Applied the Cokriging interpolation method to survey Air Quality Index (AQI) for dust TSP in Da Nang city

Nhut Nguyen Cong*, Phut Lai Van, Vuong Bui Hung

Faculty of Information Technology, Nguyen Tat Thanh University

*ncnhut@ntt.edu.vn

Abstract

Mapping to forecast the air pollution concentration in Da Nang city is an urgent issue for management agencies and researchers of environmental pollution. Although the simulation of spatial location has become popular, it uses the classical interpolation methods with low reliability. Based on the distribution of air quality monitoring stations located in industrial parks, residential areas, transport axes... and sources of air pollution, the application of geostatistical theories, this study presents the results of the Cokriging's interpolation selection which provides forecast results of air pollution distribution in Da Nang city with high reliability. In this article, we use the recorded TSP concentrations (one of major air pollution causes at large metropolis) at several observational stations in Da Nang city, employ the Cokriging interpolation method to find suitable models, then predict TSP dust concentrations at some unmeasured stations in the city. Our key contribution is finding good statistical models by several criteria, then fitting those models with high precision.

Nhận 01.08.2018
Được duyệt 10.10.2018
Công bố 25.12.2018

Keywords
Air pollution, geostatistics, Cokriging, variogram

1 Introduction

Air pollution is an issue of social concern both in Vietnam in particular and the world in general. Transportation increases, air pollution caused by industrial factories increasingly degrades environments quality, leads to severe problems in health for local inhabitants. The building of air quality monitoring stations is not essential, but also difficult because of expensive installation costs, no good information of selected areas for installation in order to achieve precise results.

According to the Center for Monitoring and Analysis Environment (Da Nang Department of Natural Resources and Environment), network quality monitoring air environment of Da Nang has 15 stations observation in the city and 9 stations in the suburban area. However, with a large area, the city needs to install more new monitoring stations. The cost to of installing a new machine costs tens of billions, and the preservation is also difficult. Therefore, the requirements are based on the remaining monitoring stations using mathematical models based to predict air pollution concentration at some unmeasured stations in the city.

Globally the use of mathematical models to solve the problems of pollution has started since 1859 by Angus Smith who used to calculate the distribution of CO_2 concentration in the city of Manchester under Gauss's mathematical methods [1].

The ISCST3 model is a Gaussian dispersion model used to assess type the impact of single sources in the industry in the USA. The AERMOD model of the US EPA is used for polluting the complex terrain. The CALPUFF model was chosen by the USA to assess the impact of industry and transport.

In Vietnam, the modelling methods used the more common, especially in the current conditions of our country. The tangled diffusion model of Berliand and Sutton was used by Anh Pham Thi Viet to assess the environmental status of the atmosphere of Hanoi in 2001 by industrial discharges [2]. In 2014, Yen Doan Thi Hai has used models Meti-lis to calculate the emission of air pollutants from traffic and industrial activities in Thai Nguyen city [3].

2 Study area

Sources of air pollution are diverse. In the Da Nang city areas, main sources of pollution pressures include traffic, construction and industrial activities, peoples daily activities and waste treatment. The study area is Da Nang city in South Central of Vietnam. It is located between $15^0$15'-$16^0$40' northing and $107^0$17'-$108^0$20' easting and the area has more than 1285 km^2 (2018). Da Nang city has more than 1.2 million people (2018). Fig. 1 shows the study area. The city has a tropical monsoon climate with two seasons: a typhoon & wet season from September to March and a dry season from April to August. Temperatures are typically high, with an annual average of 25.9^0C (78.6^0F). Temperatures are highest between June and August (with daily highs averaging 33 to 34^0C (91 to 93^0F)), and lowest between December and February (highs averaging 24 to 25^0C (75 to 77^0F)). The annual average for humidity is 81%, with highs between October and December (reaching 84%) and lows between June and July (reaching 76–77%). The main means of transport within the city are motorbikes, buses, taxis, and bicycles. Motorbikes remain the most common way to move around the city. The growing number of cars tend to cause gridlock and contribute to air pollution.

With the rapid population growth rate, the infrastructure has not yet been fully upgraded, and some people are too aware of environmental protection. So, Da Nang city is currently facing a huge environmental pollution problem. The status of untreated wastewater flowing directly into the river system is very common. Many production facilities, hospitals and health facilities that do not have a wastewater treatment system are alarming.

Fig. 2 shows the geographical location of the monitoring stations. The coordinates system used in Fig. 2 is Universal Transverse Mercator (UTM).

3 Materials and Methods

The dataset is obtained from monitoring stations in Da Nang city with these parameters NO_2, SO_2, O_3, PM_{10}, TSP. Fig. 2 shows the map of monitoring sites in Da Nang city. The dust TSP data of passive air environment measures 15 stations in March 2016, and NO_2 is secondary parameter (see Table 1). I applied a geostatistical method to predict concentrations of air pollution at unobserved areas surrounding observed ones.

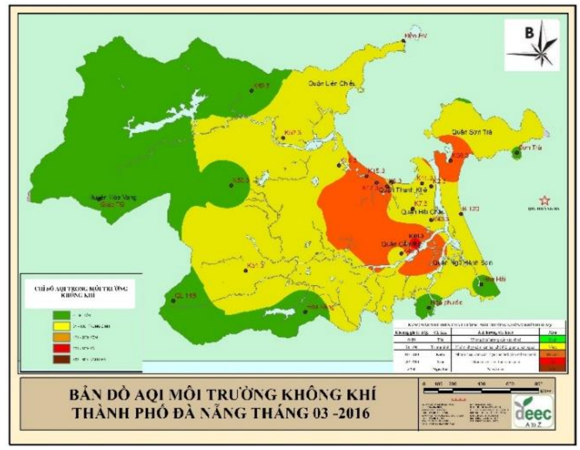

Figure 1 Passive gas monitoring map in March 2016, Da Nang city

Da Nang department of natural resources and environment

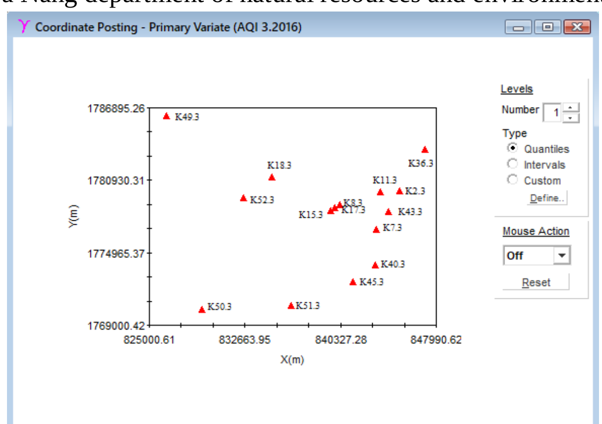

Figure 2 Map of monitoring sites in Da Nang city

Table 1 dust TSP data of passive air environment in march 2016

Station	X(m)	Y(m)	TSP (mg/m^3)	NO_2 (mg/m^3)
K2.3	845082.06	1780101.3	97.72	10.4
K7.3	843233.37	1776852.5	47.93	4.78
K8.3	840256.93	1778955.3	123.14	23.81
K11.3	843530.12	1779984.8	85.76	2.89
K15.3	839559.87	1778409	141.69	15.96
K17.3	839865.77	1778647.6	144.57	19.1
K18.3	834852.86	1781233.9	87.48	7.41
K36.3	847106.62	1783482.4	134.1	7.47
K40.3	843099.01	1773990.6	228.57	28.83
K43.3	844207.66	1778333	80.98	8.06
K45.3	841352.01	1772590.8	80.15	9.41
K49.3	826374.61	1786244.3	37.38	4.76
K50.3	829185.3	1770283.4	40.22	3.91
K51.3	836368.4	1770587.8	90.9	8.01
K52.3	832536.3	1779530.6	67.11	8.2

The main tool in geostatistics is the variogram which expresses the spatial dependence between neighbouring observations. The variogram can be defined as one-half the variance of the difference between the attribute values at all points separated by has followed [4]:

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

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

Under the second-order stationary conditions [5], one obtains:

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

and the covariance:

$$\begin{aligned} Cov[Z(s), Z(s+h)] &= E[(Z(s)-\mu)(Z(s+h)-\mu)] \\ &= E[Z(s)Z(s+h)-\mu^2] \qquad (2) \\ &= C(h) \end{aligned}$$

Then $Var[Z(s)] = C(0) = E[Z(s)-\mu]^2$

$$\gamma(h) = \frac{1}{2}E[Z(s)-Z(s+h)]^2 = C(0) - C(h)$$

The most commonly used models are spherical, exponential, Gaussian, and pure nugget effect (Isaaks & Srivastava,1989) [6]. The adequacy and validity of the developed variogram model is tested satisfactorily by a technique called cross-validation.

Crossing plot of the estimate and the true value shows the correlation coefficient R^2. The most appropriate variogram was chosen based on the highest correlation coefficient by trial and error procedure.

Kriging technique is an exact interpolation estimator used to find the best linear unbiased estimate. The best linear unbiased estimator must have a minimum variance of estimation error. We used ordinary kriging for spatial and temporal analysis, respectively. Ordinary kriging method is mainly applied for datasets without and with a trend, respectively.

The general equation of linear kriging estimator is

$$\hat{Z}(s_0) = \sum_{i=1}^{n} w_i Z(s_i) \qquad (3)$$

In order to achieve unbiased estimations in ordinary kriging the following set of equations should be solved simultaneously.

$$\begin{cases} \sum_{i=1}^{n} w_i \gamma(s_i, s_j) - \lambda = \gamma(s_0, s_i) \\ \sum_{i=1}^{n} w_i = 1 \end{cases} \qquad (4)$$

where $\hat{Z}(s_0)$ is the kriged value at location s_0, $Z(s_i)$ is the known value at location s_i, w_i is the weight associated with the data, λ is the Lagrange multiplier, and $\gamma(s_i, s_j)$ is the value of variogram corresponding to a vector with origin in s_i and extremity in s_j.

In fact, we can also use the multiple parameters in the relation to each other. We can estimate certain parameters, in addition to information that may contain enough by itself, one might use information of other parameters that have more details. Cokriging is simply an extension of auto-kriging in that it takes into account additional correlated information in the subsidiary variables. It appears more complex because the additional variables increase the notational complexity.

Suppose that at each spatial location s_i, $i = 1, 2, \ldots, n$ we observe k variables as follows:

$$\begin{array}{lll} Z_1(s_1) & Z_1(s_2) \cdots & Z_1(s_n) \\ Z_2(s_1) & Z_2(s_2) \cdots & Z_2(s_n) \\ \cdots & \cdots & \cdots \\ Z_k(s_1) & Z_k(s_2) \cdots & Z_k(s_n) \end{array}$$

We want to predict $Z_1(s_0)$, i.e. the value of variable Z_1 at location s_0.

This situation that the variable under consideration (the target variable) occurs with other variables (co-located variables) arises many times in practice and we want to explore the possibility of improving the prediction of variable Z_1 by taking into account the correlation of Z_1 with these other variables.

The predictor assumption:

$$\begin{aligned} \hat{Z}_1(s_0) = \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ji} Z_j(s_i) &= w_{11}Z_1(s_1) + \cdots \\ + w_{1n}Z_1(s_n) + w_{21}Z_2(s_1) &+ \cdots + w_{2n}Z_2(s_n) \qquad (5) \\ + \quad \cdots \quad \cdots & + \\ + w_{k1}Z_k(s_1) + \cdots &+ w_{kn}Z_k(s_n) \end{aligned}$$

We see that there are weights associated with variable Z_1 but also with each one of the other variables. We will examine ordinary cokriging, which means that $E[Z_j(s_i)] = \mu_j$ for all j and i. In vector form:

$$E[Z(s)] = \begin{pmatrix} E[Z_1(s)] \\ E[Z_2(s)] \\ \vdots \\ E[Z_k(s)] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} = \mu \qquad (6)$$

We want the predictor $\hat{Z}_1(s_0)$ to be unbiased, that is $E[\hat{Z}_1(s_0)] = \mu_1$. We take expectations of (5)

$$\begin{aligned}E[\hat{Z}_1(s_0)] &= \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ji}E[Z_j(s_i)] \\ &= w_{11}E[Z_1(s_1)] + \cdots + w_{1n}E[Z_1(s_n)] \\ &\quad + w_{21}E[Z_2(s_1)] + \cdots + w_{2n}E[Z_2(s_n)] \\ &\quad + \quad \cdots \quad \cdots \quad + \\ &\quad w_{k1}E[Z_k(s_1)] + \cdots + w_{kn}E[Z_k(s_n)]\end{aligned} \tag{7}$$

and using (6), we have

$$\begin{aligned}E[\hat{Z}_1(s_0)] &= w_{11}\mu_1 + \cdots + w_{1n}\mu_1 + w_{21}\mu_2 + \cdots + w_{2n}\mu_2 \\ &\quad + \quad \cdots \quad \cdots \quad + w_{k1}\mu_k + \cdots + w_{kn}\mu_k \\ &= \sum_{i=1}^{n} w_{1i}\mu_1 + \sum_{i=1}^{n} w_{2i}\mu_2 + \cdots + \sum_{i=1}^{n} w_{ki}\mu_k = \mu_1\end{aligned} \tag{8}$$

Therefore, we must have the following set of constraints:

$$\sum_{i=1}^{n} w_{1i} = 1, \quad \sum_{i=1}^{n} w_{2i} = 0, \cdots, \quad \sum_{i=1}^{n} w_{ki} = 0 \tag{9}$$

As with the other forms of kriging, cokriging minimizes the mean squared error of prediction (MSE):

$$\min \sigma_e^2 = E[Z_1(s_0) - \hat{Z}_1(s_0)]^2$$

or

$$\min \sigma_e^2 = E[Z_1(s_0) - \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ji}Z_j(s_i)]^2 \tag{10}$$

subject to the constraints:

$$\sum_{i=1}^{n} w_{1i} = 1, \quad \sum_{i=1}^{n} w_{2i} = 0, \cdots, \quad \sum_{i=1}^{n} w_{ki} = 0 \tag{11}$$

For simplicity, lets assume $k = 2$, in other words, we observe variables Z_1 and Z_2 and we want to predict Z_1. Therefore, from (10) (with $k = 2$) we have

$$\min \sigma_e^2 = E[Z_1(s_0) - \sum_{i=1}^{n} w_{1i}Z_1(s_i) - \sum_{i=1}^{n} w_{2i}Z_2(s_i)]^2 \tag{12}$$

From (9), we have $0 = \sum_{i=1}^{n} w_{2i} = \sum_{i=1}^{n} w_{2i}\mu_2$. Let's add the following quantities: $-\mu_1 + \mu_1 + \sum_{i=1}^{n} w_{2i}\mu_2$ on (12), we have:

$$\begin{aligned}\min \sigma_e^2 &= E[(Z_1(s_0) - \sum_{i=1}^{n} w_{1i}Z_1(s_i) - \sum_{i=1}^{n} w_{2i}Z_2(s_i) \\ &\quad -\mu_1 + \mu_1 + \sum_{i=1}^{n} w_{2i}\mu_2]^2\end{aligned} \tag{13}$$

or

$$\begin{aligned}\min \sigma_e^2 &= E[(Z_1(s_0) - \mu_1) - \sum_{i=1}^{n} w_{1i}[Z_1(s_i) - \mu_1] \\ &\quad - \sum_{i=1}^{n} w_{2i}[Z_2(s_i) - \mu_2]]^2\end{aligned} \tag{14}$$

We complete the square (14) to get:

$$\begin{aligned}&[Z_1(s_0) - \mu_1]^2 - 2\sum_{i=1}^{n} w_{1i}[Z_1(s_0) - \mu_1][Z_1(s_i) - \mu_1] \\ &-2\sum_{i=1}^{n} w_{2i}[Z_1(s_0) - \mu_1][Z_2(s_i) - \mu_2] \\ &+\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{1j}[Z_1(s_i) - \mu_1][Z_1(s_j) - \mu_1] \\ &+\sum_{i=1}^{n}\sum_{j=1}^{n} w_{2i}w_{2j}[Z_2(s_i) - \mu_2][Z_2(s_j) - \mu_2] \\ &+2[\sum_{i=1}^{n} w_{1i}[Z_1(s_i) - \mu_1]][\sum_{i=1}^{n} w_{2i}[Z_2(s_i) - \mu_2]]\end{aligned} \tag{15}$$

It can be shown that the last term of the expression (15) is equal to:

$$\begin{aligned}&2[\sum_{i=1}^{n} w_{1i}[Z_1(s_i) - \mu_1]][\sum_{i=1}^{n} w_{2i}[Z_2(s_i) - \mu_2]] \\ &= 2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{2j}[Z_1(s_i) - \mu_1][Z_2(s_j) - \mu_2]\end{aligned} \tag{16}$$

Find now the expected value of the expression (15):

$$\begin{aligned}&\min E[Z_1(s_0) - \mu_1]^2 - 2\sum_{i=1}^{n} w_{1i}E[Z_1(s_0) - \mu_1][Z_1(s_i) - \mu_1] \\ &-2\sum_{i=1}^{n} w_{2i}E[Z_1(s_0) - \mu_1][Z_2(s_i) - \mu_2] \\ &+\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{1j}E[Z_1(s_i) - \mu_1][Z_1(s_j) - \mu_1] \\ &+\sum_{i=1}^{n}\sum_{j=1}^{n} w_{2i}w_{2j}E[Z_2(s_i) - \mu_2][Z_2(s_j) - \mu_2] \\ &+2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{2j}E[Z_1(s_i) - \mu_1][Z_2(s_j) - \mu_2]\end{aligned} \tag{17}$$

We will denote the covariances involving Z_1 with C_{11}, the covariances involving Z_2 with C_{22}, and the cross-covariance between Z_1 and Z_2 with C_{12}. For example:

$$\begin{aligned}
&C[Z_1(s_0), Z_1(s_0)] = C_{11}(s_0, s_0) = C_{11}(0) = \sigma_1^2\\
&C[Z_1(s_0), Z_1(s_i)] = C_{11}(s_0, s_i)\\
&C[Z_1(s_i), Z_1(s_j)] = C_{11}(s_i, s_j)\\
&C[Z_1(s_i), Z_2(s_j)] = C_{12}(s_i, s_j)\\
&C[Z_1(s_0), Z_2(s_j)] = C_{12}(s_0, s_i)\\
&C[Z_2(s_i), Z_1(s_j)] = C_{21}(s_i, s_j)\\
&C[Z_2(s_i), Z_2(s_j)] = C_{22}(s_i, s_j)
\end{aligned} \tag{18}$$

The expectations on (17) are the covariance.
Finally, with the Lagrange multipliers we get:

$$\begin{aligned}
&\min \sigma_1^2 - 2\sum_{i=1}^{n} w_{1i}C_{11}(s_0, s_i) - 2\sum_{i=1}^{n} w_{2i}C_{12}(s_0, s_i) +\\
&\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{1j}C_{11}(s_i, s_j) + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{2i}w_{2j}C_{22}(s_i, s_j) +\\
&2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{1i}w_{2j}C_{12}(s_i, s_j) - 2\lambda_1[\sum_{i=1}^{n} w_{1i} - 1]\\
&-2\lambda_2[\sum_{i=1}^{n} w_{2i} - 0]
\end{aligned} \tag{19}$$

The unknowns are the weights $w_{11}, w_{12}, \ldots, w_{1n}$ and $w_{21}, w_{22}, \ldots, w_{2n}$ and the two Lagrange multipliers λ_1 and λ_2. We take the derivatives with respect to these unknowns and set them equal to zero.

$$\begin{aligned}
&-2C_{11}(s_0, s_i) + 2\sum_{j=1}^{n} w_{1j}C_{11}(s_i, s_j)\\
&+2\sum_{j=1}^{n} w_{2j}C_{12}(s_i, s_j) - 2\lambda_1 = 0,\ \forall i = 1, \ldots, n
\end{aligned} \tag{20}$$

$$\begin{aligned}
&-2C_{12}(s_0, s_i) + 2\sum_{j=1}^{n} w_{2j}C_{22}(s_i, s_j)\\
&+2\sum_{j=1}^{n} w_{1j}C_{21}(s_i, s_j) - 2\lambda_2 = 0,\ \forall i = 1, \ldots, n
\end{aligned} \tag{21}$$

$$\sum_{i=1}^{n} w_{1i} = 1, \sum_{i=1}^{n} w_{2i} = 0$$

Put

$$[C_{11}] = \begin{pmatrix} C_{11}(s_1, s_1) & \cdots & C_{11}(s_1, s_n) \\ \vdots & \vdots & \vdots \\ C_{11}(s_n, s_1) & \cdots & C_{11}(s_n, s_n) \end{pmatrix};$$

$$[C_{12}] = \begin{pmatrix} C_{12}(s_1, s_1) & \cdots & C_{12}(s_1, s_n) \\ \vdots & \vdots & \vdots \\ C_{12}(s_n, s_1) & \cdots & C_{12}(s_n, s_n) \end{pmatrix};$$

$$[C_{21}] = \begin{pmatrix} C_{21}(s_1, s_1) & \cdots & C_{21}(s_1, s_n) \\ \vdots & \vdots & \vdots \\ C_{21}(s_n, s_1) & \cdots & C_{21}(s_n, s_n) \end{pmatrix};$$

$$[C_{22}] = \begin{pmatrix} C_{22}(s_1, s_1) & \cdots & C_{22}(s_1, s_n) \\ \vdots & \vdots & \vdots \\ C_{22}(s_n, s_1) & \cdots & C_{22}(s_n, s_n) \end{pmatrix}$$

$$[1] = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}; \quad [0] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \quad W_1 = \begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{1n} \end{pmatrix}; \quad W_2 = \begin{pmatrix} w_{21} \\ w_{22} \\ \vdots \\ w_{2n} \end{pmatrix};$$

$$[C_{11}(s_0, s_i)] = \begin{pmatrix} C_{11}(s_0, s_1) \\ \vdots \\ C_{11}(s_0, s_n) \end{pmatrix}; [C_{12}(s_0, s_i)] = \begin{pmatrix} C_{12}(s_0, s_1) \\ \vdots \\ C_{12}(s_0, s_n) \end{pmatrix}$$

$$[1]' = (1\,1 \cdots 1)\,;\ [0]' = (0\,0 \cdots 0)$$

where the matrix [1], [0] have dimensions n × 1.
We get the following cokriging system in matrix form:

$$\begin{pmatrix} [C_{11}] & [C_{12}] & [1] & [0] \\ [C_{21}] & [C_{22}] & [0] & [1] \\ [1]' & [0]' & 0 & 0 \\ [0]' & [1]' & 0 & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ -\lambda_1 \\ -\lambda_2 \end{pmatrix} = \begin{pmatrix} [C_{11}(s_0, s_i)] \\ [C_{12}(s_0, s_i)] \\ 1 \\ 0 \end{pmatrix}$$

Put

$$G = \begin{pmatrix} [C_{11}] & [C_{12}] & [1] & [0] \\ [C_{21}] & [C_{22}] & [0] & [1] \\ [1]' & [0]' & 0 & 0 \\ [0]' & [1]' & 0 & 0 \end{pmatrix}; w = \begin{pmatrix} W_1 \\ W_2 \\ -\lambda_1 \\ -\lambda_2 \end{pmatrix};$$

$$c = \begin{pmatrix} [C_{11}(s_0, s_i)] \\ [C_{12}(s_0, s_i)] \\ 1 \\ 0 \end{pmatrix}$$

We have Gw = c

where $\forall i = 1, 2, \ldots, n$, $C_{12}(h)$ may not be the same as $C_{21}(h)$, $h = |s_i - s_j|$. This is because of definition of cross-covariance: $C_{12}(h) = \mathbb{E}\{[Z_1(s) - \mu_1][Z_2(s+h) - \mu_2]\}$ and $\hat{C}_{12}(h) = \frac{1}{N(h)}\sum Z_1(s)Z_2(s+h) - \hat{\mu}_1\,\hat{\mu}_2$, obviously, $\hat{C}_{21}(h) = \frac{1}{N(h)}\sum Z_2(s)Z_1(s+h) - \hat{\mu}_2\,\hat{\mu}_1$

is not necessarily equal to $\hat{C}_{12}$.

The Cokriging system is written as Gw = c, where the vector w, c have dimensions (2n + 2) × 1 and the matrix G has dimensions (2n + 2) × (2n + 2). The weights will be obtained by $w = G^{-1}c$.

The GS+ software (version 5.1.1) was used for geostatistical analysis in this study (Gamma Design Software, 2001) [7].

4 Results and Discussions

In order to check the anisotropy in the dust pollution TSP, the conventional approach is to compare variograms in several directions (Goovaerts,1997) [8]. In this study major angles of 0^0, 45^0, 90^0, and 135^0 with an angle tolerance of $\pm 45^0$ were used for detecting anisotropy.

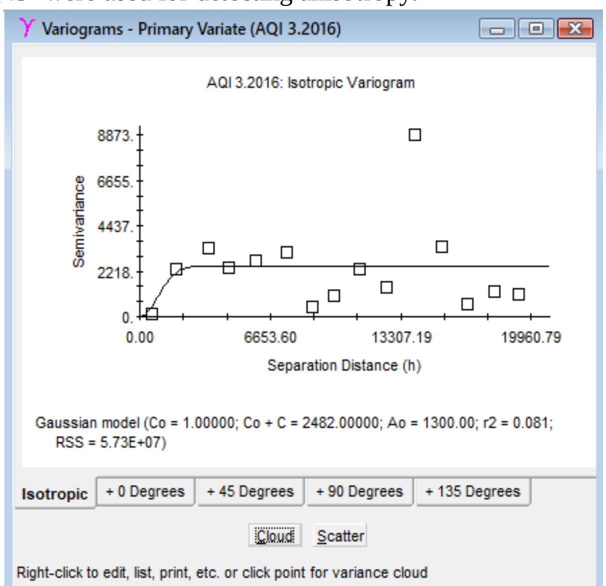

Figure 3 Isotropic variogram values of the dust TSP

Fig. 3 shows fitted variogram for spatial analysis of the dust TSP. Through Semi-variance map of parameter TSP, the model of isotropic is suitable. The variogram values are presented in Table 2.

Table 2 isotropic variogram values of the dust TSP

	Nugget	Sill	Range	r^2	RSS
Linear	2106	2499	19295	0.03	6.02E+07
Gaussian	1	2482	2252	0.081	5.73E+07
Spherical	1	2479	2930	0.078	5.76E+07
Exponetial	1	2481	3480	0.07	5.83E+07

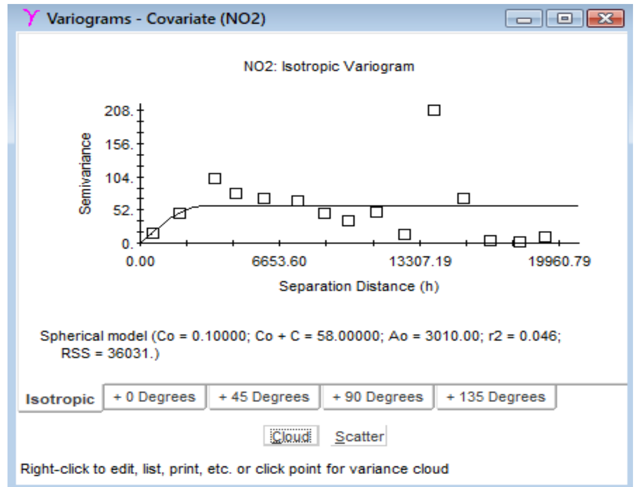

Figure 4 Isotropic variogram values of NO_2

Fig. 4 shows fitted variogram for spatial analysis of NO_2. Through Semi-variance map of parameter NO_2, the model of isotropic is suitable. The variogram values are presented in Table 3.

Table 3 Isotropic variogram values of NO_2

	Nugget	Sill	Range	r^2	RSS
Linear	54	54	19295	0.0	37749
Gaussian	1	57.8	2234	0.045	36057
Spherical	0.1	58	3010	0.046	36031
Exponetial	0.1	57.5	2760	0.041	36302

Fig. 5 shows fitted variogram for spatial analysis of TSP and NO_2.

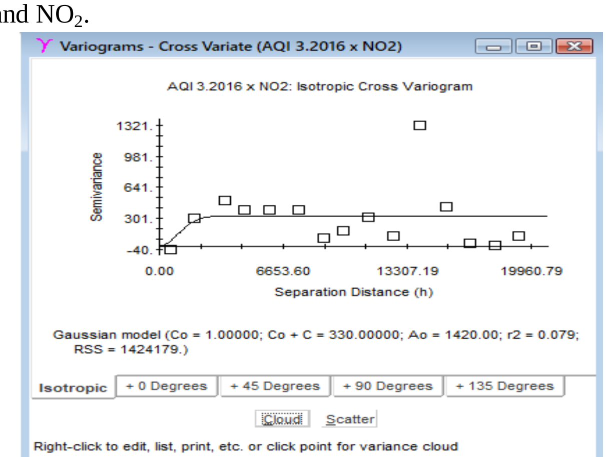

Figure 5 Isotropic variogram values of TSP and NO_2

Through Semi-variance map of these two parameters, the model of isotropic is suitable. The variograms values are presented in Table 4.

Table 4 Isotropic variogram values of tsp and NO_2

	Nugget	Sill	Range	r^2	RSS
Linear	302	302	19295	0.0	1539545
Gaussian	1	330	2460	0.079	1424179

Spherical	1	329	3270	0.076	1433748
Exponetial	1	327	3510	0.068	1452090

Model Testing: The credible result of model selection using appropriate interpolation is expressed in Table 5 by coefficient of regression, coefficient of correlation and interpolated values, in addition to the error values as the standard error (SE) and the standard error prediction (SE Prediction).

Table 5 Testing the model parameters

Coefficient regression	Coefficient correlation	SE	SE Prediction
1.026	1	0.001	0.141

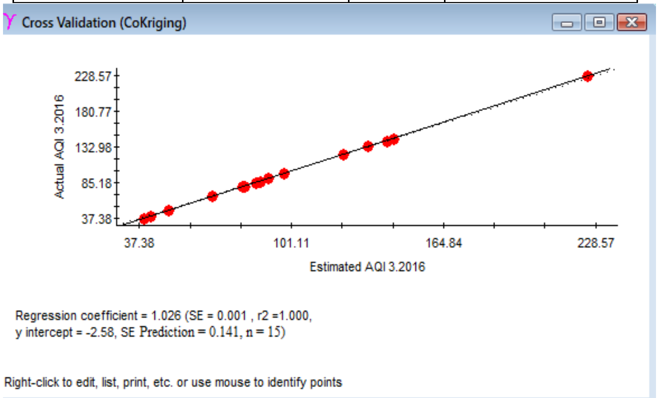

Figure 6 Error testing result of prediction TSP

Fig. 6 shows results of testing of error between real values and the estimated values by the model by cokriging method with isotropic TSP parameter and isotropic NO_2 secondary parameter. Coefficients of regression and the coefficient of correlation are close to 1, where the error values is small (close to 0) indicates that the selected model is a suitable interpolation in Fig. 7.

(CoKriging)

Record	X-Coordinate	Y-Coordinate	Actual Z	Estimated Z	Error (E-A)
1	845082.06	1780101.30	97.72	97.72	0.00
2	843233.37	1776852.54	47.93	49.25	1.32
3	840256.93	1778955.33	123.14	122.78	-0.36
4	843530.12	1779984.80	85.76	86.10	0.34
5	839559.87	1778409.00	141.69	140.87	-0.82
6	839865.77	1778647.64	144.57	143.82	-0.75
7	834852.86	1781233.91	87.48	87.77	0.29
8	847106.62	1783482.41	134.10	133.20	-0.90
9	843099.01	1773990.63	228.57	225.16	-3.41
10	844207.66	1778332.97	80.98	81.42	0.44
11	841352.01	1772590.76	80.15	80.63	0.48
12	826374.61	1786244.26	37.38	38.97	1.59
13	829185.30	1770283.42	40.22	41.73	1.51
14	836368.40	1770587.84	90.90	91.10	0.20
15	832536.30	1779530.61	67.11	67.92	0.81

Figure 7 Cross-Validation (Cokriging) of TSP

From Fig. 8 and Fig. 9, we see that, in March 2016 at K49.3 neighborhood has low pollution levels, due to transport and less population density. The process of urbanization has not developed as today. Neighborhood of K40.3 have high pollution levels, so at this point density traffic caused high proportion in pollution. This is one of the focal areas of the city. It is the intersection of districts and there are many roads with crowded transport volume. The process of urbanization is growth.

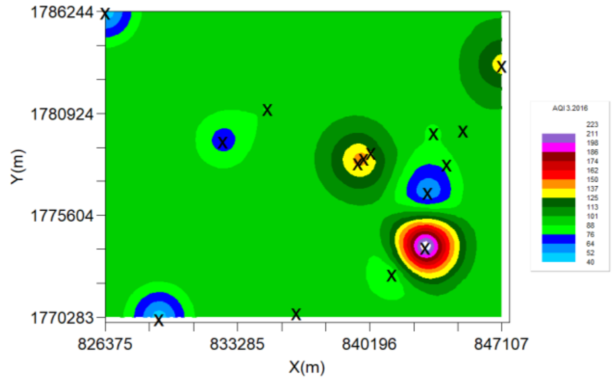

Figure 8 2D Cokriging Interpolation Map of TSP

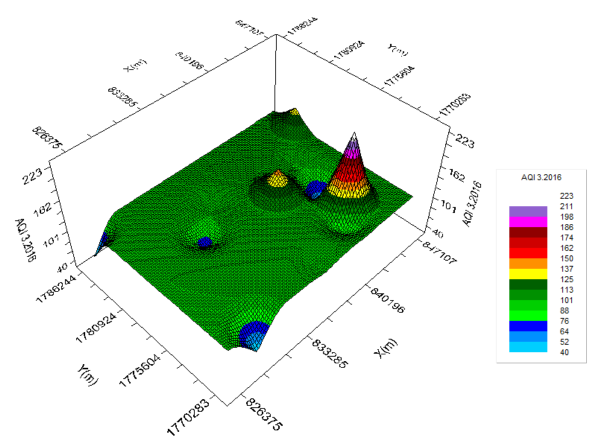

Figure 9 3D Cokriging Interpolation Map of TSP

Based on the map, we can also forecast the dust concentration in the city near the air monitoring locations and to offer solutions to overcome. The mentioned method of applied geostatistics to predict air pollution concentrations TSP in Da Nang city showed that the forecast regions closer together have the forecast deviations as small Fig. 10, meanwhile further areas contribute the higher deviation. Through this forecast case study using spatial interpolation based methods and models, we can predict air pollution levels for regions that have not been installed air monitoring sites, from which proposed measures to improve the air quality can be taken into account.

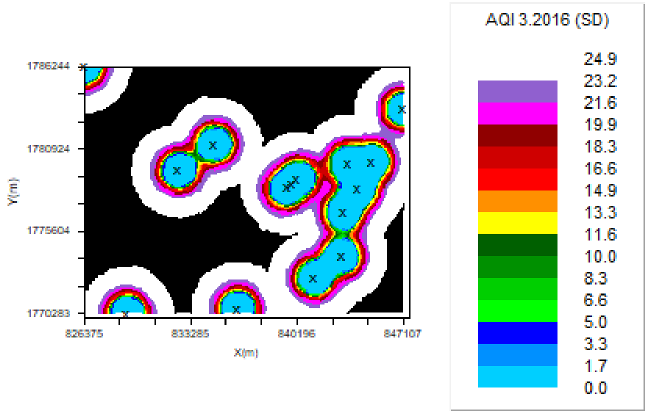

Figure 10 Estimated error by CoKriging method of TSP

As we can see from the forecast maps, forecast for the region's best results in areas affected 22990m, located outside the affected region on the forecast results can be inaccurate. If the density of monitoring stations is high and the selection of interpolation models is easier, interpolation results have higher reliability and vice versa. The middle area represents key outcomes of computation on data. The different colors represent different levels of pollution. The lowest pollution level is blue and the highest is white. Regions having the same color likely are in the same levels of pollution.

5 Conclusion

Geostatistical applications to forecast the dust TSP concentrations in Da Nang city gave the result with almost no error difference between the estimated values and the real values. Therefrom, the study showed that efficacy and rationality with high reliability of theoretical Geostatistical to building spatial prediction models are suitable. When building the model we should pay attention to the values of the model error, data characteristic of the object. We also looked at the result of the model selection which aimed to choose the most suitable model for real facts, since distinct models provide different accuracies. Therefore, experiencing the selected model also plays a very important role in the interpolation results. According to the World Meteorological Organization (WMO) and United Nations Environment Program (UNEP), the world currently has 20 types of computation models and forecasts of air pollution. The air pollution computation models include AERMOD (AMS/EPA Regulatory Model) of the US-EPA for polluting the complex terrain. For this data, we study only the key parameters of pollution, and lack of many
The paper's author expresses his sincere thank to Dr. Man NV Minh Department of Mathematics, Faculty of Science, Mahidol University, Thailand and Dr. Dung Ta Quoc Faculty of Geology and Petroleum Engineering, Vietnam.
parameters such as temperature, wind, height of site... when applying kriging interpolation to predict. In this case, the model AERMOD (US-EPA) would not be appropriate.

Air pollution simulation of Anh Pham The and Hieu Nguyen Duy is use the AERMOD model need a lot parameters like wind direction, temperature, humidity, precipitation, cloud cover... Anh Pham Thi Viet uses tangled diffusion model of Berliand and Sutton to assess the environmental status of the atmosphere of Hanoi in 2001 to several parameters such as: the level of pollution, the location coordinates, wind speed, altitude, weather [2]. In summary, previous studies to simulate air pollution needs to be more parameters related parts, while was not envisaged that the application space, the data set in this paper on the research has not performed. Within Vietnam, there are no studies that use spatial interpolation methods as in my article. Method of air pollution forecast that I present in this article reflect the spatial correlation between air monitoring stations with parameters: pollution and geographical coordinates, which previous studies have not performed.

Finally a comparison of the proposed method with several other methods can be made as follows. Polygon (nearest neighbor) method has advantages such as easy to use, quick calculation in 2D; but also possesses many disadvantages as discontinuous estimates; edge effects/sensitive to boundaries; difficult to realize in 3D. The Triangulation method has advantages as easy to understand, fast calculations in 2D; can be done manually, but few disadvantages are triangulation network is not unique. The use of Delaunay triangles is an effort to work with a "standard" set of triangles, not useful for extrapolation and difficult to implement in 3D. Local sample mean has advantages are easy to understand; easy to calculate in both 2D and 3D and fast; but disadvantages possibly are local neighborhood definition is not unique, location of sample is not used except to define local neighborhood, sensitive to data clustering at data locations. This method does not always return answer valuable. This method is rarely used.

Similarly, the inverse distance method are easy to understand and implement, allow changing exponent adds some flexibility to method's adaptation to different estimation problems. This method can handle anisotropy; but disadvantages are difficulties encountered when point to estimate coincides with data point (d=0, weight is undefined), susceptible to clustering.

Acknowledgment

Furthermore, I greatly appreciate the anonymous reviewer whose valuable and helpful comments led to significant improvements from the original to the final version of the article.

References

1. Robert Angus Smith, “On the Air of Towns”, *Journal of the Chemical Society*, 9, pp. 196-235, 1859.
2. Anh Pham Thi Viet, “Application of airborne pollutant emission models in assessing the current state of the air environment in Hanoi area caused by industrial sources”, *6th Women's Science Conference*, Ha Noi national university, pp. 8-17, 2001.
3. Yen Doan Thi Hai, “Applying the Meti-lis model to calculate the emission of air pollutants from traffic and industrial activities in Thai Nguyen city, orienting to 2020”, *Journal of Science and Technology*, Volume 106 No. 6, Thai Nguyen university, 2013.
4. S.H. Ahmadi and A.Sedghamiz, “Geostatistical analysis of Spatial and Temporal Variations of groundwater level”, *Environmental Monitoring and Assessment*, 129, 277-294, 2007.
5. R.Webster and M.A. Oliver, *Geostatistics for Enviromental Scientists*, 2nd Edition, John Wiley and Sonc LTD, The Atrium, Southern Gate, Chichester, West Sussex PO19, England, 6-8, 2007.
6. E.Isaaks and M.R. Srivastava, An introduction to *applied geostatistics*, New York: Oxford University Press, 1989.
7. Gamma Design Software, *GS+ Geostatistics for the Environmental Science*, version 5.1.1, Plainwell USA: MI, 2001.
8. P.Goovaerts, *Geostatistics for natural resources Evaluation*, New York: Oxford University Press, 1997.

Ứng dụng phương pháp nội suy Cokriging để dự báo chỉ số chất lượng không khí cho nồng độ bụi TSP thành phố Đà Nẵng

Nguyễn Công Nhựt*, Lai Văn Phút, Bùi Hùng Vương

Khoa Công nghệ thông tin, Trường Đại học Nguyễn Tất Thành, Việt Nam
*ncnhut@ntt.edu.vn

Tóm tắt Việc lập bản đồ để dự đoán nồng độ ô nhiễm không khí ở thành phố Đà Nẵng là một vấn đề cấp bách đối với các cơ quan quản lí và các nhà nghiên cứu về ô nhiễm môi trường. Mặc dù mô phỏng về vị trí không gian đã trở nên phổ biến, nó sử dụng các phương thức nội suy cổ điển với độ tin cậy thấp. Dựa trên sự phân bố các trạm quan trắc chất lượng không khí nằm trong khu công nghiệp, khu dân cư, trục giao thông ... và nguồn ô nhiễm không khí, ứng dụng các lí thuyết địa chất, nghiên cứu này trình bày kết quả lựa chọn phương pháp nội suy Cokriging dự báo ô nhiễm ở thành phố Đà Nẵng với độ tin cậy cao. Trong bài viết này, tôi sử dụng nồng độ TSP được ghi nhận (một trong những ô nhiễm không khí chính gây ra tại các đô thị lớn) tại một số trạm quan sát ở thành phố Đà Nẵng, sử dụng phương pháp nội suy Cokriging để tìm mô hình phù hợp, sau đó dự báo nồng độ bụi TSP tại một số trạm không có dữ liệu quan trắc trong thành phố. Đóng góp quan trọng của tôi là tìm kiếm các mô hình thống kê tốt theo một số tiêu chí, sau đó tìm các mô hình phù hợp với độ chính xác cao.

Từ khóa Ô nhiễm không khí, địa lí, Cokriging, variogram