EnviroInfo 2009 (Berlin)
Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools
 ISBN: 978-3-8322-8397-1

Using geostatistics and clustering to design and optmize the environmental monitoring network for Hai Duong province (Vietnam)

Vu Van Manh[1], Bui Phuong Thuy[2]

Abstract

This paper presents some mathematical modeling approaches for designing an environmental monitoring network (EMN). As an example application region, Hai Duong province was chosen. The results of this paper include the general picture of air quality in Hai Duong province, a preliminary air quality monitoring network with 20 monitoring stations and a dendrogram for optimizing the network. The works proved that geostatistics and cluster analysis can be used for designing and optimizing the EMN from scratch. The procedure to generate a monitoring network includes: i) developing an air quality index (AQI) from historical data or surveying data; ii) determining practical range (optimum distance between monitoring station) from the variogram with AQI and design of the preliminary monitoring network; iii) operate the preliminary network; iv) using monitored data of the preliminary network to generate a dendrogram for optimizing network; v) re-arrange EMN.

Keywords: *Geostatistics, Cluster Analysis, Environmental Monitoring*

1. Introduction

Environmental Monitoring Networks (EMN) has become a major of interest at all scales from local to global. They are often complex systems which usually monitor a wide range of substances and have high requirements of accuracy. The best way to obtain environmental status is to take sample at all possible locations within a specific area of interest, but in practice the design of a EMN is limited by economic and operational constraints. In such cases the spatial distribution and optimum number of monitoring stations where samples are to be taken has to be determined.

The EMNs are usually established based on experience, heuristic rules, population density, topography, etc. It is necessary to find out optimum criteria for designing optimum EMN, determine the location of monitoring stations that is representative for environmental status of a large area and the monitored data of one monitoring station can be integrated with data of other stations for interpolation and extrapolation purposes. For this purpose, geostatistics can be used as a supporting tool for traditional heuristic rules to design a new EMN. Due to the development of environment and human activities, all monitoring stations in a EMN should be checked and, if necessary, re-arranged (S.Lophaven 2004). In this paper an appropriate approach is introduced in order to combine or reduce the number of monitoring stations.

The objective of this paper is to introduce and develop such an optimum Air Quality Monitoring Network (AQMN) for Hai Duong Province (Vietnam). The SGeMS software was used to generate variograms from surveying data taken by potable sampling devices (PSDs) then common GIS software was used to generate a picture of air pollution and initiate a preliminary AQMN. After the preliminary AQMN is put

1 Dr. Vu Van Manh, Research Center for Environmental Monitoring & Modeling, Faculty of Environmental Sciences, Hanoi University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai str., Thanh Xuan dist., Hanoi city, VIETNAM. Email: fesvvm@yahoo.com
2 M.Sc. Bui Phuong Thuy, Doctoral Student of EMAU Greifswald, Friedrich-Ludwig-Jahn-Straße 17a, 17487 Greifswald. Email: phuongthuy1205@yahoo.com

into service for a certain period of time, their monitored data will be used to determine the similarity between monitoring stations then generate a dendrogram to combine, reduce or re-arrange monitoring stations for optimizing the AQMN.

2. Methodology

Approach 1. Develop air quality index (AQI)

Each monitoring station usually monitors different parameters that cause difficult to compare the environmental quality between their locations. The monitored values should be rescaled so that they have similar ranges and their squared values do not exceed the computer's precision (particularly when working with different attribute values expressed in widely different unit scales). Common Rescale formula uses the minimum and maximum data value.

In order to get the general picture of the air pollution (overall assessment), the air quality index (AQI) has been developed (N.X. Thinh & V.V. Manh 2007, 2008) as following: Because the air quality gets better at low concentration of the monitored pollutants, the monitored value X_{ij} of the pollutant i at the location A_j are standardized:

$$Z_{ij} = \frac{X_{ij} - \min\limits_{k\in\{1,2,\ldots,n\}} X_{ik}}{\max\limits_{k\in\{1,2,\ldots,n\}} X_{ik} - \min\limits_{k\in\{1,2,\ldots,n\}} X_{ik}} \qquad (1)$$

Then the air quality index AQI_j at the location A_j is to be calculated by:

$$AQI_j = \sum_{i=1}^{n} w_i * Z_{ij} \qquad (2)$$

where n is the number of samples, w_i is the relative weight of the pollutant i with $w_i > 0$ and $\Sigma w_i = 1$.
The lower the value of AQI_j, the better is the air quality at the location A_j.

Approach 2. Derive a mathematical model of spatial variability for determining optimum distance between monitoring station

In order to use spatial variability of historical data as a basis for estimating its value at other locations, we used variogram model that approximates the measured variability of the available samples. Geostatistician have developed many model types to better represent the results of variogram analysis (S.W.Houlding 2000). In this paper, the SGeMS software was used to generate variogram with spherical model (Matheron 1970) (G. Bohling 2007):

$$\begin{aligned} &\gamma(h) = C_0 && for \quad h = 0 \\ &\gamma(h) = C_0 + C\left(\frac{3a}{2h} - \frac{a^3}{2h^3}\right) && for \quad 0 < h \le a \\ &\gamma(h) = C_0 + C && for \quad h > a \end{aligned} \qquad (3)$$

Where: γ is the semi-variance to distance h between sample, C_0 is the inherent random variability, C is sill of the variograms model, a is range of influence or practical range at which monitored data become independent of one another.

Approach 3. Optimize air quality monitoring network by cluster analysis

The similarity between monitoring stations in the monitoring network is determined by calculating the average distance between each pair of monitored data of each station. The averaging is performed over all pairs (x, y) of monitored pollutants, where x is a monitored pollutant from the first station, y is a monitored pollutant from the second station. The linkage function is described by following expression:

$$D(X,Y) = \frac{1}{N_X . N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} d(x_i, y_j) \qquad (4)$$

$$x_i \in X, y_i \in Y$$

where: d(x, y) is the distance between monitored pollutants $x \in X$ and $y \in Y$

X and Y are two data sets of monitored pollutants (at each station)

N_X and N_Y are the numbers of monitored pollutants in each station X and Y respectively

The distance d(x, y) is measured by squared Euclidean metric:

$$d(x,y) = \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (5)$$

The optimization process uses an agglomerative hierarchical methods that begins with all monitoring station being separate, each forming its own cluster. In the first step, the two monitoring station which have data sets closest together are joined. In the next step, either a third monitoring station joins the first two, or two other monitoring stations join together into a different cluster. This process will continue until all clusters (monitoring stations) are join into one.

3. Design and optimization of an air quality monitoring network for Hai Duong province

3.1 Overview

The objective of this paper is – as described in the introduction – the development of the air quality monitoring network for Hai Duong province by using mathematical tools.

Hai Duong province is located in the center of the Red River Delta in the northern of Vietnam with total area of 1,662 km^2 and population of 1,698,462. In recently years, the urban development and industrialization in Hai Duong are growing very fast and, of course, causing the degradation of living environment including the air quality. The concentration and influenced areas of air pollution tend to increase year by year, especially in areas close to pollution sources (V.V. Manh, N.T.H. Hanh 2007). In order to control the air quality of Hai Duong province as well as to support the decision makers, the air monitoring network should be designed and optimized to meet these urgent needs.

3.2 Assessing air quality status of Hai Duong province by air quality index (AQI)

The first step in design a monitoring network is to collect historical data of the parameter to be monitored. In order to have this data, a survey has been carried to collect these information at 35 sampling locations that are even distributed in whole of Hai Duong province. The monitored parameters include PM_{10}, SO_2, NO_x, CO were taken by portable sampling devices (PSDs) (P.N. Ho & V.V. Manh 2007).

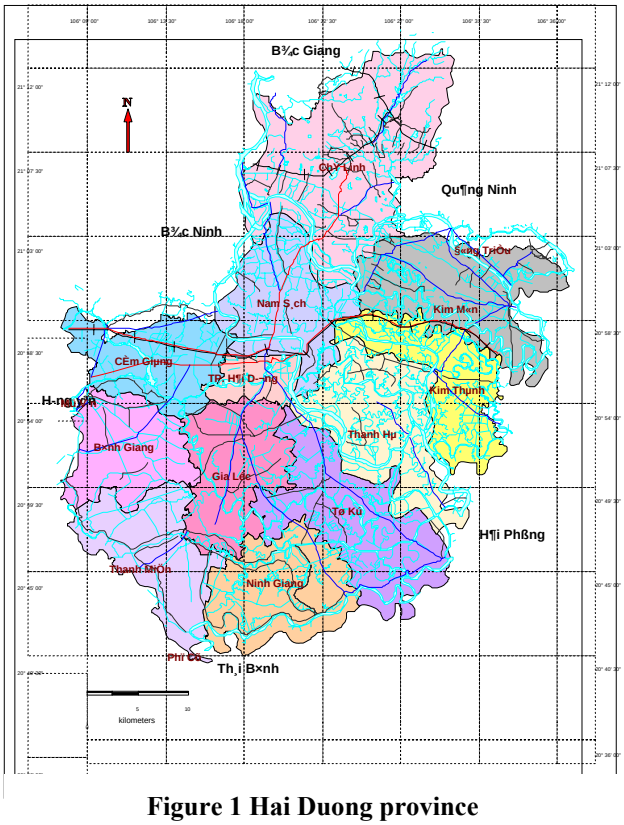

Figure 1 Hai Duong province

The result shows that almost of the surveying locations have concentration of PM10 under the Vietnamese Standard for Ambient Air Quality (TCVN 5937-2005 in 24 hours average), except at location 1 and 3 (Fig. 2). The reason is that these locations are close to National Highway No.5 that runs across Hai Duong province.

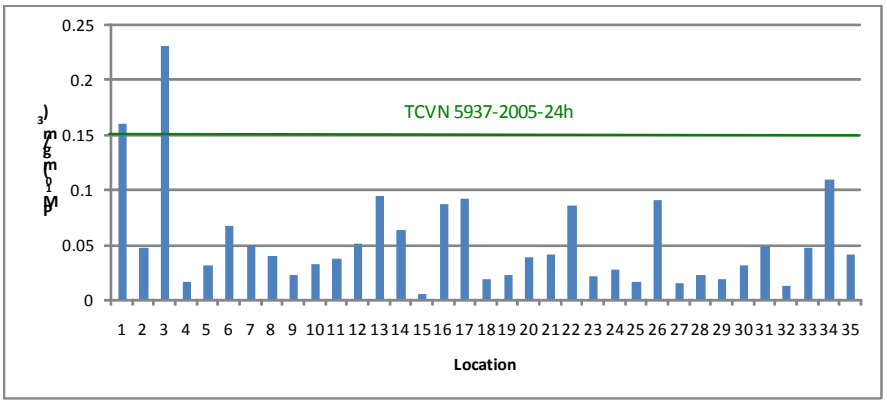

Figure 2 Concentration of PM10 (mg/m3) at 35 surveyed locations

Figure 3 shows that the concentration of SO_2 at 13 surveying locations: 1, 2, 3, 4, 7, 8, 13, 14, 28, 29, 30, 31, 34 are higher than the Vietnamese standard for SO_2 in 24 hours average. Almost of these location are close to highway and national highway. In comparison with Vietnamese standard TCVN 5937-2005 in 1 hour average, all of the surveying locations have concentration of SO_2 under the standard.

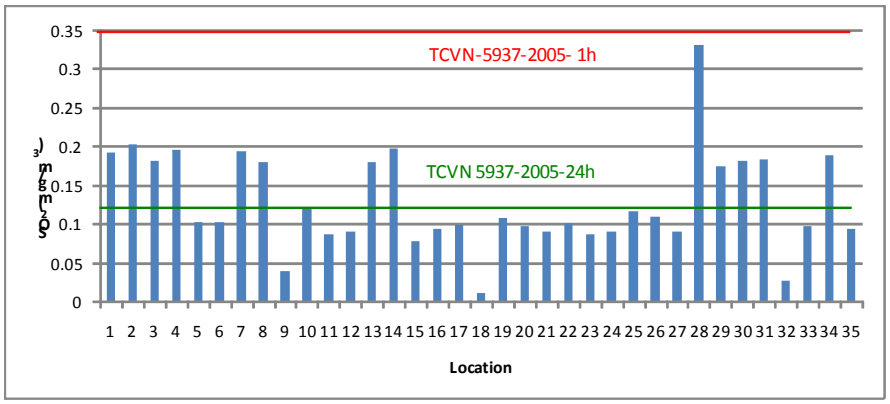

Figure 3 Concentration of SO2 (mg/m3) at 35 surveyed locations

The air at all surveying locations is not polluted by CO and NO_x. The concentrations of CO and NO_x are very low as compared with Vietnamese standards.
In order to have an overall assessment of the air quality in Hai Duong province, the AQI were used. Fig. 4 presents the 35 surveying locations and the general picture of air quality in Hai Duong province in AQI value (calculated for the air pollutants: SO_2, NO_x, CO and PM_{10}) which was interpolated by Kriging method in ArcGIS.

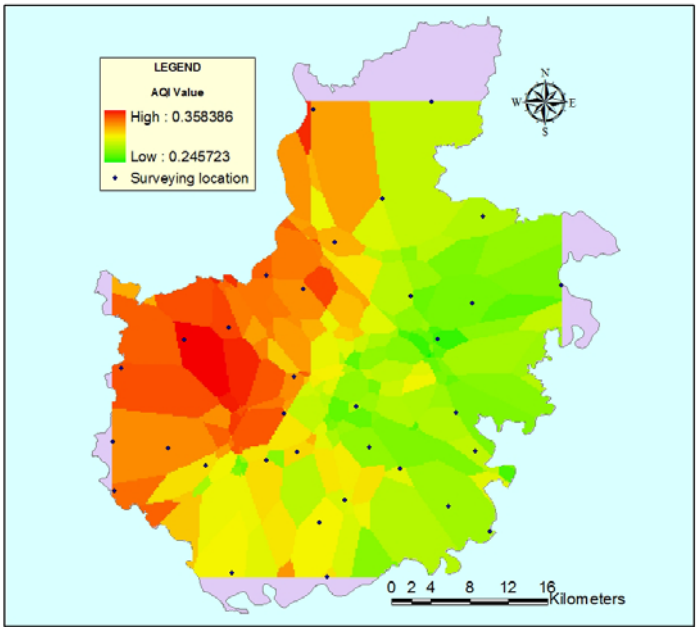

Figure 4 Spatial distribution and AQI of the 35 surveyed locations in Hai Duong 2006 (grid map)

3.3 Determining the practical range for designing the air quality monitoring network

Because γ in the variogram is the semi-variance of the data to distance h between sample so that the optimum distance between monitoring stations, when design the monitoring network, should not greater than

the practical range (or so called effective range) determined by variogram (G. Bohling 2007). In this paper, the AQIs (calculated from historical surveying data at 35 locations) were used and the variogram was calculated by SGeMS with 30 lags and leg separation of 1000 m.

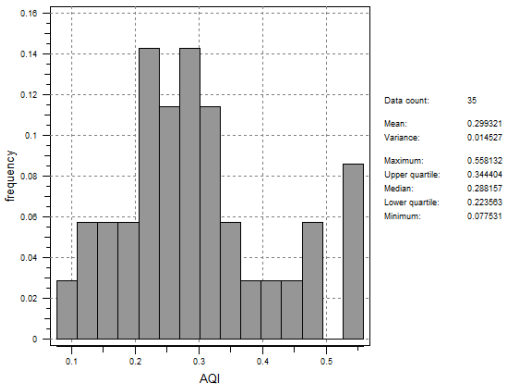

Figure 5 Histogram of AQI

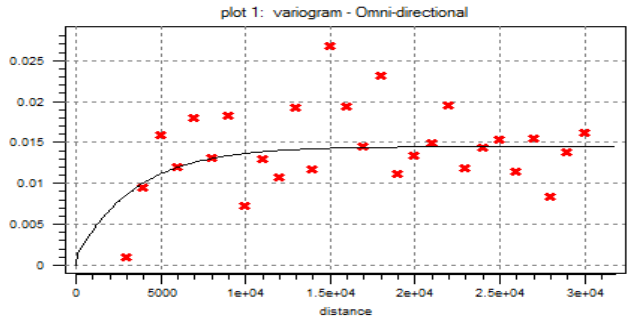

Figure 6 Variogram of AQI

From the variogram (Figure 6) we can find the practical range is of 10800 m. The next step in design the air quality monitoring network is done by a grid with cell sizes of 10800 m x 10800 m. The total amount of monitoring stations should not more than 20 station as indicated in the Figure 7.

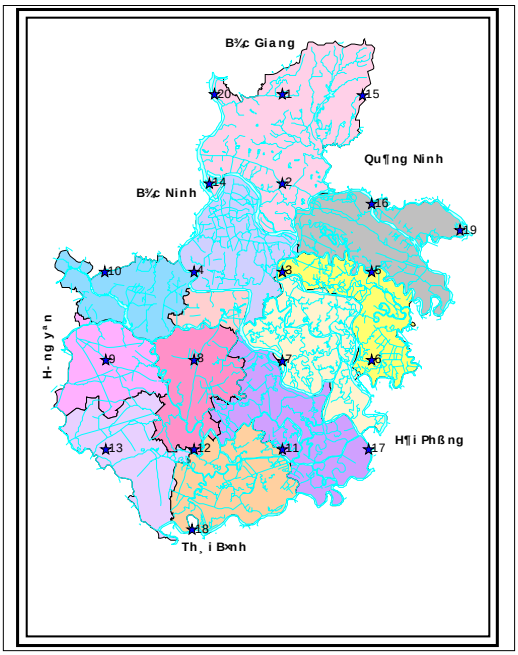

Figure 7 Preliminary air quality monitoring network of Hai Duong province

3.4 Optimisation of the air quality monitoring network for Hai Duong province

The optimization of the monitoring network can be done after a certain period of operation time of the network with enough data for the calculation. Because the preliminary air monitoring network defined in section 3.3. has not been put into fully service yet, so this paper simulated the optimization process by using data of the PSDs at these monitoring locations.

The result of cluster analysis in MiniTab with average linkage method and squared Euclidean distance, is indicated in Table 1 and dendrogram in Figure 8.

Step	Number of stations	Similarity level	Distance level	Stations joined		New station	Number of stations in New station
1	19	100.000	0.004	15	16	15	2
2	18	100.000	0.004	3	5	3	2
3	17	100.000	0.006	13	17	13	2
4	16	99.999	0.008	8	9	8	2
5	15	99.999	0.017	6	15	6	3
6	14	99.998	0.026	4	6	4	4
7	13	99.997	0.035	11	18	11	2
8	12	99.926	1.012	4	14	4	5
9	11	99.922	1.061	19	20	19	2
10	10	99.708	4.004	3	10	3	3
11	9	99.705	4.043	8	11	8	4
12	8	99.436	7.723	4	19	4	7
13	7	99.415	8.013	3	13	3	5
14	6	99.340	9.034	2	12	2	2
15	5	97.624	32.526	2	7	2	3
16	4	97.535	33.740	4	8	4	11
17	3	95.836	57.009	1	3	1	6
18	2	82.795	235.540	2	4	2	14
19	1	69.955	411.327	1	2	1	20

Table 1 of cluster analysis at 20 monitoring locations with 4 parameters CO, SO2, NOx, PM10

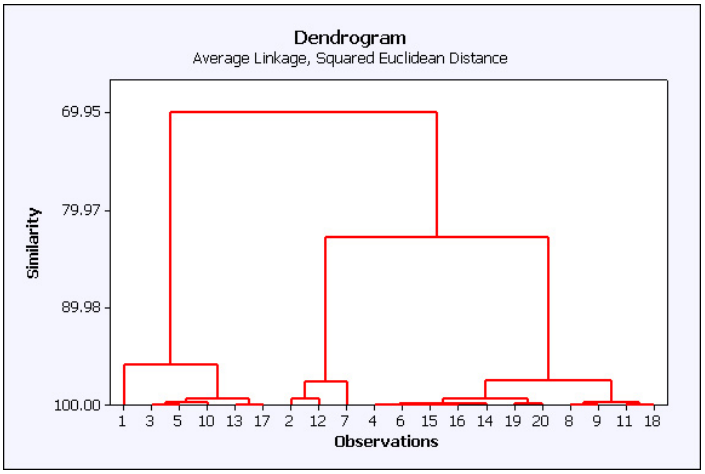

Figure 8 Similarity dendrogram of preliminary monitoring network

From this dendrogram, we can optimize the monitoring network by reduce the numbers of monitoring station to meets the need, eg. we can use station 15 instead of operate both station 15 and 16 because their monitored data has the similarity of 100%. We can do the same to select other station we want to operate.

4. Conclusion

In this paper the air quality index was introduced to rescale the variety of surveying data of Hai Duong in 2006 (differing in unit and value) to similar ranges. The air quality index can be used widely for overall assessment of environmental quality, especially for air, water and soil quality.

Also in this work, the relationship between the variance of the data and the distance between surveying locations in Hai Duong province was shown in variograms by SGeMS. The relationship resulted in the conception of the optimum distance between monitoring stations and defining cell sizes of the grid of the preliminary monitoring network for Hai Duong.

The AQI and the parameters obtained from variogram analysis were used in ArcGIS to interpolate and generate a general picture of air pollution (air quality) of Hai Duong province, and make it easier for environmental control.

The optimum monitoring network obtained for Hai Duong was done by using a dendrogram generated by MiniTab. The dendrogram is a very useful tool to help the decision maker decide which stations should or should not operate (eg. due to limited financial budget or operational constraints) and re-arrange the monitoring stations of the monitoring network.

The combination of environmental quality index, geostatistics and clustering was mentioned in this paper can be further developed and used to design and optimize other environmental monitoring networks, including water, soil, ecosystem, etc. The topography, weather, population density and other socio-economic factors should also be further studied and put in the consideration process of designing preliminary environmental monitoring network.

References

Geoff Bohling (2007): “S-GeMS Tutorial Notes”. Boise State University, Biose, Idaho.

Nguyen Xuan Thinh, Vu Van Manh (2008): “Mathematical modeling and GIS for urban environmental management – Some theoretical and technical issues and an application in Hanoi (Vietnam)”. 2nd International Conference on Asian-European Environmental Technology & Knowledge Transfer, Hefei, China, 5-6.06.2008.

Nguyen Xuan Thinh, Vu Van Manh (2007): “Two approaches for mathematical modeling in urban environmental studies – examples of models and an application to air pollution monitoring in Hanoi (Vietnam)”. 11th Workshop Modellierung und Simulation von Ökosystemen, Seebad Kölpinsee / Insel Usedom, 31.10.2007 – 02.11.2007.

Pham Ngoc Ho, Vu Van Manh et al (2007) : "Environmental Planning of Hai Duong province, period 2006-2020 ". Scientific & Technological project, Hai Duong, 2006.

Simon W. Houlding (2000): “Practical Geostatistics: Modeling and Spatial Analysis”. Springer, Germany.

S∅ren Lophaven (2004): “Design and analysis of environmental monitoring programs”. Technical University of Denmark.

Vietnamese Statistics General Department (2006): Vietnam Annual Statistical Directory. Statistical Publishing House, Hanoi.

Vu Van Manh, Nguyen Thi Hong Hanh (2007): “Using even optimization method for Assessing Air Quality of Hai Duong province”. Scientific and Technical Hydro-Meteorological Journal, Hydro-Meteorological Service, Vietnam, 8/2007.