Kriging Analysis on Hong Kong Rainfall Data

John K Y Leung  BSc(Eng)Hons MSc(CPM) MEngSc MHKIE MIEAust
Drainage Services Department, HKSAR Government

TC Law  BEng
Drainage Services Department, HKSAR Government

The Drainage Services Department has developed a computer system to automatically produce rainfall contour maps with real-time 5-minute rainfall data from the Hong Kong Observatory. These contour maps will enable the drainage engineers to qualitatively estimate the possible flooding locations and impacts caused by the rainstorm. The system used Kriging method in the interpolation of rainfall values at locations between the rain gauge stations. Presented in this paper are a brief description of the real-time rainfall data in Hong Kong, Kriging analysis, the proposed theoretical semi-variogram model, the contouring program, and some discussions on the limitations of the approach.

Keywords: Kriging Analysis, Hong Kong Rainfall, Rainfall Contours

Introduction

Rainfall is a dynamic phenomenon that varies in both space and time. Since different catchments have different drainage characteristics, the flooding effect of a rainstorm will depend very much on its spatial distribution of the rainfall intensity. For example, if the rain cell is over a floodplain, extensive flooding can be expected at the low-lying areas. However, if the rain precipitates on urban areas which are served by well designed drainage systems, the flooding situation will be under better control.

Information on the spatial distribution of a rainstorm is very useful to a drainage engineer. Together with his experience and knowledge on the catchment characteristics and the prevailing tide condition, the engineer can quickly estimate the likely flooding conditions. To provide timely information on rainfall distribution, the Drainage Services Department has developed a computer system which can produce contour maps with real-time 5-minute rainfall data from the Hong Kong Observatory (HKO). Kriging method is used for estimation of rainfall at location between the rain gauge stations.

Kriging is an interpolation technique that relies on statistical theories. It involves different stages of estimation process to improve the value of local estimates through assessing the spatial dependence of the point measurements. An additional bonus is that the method can yield estimates of errors associated with the Interpolation; something that other interpolation techniques often cannot provide. However, as this paper is intended for engineering application aspect of the Kriging method only, error estimations will not be discussed.

Rainfall Data in Hong Kong

In early 2000, there have been 179 operational rain gauge stations in Hong Kong, as summarised in the following table. Some of the gauging stations may include both ordinary and autographic (monthly) gauges at the same location.

<table>
<thead>
<tr>
<th>Type of rain gauges</th>
<th>No of stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>HKO conventional rain gauge stations</td>
<td>51</td>
</tr>
<tr>
<td>HKO Data Acquisition System rain</td>
<td>21</td>
</tr>
<tr>
<td>gauge stations - telemetered</td>
<td></td>
</tr>
<tr>
<td>HKO Weather Station rain gauges</td>
<td>18</td>
</tr>
<tr>
<td>- telemetered</td>
<td></td>
</tr>
<tr>
<td>GEO rain gauge stations - telemetered</td>
<td>86</td>
</tr>
<tr>
<td>DSD rain gauge stations - telemetered</td>
<td>3</td>
</tr>
</tbody>
</table>

The telemetered rain gauges are automatic reporting installations and can provide real-time rainfall data to HKO at every 5-min interval. Through telephone lines, the data is transmitted from HKO to the Drainage Services Department.

These automatic rain gauges are installed over Hong Kong by the Geotechnical Engineering Office, HKO, and the Drainage Services Department. Their locations are shown in Figure 1. The distribution of rain gauges is uneven, with particularly high density in Hong Kong Island for slope stability monitoring. Nevertheless, the provision of rain gauges in general is still well above the international standards. The minimum density for provision of rain gauges for ‘small mountainous Islands with very irregular precipitation’ is 25 km² per station. Urbonas et al. (1992) has suggested that a higher density is required if accurate prediction of catchment response are to be obtained for convective storm events. In the New Territories where the rain gauge density is the sparsest, we still have approximately 8 km² per station, which is much higher than the usual standards.

Figure 1 – Location of Automatic Reporting Rain Gauges

Kriging Analysis

Kriging analysis is a statistical estimation technique developed by Georges Matheron in around 1960. However, the method was named after D G Krige who was probably the first to make use of spatial correlation and BLUE (Best Linear Unbiased Estimator) in the field of mineral resources evaluation.

Kriging was originally applied to estimate the distribution of mineral contents in an ore body with information from the sampling logs. However, it can also be used to estimate any values in space from irregularly distributed sample data points. Examples include calculating the rainfall and temperature based on measurements from climatological stations, interpolating the geological strata thickness from bore logs, estimating the water table profile from piezometric measurements, etc.
The basic concept of Kriging is to weight-average the sample data and minimise the variance of error of estimation. The weights will add up to one so that the interpolated result is statistically unbiased. Moreover, the Kriging system can always provide a unique solution to the same set of data.

The appropriateness of using Kriging in rainfall analysis rests on the recognition that the spatial variation of rainfall is too irregular to be modelled by a smooth mathematical function. Instead, a stochastic surface can better describe the spatial variation. The interpolation proceeds by first exploring and modelling the stochastic aspects, i.e., the spatial variability of rainfall, with several sets of historic data. The resulting information is then used to estimate the weights for interpolation on a new event.

Luk [9] has applied Kriging method in analysing the spatial distribution of rainfall together with other techniques, namely Thiessen Polygon, Inverse Distance Weighted, Trend, and Spline. His finding was that the comparison on Kriging and Spline methods has been made with the built-in functions of the proprietary software ARC/INFO that he used. Since the parameters for both methods have not been calibrated with measured rainfall events from the concerned catchment, the results of the comparison could be influenced by the appropriateness of default parameters inside the software.

**Semi-variogram**

Semi-variogram is the core component of Kriging method. It is a function which defines the variability of data under the effect of separation distance between the data points. The sample data are evaluated pair by pair by computing the squared difference between the data values. The resulting dissimilarities are plotted against the separation distance between the sample pairs in the geographical space to form the variogram cloud. (See Figure 2) [11]

**Experimental Semi-variogram**

The experimental semi-variogram is constructed from the sample data. The semi-variance γ(h), i.e., the average dissimilarity, is obtained by grouping all the pairs of data points into specified intervals of spatial separation. The intervals can be 0-d, d-2d, 2d-3d..., etc, where d is the interval length. If there are n pairs of points within a particular interval, the experimental semi-variance value for separation distance h will be:

\[
\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} [z(x_i + h) - z(x_i)]^2
\]

where \( x_i \) and \( x_i + h \) are the locations of a pair of data points with separation distance h

\( z(x_i) \) and \( z(x_i + h) \) are the sample data values at locations \( x_i \) and \( x_i + h \)

As illustrated in Figure 2, the semi-variance usually increases when the separation distance of the pairs of data points increases. It is because points far apart will have little effect on one another while points nearby tend to have more similarities. When the separation distance is long enough, the semi-variogram will reach a sill which is theoretically the sample variance. The distance at which the semi-variance reaches the sill is called the range, beyond which the samples are no longer correlated and are considered to be independent.

**Theoretical Semi-variogram**

In reality, γ(h) is not known and could only be estimated from the sample data available. To facilitate the formulation of the semi-variogram, theoretical models are often developed to simulate the characteristics of the available data. The choice of theoretical models should not be arbitrary and must be based on the available sample information. By adopting a theoretical semi-variogram, the semi-variance at any spatial separation h can be estimated. Also, the sample value fluctuations of real data can be smoothed out. The two frequently used mathematical models are Spherical model and Exponential model.

a) Spherical model:

\[
\gamma(h) = \begin{cases} 
C_0 + C & \text{for } h > a \\
C_0 + C_1 \left[ 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right] & \text{for } 0 < h < a \\
C_0 & \text{for } h = 0
\end{cases}
\]

b) Exponential model:

\[
\gamma(h) = C_0 + C \left[ 1 - \exp \left( -\frac{|h|}{a} \right) \right]
\]

where \( C_0 \) = nugget

\( C = \text{sill above nugget} \)

\( a = \text{range} \)

**Figure 3 - Examples of Theoretical Semi-variograms**

**Ordinary Kriging**

The mean of sample values may vary from point to point. Ordinary Kriging is a special application of Kriging to account for local variations of the sample mean by limiting the domain of stationarity of the mean to the local neighbourhood centered at the location being estimated. When applying to the estimation of rainfall pattern, the rainfall value \( z^* \) at any location \( x_0 \) can be estimated with rainfall data \( z(x) \) from the n neighbouring rain gauges and combining them with their corresponding weights \( \lambda_i \).
\[ z'(x_i) = \sum_{i=1}^{n} \lambda_i \, z(x_i) \]

The weights \( \lambda_i \) should sum up to 1, i.e. \( \sum \lambda_i = 1 \). In the extreme case when all data values are equal to a constant, the estimated value should also be equal to the nearby constants. This is the non-bias condition for a zero error in expectation.

The estimation variance of error \( \sigma^2_e \) for ordinary Kriging is as follows:

\[
\sigma^2_e = \left[ \gamma(x,x) - \sum_{i=1}^{n} \lambda_i \, \gamma(x_i, x) \right] + 2 \sum_{i=1}^{n} \lambda_i \, \gamma(x_i, x) - \sum_{i=1}^{n} \lambda_i \, \gamma(x_i, x)
\]

where \( \gamma(x, x) \) is the semi-variance between points \( x \) and \( x \), and is determined from the theoretical semi-variogram relationship mentioned above.

The estimation variance of error \( \sigma^2_e \) should be the minimum so that the Kriging estimator is the best estimator. In order to minimize the estimation variance of error under the non-biased condition, Lagrangian \( L(x) \) can be used. It is a function of the data weights \( \gamma_i \), and a Lagrange parameter \( 2\mu \):

\[
L(\lambda_i, \mu) = \left[ \sum \lambda_i \, \gamma(x_i, x) \right] + \mu \left[ \sum \lambda_i - 1 \right]
\]

By minimising the estimation variance of error with the constraint on the weights \( \lambda_i \), we can obtain:

\[
\sum_{i=1}^{n} \lambda_i \, \gamma(x_i, x) + \mu = \gamma(x, x)
\]

and \( \sum \lambda_i = 1 \)

These requirements will generate \( n+1 \) number of equations which, when solved, will give the values of \( \lambda_i \). The set of equations may be written in matrix form as follows:

\[
\begin{bmatrix}
\gamma(x_1, x_1) & \gamma(x_1, x_2) & \ldots & \gamma(x_1, x_n) & 1 \\
\gamma(x_2, x_1) & \gamma(x_2, x_2) & \ldots & \gamma(x_2, x_n) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(x_n, x_1) & \gamma(x_n, x_2) & \ldots & \gamma(x_n, x_n) & 1 \\
1 & 1 & \ldots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n \\
\mu
\end{bmatrix}
= \begin{bmatrix}
\gamma(x_1, x_1) \\
\gamma(x_1, x_2) \\
\vdots \\
\gamma(x_n, x_n) \\
1
\end{bmatrix}
\]

The above matrix can be solved by Gaussian elimination, and the unknowns \( \lambda_i \) obtained. The rainfall value at location \( x_i \) will then be estimated by:

\[ z'(x_i) = \sum \lambda_i \, z(x_i) \]

**Properties of Kriging**

There are several special properties of Kriging which deserve some elaboration in the following.

**Nugget Effect**

The nugget effect accounts for the behaviour of the semi-variogram at the origin, if the intercept of the semi-variogram on the y-axis is greater than zero, a random or unstructured component of variation at zero distance is present. Ideally, the nugget effect should be zero because any two samples from the same point should have the same value. A non-zero value can indicate errors in sampling, or it may represent the sill of a very small scale structured component whose range is much less than the sample interval.

**Anisotropic: Geometric and Zonal**

A semi-variogram is said to be anisotropic when its pattern of spatial variability changes with direction, e.g. East-West against North-South. Modelling anisotropy will require functions that depend on the vector \( h \) rather than on the distance \( h = |h| \) only. There are two kinds of anisotropy, namely, the geometric and the zonal anisotropy. To test for anisotropy, the separation vector \( h \) will be decomposed into N-S and E-W components. Two sets of directional semi-variograms are constructed and compared to identify any difference between the two.

**a) Geometric Anisotropy**

An anisotropy is said to be geometric when the directional semi-variograms have the same shape and sill but different range values. (see Figure 4)

**Figure 4 - Example of Geometric Anisotropy**

**b) Zonal Anisotropy**

An anisotropy is said to be zonal if it involves sill values varying with direction as shown in Figure 5.

**Figure 5 - Example of Zonal Anisotropy**

**Cross Validation Test**

Cross validation test is used to determine the 'goodness of fit' of a theoretical semi-variogram to the experimental one. By comparing the results of cross validation tests with different semi-variogram models, the most appropriate one can be selected.

In the cross validation procedure, each rain gauge value \( z(x_i) \) is removed in turn from the data set and a value \( z'(x_i) \) at that location is estimated by Kriging using the remaining (n-1) rain gauges. The square brackets around the index \( i \) symbolise the fact that the estimation is performed at location \( x_i \) excluding the sampled value \( z(x_i) \). [11]

The difference between a data value and the estimated value can give an indication on how well the data value fits into the neighbourhood of the surrounding data values. For all the n data points, the average of the cross validation errors can be computed. If it is not far from zero, it is said to have no apparent bias; otherwise a significantly negative (or positive) average error can represent systematic over-estimation (or under-estimation respectively).
The Kriging standard deviation \( \sigma_{\|} \) represents the error predicted by the model when Kriging at location \( x_i \) (with the sample data at the location \( x_i \) being omitted). Dividing the cross validation error by \( \sigma_{\|} \) allows comparing the magnitudes of both the actual and the predicted error:

\[
\frac{1}{n} \sum_{i=1}^{n} \left( z(x_i) - z'(x_i) \right)^2 / \sigma_{\|} = 1
\]

if the mean squared standardised cross validation error is about one, then the actual estimation error is on average equal to the error predicted by the model.

Application of Kriging on Hong Kong Hourly Rainfall

Kriging analysis was carried out on the clock-hourly rainfall from the automatic rain gauges. As the gauges reported data at 5-minute intervals, each hourly data would involve the summation of 12 readings. A theoretical semi-variogram model was built up to facilitate the analysis computation. In order to define the form and the various parameters associated with the theoretical semi-variogram, the past rainfall events in 1997 and 1998 were reviewed. Since the concern of Drainage Services Department was on the behaviour of heavy rainstorms, 38 sets of rainfall data were selected for analysis. They all had hourly rainfall exceeding 30 mm at the Hong Kong Observatory station.

Theoretical Semi-variogram for Rainfall in Hong Kong

**Estimation of Sill**

Normalized experimental semi-variograms were constructed by dividing the semi-variogram values by their respective sill values. Theoretically speaking, the sill of a semi-variogram should be the sample variance. However, under situations like clustering of data points, the sample variance could be affected and was unable to reflect correctly the spatial behaviour of sample data. The deviation of sample variance from the actual sill could often be considerable and it was not desirable to directly apply the sample variance as the theoretical sill. Another method was proposed to estimate the sill by averaging the experimental semi-variance values over a pre-defined range of separation distances. The two approaches are discussed and compared in more details below.

**a) Sample Variance Method**

The first approach is to follow the theoretical definition by adopting the sample variance as the sill of the semi-variogram. For \( m \) as the mean of sample data,

\[
\text{Sill} = \frac{1}{n} \sum_{i=1}^{n} (z(x_i) - m)^2
\]

**b) Averaged \( \gamma(h) \) Method**

An alternative approach proposed is to take the sill as the average of semi-variance values \( \gamma(h) \) within a pre-determined range of separation distance from 11 km to 25 km. These upper and lower limits of separation distance were obtained through trials of different values, until the theoretical semi-variograms best fitted to the experimental semi-variograms of the selected rainfall data.

\[
\text{Sill} = \text{average} \left( \sum_{h=1}^{25} \gamma(h) \right)
\]

Figures 6 & 7 were plotted with the above two methods for comparison. The method with Averaged \( \gamma(h) \) method (Figure 7) gave better results in estimating the sill. The normalised semi-variograms generally converged to one at a separation distance of around 15 km. But for the ‘Sample variance method’, the sill did not appreciably converge to one at any separation distance. The former method with Averaged \( \gamma(h) \) was therefore adopted for estimation of the sill.
Test for Anisotropy

In order to analyze the anisotropy behaviour of Hong Kong hourly rainfall data, the normalised semi-variograms in the North-South direction and the East-West direction were constructed with the selected data sets.

From Figures 8 & 9, both ranges were approximately 15 km, which were the same as the adopted range. Since they had similar ranges, the selected sample events were not geometric anisotropic. Furthermore, as the sills of the two groups of semi-variograms were similar, the sample events were considered as not zonal anisotropic. From these observations, it was taken that the hourly rainfall data was not anisotropic in any direction, i.e. the semi-variogram obtained above were used for all directions.

Computer System for Automatic Contouring

A computer program was developed to produce rainfall contour maps at hourly intervals. The program could indicate on the map 5 rain gauges which received the heaviest hourly rainfall.

To produce the contour map, the area of Hong Kong was divided into a rectangular array of 60,000 grid points (300 x 200). The rainfall depth at each grid point was estimated with Ordinary Kriging method using the above theoretical semi-variogram model. For a particular contour level to be plotted, all the contour points falling on the grid lines would be identified and their locations obtained by linear interpolation between the two grid point coordinates.

After all the contour points were identified on the rectangular array, they would be tracked and connected to form a contour line. Leung [8] has discussed three different contour tracking algorithms, i.e. connecting to the closest point, to the link with the least change in contour direction, and to the link which can maintain the higher ground on the same side. The simplest tracking algorithm of connecting to the closest contour point was adopted in this program to reduce computation time. After the addition of contour annotation, the base map and legends, the rainfall contour map was completed. The graphical output of the program would be stored in the computer server to provide an aid to the drainage engineers for visualising the distribution of rainfall in the previous hour. A sample plot of the rainfall contour map is shown in Figure 10.

Figure 8 – Normalised Semi-variograms in N-S Direction

Figure 9 – Normalised Semi-variograms in E-W Direction

Figure 10 – Sample Isohyet Chart Produced

Limitations of Kriging Analysis on Hong Kong Rainfall Data

There are inherent limitations of the above approach in using Kriging analysis on Hong Kong rainfall data. The most profound one is on the availability of data. The hourly rainfall is calculated by adding up the twelve 5-min rainfall data within an hour. If some of the 5-min data are missing, the hourly rainfall at that gauge will be underestimated. Besides, if a rain gauge is not functioning properly, zero or inaccurate readings will be resulted. With defective data, the properties of the semi-variogram will be greatly affected, and so will be the resulting rainfall pattern.

As shown in Figure 1, the distribution of rain gauges is significantly uneven. The provision of rain gauges in Hong Kong Island is more abundant than in the New Territories. The separation distance between the Hong Kong Island rain gauges is relatively small and contributes more to the part of semi-variogram near the origin. This part of the semi-variogram may not be able to reflect the spatial relationship of rainfall data in the New Territories. Meanwhile, the separation distance between the New Territories gauges is usually larger and will dominate the upper part of the semi-variogram. As an alternative approach, rain gauges with similar separation distances could be identified and grouped together. Semi-variograms for different groups could be developed independently with the respective rainfall data to resolve the problem of uneven distribution of rain gauges.

Spatial distribution of rainfall is affected by ground topography, such as the existence of mountains, cliffs and plains. However, the semi-variance depends only on the separation distance and is unable to take into account this topographic variability. Such local variation is a common limitation and affects most surface fitting techniques. Nevertheless, if the rain gauges are strategically located and are sufficiently close to capture the variation of the rainfall pattern, the method can still be applicable because nearby data points have bigger weightings than the remote data points.

Another characteristic of rainfall pattern is that it is dependent on the type of rainstorm, e.g. monsoon rain or typhoon induced rain. For different types of rainstorms, factors other than the separation distance should also be considered. For instance, typhoon induced rainfall is affected by wind direction and could be geometric or zonal anisotropic. The 38 rainfall events examined are generally monsoon rains and do not exhibit prominent anisotropy as discussed before. The possible anisotropy effect

{433279}
Induced by typhoons has not been investigated in this study, as sufficient rainfall events with similar wind direction and wind magnitude are difficult to collect.

**Conclusion**

Presented in this paper are the formulation of a theoretical semi-variogram for hourly rainfall in Hong Kong. On the basis of this result, a computer system has been developed by the Drainage Services Department to automatically generate contour maps of rainfall distributions at hourly intervals. This enables the drainage engineers to qualitatively estimate the possible flooding locations and impacts within a relatively short time.

Kriging is an interpolation technique that relies on statistical theory. In principle, Kriging is an ideal interpolator. The steps in the estimation process are presented in this paper and it has been demonstrated how the spatial dependence of point measurements can improve the value of local estimates.

Nevertheless, there are inherent limitations on using Kriging analysis on rainfall data. The results produced are used for visual presentation of rainfall spatial pattern only. They are not suitable for estimation of point-rainfall at specific locations, nor they be used for meteorological interpretation of rainfall characteristics. If sufficient rain gauges are provided to capture the spatial variation of rainfall, eg at the change of ground topography, the accuracy on estimated pattern of rainfall could further be improved.

**Acknowledgement**

The authors would like to acknowledge the Director of Drainage Services for permission to publish this paper. However, the views expressed are those of the authors and do not necessarily reflect in any way the official views or policy of the Government of Hong Kong SAR.

The authors would also acknowledge the Director of Hong Kong Observatory and the Director of Civil Engineering for the use of rainfall data in the study. They wish to thank the support and advice from their colleagues and friends in the writing of this paper. They are particularly in debt to Mr Y H Chan of the University of Hong Kong, Ir Dr K C Luk of Maunsell Consultants (Asia) Limited, and their colleague Ir Dr Kermen Mok for their advice and sharing with the authors their knowledge and experience on the subject.

**References**

7. Law, G.T.C., Kriging Analysis on Hong Kong's Rainfall, BSc(Eng) Project, Department of Civil Engineering, University of Hong Kong (1999).

**John K Y Leung**

Ir John K Y Leung graduated from the Hong Kong Polytechnic and the Hong Kong University with Higher Diploma in Structural Engineering and Honours Degree in Civil Engineering respectively. He has obtained Masters Degree in Construction Project Management from the Hong Kong University and later Masters Degree in Water Engineering from the University of New South Wales, Australia. Since 1989, he has served in the Drainage Services Department of the HKSAR Government on operation and management of drainage and sewerage systems. His main interests are in flood control methodology and catchment management. He is now the senior engineer of the flood control unit.

**T C Law**

Mr T C Law obtained his Degree of Bachelor of Engineering with first class honours from the University of Hong Kong in 1989. Upon graduation, he joined the Government of Hong Kong Special Administrative Region as a Civil Engineering Graduate to receive his practical training. He is currently pursuing his Master of Science (Civil Engineering) Degree in the University of Hong Kong on a part-time basis.