



## Application of Regression Kriging to Air Pollutant Concentrations in Japan with High Spatial Resolution

Shin Araki<sup>1,2\*</sup>, Kouhei Yamamoto<sup>3</sup>, Akira Kondo<sup>2</sup>

<sup>1</sup> Otsu City Public Health Center, Goryocho 3-1, Otsu, Shiga 520-8575, Japan

<sup>2</sup> Graduate School of Engineering, Osaka University, Yamadaoka 2-1, Suita, Osaka 565-0871, Japan

<sup>3</sup> Graduate School of Energy Science, Kyoto University, Yoshidahonmachi, Sakyo, Kyoto 606-8501, Japan

---

### ABSTRACT

The application of regression kriging to air pollutants in Japan was examined for the purpose of providing a practical method to obtain a spatial distribution with sufficient accuracy and a high spatial resolution of  $1 \times 1$  km. We used regulatory air monitoring data from the years 2009 and 2010. Predictor variables at  $1 \times 1$  km resolution were prepared from various datasets to perform regression kriging. The prediction performance was assessed by indicators, including root mean squared error (RMSE) and  $R^2$ , calculated from the leave-one-out cross validation results, and was compared to the results obtained from a linear regression method, often referred to as land use regression (LUR). Regression kriging well-explained the spatial variability of  $\text{NO}_2$ , with  $R^2$  values of 0.77 and 0.78. Ozone ( $\text{O}_3$ ) was moderately explained, with  $R^2$  values of 0.52 and 0.66. The reason for this difference in performance between  $\text{NO}_2$  and  $\text{O}_3$  might be the characteristics of these pollutants - primary or secondary. Regression kriging outperformed the linear regression method with regard to RMSE and  $R^2$ . The performance of regression kriging in this work was comparable to that found in previous studies. The results indicate that regression kriging is a practical procedure that can be applied for the prediction of the spatial distribution of air pollutants in Japan, with sufficient accuracy and a high spatial resolution.

**Keywords:** Spatial distribution; Geostatistics; Air quality; Ozone;  $\text{NO}_2$ .

---

### INTRODUCTION

Spatial distribution of air pollutants with sufficient accuracy and high spatial resolution is the key piece of information for assessing the risks to human health or evaluating the air quality policy quantitatively. A regression method has been used successfully for predicting the spatial distribution of the long-term mean concentrations of air pollutants mainly for epidemiological studies (e.g., Briggs *et al.*, 2000; Ross *et al.*, 2007). This approach, often referred to as land use regression (LUR), uses measured data of air pollutants and develops linear regression models using predictor variables that may influence the concentrations. The regression model obtained is then applied to unmonitored locations in the domain in order to predict the concentrations. In the case of  $\text{NO}_2$ , for instance, predictor variables such as land use, traffic intensity and population are used to construct a regression model (e.g., Briggs *et al.*, 2000; Hoek *et al.*, 2008). This technique is

significantly less complicated in terms of computation compared to widely-used chemical transport models which simulate physical and chemical processes including emission, transportation, transformation and deposition. The predictions of the regression method have been reported to correlate better with monitored concentrations than chemical transportation models for long-term statistics such as annual mean (Cyrus *et al.*, 2005; Marshall *et al.*, 2008). On the other hand, application of the regression method to finer temporal resolution (e.g., hourly mean) is difficult because the predictor variables at such high temporal resolution are usually unavailable.

A few studies have applied regression kriging for the prediction of the spatial distribution of air pollutants (e.g., Beelen *et al.*, 2009; Pearce *et al.*, 2009). Regression kriging is a geostatistical approach - ordinary kriging with the residuals of regression analysis. In this method, a variable of interest is modelled as a sum of the deterministic and stochastic components separately (Hengl, 2007). The deterministic component is estimated by a regression analysis, which is equivalent to the regression method described above. The residuals of the regression model are interpolated using ordinary kriging and added to the estimated deterministic part (Hengl, 2007). There are other terms such as universal kriging and kriging with external

---

\* Corresponding author.

Fax: 81-77-522-1097

E-mail address: araki@ea.see.eng.osaka-u.ac.jp

drift (KED), but these methods including regression kriging are basically the same technique (Hengl *et al.*, 2003). Some studies compared regression kriging with the regression method and reported that the prediction performance of regression kriging was better (Beelen *et al.*, 2009; Pearce *et al.*, 2009; Mercer *et al.*, 2011). This method has been successfully applied at the EU-wide scale at  $1 \times 1$  km resolution for the prediction of annual means of air pollutant concentrations (Beelen *et al.*, 2009), which would be computationally demanding for chemical transportation models. Taking account of these facts, regression kriging is a reasonable and practical method for the prediction of the spatial distribution of air pollutants for long-term statistics.

In Japan, air pollution is an unsolved problem and an issue of social concern: thus, the spatial distribution of air pollutants is needed. As regression kriging includes an empirical model in contrast to the chemical transportation models, limitations in transferability may be inevitable to some extent (Hoek *et al.*, 2008). To our knowledge, however, no study has applied regression kriging to predict the spatial distribution of air pollutants in Japan or even in Asian countries to date.

In this study, we examined the possibility of applying regression kriging to air pollutants in Japan in order to provide a practical method to obtain a spatial distribution with sufficient accuracy and high spatial resolution. We validate the performance of regression kriging and discuss the possibility of improvement.

## METHODS

### Study Area

Japan consists of four major islands stretching from north to south and has an area of 377914 km<sup>2</sup> and a population of 127 million. The study area included the main islands of Japan (129.1–145.8°E, 31.0–45.5°N) but remote or small islands were excluded. We constructed a prediction grid at  $1 \times 1$  km resolution only over land areas. The concentrations of air pollutants were predicted at each cell of the grid.

### Monitored Air Quality Data

The air quality data for the year 2009 and 2010 was obtained from the database of the regulatory monitoring network, which is maintained by the local government according the national law. The monitored data is collected and summarized by the Japanese Ministry of Environment. There are two types of monitoring station: road side and general environment. The road side stations are located at crossroads or road sides and monitor air pollutants from automobile traffic, whereas the general environment stations are located where they are not affected by specific emission sources. We used only the general environment station data, because of the difficulty to model the small spatial variation near the road side with our prediction grid at  $1 \times 1$  km resolution. The number of general environment stations in Japan in the year 2010 was reported to be 1503 according to the Japanese Ministry of Environment.

The objective pollutants were NO<sub>2</sub> and O<sub>3</sub>. Regarding NO<sub>2</sub>, we predicted the spatial distribution of the annual 98

percentile daily mean concentrations, which is the air quality standard for NO<sub>2</sub> in Japan. We used only the data from monitoring stations with annual data coverage of more than 6000 hours, because conformity to the air quality standards is evaluated only for those stations. As for O<sub>3</sub>, the air quality standard in Japan is set as an hourly mean. Some variables including meteorological parameters with the temporal resolution of one hour are necessary for the application of regression kriging to hourly mean concentrations of O<sub>3</sub>. Because the meteorological data with such high temporal resolution which covers the main islands of Japan is not publicly available, a meteorological model is required to obtain those variables. This would be computationally demanding and reduce the efficiency that is one of the advantages of regression kriging. Besides, it would be rational to apply a chemical transport model instead of regression kriging because that is designed to be driven by a meteorological model. For these reasons, instead of hourly mean values, we predicted the annual 26th highest daily maximum 8 hour mean concentrations which is the air quality standard in the EU. Also, we can compare our result to the previous studies (De Smet *et al.*, 2010, 2011) that applied regression kriging to the EU standard. The daily maximum 8 hour mean concentrations were calculated only for the day when the valid operation time was more than 20 hours, and the stations with more than 250 daily maximum 8 hour means available for a year were provided for the analysis. The annual statistics for NO<sub>2</sub> and O<sub>3</sub> were calculated for a fiscal year in accordance with the report of the Japanese Ministry of Environment.

### Datasets

The datasets used to construct the  $1 \times 1$  km resolution predictor variables are presented in Table 1 and described in detail below. The selection of datasets was made principally in consideration of the key factors in the spatial distribution of air pollutants including emission, transportation, transformation and deposition. The accessibility and usability were also considered. If necessary, we spatially aggregated, disaggregated or resampled the original data into a  $1 \times 1$  km prediction grid and/or calculated the annual means for a fiscal year from the data with finer temporal resolution (e.g., monthly).

Vegetation is one of the main emission sources of volatile organic compounds (VOCs), which are the precursors of O<sub>3</sub>. Normalized Difference Vegetation Index (NDVI) indicates the amount of vegetation. Annual mean NDVI were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) Terra data (MOD13A3) downloaded through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), and USGS/Earth Resource Observation and Science (EROS) Center, and provided for a predictor variable representing the emission of VOCs.

For one of the predictor variables for O<sub>3</sub>, we used the amount of UV which promotes O<sub>3</sub> formation through a photochemical reaction. We obtained half monthly mean UV-A (wave length: 320–400 nm) from Japan Aerospace Exploration Agency (JAXA) Satellite Measurements for

**Table 1.** Summary of the data used in this study.

Description	Source	Field	Spatial scale	Time period
Measured air quality data	Ministry of Environment	NO <sub>2</sub> , O <sub>3</sub>	point	2009–2010
MODIS Land Product	NASA/USGS	NDVI	1 km	2009–2010
JASMES Products	JAXA/Tokai University	UV-A	1km	2009–2010
TRMM Product	JAXA/NASA/NICT	Precipitation	0.25 deg	2009–2010
GLDAS Product	NASA	Temperature	0.25 deg	2009–2010
		Wind speed		
ASTER GDEM	MTI and NASA	Altitude	1 arcsec	2011 (ver.2)
National Census 2005	Statistics Bureau of Japan	Population	500 m	2005
Global Map Japan	Geographical Information Authority of Japan	Land use	30 arcsec	2006 (ver. 1.1)
		Road lines	Vector	2011 (ver. 2)
		Coast lines	Vector	2011 (ver. 2)
National Land Numerical Information	Ministry of Land, Infrastructure, Transportation and Tourism	Road length	1 km	2010

Environmental Studies (JASMES) products courtesy of JAXA/Tokai University.

As for the meteorological parameters, we obtained TRMM Product 3B43 precipitation data (JAXA/NASA/NICT) through Goddard Earth Science (GES) Data and Information Service Center (DISC), and modelled monthly mean surface temperature and wind velocity data from GLDAS NOAH025\_M (Rodell *et al.*, 2004) through Goddard Earth Science (GES) Data and Information Service Center (DISC).

The altitude data was obtained from ASTER Global Digital Elevation Model (GDEM) version 2, which is a product of METI and NASA. The population data was obtained from the National Census of the year 2005 through the Statistics Bureau of Japan. As these two variables are skewed to zero, we transformed them to a natural logarithmic scale before analysis. Build-up area ratio in a grid cell was calculated from land use data obtained from Global Map Japan version 1.2.1 downloaded from Geospatial Information Authority of Japan (GSI).

Transport is an important emission source for NO<sub>x</sub> (NO + NO<sub>2</sub>). The concentrations of NO<sub>2</sub> decrease with distance from roads, whereas the concentrations of O<sub>3</sub> tend to decrease towards roads because it is scavenged by NO. Hence, the distance to a road was provided as a predictor variable. The road network data was obtained from Global Map Japan version 2 downloaded from GSI. In this data, road types are classified into three categories; primary, secondary and highway. The distance to a road was calculated for each grid cell centroid for each of these three categories. Likewise road length was obtained from the National Land Numeric Information Data downloaded through the Japanese Ministry of Land, Infrastructure, Transportation and Tourism. This road length data is classified into 10 categories depending on the road width. We reclassified them into three new categories; Road A (road width ≥ 19.5 m), Road B (13 ≤ road width < 19.5 m) and Road C (5.5 m ≤ road width < 13 m).

When typical land and sea breezes dominated, polluted air parcels are transported from industrial or urban areas in coastal regions to inland areas and O<sub>3</sub> concentrations increase via a photochemical reaction during the transportation

(Kannari, 2006). Therefore, we used distance to coastline as a predictor variable for O<sub>3</sub>. This distance was calculated for each grid cell centroid as the nearest straight-line distance to coastline, which was obtained from Global Map Japan version 2.

### Regression Kriging

Linear regression models were constructed prior to ordinary kriging on its residuals. Candidates for predictor variables of linear regression models for each pollutant are presented in Table 2 with the pre-specified direction of effect according to the physical or chemical relationship between the pollutants and the predictor variables (Beelen *et al.*, 2009). Linear regression models were developed using backward stepwise procedure to select the significant variables (Hengl, 2007). The selected variables that had coefficients that conformed to the pre-specified direction of effect were retained in the final linear regression models, but others were discarded (Beelen *et al.*, 2009). Ordinary kriging was performed on the residuals of the final linear regression models developed as above. The concentrations of pollutants were transformed to a natural logarithmic scale before analysis, and the predictions were back transformed after analysis. This procedure has the advantage that predicted concentrations are positive, which is found not to be the case when analyses are performed without transformation (Beelen *et al.*, 2009).

### Validation

The validation was carried out by leave-one-out cross validation. This procedure has been applied in many previous studies for model validations (e.g., Pearce, *et al.*, 2009; De Smit *et al.*, 2010). In this method, one measurement point is removed and then the concentration at that point is predicted by using the rest of the points. This procedure is repeated for all measurement points. The predicted and measured values are compared at all points. From the result of the cross validation, we computed root mean squared error (RMSE) and R<sup>2</sup> between the predicted and measured values as indicators of the prediction performance. RMSE should be as small as possible.

Data analysis was carried out using R statistical software

**Table 2.** Predictor variables and predefined directions of effect<sup>a</sup>.

Predictor variables	Unit	NO <sub>2</sub>	O <sub>3</sub>
NDVI	None		+
Built-up area ratio <sup>b</sup>	None	+	
Population	Person	+	
UV-A	W/m <sup>2</sup>		+
Precipitation	mm/hr	–	
Temperature	°C		+
Wind speed	m/sec	–	–
Altitude	m	±	±
Distance to coastline	km		+
Distance to highway	km	–	+
Distance to primary road	km	–	+
Distance to secondary road	km	–	+
Road length A	m/km <sup>2</sup>	+	–
Road length B	m/km <sup>2</sup>	+	–
Road length C	m/km <sup>2</sup>	+	–

<sup>a</sup>+: positive direction, –: negative direction

<sup>b</sup> ratio of land use type

2.15.1 (R Core Team, 2012) with the raster package (Hijmans *et al.*, 2012) for the integration and construction of the grid data of predictor variables and with the gstat package (Pebesma *et al.*, 2004) for the performance of regression kriging.

## RESULTS AND DISCUSSION

### NO<sub>2</sub>

Linear regression methods moderately explained the variability of the 98 percentile daily mean concentrations of NO<sub>2</sub> for both years with R<sup>2</sup> values of 0.53 and 0.58. The retained predictor variables and their coefficients are given in Table 3. Most of the potential predictor variables were retained in the models except for distance to primary road and road length C. The retained variables were the same in both years and their coefficients were similar in general to the corresponding ones in the other year.

The validation results of regression kriging are given in Table 4 with those of linear regression methods. Regression kriging well explained the spatial variation of NO<sub>2</sub>

concentrations with R<sup>2</sup> values of 0.79 and 0.78, which were much higher than those for linear regression methods. RMSE values were less than 5 ppb and smaller than those for the linear regression methods. Figs. 1(a) and 1(b) are the scatter plots of the predicted and observed concentrations of NO<sub>2</sub> from cross validation results in the year 2009 and 2010 respectively. These plots show good correlation between the predicted and observed values, but some over predictions are found at the higher concentrations. These overestimations generally occurred at locations at short distance to highway and long road length A and B that were the retained predictor variables in the regression models. These variables represent basically the same emission source from automobile traffic, and simultaneous selection of these variables in the regression models might bring some positive bias to the predictions.

There was no prediction value above the air quality standard in Japan of 60 ppb either in the year 2009 or 2010. Fig. 2 illustrates the spatial distribution of NO<sub>2</sub> in the year 2010 predicted by regression kriging. Relatively high concentration areas are located in urban areas such as metropolitan Tokyo and along major highways, which generally reflects the distribution of the selected predictor variables in the final regression model.

### O<sub>3</sub>

Linear regression methods poorly explained the spatial variation of the 26th highest daily maximum 8-hour mean concentrations of O<sub>3</sub> in the year 2009 with an R<sup>2</sup> value of 0.23, whereas they moderately explained spatial variation in the year 2010 with an R<sup>2</sup> value of 0.40. The retained predictor variables were not consistent for each year; UV-A and altitude were only included in 2009, while distance to primary road was retained only in the year 2010 (Table 5). Temperature, wind velocity, distance to coastline and road length B were included in the models for both years.

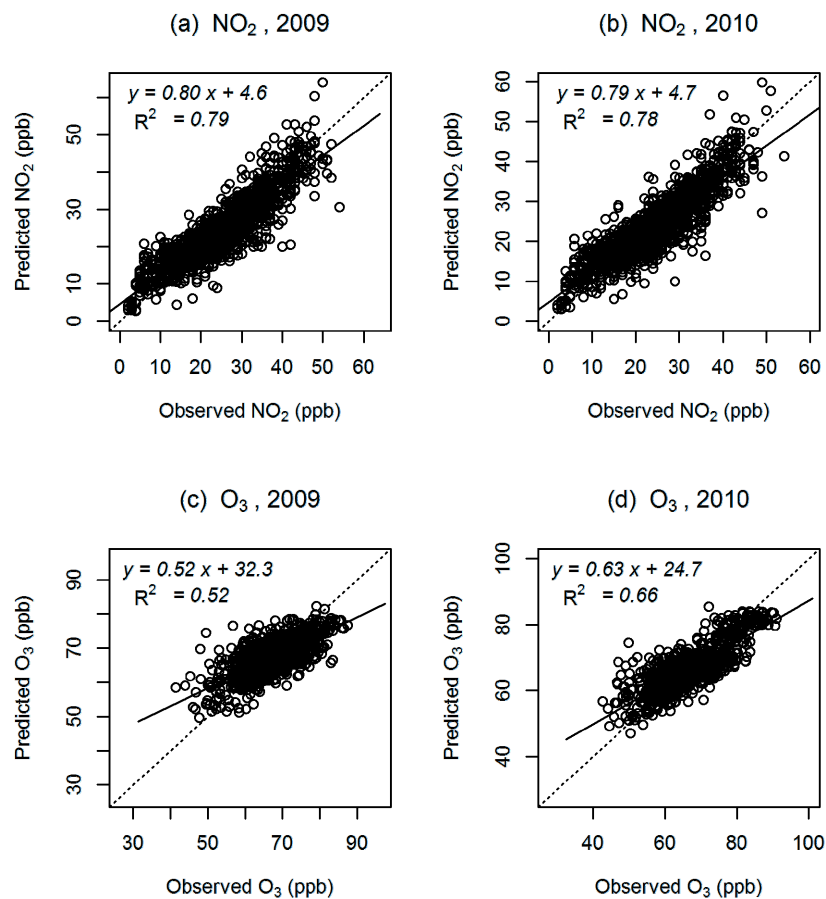
The validation results for regression kriging of O<sub>3</sub> are presented in Table 4 with those for regression methods. Regression kriging moderately explained the spatial variation of O<sub>3</sub> with R<sup>2</sup> values of 0.52 and 0.66, which were much higher than those for the linear regression methods. RMSE values were 4.9 and 4.8 ppb, marginally smaller than those for the linear regression methods. Figs. 1(c) and 1(d) show the scatter plots of predicted and observed concentrations

**Table 3.** Result of regression models for NO<sub>2</sub> (ppb).

Predictor variables	2009		2010	
	Parameter estimate	Standard error	Parameter estimate	Standard error
Intercept	3.252	0.098	3.332	0.096
Built-up area ratio	0.213	0.026	0.200	0.024
Population	0.072	0.008	0.080	0.007
Precipitation	–1.438	0.303	–2.761	0.265
Wind speed	–0.068	0.012	–0.047	0.011
Altitude	–0.053	0.009	–0.046	0.009
Distance to highway	–0.023	0.001	–0.022	0.001
Distance to secondary road	–0.016	0.006	–0.020	0.006
Road length A	0.078	0.022	0.075	0.021
Road length B	0.058	0.008	0.060	0.008
R <sup>2</sup>	0.53		0.58	

**Table 4.** Validation results for linear regression methods and regression kriging.

Air pollutants	Year	Linear regression methods		Regression kriging	
		RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>
NO <sub>2</sub> (ppb)	2009	7.2	0.53	4.6	0.79
	2010	6.5	0.58	4.6	0.78
O <sub>3</sub> (ppb)	2009	6.2	0.23	4.9	0.52
	2010	6.2	0.40	4.8	0.66



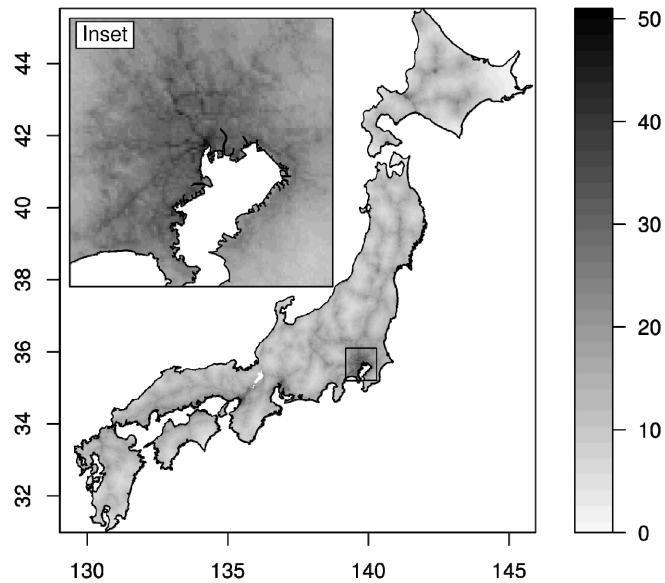
**Fig. 1.** Scatter plot of the predicted and observed concentrations of the annual 98 percentile daily mean NO<sub>2</sub> in the years a) 2009 and b) 2010, and the annual 26th highest daily maximum 8-hour mean O<sub>3</sub> in the years c) 2009 and d) 2010. The solid lines represent regression lines and dotted lines are 1:1 lines.

of O<sub>3</sub> from cross validation results in the year 2009 and 2010 respectively. These plots indicate overestimation for the higher values and underestimation for the lower values. This was caused by a typical smoothing effect natural to spatial interpolation methods. Closer regression would reduce this effect (De Smet *et al.*, 2011), which is supported by the result for NO<sub>2</sub> with high correlation as shown in Figs. 1(a) and 1(b) where over/under predictions are not significant. However, remarkable over predictions at relatively lower measured values are not fully explained by this effect. This might be the influence of a near-by stationary source of NO that scavenges O<sub>3</sub> around these measurement points, which we were not able to model because such emission sources were not considered in our study.

For reference, we compared our prediction values to the EU standard of 120 µg/m<sup>3</sup>, which is transformed to 60 ppb

assuming 1 ppb is 2 µg/m<sup>3</sup> (this will also apply for O<sub>3</sub> unit conversion where necessary). The percentages of the grid cells with predicted values above 60 ppb were 74% and 65% in the year 2009 and 2010 respectively. The population of areas above the standard value was calculated to be 94% and 91% in the years 2009 and 2010 respectively using the population grid data prepared as one of the predictor variables. The reported values of this percentage for Greece, which is located at a similar latitude of Japan, are 85 and 59% in the years 2008 and 2009 respectively (De Smet *et al.*, 2010, 2011).

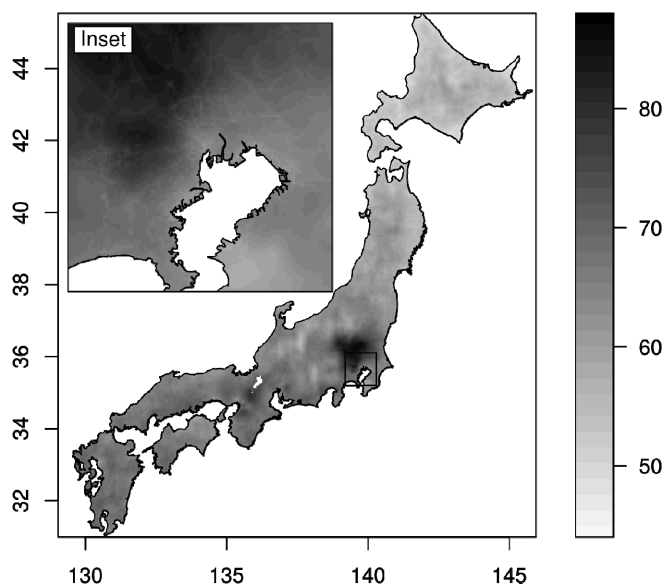
Fig. 3 illustrates the spatial distribution of O<sub>3</sub> in the year 2010 predicted by regression kriging. High concentration areas are generally located in inland areas of large cities, broadly reflecting the distance to coastline, which was a retained predictor variable in the regression model. This



**Fig. 2.** The spatial distribution of the annual 98th percentile daily mean concentrations of NO<sub>2</sub> for the year 2010 predicted using regression kriging at 1 × 1 km resolution (ppb).

**Table 5.** Result of regression models for O<sub>3</sub> (ppb).

Predictor variables	2009		2010	
	Parameter estimate	Standard error	Parameter estimate	Standard error
Intercept	3.936	0.061	3.845	0.025
UV-A	0.018	0.007		
Temperature	0.016	0.002	0.032	0.002
Wind speed	-0.038	0.004	-0.045	0.004
Altitude	0.002	0.004		
Distance to coastline	0.001	0.000	0.003	0.000
Distance to primary road			0.008	0.002
Road length B	-0.009	0.002	-0.004	0.002
R <sup>2</sup>	0.23		0.40	



**Fig. 3.** The spatial distribution of the annual 26th highest daily maximum 8-hour mean concentrations of O<sub>3</sub> for the year 2010 predicted using regression kriging at 1 × 1 km resolution (ppb).

spatial feature is remarkable for inland areas of metropolitan Tokyo (inset of Fig. 3), suggesting that typical land and sea breezes contributed to high concentrations of O<sub>3</sub> in these areas.

### Prediction Performance

The retained predictor variables for NO<sub>2</sub> in each year were equivalent and their coefficients were broadly similar (Table 3). Also, R<sup>2</sup> values were moderately high for both years (Table 3). These facts demonstrate the robustness of the linear regression model for NO<sub>2</sub> developed in this study. However, the retained variables for O<sub>3</sub> were different in each year (Table 5) and R<sup>2</sup> values were lower than those for NO<sub>2</sub> (Table 4). The reason for this might be the difference in the characteristics of these pollutants. NO<sub>2</sub> is a primary pollutant, whereas O<sub>3</sub> is a secondary pollutant which is formed through a complicated photochemical reaction involving NO<sub>2</sub> and VOCs. Although UV-A and temperature were considered in our study as predictor variables related to secondary pollutant formation, these variables did not sufficiently contribute to the prediction performance. This resulted in a poorer performance using the linear regression methods with O<sub>3</sub> than with NO<sub>2</sub>.

Regarding the comparison of regression kriging to linear regression methods, RMSE values of regression kriging were smaller and R<sup>2</sup> values were significantly higher for all cases. Taking account of these validation results, regression kriging outperformed the linear regression methods, which is consistent with the reported results (Beelen *et al.*, 2009; Pearce *et al.*, 2009; Mercer *et al.*, 2011) as described earlier.

We compare our result to those from previous studies. Beelen *et al.* (2009) applied regression kriging to the annual mean concentrations of NO<sub>2</sub> at the EU-wide scale and reported an R<sup>2</sup> value of 0.61. Mercer *et al.* (2011) predicted 2-week average NO<sub>x</sub> concentrations at the urban scale in Los Angeles, US and reported R<sup>2</sup> values of 0.72–0.75. These values are similar to the R<sup>2</sup> values of 0.79 and 0.78 in this study (Table 4).

We applied regression kriging to predict 26th highest maximum daily 8-hour mean concentrations of O<sub>3</sub> with RMSE values of 4.9 ppb and 4.8 ppb and R<sup>2</sup> values of 0.52 and 0.66 (Table 4). Beelen *et al.* (2009) predicted the annual mean concentrations of O<sub>3</sub> at the EU-wide scale and the reported R<sup>2</sup> value was 0.70. De Smet *et al.* (2010, 2011) estimated 26th highest daily maximum 8-hour mean for rural and urban areas separately and reported RMSE values of 4.1–4.7 ppb and R<sup>2</sup> values of 0.56–0.69. These RMSE and R<sup>2</sup> values from previous studies were similar to those of our study. Overall, the performance of regression kriging in our study was comparable to the performance reported in previous studies.

### Further Development

In this study, we implemented predictor variables representing NO<sub>2</sub> emissions from automobile traffic, whereas variables representing stationary sources such as power plants were not included. The NO<sub>x</sub> emissions from automobile traffic accounts for 40% of all NO<sub>x</sub> emissions in Japan (Kannari *et al.*, 2007). Thus, more than half of NO<sub>2</sub>

emissions were not included in our linear regression models. Regarding VOCs emissions, we implemented NDVI as a predictor variable representing biogenic VOCs emissions, whereas anthropogenic VOCs were not considered. Kannari *et al.* (2007) estimated biogenic and anthropogenic VOCs emissions in Japan as 1.5 and 2.1 (Tg/year) respectively. Thus, more than half of the emissions of VOCs were not considered in this study. Therefore, providing more variables representing other emission sources should be considered in order to improve prediction performance. From a practical point of view, however, too many predictor variables would reduce the efficiency of this procedure. Applying emission inventory at sufficiently high spatial resolution is a good alternative to this, where the quantitative relationship between the classes of emission sources can be considered. Also, the traffic related variables in our regression models were proxies of emissions from automobile traffic and represented basically the same emission source. This could result in biased predictions when more than two of these variables are simultaneously retained in the regression models. In this regard, introducing emission inventory would improve the predictions.

Meteorological parameters such as temperature and wind speed were significant variables for the regression models in our study (Tables 3 and 5). However, the spatial resolution of these variables is 0.25 deg, which is much coarser than that of our prediction grid resolution of 1 km. Therefore, higher spatial resolution of meteorological parameters could improve the performance of the prediction.

## CONCLUSIONS

This study aimed to develop a practical procedure to provide spatial distribution of air pollutants in Japan with sufficient accuracy at 1 × 1 km resolution. We applied regression kriging to the concentrations of NO<sub>2</sub> and O<sub>3</sub> with predictor variables representing the key factors in the spatial distribution of air pollutants including emission, transportation, transformation and deposition in two consecutive years. The performance of regression kriging assessed by cross validation was satisfactory, outperformed the conventional regression method and was comparable to previous studies. This brought us to the conclusion that regression kriging is a practical procedure that can be applied to the prediction of the spatial distribution of air pollutants in Japan with sufficient accuracy and high spatial resolution. Although further improvement is required, distribution maps produced using regression kriging can be a basis for risk assessment or policy evaluation.

## REFERENCES

- Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K. and Briggs, D.J. (2009). Mapping of Background Air Pollution at a Fine Spatial Scale across the Europe Union. *Sci. Total Environ.* 407: 1852–1867.
- Briggs, D., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S. and Smallbone, K. (2000). A Regression Based Method for Mapping Traffic-related Air Pollution:

- Application and Testing in Four Contrasting Urban Environments. *Sci. Total Environ.* 253: 151–167.
- Cyrus, J., Hochadel, M., Gehring, U., Hoek, G., Diegmann, V., Brunekreef, B. and Heinrich, J. (2005). GIS-based Estimation of Exposure of Particle Matter and NO<sub>2</sub> in an Urban Area: Stochastic versus Dispersion Modelling. *Environ. Health Perspect.* 113: 987–992.
- De Smet, P., Horálek, J., Coňková, M., Kurfürst, P., de Leeuw, F. and Denby, B. (2010). European Air Quality Maps of Ozone and PM<sub>10</sub> for 2008, ETC/ACC Technical Paper, European Topic Centre on Air and Climate Change.
- De Smet, P., Horálek, Kurfürst, P., Schreiberová, M. and de Leeuw, F. (2011). European Air Quality Maps of Ozone and PM<sub>10</sub> for 2009, ETC/ACC Technical Paper, European Topic Centre on Air and Climate Change.
- Hengl, T., Geuvelink, G.B.M. and Stein, A. (2003). Comparison of Kriging with External Drift and Regression-kriging, Technical Report, International Institute for Geo-information Science and Earth Observation (ITC) Enschede, URL [http://www.itc.nl/library/Academic\\_output/](http://www.itc.nl/library/Academic_output/).
- Hengl, T. (2007). A Practical Guide to Geostatistical Mapping of Environmental Variables. EUR 229404 EN. Office for Official Publications of the European Communities, Luxembourg.
- Hijmans, R.J. and van Etten J. (2012). Raster: Geographic Analysis and Modelling with Raster Data. R Package Version 2.0-12, <http://CRAN.R-project.org/package=raster>.
- Hoek, G., Beelen, R., de Hoogh K., Vienneau, D., Gulliver, J., Fischer, P. and Briggs, D. (2008). A Review of Land-use Regression Models to Assess Spatial Variation of Outdoor Air Pollution. *Atmos. Environ.* 42: 7561–7578.
- Kannari, A. (2006). An Analysis of Weekend Effects on Photochemical Oxidant Concentration in the Kanto and Kansai Regions: Part2 Verification of Dynamically Changing Ozone Formation Regimes. *J. Japan Soc. Atmos. Environ.* 41: 220–233.
- Kannari, A., Tonooka, Y., Baba, T. and Murano, K. (2007). Development of Multiple-species 1 km × 1 km Resolution Hourly Basis Emissions Inventory for Japan. *Atmos. Environ.* 41: 3428–3439.
- Marshall, J.D., Nethery, E. and Brauer, M. (2008). Within-urban Variability in Ambient Air Pollution: Comparison of Estimation Methods. *Atmos. Environ.* 42: 1359–1369.
- Mercer, L.D., Szpiro, A.A., Sheppard, L., Lindström, J., Adar, S.D., Allen, R.W., Avol, E.L., Oron, A.P., Larson, T., Liu, L.J.S. and Kaufman, J.D. (2011). Comparing Universal Kriging and Land-use Regression for Predicting Concentrations of Gaseous Oxides of Nitrogen (NO<sub>x</sub>) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos. Environ.* 45: 4414–4420.
- Pearce, J.L., Rathbun, S.L., Aguilar-Villalobos, M. and Naeher, L.P. (2009). Characterizing the Spatiotemporal Variability of PM<sub>2.5</sub> in Cusco, Peru Using Kriging with External Drift. *Atmos. Environ.* 43: 2060–2069.
- Pebesma, E.J. (2004). Multivariable Geostatistics in S: the Gstat Package. *Comput. Geosci.* 30: 683–691.
- R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, URL <http://www.R-project.org/>.
- Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmanm, D. and Toll, D. (2004). The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.* 85: 381–393.
- Ross, Z., Jerret, M., Ito, K., Tempalski, B. and Thurston, G.D. (2007). A Land Use Regression for Predicting Fine Particle Matter Concentration in the New York City Region. *Atmos. Environ.* 41: 2255–2269.

Received for review, January 17, 2014

Revised, May 11, 2014

Accepted, May 20, 2014