Applying CoKriging Method for Air Pollution Prediction PM10 in Binh Duong Province

Nguyen Cong Nhut(✉)

Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
ncnhut@ntt.edu.vn

Abstract. Geostatistics is a branch of statistics that focuses on spatial data sets. Geostatistics is used for prediction in many fields such as ore mining, petroleum geology, climate, geography, environment. CoKriging can be seen as a point interpolation, which requires a point map as input and which returns a raster map with estimations and optionally an error map. CoKring is a prediction method for problems with two or more parameters, including one main parameter. Tthe costs of installing new observational stations to observe metropolitan air pollution sources, as PM10 (Particulate Matter), CO (Cacbon monoxit), SO_2 (Dioxit Sunfua) and NO_2 (oxit Nito) concentrations are high. In this study, analysis of air pollution of 16 stations monitored in 2018 year was carried out. Geostatistics have been used by many researchers to effectively predict air pollution, ore mining, and groundwater levels.

Keywords: Air pollution · Cokriging · Prediction · PM10

1 Introduction

The information from the monitoring center - technical resources and environment of Binh Duong province, network quality monitoring air environment of Binh Duong has 16 stations. This number is still too small compared to a province with a large population and large area. However, the investment cost for a monitor is too expensive, and the preservation in tropical climates like Vietnam is also very difficult. The problem is based on existing stations, forecasting for areas that have not been installed.

Vietnam is a developing country with many effective activities for economic growth. However, in parallel with economic development, development activities are also sources of emissions causing environmental pollution in general and air environment in particular. In which, the main sources of air pollution include: transportation; Industrial production; construction and people living; agriculture and craft villages; landfilling and waste treatment. The main air pollutants include: total suspended dust (TSP), PM10 dust (dust $\leq$ 10 µm), lead (Pb), ozone (O_3); inorganic

In this article, author use the recorded PM10 concentrations at several observational stations in Binh Duong province, employ the CoKriging interpolation method to find suitable models, then predict PM10 concentrations at some unmeasured stations in the city. From the data set, author found the best forecast model with the smallest forecast error to predict PM10 concentration.

P. Cong Vinh and A. Rakib (Eds.): ICCASA 2021, LNICST 409, pp. 323–335, 2021.
https://doi.org/10.1007/978-3-030-93179-7_25

substances such as carbon monoxide (CO), sulfur dioxide (SO_2), nitrous oxide (NOx), hydrochloride (HCl), hydrofluoride (HF)…; organic substances such as hydrocarbons (CnHm), benzene (C_6H_6)…; unpleasant odors such as ammonia (NH_3), hydrosulfide (H_2S)…; heat, noise.

The study area is Binh Duong in Southeastern of Vietnam. It is located between 10° 51′46″–11°30′ northing and 106°20′–106°58′ easting and has the following administrative boundaries: the East borders on Dong Nai province; the West borders on Tay Ninh province and Ho Chi Minh City; to the South, it borders on Ho Chi Minh City; the North borders on Binh Phuoc province. And the area has more than 2694.4 km^2. Binh Duong has a fairly flat terrain. Binh Duong is located in a tropical monsoon climate, subequatorial in nature. There are two seasons in a year, the rainy season from May to November, the dry season from December to April of the following year. Figure 1 shows the study area.

Fig. 1. Air monitoring network in Binh Duong province.

Figure 2 shows the geographical location of the monitoring with pollution levels.

2 Materials and Methods

Dust is a common name for solid and liquid particles, a few micrometres to half a millimeter in diameter, settle on their own by their weight but can remain suspended in the air for a while. PM10 dust: is dust particles with kinetic diameter ≤ 10 µm. Figure 2 shows the air quality monitoring stations PM10 in Binh Duong (see Table 1). Physical noises are vibrations of sound waves of varying magnitudes and frequencies,

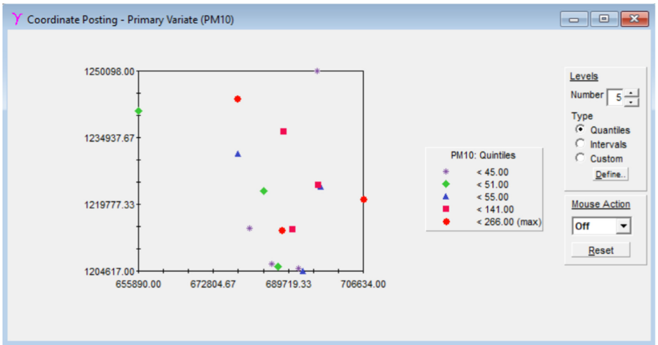


Fig. 2. Map of air quality monitoring stations with pollution levels.

arranged out of order and propagated in an elastic medium. The unit of noise measurement is dB (decibel). The sources of noise are from objects with high amplitude fluctuations, in excess of the hearing threshold (70 dB), for example: heavy machinery in works, lightning, loud singing, etc. The noise affects the ears, then affects the central nervous system, then the cardiovascular system, the stomach and other organs, and then the hearing organs. To predict for areas that have not yet installed monitors, the author uses the CoKriging method.

Table 1. PM10 and dB data of air pollution stations in Binh Duong.

Stations	X	Y	PM10 (mg/m^3)	dB (A)
N	696.193	1.250.098	85	52.3
NT	655.890	1.241.092	92	58.1
DT1	693.004	1.204.617	105	63.5
DT2	686.008	1.206.206	93	66.1
DT3	681.042	1.214.323	114	59.6
DT4	690.640	1.214.132	162	71.4
DT5	687.438	1.205.692	160	66.3
DT6	692.090	1.205.134	259	65.0
GT1	688.151	1.213.933	500	76.5
GT2	706.634	1.220.991	154	75.9
GT3	678.398	1.231.367	175	71.4
CN1	684.214	1.222.794	219	67.6
CN2	697.081	1.223.852	162	65.0
CN3	696.411	1.224.187	513	64.4
CN4	678.238	1.243.845	34	66.0
CN5	688.598	1.236.343	47	65.7

The variogram $\gamma(h)$ can be defined as one-half the variance of the difference between the attribute values at all points separated by has followed [1] and [2]

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

where Z(s) indicates the value of the variable, and N(h) is the total number of pairs are separated by a distance h.

The second-order stationary conditions [3] and [4] one obtains

$$E[Z(s)] = \mu,$$

And the covariance

$$Cov[Z(s), Z(s+h)] = E[(Z(s) - \mu)(Z(s+h) - \mu)] = C(h). \tag{2}$$

The most common models are spherical, exponential, Gaussian, and pure nuclear effects [5] and [2]. Cross-validation technique is used to check the completeness and validity of the model.

The correlation coefficient r^2 shows the estimate and the true value. The most appropriate variogram was chosen based on the highest correlation coefficient.

Cokring: Recall that parameter d = 2 is the dimension of our geological domain D of interest, $n > 2$ is the number of sampling places/locations $s_i \in D$, and $k > 1$ is the number of pollutants under investigation.

Generally, suppose that at each spatial location s_i we observe $k > 1$ variables Z_j, described by a data matrix M of size $k \times n$ as follows:

$$\mathbf{M} = \begin{pmatrix} Z_1(\boldsymbol{s}_1) & Z_1(\boldsymbol{s}_2) \cdots Z_1(\boldsymbol{s}_i) \cdots Z_1(\boldsymbol{s}_n) \\ \cdots & \cdots \quad \cdots \quad \cdots \\ Z_j(\boldsymbol{s}_1) & Z_j(\boldsymbol{s}_2) \cdots Z_j(\boldsymbol{s}_i) \cdots Z_j(\boldsymbol{s}_n) \\ \cdots & \cdots \quad \cdots \quad \cdots \\ Z_k(\boldsymbol{s}_1) & Z_k(\boldsymbol{s}_2) \cdots Z_k(\boldsymbol{s}_i) \cdots Z_k(\boldsymbol{s}_n) \end{pmatrix}, \tag{3}$$

for j = 1, 2, …, k, and i = 1, 2, …, n.

We predict $Z_1(\boldsymbol{s}_0)$, the value of variable Z_1 at unobserved location $\boldsymbol{s}_0$.

Given the fact that the *target variable* Z_1 occurs with other variables (call *co-located variables*), we explore the possibility of improving the prediction of variable Z_1 by taking into account the correlation of Z_1 with the other variables.

Definiton 1 [6, 7]**.** *The cokriging model of prediction takes the form*

$$\begin{aligned} \hat{Z}_1(\boldsymbol{s}_0) &= \sum_{j=1}^{k} \sum_{i=1}^{n} w_{ji} Z_j(\boldsymbol{s}_i) \\ &= w_{11} Z_1(\boldsymbol{s}_1) + \cdots + w_{1n} Z_1(\boldsymbol{s}_n) + \cdots + w_{k1} Z_k(\boldsymbol{s}_1) + \cdots + w_{kn} Z_k(\boldsymbol{s}_n). \end{aligned} \tag{4}$$

Assuming that the sampling area is relatively homogeneous, i.e. distinct sampling points $\boldsymbol{s}_i$ have different values $Z_j(\boldsymbol{s}_i)$ but their expectation are the same, we denote

$$\mu_j = \mathbf{E}[Z_j(\boldsymbol{s}_i)] = \mathbf{E}[Z_j(\boldsymbol{s})],$$

for each pollutant $j = 1, \ldots, k$; for all $i = 1, \ldots, n$, and for any sampling point $\boldsymbol{s} \in D$. We will examine *ordinary co-kriging* (the extension of ordinary kriging of a single variable to two or more variables). The expectation vector of k variables Z_j then is

$$\mathbf{E}[\mathbf{Z}(\boldsymbol{s})] = \begin{pmatrix} \mathbf{E}[Z_1(\boldsymbol{s})] \\ \mathbf{E}[Z_2(\boldsymbol{s})] \\ \vdots \\ \mathbf{E}[Z_k(\boldsymbol{s})] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} = \mu.$$

We want the predictor $\hat{Z}_1(\boldsymbol{s}_0)$ to be unbiased, that is $\mathbf{E}[\hat{Z}_1(\boldsymbol{s}_0)] = \mu_1$, where

$$\begin{aligned} \mathbf{E}[\hat{Z}_1(\boldsymbol{s}_0)] &= \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ji}\mathbf{E}[Z_j(\boldsymbol{s}_i)] = \sum_{i=1}^{n} w_{1i}\mu_1 + \cdots + \sum_{i=1}^{n} w_{ki}\mu_k \\ &= w_{11}\mathbf{E}[Z_1(\boldsymbol{s}_1)] + \cdots + w_{1n}\mathbf{E}[Z_1(\boldsymbol{s}_n)] + \cdots + w_{k1}\mathbf{E}[Z_k(\boldsymbol{s}_1)] + \cdots + w_{kn}\mathbf{E}[Z_k(\boldsymbol{s}_n)]. \end{aligned} \tag{5}$$

Definiton 2 [6, 7]**.** *The mean squared error (MSE) of prediction of Z_1 is given by.*

$$\mathbf{E}\left[\{Z_1(\boldsymbol{s}_0) - \hat{Z}_1(\boldsymbol{s}_0)\}^2\right] = \Sigma_1^2. \tag{6}$$

Therefore, to get $\mathbf{E}[\hat{Z}_1(\boldsymbol{s}_0)] = \mu_1$ we must have the followings

$$\sum_{i=1}^{n} w_{1i} = 1, \ \sum_{i=1}^{n} w_{2i} = 0, \ \cdots, \ \sum_{i=1}^{n} w_{ki} = 0. \tag{7}$$

Co-kriging for Two Pollutants

Let's assume $k = 2$, in other words, we observe variables Z_1 and Z_2 (e.g. PM_{10} and dB in our sample data) and we want to predict $Z = Z_1$.

Lemma 1 [6, 7]**.** The variance $\Sigma_1^2 = \mathbf{E}[\{Z_1(\boldsymbol{s}_0) - \hat{Z}_1(\boldsymbol{s}_0)\}^2]$ has expansion.

$$\begin{aligned} \Sigma_1^2 = {} & \mathbf{E}[\{Z(\boldsymbol{s}_0) - \mu_1\}^2] - 2\sum_{i=1}^{n} w_{1i}\mathbf{E}[(Z_1(\boldsymbol{s}_0) - \mu_1)\ (Z_1(\boldsymbol{s}_i) - \mu_1)] \\ & - 2\sum_{i=1}^{n} w_{2i}\mathbf{E}[(Z_1(\boldsymbol{s}_0) - \mu_1)\ (Z_2(\boldsymbol{s}_i) - \mu_2)] + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{1j}\mathbf{E}[(Z_1(\boldsymbol{s}_i) - \mu_1)(Z_1(\boldsymbol{s}_j) - \mu_1)] \\ & + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{2i}w_{2j}\mathbf{E}[(Z_2(\boldsymbol{s}_i) - \mu_2)(Z_2(\boldsymbol{s}_j) - \mu_2)] + 2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{2j}\mathbf{E}[(Z_1(\boldsymbol{s}_i) - \mu_1)(Z_2(\boldsymbol{s}_j) - \mu_2)]. \end{aligned} \tag{8}$$

Proof. See in Sect. 5.4 [8] or [3].

Theorem 1 [6, 7]**.** Consider the following optimization model min Σ_1^2 where.

$$\Sigma_1^2 = \mathbf{E}[\{Z(\boldsymbol{s}_0) - \sum_{i=1}^{n} w_{1i}Z_1(\boldsymbol{s}_i) - \sum_{i=1}^{n} w_{2i}Z_2(\boldsymbol{s}_i)\}^2].$$

The above optimization model is converted into a kriging system of linear equations

$$\mathbf{G}\boldsymbol{w} = \boldsymbol{c}, \tag{9}$$

where vectors $\boldsymbol{w}$,$\boldsymbol{c}$ have dimensions (2n + 2) × 1 and **G** has dimension (2n + 2) × (2n + 2).

If the multivariate covariance model is strictly positive definite and if there is no data duplication, the optimal weights will be obtained via the vector

$$\boldsymbol{w} = \mathbf{G}^{-1}\boldsymbol{c}.$$

Proof. See (Paul, 2011) or (Webster, 2007, Sect. 10.4).

Algorithm for Coping with Unmeasured Data Points:
For this realistic data, we got m = 16 good data points. The main idea of the linear kriging in (4) is that the predicted value $\hat{Z}_1(\boldsymbol{s}_0) = \sum_{j=1}^{2}\sum_{i=1}^{16} w_{ji}Z_j(\boldsymbol{s}_i) = w_{11}Z_1(\boldsymbol{s}_1) + \cdots + w_{116}Z_1(\boldsymbol{s}_{16}) + w_{21}Z_2(\boldsymbol{s}_1) + \cdots + w_{216}Z_2(\boldsymbol{s}_{16})$, $w_{ji} \geq 0$ at certain unknown point s_0 in fact is just a convex combination of 16 observed value at $Z(s_i)$ at m = 16 monitoring points, where the weights wi fulfill $\sum_{i=1}^{16} w_{1i} = 1$, $\sum_{i=1}^{16} w_{2i} = 0$.

Algorithm 1 Progressively Imputation

INPUT: a finite set V_0 of all n designed monitoring sites,
$V = \{s_1, s_2, \cdots, s_m\}$ of observable sites ($m \leq n$), $V \subseteq V_0$,
a $k \times m$ data matrix M of k observable factors (monitored at m sites), modified from data matrix (3).
OUTPUT: the fully updated set V of n known sites with available data, from which observation $Z(s)$ at any location $s \in$ CH(V) (the convex hull of V) is estimated by Equation (4)
If $m = n$ stop, else proceed to next step.
while $m < n$ do
1. Select a site $s_0 \in V_0 \backslash V$ (an unmeasured site)
2. Set up the system (9) from data M and s_0
3. Compute the weight vector $w = \mathrm{G}^{-1} c$
4. Calculate the predicted value at s_0
5. Update $V := V \cup \{s_0\}$; update $m = |V|$; update data matrix M;
end while
return The full network V of all observable sites.

3 Results and Discussions

Anisotropy was tested by comparing variations in several directions 0°, 45°, 90°, and 135° with an angle tolerance of ±45° used to detect anisotropy.

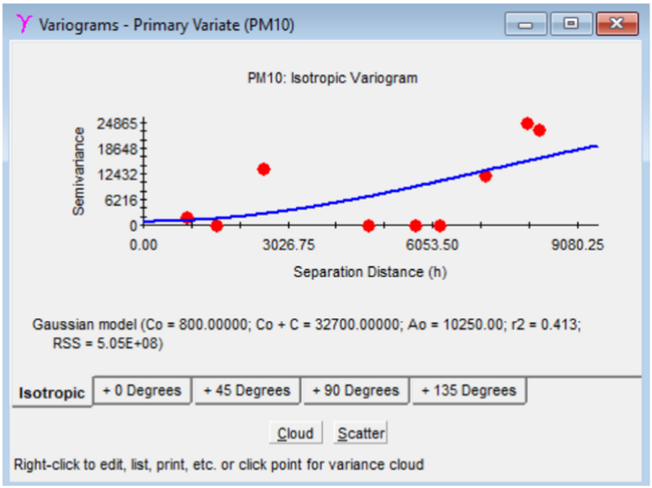


Fig. 3. Fitted variogram for the parameter PM10.

Figure 3 and 4 show fitted variogram for the parameters of PM_{10} and dB, respectively; and Fig. 5 shows fitted variogram for the parameters of both PM_{10} and dB, allfound by the isotropic-based Gaussian model, see Table 2.

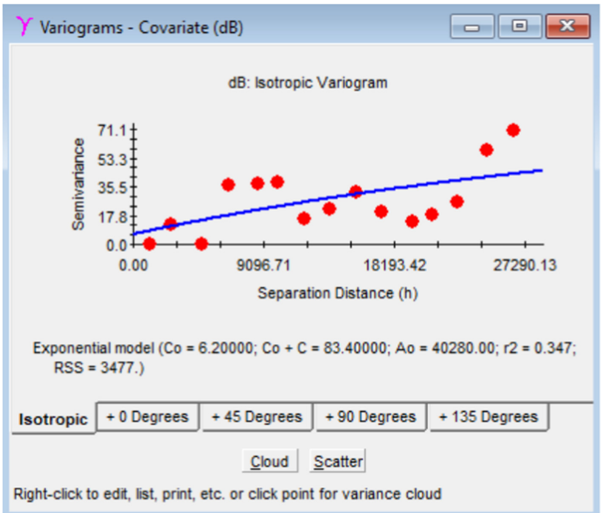


Fig. 4. Fitted variogram for the parameter dB.

Figure 3 shows fitted variogram for the parameter PM10. Gaussian model shows the best fitted omnidi-rectional variogram of air pollution obtained based on cross-validation. The variogram values are presented in Table 2.

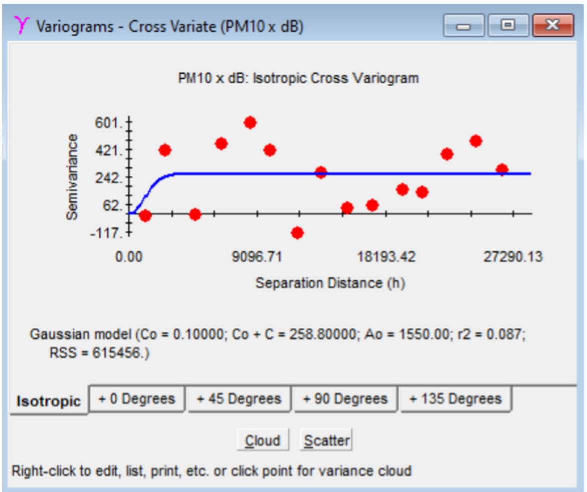


Fig. 5. Fitted variogram for two the parameter.

Table 2. Isotropic variograms values of PM_{10}, dB and two parameters

Data set and model	Estimates of parameters			RSS[a]	r^2	$\frac{C}{C_0+C}$
	Nugget (m)	Sill (m)	Range A			
	C_0	$C_0 + C$	(m)			
PM_{10} (n = 16)						
Linear	10	2.735	2.1482	5.66e08	0.351	11374.376
Gaussian	**800**	**32700**	**17754**	**5.05e08**	**0.413**	**0.976**
Spherical	10	26130	21100	5.79e08	0.335	1
Exponential	10	31120	43200	6.04e08	0.309	1
dB (n = 16)						
Linear	7.4884	45.936	26391	3390	0.364	0.837
Gaussian	0.1	32.96	9058.6	3789	0.289	0.997
Spherical	0.1	32.5	10590	3832	0.289	0.997
Exponential	**6.2**	**83.4**	**120840**	**3477**	**0.347**	**0.926**
PM_{10} and dB (n = 16)						
Linear	207.33	277.15	26390.5	658459	0.01	0.252
Gaussian	**0.1**	**258.8**	**2684.7**	**615456**	**0.087**	**1**
Spherical	0.1	256.4	3020	620895	0.092	1
Exponential	0.1	258.5	4500	631622	0.058	1

RSS[a] is the sum of squares of the residuals from the fitted function.

Thus, from the stations s_1, s_2,..., s_{16} we find the best interpolation model based on RSS ($RSS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Z_i - \widehat{Z}_i)^2$), r^2 and $C/(C_0 + C)$. Using the models found, we forecast for stations, where missing data occurred.

RSS represents the fit of the model to the data set, the lower the RSS indicates the better the model fits the data. r^2 is the fit of the model to the data; this value is not as strong as the Residual SS value.

Selecting the Best-Fit Model [7]**:** The model that uses interpolation is called optimal if it has the lowest error forecast. Few following statistics (integrated in GS+) can be used to explain the output of the model.

First, the residual sum of squares (RSS): a small RSS indicates a tight fit of the model to the data.

Second, the coefficient of determination, r^2: not strong for fitting the model as RSS, but used to look at the impact of change in the model parameters.

Third, the proportion $C/(C_0 + C)$ statistic provides a measure of the proportion of sample variance ($C_0 + C$) that is explained by spatially structured variance C.

This value will be 1.0 for a variogram with no nugget variance (where the curve passes through the origin).

Conversely, it will be 0 where there is no spatially dependent variation at the range specified, i.e. where there is a pure nugget effect.

Correlation coefficients are used to measure how strong a relationship is between two variables ($r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$).

Model checking: Selecting the right model for the data based on the results: regression coefficient, correlation coefficient and interpolation values, besides error values are standard error (SE) and standard error prediction (SE prediction) (Table 3).

Table 3. Result testing parameters of the model.

Coefficient regression	Coefficient correlation	SE	SE Prediction
1.044	0.994	0.021	5.440

Figure 6 shows the results of checking the error between the actual value and the estimated value using the cokriging method. The regression coefficient and the correlation coefficient are approximately equal to 1, the standard error is small (close to zero). Thus, the selected model is a suitable interpolation (Fig. 7).

From Fig. 8 and Fig. 9 show estimated errors of predicting PM10 concentrations 2D and 3D in Binh Duong by the discussed cokriging method. This shows that forecast areas close together have small forecast errors. Using the CoKriging interpolation method, we can predict air pollution levels for areas that have not yet installed monitoring points.

Based on the forecast map in Figs. 8 and 9, we can predict the air pollution concentration at locations around the air monitoring stations. Positions in the same color have a small prediction error. Through the use of CoKriging spatial interpolation

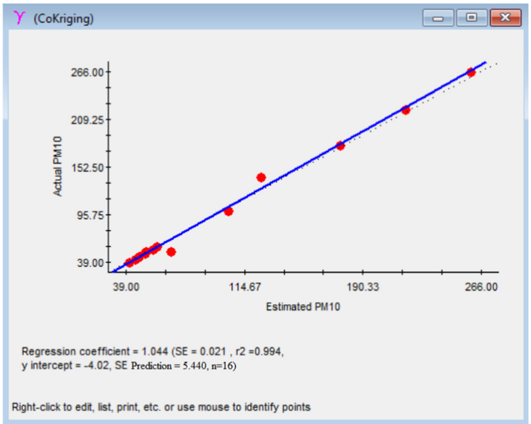


Fig. 6. Error testing result of prediction PM10.

(CoKriging)

Record	X-Coordinate	Y-Coordinate	Actual Z	Estimated Z	Error (E-A)
1	696193.00	1250098.00	39.00	40.67	1.67
2	655890.00	1241092.00	46.00	47.33	1.33
3	693004.00	1204617.00	52.00	51.99	-0.01
4	686008.00	1206206.00	43.00	44.53	1.53
5	681042.00	1214323.00	45.00	46.55	1.55
6	690640.00	1214132.00	101.00	104.37	3.37
7	687438.00	1205692.00	51.00	50.76	-0.24
8	692090.00	1205134.00	45.00	46.43	1.43
9	688151.00	1213933.00	266.00	258.97	-7.03
10	706634.00	1220991.00	178.00	175.43	-2.57
11	678398.00	1231367.00	55.00	55.85	0.85
12	684214.00	1222794.00	50.00	51.11	1.11
13	697081.00	1223852.00	52.00	67.63	15.63
14	696411.00	1224187.00	141.00	125.30	-15.70
15	678238.00	1243845.00	221.00	217.43	-3.57
16	688598.00	1236343.00	58.00	58.89	0.89

Fig. 7. Cross-Validation (CoKriging) of PM10.

model, we can predict air pollution levels for areas that have not yet installed air monitors. The higher the density of monitoring stations, the easier it is to choose the interpolation model and the more reliable the interpolation results and vice versa.

The prediction results of the model also have some limitations in terms of forecasting error. The reason is that in addition to the PM10 pollutant, there are other pollutants in the air such as TSP, CO, SO, NO2, PM25.

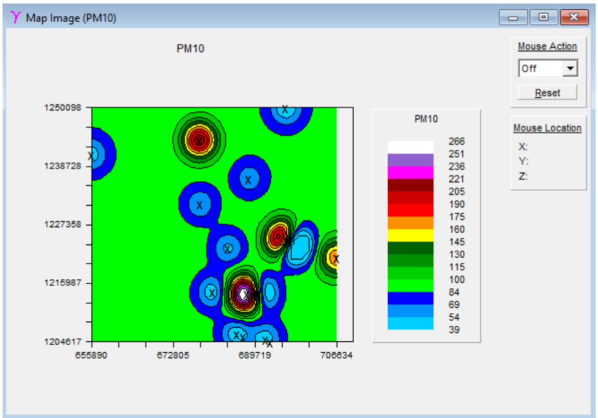


Fig. 8. The Cokriging interpolation method in 2D space.

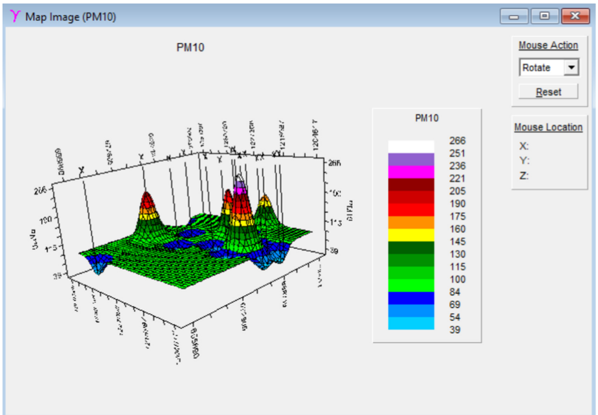
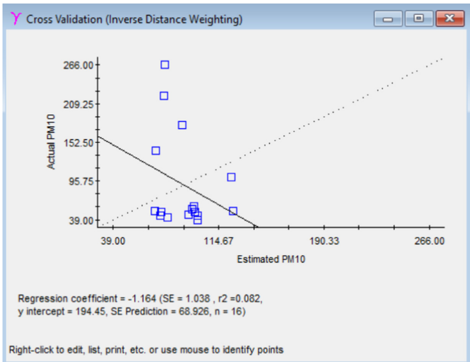


Fig. 9. The Cokriging interpolation method in 3D space.

4 Conclusion

CoKriging application to forecast PM10 concentration results in low error between estimated value and actual value. Therefore, the study shows that the efficiency and high reliability of geostatistics to build spatial prediction models. There is no way we can find the best model, only by experiment. The author must test all models, change the model parameters and finally choose the model that fits the data, the model with the lowest prediction error. When building the model, we should pay attention to the outliers, which will determine the error of the model, so we have to deal with them.

The following are the predicted results using other methods. Figure 10 shows the results of PM10 concentration prediction by the distance inverse method, which results in high predictive error 68,926. Figure 11 shows the results of PM10 concentration prediction by Kriging method giving high predictive error 71,912. The prediction error by CoKriging method in this paper is 5,440. Thus, the Cokriging method used in this paper to predict PM10 concentrations is consistent with the data.

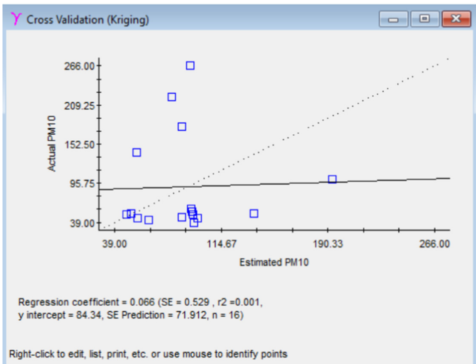


Cross Validation (Inverse Distance Weighting)

Record	X-Coordinate	Y-Coordinate	Actual Z	Estimated Z	Error (E-A)
1	696193.00	1250098.00	39.00	99.42	60.42
2	655890.00	1241092.00	46.00	92.85	46.85
3	693004.00	1204617.00	52.00	68.71	16.71
4	686008.00	1206206.00	43.00	78.26	35.26
5	681042.00	1214323.00	45.00	99.65	54.65
6	690640.00	1214132.00	101.00	123.95	22.95
7	687438.00	1205692.00	51.00	73.56	22.56
8	692090.00	1205134.00	45.00	72.44	27.44
9	688151.00	1213933.00	266.00	75.77	-190.23
10	706634.00	1220991.00	178.00	88.72	-89.28
11	678398.00	1231367.00	55.00	95.74	40.74
12	684214.00	1222794.00	50.00	97.37	47.37
13	697081.00	1223852.00	52.00	125.11	73.11
14	696411.00	1224187.00	141.00	69.30	-71.70
15	678238.00	1243845.00	221.00	75.56	-145.44
16	688598.00	1236343.00	58.00	97.11	39.11

Fig. 10. Inverse distance weighting method.

Up to now, there has been no research on air pollution forecasting using CoKriging spatial interpolation method in the country. Domestic studies only stop at the level of data simulation [9, 10] or Kriging interpolation as in the article: [11, 12]. The data simulation method leads to high errors, the geographical spatial correlation has not been evaluated, and the relationship between the secondary parameters affecting the main parameters to be forecasted. The Kriging method only provides a predictive model for data with exactly one pollutant, when in fact many pollutants are related to each other.



Cross Validation (Kriging)

Record	X-Coordinate	Y-Coordinate	Actual Z	Estimated Z	Error (E-A)
1	696193.00	1250098.00	39.00	95.18	56.18
2	655890.00	1241092.00	46.00	86.49	40.49
3	693004.00	1204617.00	52.00	50.62	-1.38
4	686008.00	1206206.00	43.00	62.89	19.89
5	681042.00	1214323.00	45.00	97.96	52.96
6	690640.00	1214132.00	101.00	193.29	92.29
7	687438.00	1205692.00	51.00	47.00	-4.00
8	692090.00	1205134.00	45.00	55.27	10.27
9	688151.00	1213933.00	266.00	92.24	-173.76
10	706634.00	1220991.00	178.00	86.40	-91.60
11	678398.00	1231367.00	55.00	93.77	38.77
12	684214.00	1222794.00	50.00	94.23	44.23
13	697081.00	1223852.00	52.00	137.61	85.61
14	696411.00	1224187.00	141.00	54.53	-86.47
15	678238.00	1243845.00	221.00	79.13	-141.87
16	688598.00	1236343.00	58.00	93.51	35.51

Fig. 11. Kriging method

Acknowledgment. The paper's author expresses his sincere thank to the scientists and the referees for their comments. The author thanks the anonymous reviewer whose helpful and valuable comments led to significant improvements to the final version of the paper. The author thanks the editor, the reviewer for facilitating the author's article to be published in the journal.

References

1. Ahmadi, S.H., Sedghamiz, A.: Geostatistical analysis of spatial and temporal variations of groundwater level. Environ. Monit. Assess. **129**, 277–294 (2007)
2. Kitadinis, P.K.: Introduction to Geostatistics: Applications to Hydrogeology. Cambridge University Press, Cambridge (2003)
3. Webster, R., Oliver, M.A.: Geostatistics for Enviromental Scientists, 2nd edn. Wiley, Chichester (2007). 6–8
4. Gentile, M., Courbin, F., Meylan, G.: Interpolating point spread function anisotropy. Astronomy and Astrophysics manuscript no. psf˙interpolation, 10 October 2012 (2012)
5. Isaaks, E., Srivastava, M.R.: An Introduction to Applied Geostatistics. Oxford University Press, New York (1989)
6. Nguyen, M.V.M., Nguyen, N.C.: Analyzing incomplete spatial data in air pollution prediction. Southeast-Asian J. Sci. **6**(2), 111–133 (2018)
7. Nguyen, C.N., Nguyen, M.V.M., Vo, L.P.: Co-kriging method for air pollution prediction: a case study in saigon. Thailand Stat. J. (2020). ISSN 2351–0676
8. Chiles, P., Delfiner, P.: Geostatistics - Modeling Spatial Uncertainty, 2nd edn. Wiley, Hoboken (2011)
9. Anh, P.T.V.: Application of airborne pollutant emission models in assessing the current state of the air environment in Hanoi area caused by industrial sources. In: 6th Women's Science Conference, pp. 8–17. Ha Noi National University (2001)
10. Yen, D.T.H.: Applying the Meti-lis model to calculate the emission of air pollutants from traffic and industrial activities in Thai Nguyen city, orienting to 2020. J. Sci. Technol. **106**(6) (2013)
11. Hà, P.T.S., Sơn, L.M.: Ứng dụng phương pháp nội suy Kiging khảo sát sự phân bố tầng đất yếu tuổi Holocen ở khu vực nội thành Thành phố Hồ Chí Minh. Tạp chí phát triển KH& CN Tập **10**(02), 43–53 (2007)
12. Trần, Q.Â.: Ứng dụng Variogram của địa thống kê để nghiên cứu một vài đặc điểm biến đổi hàm lượng đồng, Tạp chí địa chất số 190–191, Cục địa chất & khoáng sản Việt Nam (1989)