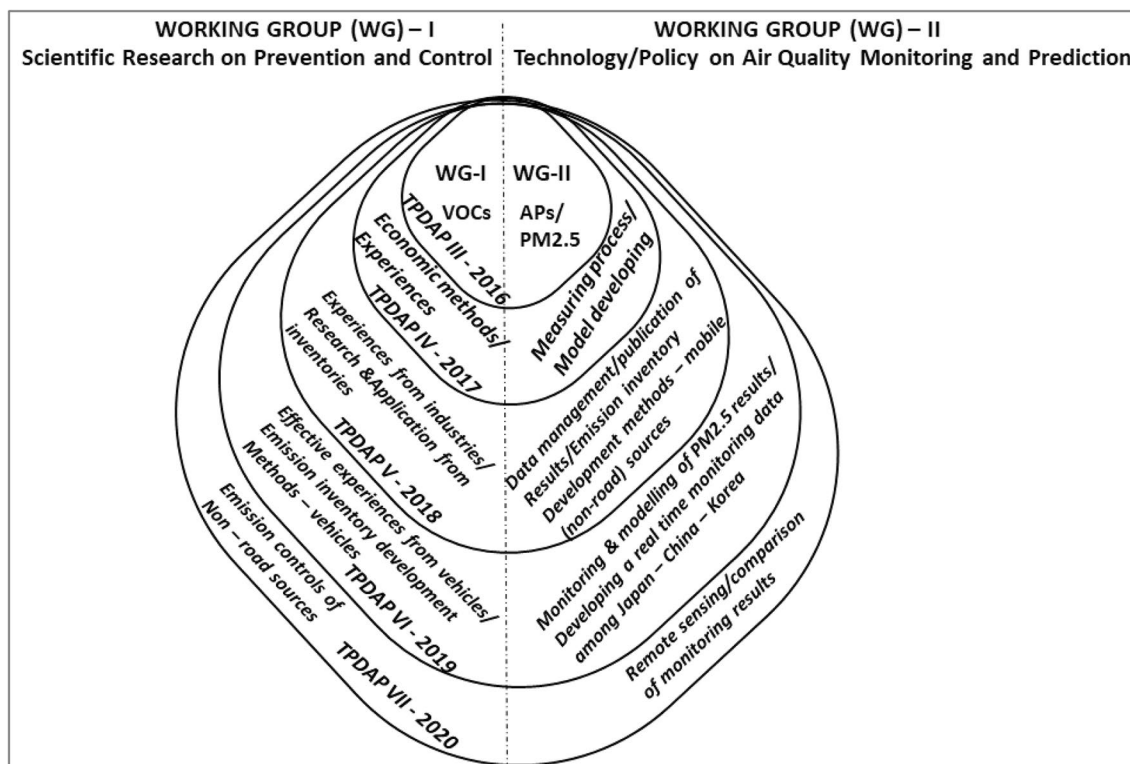


Figure 1. Tripartite Policy Dialogue on Air Pollution working groups one and two.

Country	Policy	Targeted sources	Impact	Research
Japan	Since the introduction of the "Comprehensive Policy Effort on PM _{2.5} " in 2013, policies and measures were introduced that fall under three categories: -Domestic measures -International cooperation -Reinforcing scientific knowledge	-Stationary -Vehicles -Vessels -Open burning -NH ₃ -Regional	Since 2013, the annual average of PM _{2.5} concentration has decreased, with the number of occasions issuing public alerts on PM _{2.5} decreasing from 37 in 2013 to 2 in 2017	Declining PM _{2.5} as a result of policy implementation: 2010–2018 declining PM _{2.5} ²² 2003–2018 declining PM _{2.5} ²³
China	Since the introduction of the "Action Plan on the Prevention and Control of Air Pollution" in 2013, significant policies and measures have been introduced which fall under five categories: -Rule of law -Scientific and Technological support -Comprehensive emission reduction -Innovative management -Social participation	-Industry -Energy sector -Vehicles -Non-point source pollution	The average concentration of PM _{2.5} in 74 cities decreased by 42% from 2013 to 2018. These cities are applying Ambient Air Quality Standard	Declining PM _{2.5} 2013–2018 ²⁴ Declining PM _{2.5} 2019–2020 ²⁵
Korea	Since the revision of the "Air Quality Preservation Act" in 2013 and the introduction of "Comprehensive Plan on Fine Dust Management Implementation" in 2017, strategic policies and measures have been implemented which fall under four categories: -Reduction of domestic emissions -Public health -International cooperation -Solid foundation and communication	-Industry -Transport -Power generation -Agriculture and daily surrounding -Indoor air quality -Regional	PM _{2.5} reduction from 2010 to 2019 has shown minimal changes each year, with the PM _{2.5} level decreasing by 2 µg/m ³ in 2018 from the level of 2017. This is due to the implementation of the comprehensive plan in 2017	PM _{2.5} decrease 2003–2017 ²⁶ PM _{2.5} decrease 2010–2020 ²⁷

Table 1. Significant policies implemented by China, Japan, and Korea to improve air quality.

factors had a strong positive and negative correlation, with others having a weak positive and negative correlation. Average temperature, maximum temperature, minimum temperature, average humidity, and minimum humidity had a strong negative correlation, with average local pressure and average sea level pressure having a strong correlation. On the other hand, the southern oscillation index, rain maximum 10 min, average wind speed, and sunlight time had a weak negative correlation, with maximum wind speed and maximum instantaneous wind speed having a weak positive correlation. The other factors had mixed results, with some stations having positive, negative, or no correlation.

Year	PM _{2.5}			NO ₂		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
2013	6	24	12.6	2	26	6.3
2014	11	24	16.2	1	21	6.3
2015	8	19	13.7	2	21	6.1
2016	11	21	14	2	20	5.8
2017	10	19	12.9	2	20	6
2018	7	19	11.8	2	19	5.6
2019	8	18	11.8	0	18	5.8
2020	8	14	10.4	1	17	4.7
2021	7	14	9.7	1	15	4.4

Table 2. Annual descriptive statistics for PM_{2.5} and NO₂.

The RF model feature selection results at Tsushima, Goto, Daitou, Inasa, Obama, and Yoshii for PM_{2.5} and NO₂ are shown in (Fig. 5). These stations were selected based on their location, which are representative of the study area. Among the monitoring stations, the most important predicting factors for PM_{2.5} were Spring, maximum instantaneous wind speed one day, maximum instantaneous wind speed, humidity, sunlight time one day, and southern oscillation index. At the observation tower of Tsushima, Goto, Inasa, and Yoshii, Spring is among the three major predicting factors of PM_{2.5}, with Spring being the primary predictor at Goto. Maximum instantaneous wind speed was among the three most important predictors for Goto, Daitou, Inasa, and Yoshii and was the primary predictor at Tsushima. Maximum instantaneous wind speed was the main predictor for Daitou, Inasa, and Yoshii and the second most important predictor for Obama. Humidity was the main predictor of Obama and the third predictor at Goto. Southern oscillation index was only a significant predictor at Tsushima.

On the other hand, for NO₂, the three most important predictors that were among the stations were average wind speed, minimum temperature one day, maximum instantaneous wind speed, average sea level pressure one day, average temperature one day, average sea level pressure three days, southern oscillation index seven days, average temperature, maximum temperature one day and average local pressure one day. Average wind speed was among the three main predictors in Tsushima, Goto, Daitou, Inasa, and Yoshii, and the main predictor in Goto. Minimum temperature was the major predictor in Tsushima and Daitou. In Goto and Daitou, maximum instantaneous wind speed was the second and third major predictor, respectively. Sea level pressure one day was the main predictor in Inasa and Yoshii. Average temperature one day, average sea level pressure three days, southern oscillation index seven days, average temperature, maximum temperature one day, and average local pressure one day were among the three most important factors in only one of the stations. The feature selection result from RF indicated that the factors influencing PM_{2.5} and NO₂ concentrations in Nagasaki Prefecture vary depending on the location of the monitoring stations.

Tables 3 and 4 show the results of the random forest models for each of the 18 stations for PM_{2.5} and NO₂, respectively. Model accuracy was evaluated using R² and MSE. In the case of PM_{2.5}, the result indicated that the accuracy estimates for the 18 stations are varied with R² values in the range of 0.41–0.53 and MSE of 22.7–37.6 for the training dataset and R² values ranging from 0.16 to 0.33 and MSE 32.7–51.3 for the test dataset. The low values of R², high values for MSE, and the high difference of R² between the trained model and the test model indicated that the RF models constructed with these factors could not be used to predict PM_{2.5} concentrations.

Whereas, in the case of NO₂, R² was higher (Test: 0.354–0.735) than the R² values of PM_{2.5}; thus, the factors used in this study are a better predictor of NO₂ concentration than PM_{2.5}. However, the R² values for most of the NO₂ test models are still low to be used to predict NO₂ concentrations, except for the results of Isahaya, which can be considered acceptable.

Trend and Forecast analysis. At a 0.05 significance level, the Mann–Kendall test determined that PM_{2.5} and NO₂ in most of the monitoring stations had a monotonic trend and a negative slope (Table 5). The stations with no monotonic trend for PM_{2.5} were Shimabara, Oomura, Kawadana, Togitsue, MatsuuraShimachi, and Tsushima, and for NO₂ were Yukiura, Tsushima, Iki, Obama, and Muramatsu. The stations that had the most significant magnitude of reduction for PM_{2.5} were Daitou (−1.278 µg/m³), Fukuishi Jihai (−1.178 µg/m³), and Kogakura (−1.01 µg/m³), while Goto (−0.43 µg/m³) had the lowest. For NO₂, the most significant magnitude of reduction was observed in Fukuishi Jihai (−0.78 ppb), Higashi Nagasaki (−0.50 ppb), and Kogakura (−0.49 ppb), and the lowest in Matsuura Shisamachi (−0.12 ppb). Figures S1 and S2 represent the data decomposition and trend for six monitoring stations.

The results of Holt–Winters and ARIMA forecast analysis are presented in Table 5, which indicated, based on the MAPE and RMSE, that model suitability to forecast PM_{2.5} and NO₂ varies depending on the location of the monitoring station. For PM_{2.5}, Isahaya, Kawadana, Matsuura Shisamachi, Tsushima, Iki, Obama, and Inasa ARIMA gave better results; ETS and Holt–Winters gave better results in the other monitoring stations. In the case of NO₂, ETS and Holt–Winters gave better results in Daitou, Isahaya, Shimabara, Oomura, Matsuura Shisamachi, Tsushima, Iki, Goto, Inasa, Fukuishi Jihai, and Yoshii with ARIMA providing better results in the other stations. The highest MAPE for Holt–Winters for PM_{2.5} was in Shimabara (23.566), and the lowest was in Higashi Nagasaki (11.98). For ETS, the highest MAPE was in Yoshii (26.59), the lowest was in Muramatsu (14.25), and

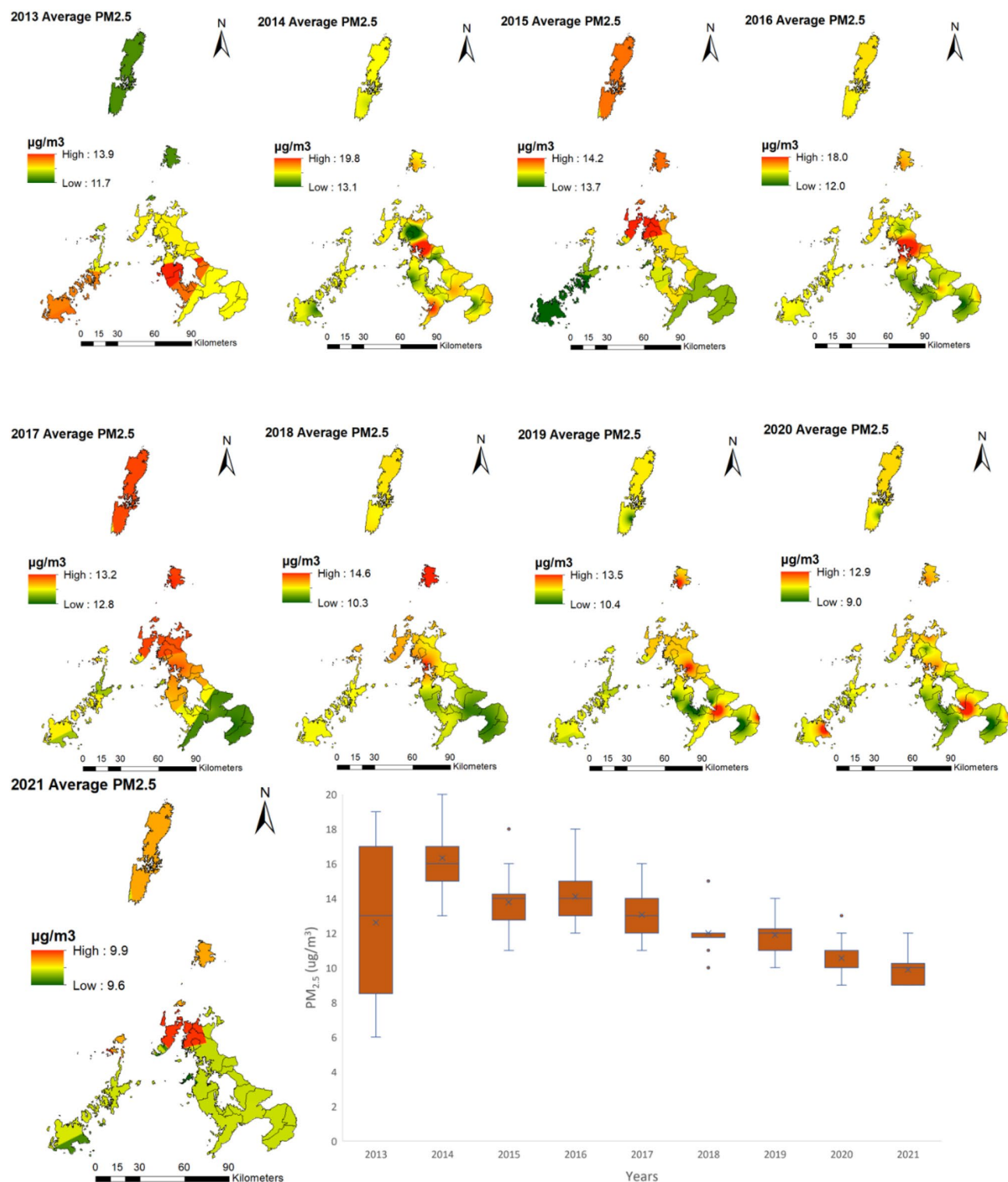


Figure 2. Spatiotemporal distribution of $PM_{2.5}$ from 2013 to 2021 in Nagasaki Prefecture, Japan. Created with ArcMap 10.7 (<https://www.arcgis.com/index.html>).

for ARIMA, the highest MAPE was in Kawadana (23.99), and the lowest was in Tsushima (12.63). For NO_2 , the highest RMSE for Holt-Winters was in Iki (43.77), the lowest was in Tsushima (10.18), and for ETS, the highest was in Oomura (15.84), and the lowest was in Fukuishi Jihai (7.83), and for ARIMA the highest was in Obama (32.75), and the lowest was in Kawadana (8.61).

Figure 6 shows the forecast results for $PM_{2.5}$ and NO_2 for six monitoring stations using the best model, ETS, Holt-Winters, or ARIMA (Table 5). $PM_{2.5}$ for Tsushima, Inasa, Yoshii and Obama was forecasted with ARIMA, while Goto and Daitou were forecasted with Holt winters. The forecast of $PM_{2.5}$ produced by ARIMA in Inasa

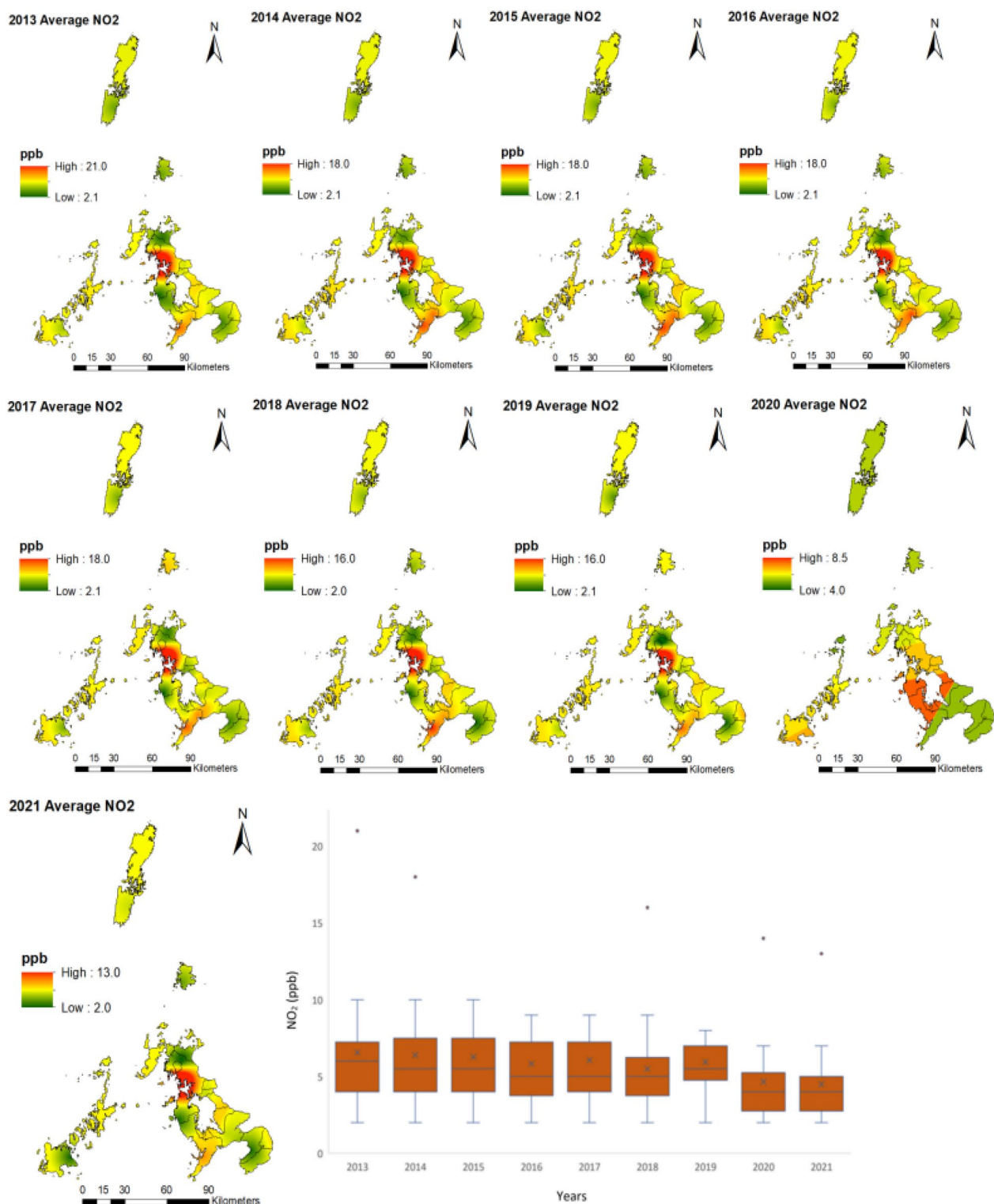


Figure 3. Spatiotemporal distribution of NO₂ from 2013 to 2021 in Nagasaki Prefecture, Japan. Created with ArcMap 10.7 (<https://www.arcgis.com/index.html>).

and Obama tends to converge to the mean. In Tsushima, the ARIMA model; in Goto, the Holt-Winters model; and in Daitou and Yoshii, the ETS model was able to replicate the trend and the seasonal components of the data for PM_{2.5}. For NO₂, Tsushima and Goto were forecasted with Holt-Winters, Obama was forecasted with ARIMA, and the other stations were forecasted with ETS. For NO₂ ETS, Holt-Winters and ARIMA were able to replicate the data's trend and seasonal components. However, the tendency of the data to converge towards the mean was not observed in the case of NO₂.

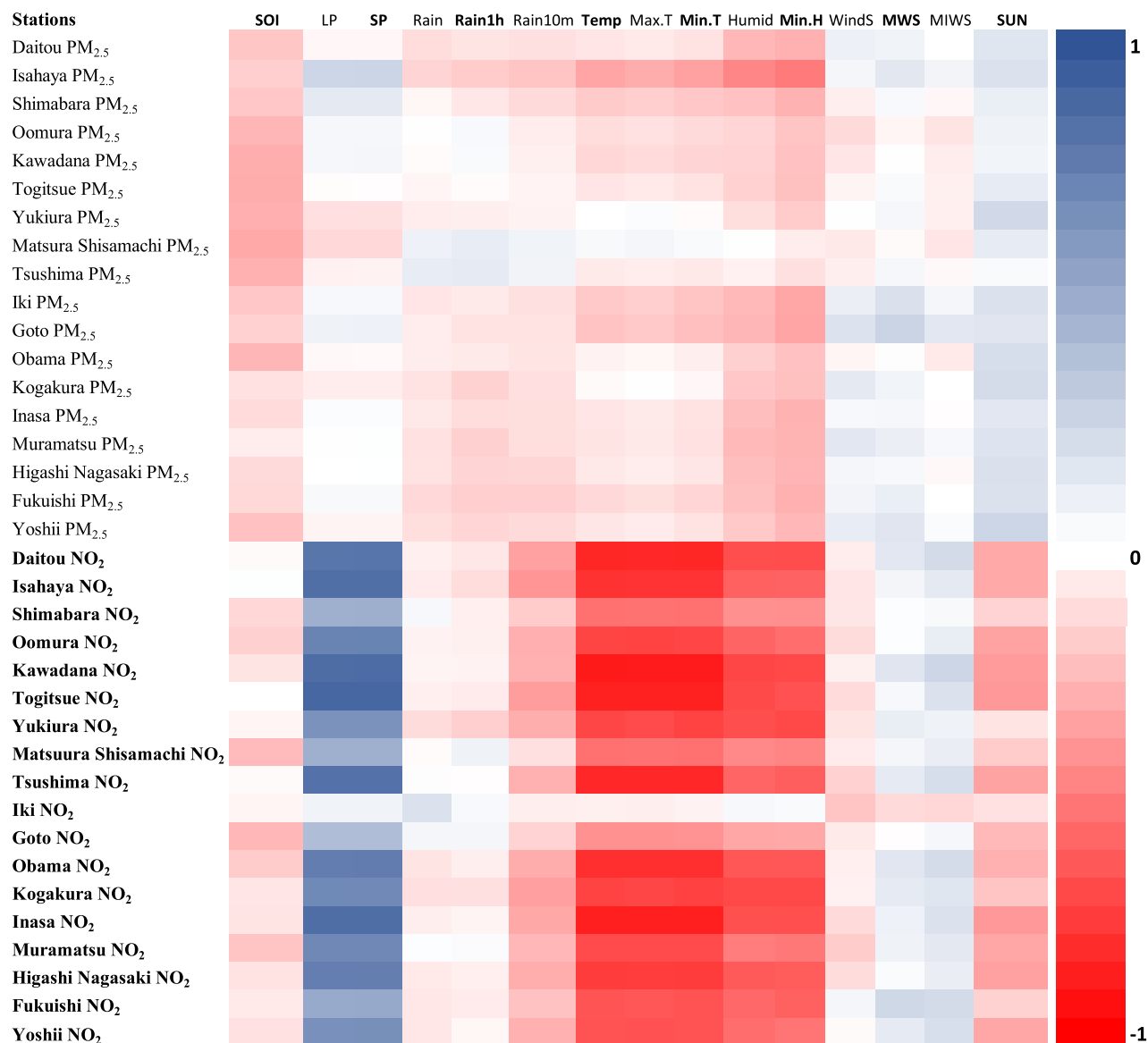


Figure 4. Heatmap represents the correlation between climatic factors and PM_{2.5} and NO₂ air pollution data for 18 monitoring stations in Nagasaki Prefecture, Japan. * SOI= Souther Oscillation Index, LP= Average local pressure, SP= Average sea level pressure, Rain= Daily Precipitation, Rain1h= Maximum 1-h precipitation, Rain10m= Maximum 10 min precipitation, Temp= Average Temperature, Max.T= Maximum Temperature, Min.T= Minimum Temperature, Humid= Average humidity, Min.H= Minimum humidity, WindS= Average wind Speed, MWS= Minimum wind speed, MIWS= Maximum instantaneous wind speed, SUN= Sunlight time.

For PM_{2.5}, in general, the future forecast indicates a negative trend. However, the future concentration of PM_{2.5} will remain above the 2021 WHO recommendations (5 µg/m³). Also, in most stations, the future concentrations will stay above or below the 2005 WHO recommendations (10 µg/m³), depending on the season, except for Daitou, which shows that the future concentrations will decline below the 2005 recommendations. For NO₂, the forecast shows a very slight declining trend for the majority of the stations. Compared to the other stations in Yoshii, the decline is more consistent. For NO₂, all the future forecasts are below the 2005 WHO recommendations (21 ppb). Also, for most stations, the NO₂ concentration will be below or above the 2021 WHO recommendations (5 ppb), depending on the season. Yoshii and Obama are exceptions, as NO₂ concentration levels are below the 2021 recommendations.

Discussion

Acknowledging that air pollution has adverse health effects, even at the lowest observed levels, is crucial for reconsidering current legislation and regulation. Thus, reducing the health impacts caused by the average annual exposure to NO₂ and PM_{2.5} needs to be prioritized to address known inequities owing to economic activities, socioeconomic conditions, and increased vulnerability of the residential population⁸. Although regulation and legislation in Japan have effectively reduced PM_{2.5} and NO₂ concentrations over the years, and it is considered

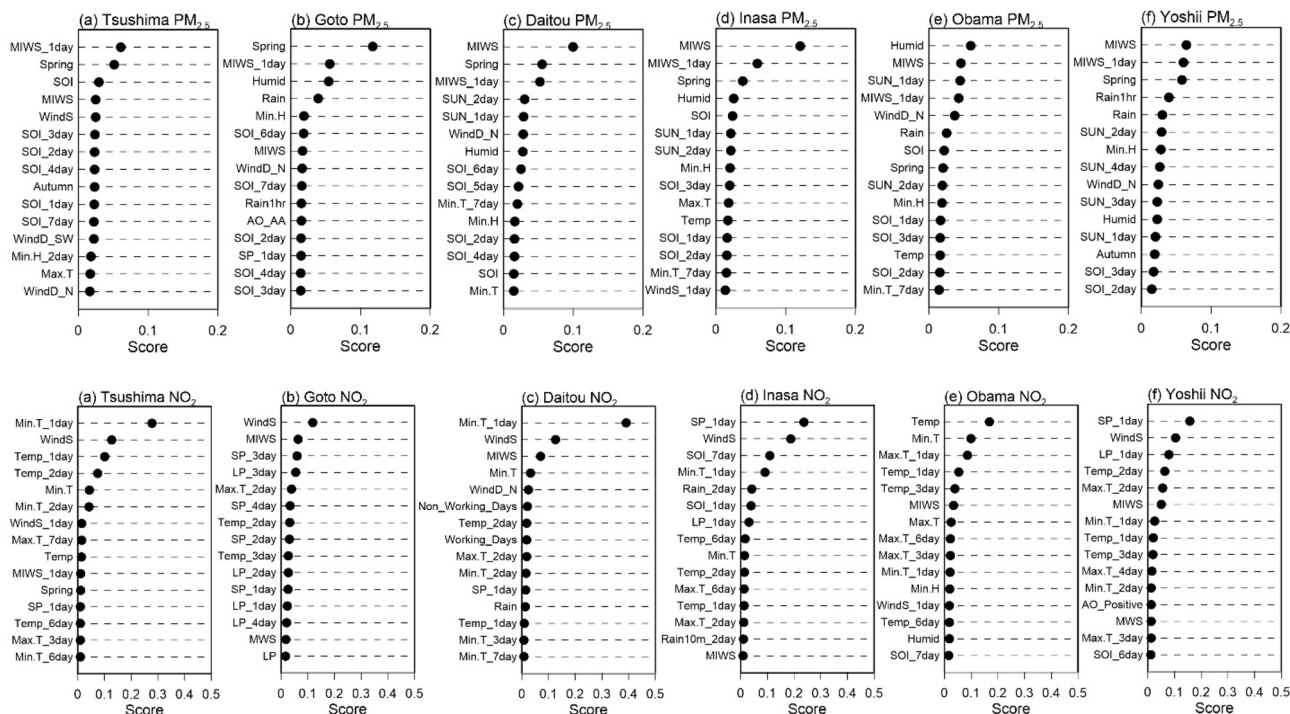


Figure 5. RF feature selection for PM_{2.5} and NO₂ of different monitoring stations in Nagasaki Prefecture, Japan.

Point (PM _{2.5})	Train R ²	Test R ²	Train MSE	Test MSE
Daitou	0.53	0.33	33.2	51.3
Isahaya	0.49	0.29	29.6	43.1
Shimabara	0.49	0.31	29.4	40.1
Oomura	0.50	0.28	24.5	32.7
Kawadana	0.50	0.27	23.6	33.0
Togitsue	0.49	0.22	23.5	35.9
Yukiura	0.47	0.18	24.0	35.1
Matsuura Shisamachi	0.47	0.25	28.8	37.1
Tsushima	0.43	0.16	35.2	61.5
Iki	0.43	0.18	37.6	61.2
Goto	0.41	0.16	32.7	51.1
Obama	0.47	0.27	22.7	29.4
Kogakura	0.47	0.24	29.7	39.2
Inasa	0.46	0.21	30.3	43.4
Muramatsu	0.48	0.23	24.6	36.1
Higashi Nagasaki	0.48	0.24	25.5	36.3
Fukuishi (Jihai)	0.47	0.23	36.0	53.1
Yoshii	0.46	0.23	26.7	32.1

Table 3. Result of the random forest model for PM_{2.5} using hyperparameters of the optimum value.

one of the industrialized countries with low levels of these pollutants, the results indicate that the average annual concentrations of PM_{2.5} and NO₂ exceeded the 2021 pollution concentration guidelines of WHO. In particular, PM_{2.5} annual average concentration exceeded the 2005 and 2021 pollution guidelines. The difficulty of regulating and reducing PM_{2.5} concentration in Nagasaki Prefecture is due to the long-range transport of PM_{2.5} from East Asia and Eurasia²⁸.

As a result of the long-range transport characteristic of PM_{2.5}, its spatial distribution and concentration vary throughout the study period as it is affected by climatic and temporal factors. For instance, Pearson's correlation and the random forest feature selection indicated that the most important factors influencing PM_{2.5} were Spring, maximum instantaneous wind speed, humidity, sunlight time one day, and southern oscillation index. In Spring, PM_{2.5} concentrations are higher than in other seasons (Table S2). This is due to the changes in meteorological

Point (NO ₂)	Train R ²	Test R ²	Train MSE	Test MSE
Daitou	0.709	0.564	4.865	7.979
Isahaya	0.803	0.735	2.458	3.393
Shimabara	0.725	0.542	1.411	2.181
Oomura	0.717	0.608	2.348	3.772
Kawadana	0.745	0.58	1.124	1.678
Togitsue	0.807	0.63	1.753	3.451
Yukiura	0.541	0.359	0.38	0.511
Matsuura Shisamachi	0.586	0.423	0.969	1.524
Tsushima	0.711	0.587	1.362	2.066
Iki	0.671	0.495	2.284	3.255
Goto	0.602	0.357	3.341	3.577
Obama	0.656	0.524	0.403	0.674
Kogakura	0.665	0.499	4.913	7.814
Inasa	0.51	0.354	9.396	8.8
Muramatsu	0.525	0.43	4.121	2.95
Higashi Nagasaki	0.735	0.654	3.156	4.635
Fukuishi(Jihai)	0.673	0.617	9.277	10.43
Yoshii	0.669	0.566	0.95	1.35

Table 4. Result of the random forest model for NO₂ using hyperparameters of the optimum value.

conditions, especially wind direction, which affects the long-range transport of PM_{2.5} from East Asia²⁰. Maximum instantaneous wind speed was negatively correlated in some stations, showing that horizontal dispersion plays a role in Nagasaki Prefecture. However, maximum wind speed was positively correlated, indicating that PM_{2.5} pollutants are being transported from other areas²⁹. This result is further reinforced by Fig. 2, which suggests that from 2014 through 2021, the highest concentration of PM_{2.5} are located in the westernmost part of Nagasaki Prefecture. Several studies have indicated that PM_{2.5} is transported to Nagasaki Prefecture from East Asia; thus, the proximity of the westernmost part of Nagasaki's Prefecture to East Asia, its downwind location, and the change of wind direction in Spring are the main reasons for high PM_{2.5} concentrations detected during the study period. The less affected areas are those located in the easternmost part, which is further away from East Asia. The wide distribution of PM_{2.5} and its spatial variability thought the study period makes it difficult to regulate and identify specific hotspots. Its wide distribution is also a cause for concern as it has health implications for many of the resident population in Nagasaki Prefecture. However, during the study period, as indicated by Table 2, Figs. 2, 3, and 5, there has been a decline in PM_{2.5} concentrations. This decline in PM_{2.5} in Nagasaki Prefecture is related to the decrease in PM_{2.5} concentrations in China and Korea. This reduction can be attributed to the changes in policy, technology, social, environmental, and economic factors in Japan, Korea, and China. For instance, the changes in environmental policies and the tri-national cooperation between these countries have generated positive results in reducing PM_{2.5} (see Section "Policy background" Table 1). Also, the restrictions on social and economic activities imposed due to the COVID-19 pandemic resulted in a notable reduction of PM_{2.5} in 2020 and 2021. Although PM_{2.5} shows a declining trend, better local and regional strategies are needed to reduce PM_{2.5} further as the pollution levels are above the WHO guidelines amidst local and tri-national efforts.

As for NO₂, the average annual concentration is below the 2005 pollution guidelines. However, the results indicated that the hotspots identified are above the WHO 2021 pollution concentration guidelines. NO₂ pollution concentrations are also influenced by climatic and temporal factors, as indicated by Pearson's correlation and random forest feature selection analysis. For NO₂, average wind speed was negatively correlated due to the dilution and dispersion of pollutants. However, maximum wind speed and maximum instantaneous windspeed were positively correlated, which can be attributed to the notion that the NO₂ plum is buoyant, but at higher wind speeds, the plum is brought down to ground level³⁰. Temperature was negatively correlated with NO₂; temperature is known to promote air convection, leading to pollution dispersion and dilution³¹. Average local pressure and average sea level pressure were positively correlated due to the low atmospheric boundary layer, which accompanies high pressure and prevents air pollutants' vertical dispersion²⁹. Sunlight time was negatively correlated to NO₂; this can be attributed to the photochemical reactions of solar radiation, which reduced NO₂ concentration. Amidst the influence of climatic factors on NO₂, its spatial distribution remained constant throughout the study period with consistent hotspot areas, except for 2020, where the pollution concentration was the lowest and more dispersed with no visible hotspot. From 2013 to 2019 and 2021, hotspots were located in Nagasaki's Prefecture major cities, Nagasaki, Sasebo, Isahaya, and Oomura. Nagasaki and Sasebo are the two largest cities in Nagasaki Prefecture with the highest concentrations of NO₂ throughout the study period; this is because of the economic activities in the area associated with shipbuilding, power plants, machinery, and heavy industries and also the burning of fossil fuels, especially from the transport sector. The lowest concentrations of NO₂ were in 2020–2021, as indicated by Table 2 and Fig. 3; this remarkable reduction of NO₂ can be attributed to the restrictions imposed by the Japanese government on social and economic activities due to the COVID-19 pandemic. The reduction of NO₂ in 2018–2019 could be due to the decommissioning of the Ainoura Power Station, a crude oil-fired power

Location	Seasonal Sen slope	Correlated seasonal Mann-Kendal	MAPE trend STL	ADF test	Models	MAPE	RMSE
Daitou_PM _{2.5}	-1.278	0.001	0.1321	0.01	ETS(M,A,M)	15.1101	1.8093
Isahaya_PM _{2.5}	-0.697	0.012	0.1820	0.01	ARIMA(1,1,1)(2,0,0)[12]	16.6838	2.2761
Shimabara_PM _{2.5}	-0.626	0.151	0.1413	0.01	Holt-Winters' additive method	23.5660	2.7902
Oomura_PM _{2.5}	-0.557	0.178	0.1515	0.04	ETS(M,A,A)	22.0190	2.4334
Kawadana_PM _{2.5}	-0.467	0.204	0.1516	0.03	ARIMA(1,0,1) with non-zero mean	23.9933	2.4873
Togitsue_PM _{2.5}	-0.581	0.081	0.1360	0.01	Holt-Winters' additive method	21.0916	2.3016
Yukiura_PM _{2.5}	-0.574	0.045	0.1334	0.01	ETS(M,Ad,A)	21.9601	2.2308
Matsuura Shisamachi_PM _{2.5}	-0.605	0.188	0.1602	0.02	ARIMA(2,0,0)(0,0,1)[12] with non-zero mean	18.1827	2.0515
Tsushima_PM _{2.5}	-0.581	0.168	0.1701	0.01	ARIMA(0,1,1)(2,1,0)[12]	12.6324	1.5047
Iki_PM _{2.5}	-0.718	0.005	0.1547	0.01	ARIMA(0,0,0)(2,1,0)[12] with drift	19.8552	2.5151
Goto_PM _{2.5}	-0.429	0.014	0.1470	0.01	Holt-Winters' additive method	18.1000	2.2244
Obama_PM _{2.5}	-0.542	0.064	0.1347	0.02	ARIMA(0,1,1)(1,0,0)[12]	17.3931	1.7036
Kogakura_PM _{2.5}	-1.014	0.001	0.1138	0.01	Holt-Winters' additive method	12.5448	1.7062
Inasa_PM _{2.5}	-0.824	0.003	0.1173	0.01	ARIMA(1,1,1)(1,0,0)[12]	15.9630	1.9276
Muramatsu_PM _{2.5}	-0.865	0.002	0.1227	0.01	ETS(M,Ad,A)	14.2451	1.6485
Higashi Nagasaki_PM _{2.5}	-0.833	0.003	0.1108	0.01	Holt-Winters' additive method	11.9808	1.5653
Fukuishi Jihai_PM _{2.5}	-1.178	0.002	0.1400	0.01	ETS(M,A,M)	14.8669	1.8090
Yoshii_PM _{2.5}	-0.559	0.003	0.1414	0.01	ETS(M,A,A)	26.5897	2.5581
Daitou_NO ₂	-0.414	0.003	0.2773	0.01	ETS(A,A,A)	8.6719	0.7435
Isahaya_NO ₂	-0.218	0.028	0.4277	0.01	Holt-Winters' additive method	34.2538	1.3374
Shimabara_NO ₂	-0.162	0.032	0.1629	0.01	Holt-Winters' additive method	30.5764	1.4037
Oomura_NO ₂	-0.262	0.043	0.2977	0.01	ETS(A,A,A)	15.8421	0.8721
Kawadana_NO ₂	-0.179	0.006	0.3297	0.01	ARIMA(1,0,0)(2,1,1)[12] with drift	8.6144	0.4449
Togitsue_NO ₂	-0.194	0.007	0.3879	0.01	ARIMA(0,0,0)(0,1,1)[12] with drift	13.2104	0.8717
Yukiura_NO ₂	-0.016	0.332	0.1968	0.01	ARIMA(0,0,1)(2,1,0)[12]	20.5383	0.4300
Matsuura Shisamachi_NO ₂	-0.123	0.026	0.2012	0.01	Holt-Winters' additive method	16.8278	0.5062
Tsushima_NO ₂	-0.032	0.394	0.3543	0.01	Holt-Winters' additive method	10.1772	0.4442
Iki_NO ₂	-0.035	0.633	0.2060	0.01	Holt-Winters' additive method	43.7718	1.3767
Goto_NO ₂	-0.250	0.034	0.2290	0.01	Holt-Winters' additive method	26.8487	0.6317
Obama_NO ₂	-0.085	0.060	0.2679	0.01	ARIMA(1,0,0)(2,1,0)[12]	32.7484	0.6606
Kogakura_NO ₂	-0.488	0.002	0.2506	0.01	ARIMA(1,0,0)(2,1,0)[12] with drift	21.6239	1.4213
Inasa_NO ₂	-0.247	0.004	0.2993	0.01	ETS(A,A,A)	14.3113	0.8219
Muramatsu_NO ₂	-0.145	0.296	0.2812	0.01	ARIMA(0,1,0)(0,1,2)[12]	21.0657	0.9726
Higashi Nagasaki_NO ₂	-0.501	0.003	0.2849	0.01	ARIMA(0,1,1)(0,1,1)[12]	10.7545	0.6807
Fukuishi Jihai_NO ₂	-0.776	0.002	0.1526	0.01	ETS(M,A,M)	7.8342	1.2311
Yoshii_NO ₂	-0.242	0.002	0.2727	0.01	ETS(A,A,A)	13.5204	0.3613

Table 5. PM_{2.5} and NO₂ Sen slope, Mann Kendal and Forecast of the monitoring stations in Nagasaki Prefecture, Japan.

plant. The more gradual decrease of NO₂ during the study period, as indicated by the trend and forecast analysis, can be attributed, among other factors, to the stricter vehicle emission regulations implemented²² and also the regulation of emissions from stationary sources such as fossil fuel powerplants, electric and industrial boilers.

Pearson's correlation and the random forest feature selection identified major factors influencing PM_{2.5} and NO₂ and provided a good indication of the complex relationship between the significant climatic and temporal factors and PM_{2.5} and NO₂ pollutants in Nagasaki Prefecture. However, the results indicate that the correlation and factors of importance that influence PM_{2.5} and NO₂ vary depending on the monitoring station. These differences observed in terms of correlation, factors of importance, trend, and model performance among the 18 stations can be attributed to the varying unique characteristics of climatic, environmental, social, and economic factors in each location, which affect PM_{2.5} and NO₂ concentrations. For instance, in the case of Goto, the major predictor of PM_{2.5} is Spring and humidity in Obama. This difference can be attributed to the location of these monitoring stations. Goto is located in the westernmost part of Nagasaki Prefecture, which is the area most affected by the long-range transport of PM_{2.5} from East Asia in Spring, as opposed to Obama, which is located in the easternmost part of Nagasaki Prefecture, which is the least affected by the seasonal changes. Although RF

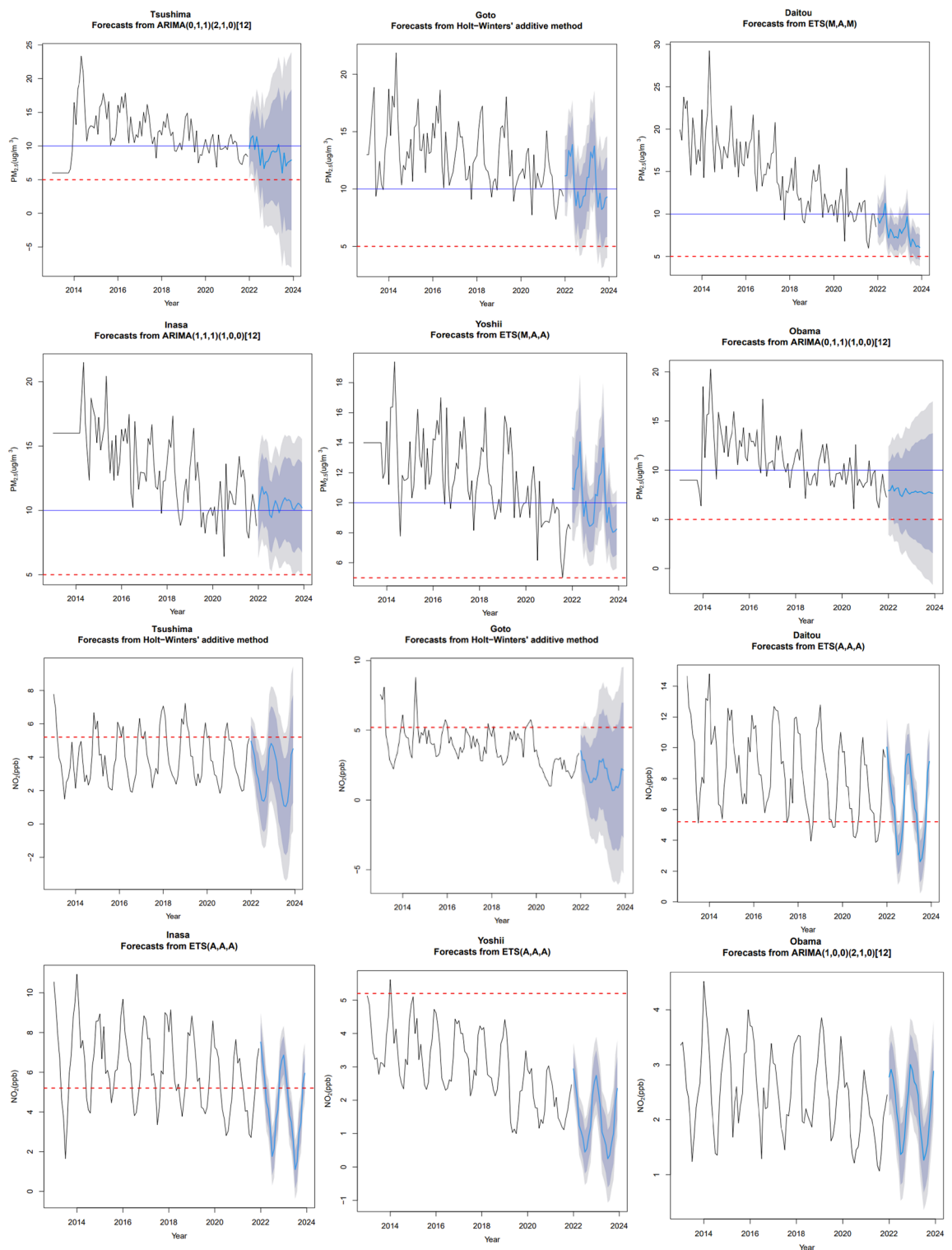


Figure 6. Models and Forecast of PM_{2.5} and NO₂ for different monitoring stations in Nagasaki Prefecture, Japan. Blue and red lines represent the WHO PM_{2.5} and NO₂ recommendations for 2005 and 2021, respectively.

was able to identify the major factors influencing PM_{2.5} and NO₂, the model's prediction of PM_{2.5} and NO₂ can be further improved by including not only climatic and temporal factors but emission sources and factors related to human activities such as economic development, transportation, and energy utilization³². And in the case of PM_{2.5}, including emission sources and human activity factors from China and Korea can improve the model's predictive capabilities. Therefore, even though this study has generated valuable information on the spatiotemporal distribution, trend, influencing factors, and forecast of PM_{2.5} and NO₂ in Nagasaki Prefecture, additional studies are needed to evaluate further the influence of social, environmental, economic, and technological factors

affecting the spatiotemporal distribution and trend of $PM_{2.5}$ and NO_2 in Nagasaki Prefecture. And also to assess the differences that exist (e.g., trend, influencing factors, etc.) among the monitoring stations.

Materials and methods

Study site. Nagasaki Prefecture is located on the island of Kyushu (Fig. 7). The prefecture has an area of approximately 4,105 km² with a population of 1,377,187. Nagasaki borders Saga Prefecture on the east and is surrounded by the Tsushima Straits, the Ariake Bay, and the East China Sea. Nagasaki air pollution is relatively low but is influenced by transboundary air pollution from Asia and Eurasia^{28,33,34}. Studies conducted in Nagasaki have demonstrated that air pollution has adverse health effects, especially in children^{14,16}. Moreover, 8.3 and 29.6% of the population in Nagasaki Prefecture are less than or equal to 10 and more than or equal to 65 years of age, respectively. Therefore, they are considered vulnerable to air pollutants³⁵. Although up until March 2012, Nagasaki Prefecture had no $PM_{2.5}$ monitoring stations, the first two stations to record $PM_{2.5}$ concentration were installed in Isahaya and Iki.

Air pollution datasets. The monitoring station network in Nagasaki Prefecture has increased from four monitoring stations in March 2012 to 18 monitoring stations. The recorded data of these eighteen stations are available from the Nagasaki Prefecture Atmospheric Environment Information (<http://www.pref.nagasaki.jp/>). $PM_{2.5}$ and NO_2 data is collected daily at one-hour intervals, from which monthly and annual averages are calculated. Monitoring stations are located at municipal offices, elementary schools, and towers. We selected the 2013–2021 $PM_{2.5}$ and NO_2 datasets because measurements of these pollutants were collected at each of the 18 monitoring stations for each year of the study period. For this study, we calculated the monthly mean concentration of $PM_{2.5}$ and NO_2 for each of the 18 monitoring stations. In Japan, NO_2 concentrations are given in parts per billion (ppb) as opposed to $PM_{2.5}$ measurements, which are given in micrograms per cubic meter ($\mu\text{g}/\text{m}^3$). Climatic data of Nagasaki from 2013–2021 were collected from the Japan Meteorological Agency website (Table 6).

Data processing and ordinary kriging. For the 18 stations, we calculated the annual average for $PM_{2.5}$ and NO_2 from the daily data collected in each monitoring station from 2013 to 2021. This resulted in nine datasets for $PM_{2.5}$ and nine for NO_2 , which were used to implement ordinary kriging to predict the spatiotemporal distribution. The dataset for each year was divided into four seasons, Spring: March to May, Summer: June to

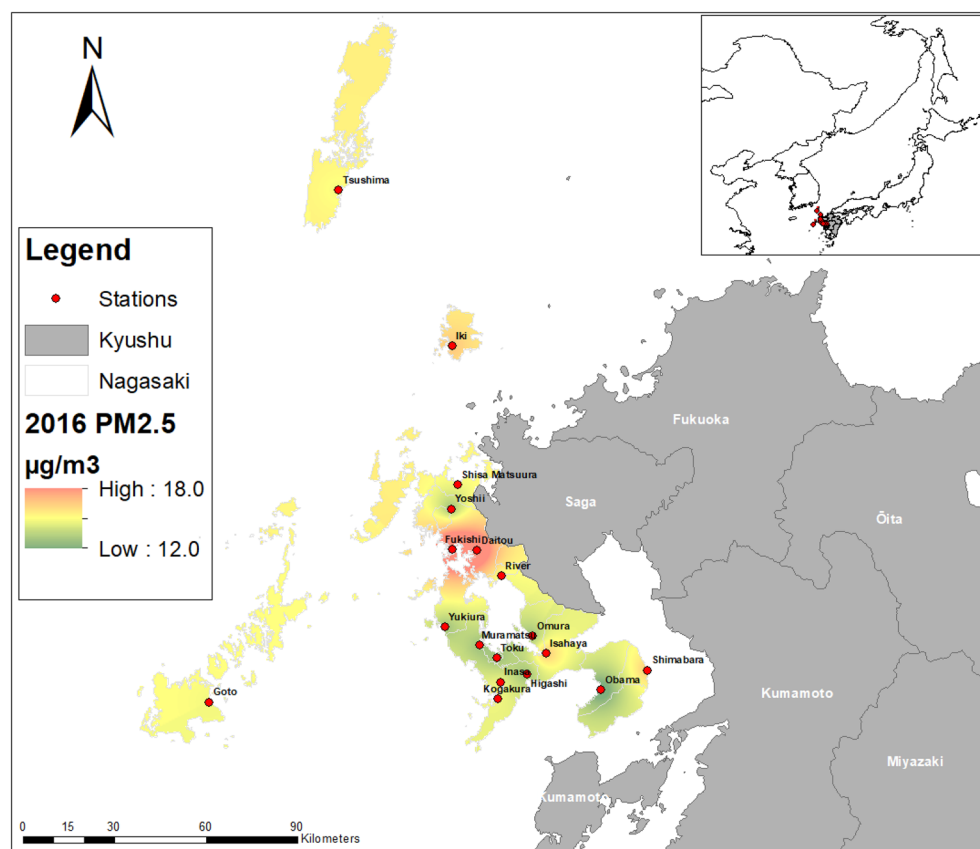


Figure 7. Study site and air pollution monitoring stations in Nagasaki Prefecture, Japan. Created with ArcMap 10.7 (<https://www.arcgis.com/index.html>).

Type	Parameter	Symbol	Unit (Daily)	Location
Air pollutant	PM _{2.5}	PM _{2.5}	µg/m ³	18 Point (Fig. 7)
	NO ₂	NO ₂	ppb	18 Point (Fig. 7)
Climatic factors	Average local pressure	LP	hPa	Data from
	Average sea level pressure	SP	hPa	Japan Meteorological
	Daily Precipitation	Rain	mm	Agency in Nagasaki ¹⁾
	Maximum 1-h Precipitation	Rain1h	mm	
	Maximum 10 min Precipitation	Rain10m	mm	
	Average temperature	Temp	°C	
	Maximum temperature	Max.T	°C	
	Minimum temperature	Min.T	°C	
	Average humidity	Humid	%	
	Minimum humidity	Min.H	%	
	Average wind speed	WindS	m/s	
	Maximum wind speed	MWS	m/s	
	Maximum instantaneous wind speed	MIWS	m/s	
	Sunlight time	SUN	h	
Temporal factors	Wind direction	WindD	8 Directions	Nagasaki ¹⁾
	Seasons	Season	4 Seasons	Japan
	Japan Day	Day	Holiday or not	Japan
	Arctic Oscillation index	AOI	6 Rank	NOAA CPC ²⁾
	Southern Oscillation Index	SOI	Positive or negative	Queensland ³⁾

Table 6. Parameters for the random forest model. ^awww.jma.go.jp. ^bwww.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml. ^cwww.longpaddock.qld.gov.au/soi. *Time lag data (1–7 days past each day) was created for the climatic variables.

August, Autumn: September to November, and Winter: December to February, and summary statistics were calculated (Table S2).

Ordinary kriging. Ordinary kriging (OK) interpolation is suitable for PM_{2.5} and NO₂ concentration mapping as it is a commonly used geostatistical estimator in air pollution interpolation and is often referred to as the unbiased estimator³⁶. Ordinary kriging models the unsampled value $z^*(x_0)$ as a combination of neighboring observations n , Eq. (1),³⁷:

$$z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (1)$$

where $z^*(x_0)$ estimate value at x_0 , $Z(x_i)$ measure value at x_i and λ_i weight is assigned for the residual of $Z(x_i)$.

Semivariogram. We derived the experimental semi-variogram for the 18 datasets to determine the spatial autocorrelation and the spatial structure of data points. The semi-variograms are expressed as a function of the distance between data points and explain the measured points' spatial relationship, Eq. (2),³⁸.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^N [Z(x_i + h) - Z(x_i)] \quad (2)$$

where $\gamma(h)$ quantity function of increment h , $N(h)$ numbers of pairs separated by the vector h , $Z(x_i)$ is the sampled values at location x_i and $Z(x_i + h)$ sampled measurements at location $X_i + h$.

In this study, we fitted the experimental semi-variogram to two theoretical semi-variogram models: exponential and spherical, two of the most commonly used models³⁹. The parameters determined were: range (a) the distance up until which the regionalized variable is auto-correlated, partial sill (c) which is the spatially structured part of the residuals, and the nugget (c_0) the non-spatial variability⁴⁰. The spherical and exponential models are defined by Eqs. (3) and (4)⁴¹.

$$\text{Exponential Model} \\ \gamma(h) = c_0 + c \left[1 - \exp\left(-\frac{3h}{a}\right) \right] \quad (3)$$

Spherical Model

$$\begin{cases} \gamma(h) = c_0 + c \left[1.5 \left(\frac{h}{a} \right) - 0.5 \left(\frac{h}{a} \right)^3 \right], & h \leq a \\ \gamma(h) = c_0 + c, & h > a \end{cases} \quad (4)$$

Cross validation. The model's prediction ability of the unsampled PM_{2.5} and NO₂ locations was conducted using cross-validations to calculate the mean error (ME), standard mean error (SME), root mean square error (RMSE), root mean square standard error (RMSSE) and average standard error (ASE). We analyzed the cross-validation results for both spherical and exponential models; the model with better results was selected for interpolating PM_{2.5} and NO₂ (Table S1). The RMSE and RMSSE are defined by Eqs. (5) and (6), respectively⁴².

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N [Z(x_i) - Z^*(x_i)]^2} \quad (5)$$

$$\text{RMSSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N [Z_1(x_i) - Z_2(x_i)]^2} \quad (6)$$

where N number of validation points, Z(x_i) measured value and Z*(x_i) standard values being Z₁(x_i) and Z₂(x_i).

A RMSE closer to 0 and a RMSSE closer to 1 depict that the parameters and fitting model are excellent and the kriging estimators are robust.

Pearson's correlation and random forest. Pearson's correlation analysis was conducted for each of the 18 monitoring stations to produce a heatmap depicting the correlation between major climatic and temporal factors and PM_{2.5} and NO₂ pollutants (Table 6). We then used random forest (RF) to identify the most important climatic and temporal factors influencing PM_{2.5} and NO₂ in each of the 18 stations⁴³. The factors identified were then used to construct the RF models for each of the 18 stations to make PM_{2.5} and NO₂ predictions. For each of the 18 stations, the random forest models were trained with 80% and validated with 20% of the respective monitoring station data. The RF model was then evaluated using the root mean square error (R²) and the mean, standard error (MSE). Random forest modeling is a type of ensemble learning method used for classification and regression analysis. It is well known to have advantages in terms of accuracy, robustness, and computational efficiency compared to other models⁴⁴. The RF model was constructed using the open-source machine learning library scikit-learn⁷² on Python. Next, the categorical factors were converted into dummy or indicator factors using the Python Pandas method (get dummies) (Tables 6 and 7). Furthermore, 7-day time lag data was added to the climatic factors to confirm the influence of past dependent factors. Finally, hyperparameters were determined with the ranges and steps indicated in (Table 8) using the grid-search technique for optimal model construction.

Rank	AOI
-AAA	< -2
-AA	-2 to -1
-A	-1-0
+A	0-1
+AA	1-2
+AAA	2 <

Table 7. Categorical rank from the continuous data of AOI.

Parameter	Range	Step
Max_depth	3-7	1
Min_sample_leaf	2-30	2
n_estimators	10-150	10
Mas_features	Auto	

Table 8. Hyperparameters of the random forest grid search.

Trend and forecast analysis. R statistical software was used to conduct the trend and forecast analysis (R Core Team, 2022). For the trend analysis, monthly mean concentrations for PM_{2.5} and NO₂ were utilized. The `cmk.test` function was used to conduct the Mann–Kendall test for trend detection. Equation (7) gives Mann–Kendall Statistics S , Variance $V(S)$, and standardized test statistics Z ^{45,46}.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn}(x_j - x_i), \quad (7)$$

$$\operatorname{sgn}(x_j - x_i) = \begin{cases} +1, & > (x_j - x_i) \\ 0, & = (x_j - x_i) \\ -1, & < (x_j - x_i) \end{cases}$$

$$V(S) = \frac{1}{18} \left[n(n-1)(2n+5) - \sum_{p=1}^q t_p(t_p-1)(2t_p+5) \right],$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{\operatorname{var}(s)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{\operatorname{var}(S)}} & \text{if } S < 0 \end{cases}$$

where x_j and x_i time series and n number of data points in the time series. Where t_p number of ties up to sample p . A positive Z value signifies a rising trend, a negative Z signifies a descending trend for the data period.

The `sens.slope` function was used to calculate the Sen's slope which indicated the magnitude of the trend. Equation (8) gives the slope for all data pairs and Eq. (9) the median of the n values of T_i , Sen's slope estimator (Q_i)⁴⁷.

$$T_i = \frac{x_j - x_k}{j - k} \quad (8)$$

where T_i slope and x_j and x_k data values at time j and k .

$$Q_i = \begin{cases} T_{\frac{n+1}{2}}, & n \text{ is odd} \\ \frac{1}{2} \left(T_{\frac{n}{2}} + T_{\frac{n+2}{2}} \right), & n \text{ is even} \end{cases} \quad (9)$$

A positive Q_i signifies a rising trend; a negative Q_i signifies a declining trend over time.

Both Mann–Kendall and Sen's slope consider the seasonality of the data. The `tsfresh` package in R was used to do the correlated seasonal Man-Kendall test and the seasonal Sen's slope tests⁴⁸. Both functions do not operate on missing data; therefore, the `tsfresh` function in the forecast package was used^{49,50}. To obtain the trend of PM_{2.5} and NO₂, we decompose the time series data into a trend, seasonal and irregular components by using the `stl` (seasonal decomposition of time series by LOESS) function developed by William Cleveland^{51,52}. The `stl` function from the `stats` package was used to fit the loess to the data and the `tsfresh` function in the forecast package was used to identify and replace outliers and missing values before applying the `stl` function. Then the `stl` function from the `stats` package was used to fit the loess to the data. The mean absolute percentage error (MAPE) and root mean square error was computed to determine if the component after the LOESS decomposition had satisfactorily captured the PM_{2.5} and NO₂ data information. The goodness of fit of the trend line was determined by checking the residuals; this was done by using the `checkresiduals` function from the forecast package.

Both exponential smoothing and ARIMA models were evaluated for the forecast analysis. These methods have been used to perform air pollution forecast analysis and, in some cases, have performed better than deep learning methods. First, the Augmented Dickey-Fuller test (ADF Test) was performed to ensure the stationarity of the time-series data⁵³. Once stationarity was confirmed, the two models were trained and tested with the 2013–2019 and the 2020–2021 pollutants datasets, respectively. Next, validation was performed using the test set whereby the mean absolute percentage error (MAPE) and root mean square error were computed to determine if the EST and ARIMA had satisfactorily captured the information of the PM_{2.5} and NO₂ data. The models with the lowest AIC were then used to do the forecasting of both PM_{2.5} and NO₂.

Exponential smoothing (ES) forecasting methods and models. Brown, Winter and Holt introduced the exponential smoothing⁵⁴. Gardner⁵⁵ extensively reviews the various ES methods. The exponential smoothing forecasting formulation consists of the forecast method and the statistical model. The forecast method uses an algorithm to produce a point forecast which is a prediction of a single value whereas the forecast statistical model is a process which generated an entire probability distribution with several values which when averaged generates a point forecast and provides prediction intervals with a level of confidence⁵⁴.

The exponential smoothing forecasting method is based on the idea that the forecast produced are weighted averages of past observations, with the weight associated to each observation exponentially decreasing as the

observation gets older^{54,55}. Model formulations are of component (recursive) form and error correction form⁵⁴. The error correction form is derived from the rearrangement of the equation in the component form. This error correction form uses the state space approach to exponential smoothing method since it consists of a measurement (observed) equation and a state (transition) equation. These two equations with their error distribution constitute a specified statistical model known as state space model. Since all observations and state variables use the same error process it is called "single source of error" (SSOE) or "innovation" and more specifically known as "innovation state space model. The single source of error (SSOE) was formulated by Snyder⁵⁶.

Pegels provided classification of the trend and the seasonal patterns depending on whether they are additive (linear) or multiplicative (nonlinear)⁵⁶. The family of exponential smoothing forecasting methods can be systematically described as a combination of level, trend, and seasonality^{54,58,59}. Each one can be of either an additive character or multiplicative character. The trend component can be classified as having no trend, additive trend, additive damped trend, multiplicative trend, and multiplicative damped trend^{55,57-59}. The simplest classification is the single exponential smoothing (SES) method, which considers only the constant level model and uses data with no trend or seasonality. This method consists of a forecast and smoothing equations for the level. The Holt linear trend method, also known as Double Exponential Smoothing (DES), consist of a forecast equation, and two smoothing equations: a level equation and a trend equation.

The Holt-Winters seasonal method, also known as the triple exponential smoothing (TES), consist of a forecast equation and three smoothing equations: a level equation, trend equation, and seasonality equation. The family classification of exponential smoothing generates a combination of 15 exponential smoothing methods with different components^{58,59}. Rearranging the terms in the different components for each of the 15 exponential smoothing methods (i.e., level component, trend component, and seasonal component), generate an error correction form model for each of the 15 methods with each having an additive or a multiplicative error model thus producing a total of 30 error models. These error-correction form models, also known as "innovative" state space models, are labeled as ETS (; ;), representing Error, Trend, and Seasonal. The forecast equation is the measurement equation, and the smoothing equations becomes the state equation, with both having the same source of error⁵⁴.

Of the 15 exponential smoothing methods, six were considered here. These are the Holts linear trend method, Holt linear damped trend, Holt-Winters additive, Holt-Winters additive damped, Holt-Winters multiplicative, and Holt-Winters multiplicative damped component. These six methods are converted to their error correction components form with their respective additive and multiplicative error correction model, yielding 12 error correction models. These 12 error correction models were used for model selection (Table S3).

The innovative state space model forecasting for each univariate time series was generated using the ets() function in the forecast package in R^{49,58}. Two procedures using ets() function for model selection were used: the automatic selection and the manual selection of a model. The automatic selection of the ETS models provides options for which models to be evaluated and selects the most appropriate model given the data. The model option used was model="ZAZ" where the first Z represents either additive or multiplicative error, the second Z represents automatic selection in which the choices are no seasonality, additive seasonality, or multiplicative seasonality, and the A represents an additive trend. Based on this "ZAZ" option, 12 models were evaluated from the 30 error models available. Selection of the best-fitted model from the 12 models was based on the minimization of the corrected Akaike Information Criterion (AIC), which avoids over-fitting by considering both goodness of fit and model complexity.

The manual model selection procedure used was selecting the hw() function from the forecast package in R. This function selects the Holt-Winters additive model, which corresponds to the ETS(A;A;A) in the ets () function which stands for additive error, additive trend, and additive seasonality⁵⁵. The Ljung-Box Q test was used for residual diagnostics to determine whether the residuals were white-noise sequences. The Box.test was used from the stata package in R.

The ARIMA models. Slutsky, Walker, Yaglom, and Yule first articulated autoregressive (AR) and moving average (MA) models. Box & Jenkins integrated the existing knowledge formulating ARIMA, known as the Box-Jenkins approach⁶⁰. An autoregressive model (AR) assumes the forecasted value is a linear combination of the past values of the variable, and moving average models (MA) assumes a linear combination of past forecasting errors. Combining these two models, AR and MA, produces an ARMA model. If the time series is non-stationary then the series are differenced to create stationarity before modeling, then the I is introduced in the ARMA. The I in an ARIMA model represents the integration parameter produced by differencing. The non-seasonal ARIMA models have parameters p, d, and q. The p represents the lag order of the autoregression, the d is the order of the differencing, and the q is the order of the moving average for the non-seasonal part. The seasonal ARIMA, also known as SARIMA, incorporates an additional set of terms, like the ARIMA models, that considers the seasonal effects. The seasonal parameters incorporated are P, D, Q, and m. The P, D, and Q represent the lag order of the autoregression, the order of the differencing, and the order of the moving average for the seasonal part, and the m represents the number of periods in each season. Box et al. and Chatfield^{61,62} expressed the AR(p), MA(p), and ARMA (p, q), mixed seasonal ARMA(p,q)(P,Q)_m, ARIMA(p,d,q), and mixed seasonal ARIMA -SARIMA(p,d,q)(P,D,Q)_m models.

ARIMA forecast was done using the automatic ARIMA algorithm for model selection using the auto.arima() function from the forecast package in the R program^{49,50}. Two procedures were used in the automatic selection. The first method used the default settings (restricted models), and the second was full model selection. Sometimes running the full model selection will produce a different optimal model⁵⁴.

Data availability

The data supporting this research's findings is provided in the supplementary materials. The raw data used for the study is also available from the corresponding author upon reasonable request.

Received: 1 September 2022; Accepted: 10 January 2023

Published online: 16 January 2023

References

- Xing, Y. F., Xu, Y. H., Shi, M. H. & Lian, Y. X. The impact of PM_{2.5} on the human respiratory system. *J. Thorac. Dis.* **8**, E69–E74. <https://doi.org/10.3978/j.issn.2072-1439.2016.01.19> (2016).
- Fann, N. & Risley, D. The public health context for PM_{2.5} and ozone air quality trends. *Air Qual. Atmos. Health* **6**, 1–11. <https://doi.org/10.1007/s11869-010-0125-0> (2013).
- Liu, Y., Wu, J. & Yu, D. Characterizing spatiotemporal patterns of air pollution in China: A multiscale landscape approach. *Ecol. Ind.* **76**, 344–356. <https://doi.org/10.1016/j.ecolind.2017.01.027> (2017).
- Faustini, A., Rapp, R. & Forastiere, F. Nitrogen dioxide and mortality: Review and meta-analysis of long-term studies. *Eur. Respir. J.* **44**(3), 744–753 (2014).
- Bran, S. H. & Srivastava, R. Investigation of PM_{2.5} mass concentration over India using a regional climate model. *Environ. Pollut.* **224**, 484–493. <https://doi.org/10.1016/j.envpol.2017.02.030> (2017).
- Gautam, S., Yadav, A., Tsai, C.-J. & Kumar, P. A review on recent progress in observations, sources, classification and regulations of PM_{2.5} in Asian environments. *Environ. Sci. Pollut. Res.* **23**, 21165–21175. <https://doi.org/10.1007/s11356-016-7515-2> (2016).
- Li, T.-C. *et al.* Clustered long-range transport routes and potential sources of PM_{2.5} and their chemical characteristics around the Taiwan Strait. *Atmos. Environ.* **148**, 152–166. <https://doi.org/10.1016/j.atmosenv.2016.10.010> (2017).
- Hoffmann, B. *et al.* WHO Air Quality Guidelines 2021-Aiming for healthier air for all: A joint statement by medical, public health, scientific societies and patient representative organisations. *Int. J. Public Health* **66**, 1604465. <https://doi.org/10.3389/ijph.2021.1604465> (2021).
- Araki, S. Application of regression kriging to air pollutant concentrations in Japan with high spatial resolution. *Aerosol Air Qual. Res.* <https://doi.org/10.4209/aaqr.2014.01.0011> (2015).
- Jassen, S., Dumont, G., Fierens, F. & Mensink, C. Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmos. Environ.* **42**, 4903 (2009).
- Lang, P. E., Carslaw, D. C. & Moller, S. J. A trend analysis approach for air quality network data. *Atmos. Environ.* <https://doi.org/10.1016/j.aeoa.2019.100030> (2019).
- Kashima, S., Yorifuji, T., Tsuda, T. & Doi, H. Application of land use regression to regulatory air quality data in Japan. *Sci. Total Environ.* **407**, 3055–3062. <https://doi.org/10.1016/j.scitotenv.2008.12.038> (2009).
- Shimadera, H., Kojima, T. & Kondo, A. Evaluation of air quality model performance for simulating long-range transport and local pollution of PM_{2.5} in Japan. *Adv. Meteorol.* <https://doi.org/10.1155/2016/5694251> (2016).
- Kim, Y. *et al.* Respiratory function declines in children with asthma associated with chemical species of fine particulate matter (PM_{2.5}) in Nagasaki Japan. *Environ. Health* **20**, 110. <https://doi.org/10.1186/s12940-021-00796-x> (2021).
- Ng, C. F. S. *et al.* Associations of chemical composition and sources of PM_{2.5} with lung function of severe asthmatic adults in a low air pollution environment of urban Nagasaki, Japan. *Environ. Pollut.* **252**, 599–606. <https://doi.org/10.1016/j.envpol.2019.05.117> (2019).
- Nakamura, T. *et al.* Association between Asian dust exposure and respiratory function in children with bronchial asthma in Nagasaki Prefecture, Japan. *Environ. Health Prev. Med.* **25**, 8. <https://doi.org/10.1186/s12199-020-00846-9> (2020).
- Chandra, I. *et al.* New particle formation under the influence of the long-range transport of air pollutants in East Asia. *Atmos. Environ.* **141**, 30–40. <https://doi.org/10.1016/j.atmosenv.2016.06.040> (2016).
- Irei, S., Takami, A., Hara, K. & Hayashi, M. Evaluation of transboundary secondary organic aerosol in the urban air of western Japan: Direct comparison of two site observations. *ACS Earth Space Chem.* **2**, 1231–1239. <https://doi.org/10.1021/acsearthspacechem.8b00106> (2018).
- Kubo, T., Bai, W., Nagae, M. & Takao, Y. Seasonal fluctuation of polycyclic aromatic hydrocarbons and aerosol genotoxicity in long-range transported air mass observed at the western end of Japan. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph17041210> (2020).
- Wang, J. & Ogawa, S. Effects of meteorological conditions on PM_{2.5} concentrations in Nagasaki, Japan. *Int. J. Environ. Res. Public Health* **12**, 9089–9101. <https://doi.org/10.3390/ijerph120809089> (2015).
- Japan, Tripartite Policy Dialogue on Air Pollution: Air Quality Policy Report. The cooperation progress and outcomes (2019).
- Ito, A., Wakamatsu, S., Morikawa, T. & Kobayashi, S. 30 years of air quality trends in Japan. *Atmosphere* <https://doi.org/10.3390/atmos12081072> (2021).
- Yamagami, M. *et al.* Trends in PM_{2.5} concentration in Nagoya, Japan, from 2003 to 2018 and Impacts of PM_{2.5} countermeasures. *Atmosphere* <https://doi.org/10.3390/atmos12050590> (2021).
- Zhai, S. *et al.* Fine particulate matter (PM_{2.5}) trends in China, 2013–2018: Separating contributions from anthropogenic emissions and meteorology. *Atmos. Chem. Phys.* **19**, 11031–11041. <https://doi.org/10.5194/acp-19-11031-2019> (2019).
- Liu, H., Yue, F. & Xie, Z. Quantify the role of anthropogenic emission and meteorology on air pollution using machine learning approach: A case study of PM_{2.5} during the COVID-19 outbreak in Hubei Province, China. *Environ. Pollut.* **300**, 118932. <https://doi.org/10.1016/j.envpol.2022.118932> (2022).
- Kim, Y., Yi, S. M. & Heo, J. Fifteen-year trends in carbon species and PM_{2.5} in Seoul, South Korea (2003–2017). *Chemosphere* **261**, 127750. <https://doi.org/10.1016/j.chemosphere.2020.127750> (2020).
- Uhm, J.-H. *et al.* Status of ambient PM_{2.5} pollution in the Seoul Megacity (2020). *Asian J. Atmos. Environ.* **15**, 95–106. <https://doi.org/10.5572/ajae.2021.022> (2021).
- Ikeda, K. & Tanimoto, H. Exceedances of air quality standard level of PM_{2.5} in Japan caused by Siberian wildfires. *Environ. Res. Lett.* **10**, 105001. <https://doi.org/10.1088/1748-9326/10/10/105001> (2015).
- Wu, J., Zhu, J., Li, W., Xu, D. & Liu, J. Estimation of the PM_{2.5} health effects in China during 2000–2011. *Environ. Sci. Pollut. Res. Int.* **24**, 10695–10707. <https://doi.org/10.1007/s11356-017-8673-6> (2017).
- Jones, A. M., Harrison, R. M. & Baker, J. The wind speed dependence of the concentrations of airborne particulate matter and NO_x. *Atmos. Environ.* **44**, 1682–1690. <https://doi.org/10.1016/j.atmosenv.2010.01.007> (2010).
- Luo, J. *et al.* Spatiotemporal pattern of PM_{2.5} concentrations in mainland China and analysis of its influencing factors using geographically weighted regression. *Sci. Rep.* **7**, 40607. <https://doi.org/10.1038/srep40607> (2017).
- Sun, R., Zhou, Y., Wu, J. & Gong, Z. Influencing factors of PM_{2.5} pollution: Disaster points of meteorological factors. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph16203891> (2019).
- Kaneyasu, N. *et al.* Impact of long-range transport of aerosols on the PM_{2.5} composition at a major metropolitan area in the northern Kyushu area of Japan. *Atmos. Environ.* **97**, 416–425. <https://doi.org/10.1016/j.atmosenv.2014.01.029> (2014).

34. Aikawa, M. *et al.* Field survey of trans-boundary air pollution with high time resolution at coastal sites on the sea of Japan during winter in Japan. *Environ. Monit. Assess.* **122**, 61–79. <https://doi.org/10.1007/s10661-005-9165-6> (2006).
35. Nathaniel, M. Who is at risk?. *Environ. Health Perspect.* **119**, 177 (2011).
36. Jerrett, M. *et al.* A review and evaluation of intraurban air pollution exposure models. *J. Expo. Anal. Environ. Epidemiol.* **15**, 185–204. <https://doi.org/10.1038/sj.jea.7500388> (2005).
37. Saito, H., McKenna, S. A., Zimmerman, D. A. & Coburn, T. C. Geostatistical interpolation of object counts collected from multiple strip transects: Ordinary kriging versus finite domain kriging. *Stoch. Env. Res. Risk Assess.* **19**, 71–85. <https://doi.org/10.1007/s00477-004-0207-3> (2005).
38. Yang, Y., Zhu, J., Tong, X. & Wang, D. IFIP International Federation for Information Processing, Volume 293. Computer and Computing Technologies in Agriculture II, Volume 1 (eds. Li, D. & Chunjiang, Z.) 125–134 (Springer, Boston, 2009).
39. Kholghi, M. & Hosseini, S. M. Comparison of groundwater level estimation using neuro-fuzzy and ordinary kriging. *Environ. Model. Assess.* **14**, 729. <https://doi.org/10.1007/s10666-008-9174-2> (2008).
40. Adhikary, P. P., Chandrasekharan, H., Chakraborty, D. & Kamble, K. Assessment of groundwater pollution in West Delhi, India using geostatistical approach. *Environ. Monit. Assess.* **167**, 599–615. <https://doi.org/10.1007/s10661-009-1076-5> (2010).
41. Jang, C.-S., Chen, S.-K. & Cheng, Y.-T. Spatial estimation of the thickness of low permeability topsoil materials by using a combined ordinary-indicator kriging approach with multiple thresholds. *Eng. Geol.* **207**, 56–65. <https://doi.org/10.1016/j.enggeo.2016.04.008> (2016).
42. Chabala, L. M., Mulolwa, A. & Lungu, O. Application of ordinary kriging in mapping soil organic carbon in Zambia. *Pedosphere* **27**, 338–343. [https://doi.org/10.1016/s1002-0160\(17\)60321-7](https://doi.org/10.1016/s1002-0160(17)60321-7) (2017).
43. Kaminska, J. A. The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: A case study in Wrocław. *J. Environ. Manag.* **217**, 164–174. <https://doi.org/10.1016/j.jenvman.2018.03.094> (2018).
44. Yu, R., Yang, Y., Yang, L., Han, G. & Move, O. A. RAQ-A random forest approach for predicting air quality in urban sensing systems. *Sensors* <https://doi.org/10.3390/s16010086> (2016).
45. Kendall, M. *Rank Correlation Methods* 4th edn. (Charles Griffin, 1975).
46. Mann, H. Nonparametric tests against trend. *Econometrica* **13**(3), 245–259 (1945).
47. Sen, P. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **63**(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934> (1968).
48. Pohlert, T. Trend: Non-parametric trend tests and change-point detection. R package version 1.1.4. <https://CRAN.R-project.org/package=trend> (2020).
49. Hyndman, R. J. & Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **26**(3), 1–22. <https://doi.org/10.18637/jss.v027.i03> (2008).
50. Hyndman, R. *et al.* Forecast: Forecasting functions for time series and linear models. R package version 8.16. <https://pkg.robjhyndman.com/forecast/> (2022).
51. Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. J. STL: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **6**(1), 3–33 (1990).
52. Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **74**(368), 829–836 (1979).
53. Trapletti, A. & Hornik, K. tseries: Time series analysis and computational finance. R package version 0.10-48 (2020).
54. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: Principles and Practice* 2nd edn. (OTexts, 2018).
55. Gardner, E. S. Exponential smoothing: The state of the art. *J. Forecast.* **4**(1), 1–28 (1985).
56. Snyder, R. D. Recursive estimation of dynamic linear models. *J. R. Stat. Soc. Ser. B (Methodological)* **47**(2), 272–276 (1985).
57. Pegels, C. C. Exponential forecasting: Some new variations. *Manag. Sci.* **15**(5), 311–315 (1969).
58. Hyndman, R. J., Koehler, A. B., Snyder, R. D. & Grose, S. A state space framework for automatic forecasting using exponential smoothing methods. *Int. J. Forecast.* **18**(3), 439–454 (2002).
59. Taylor, J. W. Exponential smoothing with a damped multiplicative trend. *Int. J. Forecast.* **19**(4), 715–725 (2003).
60. De Gooijer, J. G. & Hyndman, R. J. 25 years of time series forecasting. *Int. J. Forecast.* **22**(3), 443–473. <https://doi.org/10.1016/j.ijforecast.2006.01.001> (2006).
61. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. *Time Series Analysis: Forecasting and Control* 5th edn. (John Wiley & Sons, 2016).
62. Chatfield, C. *Time Series Forecasting* (Chapman and Hall, 2000).

Acknowledgements

The authors want to thank Nagasaki Prefecture for making the air pollution data available, which was indispensable for this research.

Author contributions

S.K. and S.D.C. conceptualize and design the research. All the authors contributed to data collection, analyzing, writing, and editing of the manuscript. Finally, all authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-27936-2>.

Correspondence and requests for materials should be addressed to S.D.C. or S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023