GEOSPATIAL APPROACH FOR HYDROMET NETWORK DESIGN

Dr. Vijay Kumar
Scientist 'F'
Ministry of Earth Sciences, New Delhi
vijay.kumar66@nic.in; vijay.moes@gmail.com

1.0 INTRODUCTION

A hydrometric network can be defined as a group of data collection activities for different components of the hydrological cycle that are designed and operated to address a single objective or a set of compatible objectives (WMO, 1994). Hydrometric information collected over a basin constitutes the fundamental inputs for the design of various water resources projects such as the design of reservoirs, water distribution systems, irrigation networks, study of climate change impact on water resources, etc. Depending upon the purpose of the hydrometric network, it can be called a surface water network (usually including precipitation network), a groundwater network, or a water quality network. The increasing pressure on water resources due to growing population water needs calls for a more adequate hydrometric network design to provide appropriate information for better water resources planning and management.

Accurate and complete rainfall data is indispensable for planning, designing and operating water projects. Modern rainfall network established to monitor hydrological features should provide the necessary and real-time information for purposes such as management of water resources, reservoir operation and flood forecast and control. The rain gauge densities and distributions have to be sufficient to permit the valid rainfall information reflecting the spatial and temporally varying features in a river basin. However, most rainfall networks employed for hydrological purposes are subdued and largely inadequate; and most catchments of the world are ungauged or poorly gauged. Instead of requiring accuracy, rainfall network design has been strongly influenced by non-hydrological factors, such as convenience and cost. Inadequate rainfall information increases the risk of water projects, loss of life and damage of property. Quantification of rainfall data is essential for providing information of rainfall in time and space. The objective of rainfall network design is to meet the need for information as efficiently and economically as possible. Thus, the rainfall network design should involve the analysis of the number and location of stations necessary for achieving the required accuracy.

The problem of optimization of rain gauges networks could be considered rather out-of-date, as nowadays weather radar provides an estimation of rainfall rates with excellent spatial and temporal resolution. However, a complete coverage by weather radars is still limited to some western countries and obviously rainfall amounts need to be estimated by different scientific and practical reasons all over the globe (including the oceans). Rainfall estimation using satellite data (mainly thermal infrared images) gives satisfactory results. Satellite rainfall estimation algorithms must be calibrated and validated using rain gauges networks which are usually very scarce in important parts of the world such as tropical and subtropical Africa. In that area the optimization of the rain gauge networks is still a serious problem. Rainfall estimations using satellite data are done generally on a 10-day and a monthly basis, thus, rain gauge network optimisation using monthly climatological variograms is of great practical interest. For rainfall events of shorter duration (from 1 day to minutes), the spatial variability increases and the spatial continuity is a function mainly of the duration of the event.

The optimisation of a rain gauge network for monthly measurements does not imply that it is optimal for daily measurements; generally, not because the disposition of the rain gauges, but because of the number of rain gauges. To characterise rainfall patterns of great variability and intermittency a much denser network is necessary.

Recent reviews have noted a marked decline in the amount of data being collected in many parts of the world. The reasons are varied but include the problems of insufficient funding, inadequate institutional frameworks, a lack of appreciation of the worth of long-term hydrological data, and in a few cases, the turmoil caused by wars and other disasters. Even where monitoring programmes are operating, they may fail to provide adequate information to support effective management because:

- The objectives are not properly defined;
- Data are poorly archived and not readily accessible in the type of formats that can be used to support management and inform other stakeholders.

2.0 GEOSPATIAL APPROACH FOR HYDROMET NETWORK DESIGN

Geospatial approach is an approach to applying statistical analysis and other informational techniques to data which has a geographical or geospatial aspect. Geostatistics is a collection of statistical methods for the analysis and estimation of spatial data for use primarily in the earth sciences. According to Matheron (1963) "Geostatistics is the application of the formalism of random functions to the reconnaissance and estimation of natural phenomena". It can be described as a systematic approach for making inferences about quantities that vary in space. Geostatistical techniques incorporate the spatial characteristics of actual data into statistical estimation processes. Geostatistics has perhaps been most clearly described by ASCE (1990), as follows:

Geostatistics...provides the statistical tools for (1) Calculating the most accurate (according to well-defined criteria) predictions, based on measurements and other relevant information; (2) quantifying the accuracy of these predictions; and (3) selecting the parameters to be measured, and where and when to measure them, if there is an opportunity to collect more data.

Geostatistics is based on the Theory of Regionalized Variables. When a variable is distributed in space, it is said to be "regionalized" (Journel and Huijbregts, 1978). A Regionalized Variable (ReV) is defined by Matheron (1963) as the variable that spreads in space and exhibits certain spatial structure. Such variables show a complex behaviour. Their variations in space are erratic and often unpredictable from one point to another; however, these are not completely random as these exhibit some spatial correlation.

Annual precipitation is a typical regional variable, in that it usually shows high variability over a large area, smaller irregularities at a more local scale, and similar values for neighbouring stations. All the parameters generally used in groundwater hydrology, such as transmissivity, hydraulic conductivity, piezometric heads, vertical recharge etc. can be called regionalized variables.

The main features of linear geostatistics, which is the most popular branch of geostatistics, are: (1) It uses the spatial-correlation structure of spatial functions; (2) its estimates are calculated by weighting the measurements with coefficients that are determined from the minimization of the mean square error, subject to un-biasedness conditions; and (3) it can process measurements averaged over different volumes and sizes.

Semivariogram

Kriging uses the semivariogram to measure the spatial correlated components, a component that is also called spatial dependence or spatial autocorrelation. The semivariogram is half of the arithmetic mean of the squared difference between two experimental measures, ($Z(x_i)$ and $Z(x_i+h)$), at any two points separated by the vector h. Before values of any parameter can be estimated with kriging, it is necessary to identify the spatial correlation structure from the semivariogram, which shows the relationship between semivariance and the distance between sample pairs.

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

where $\gamma^*(h)$ = estimated value of the semivariance for lag h; N(h) is the number of experimental pairs separated by vector h; $z(x_i)$ and $z(x_i+h)$ = values of variable z at x_i and x_i+h, respectively; x_i and x_i+h = position in two dimensions. A plot of $\gamma^*(h)$ versus the corresponding value of h, also called the semivariogram, is thus a function of the vector h, and may depend on both the magnitude and the direction of h. A sample plot of semivariogram is shown in Fig. 1.

The distance at which the variogram becomes constant is called the range, *a*. It is considered that any data value Z(x) will be correlated with any other value falling within a radius, *a* and thus range corresponds to the zone of influence of the RV. The value of the semivariogram at a distance equal to the range is called the sill. Semivariograms may also increase continuously without showing a definite range and sill. The value of the semivariogram at extremely small separation distance is called the nugget effect. . Ideally, the experimental semivariogram should pass through the origin when the sample distance is zero. But, many regionalised variables show nugget effect. It could be caused by sampling errors or short scale variability of the property which cannot be detected at the scale of sampling.

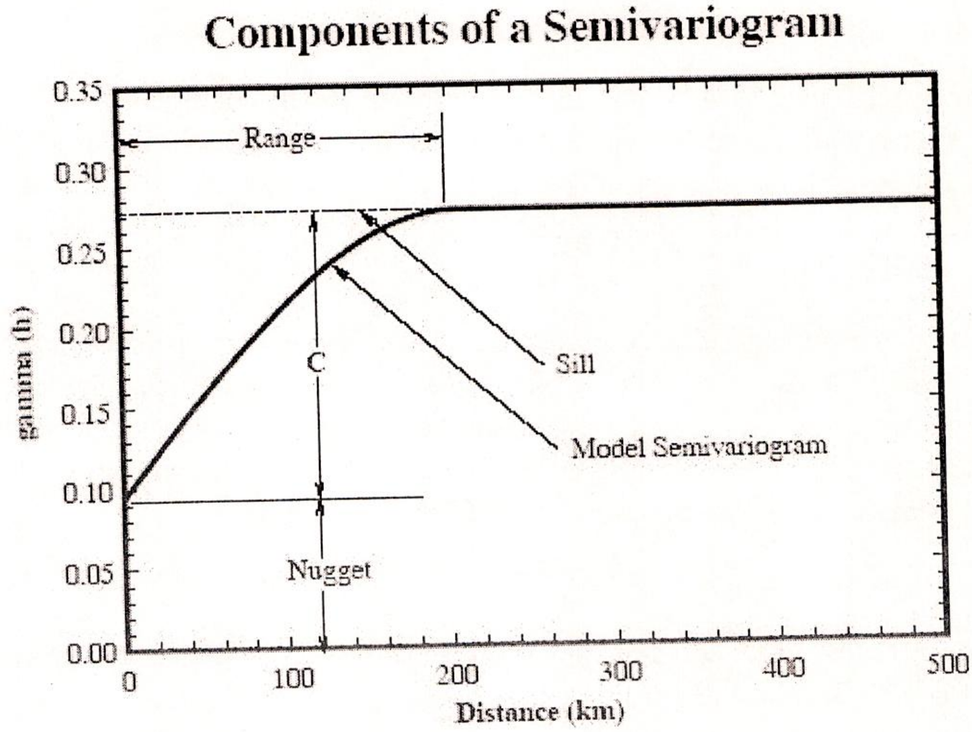

Fig. 1 Sample Plot of Semivariogram

The semivariogram, given in Eq. 1 is also termed the true semivariogram of the ReV. As only one realization of the RF is available, the true semivariogram can only be estimated and this estimate is known as the experimental semivariogram. If the sampling is done on a regular grid, the $\gamma^*(h)$ may be estimated for values of h, known as lag distance or lag increment which are multiples of the grid spacing. This situation is rare in practice, particularly in the context of groundwater and the chance of finding pairs at exactly same specified distance h is very small. To overcome this, a tolerance, δh is placed on the distance. Every pair of observations that are separated by a lag $h \pm \delta h/2$ are then used to estimate $\gamma^*(h)$.

The above procedure is used for calculating the isotropic experimental semivariogram, also known as omni-directional semivariogram. In this case, it is assumed that the variation is the same in every direction. To find the anisotropies, the semivariograms are calculated in different directions. To do this, a tolerance, $\delta\theta$, is placed on the directional angle. The tolerance in direction and distance are represented in Fig. 2.

Structural Analysis

The observed data is used to calculate the experimental semivariogram. A mathematical function used to approximately represent this semivariogram is known as the theoretical semivariogram. Some of the theoretical semivariogram models are (Fig. 3).

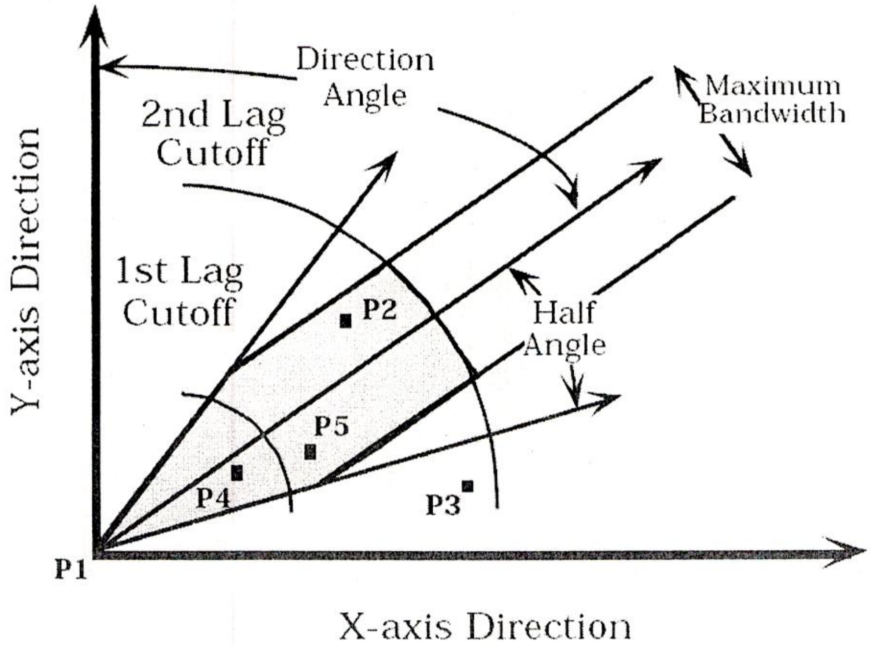

Fig. 2

Spherical model: -
$$\gamma(h)=\begin{cases} C_0[1-\delta(h)]+C\left[\dfrac{3}{2}\dfrac{h}{a}-\dfrac{1}{2}\dfrac{h^3}{a^3}\right] & h\le a \\ C_0+C & h>a \end{cases} \quad (2)$$

Exponential model: -
$$\gamma(h)=C_0[1-\delta(h)]+C\left[1-\exp\left(-\frac{h}{a}\right)\right] \quad (3)$$

Gaussian model: -
$$\gamma(h)=C_0[1-\delta(h)]+C\left[1-\exp\left(-\frac{h^2}{a^2}\right)\right] \quad (4)$$

Linear model: -
$$\gamma(h)=C_0[1-\delta(h)]+bh \quad (5)$$

where, δ(h) is the Kronecker delta = $\begin{cases}1 & h=0\\ 0 & h\neq 0\end{cases}$, C_0 is the Nugget effect, C_0+C is the Sill, *a* is the Range and *b* is the slope.

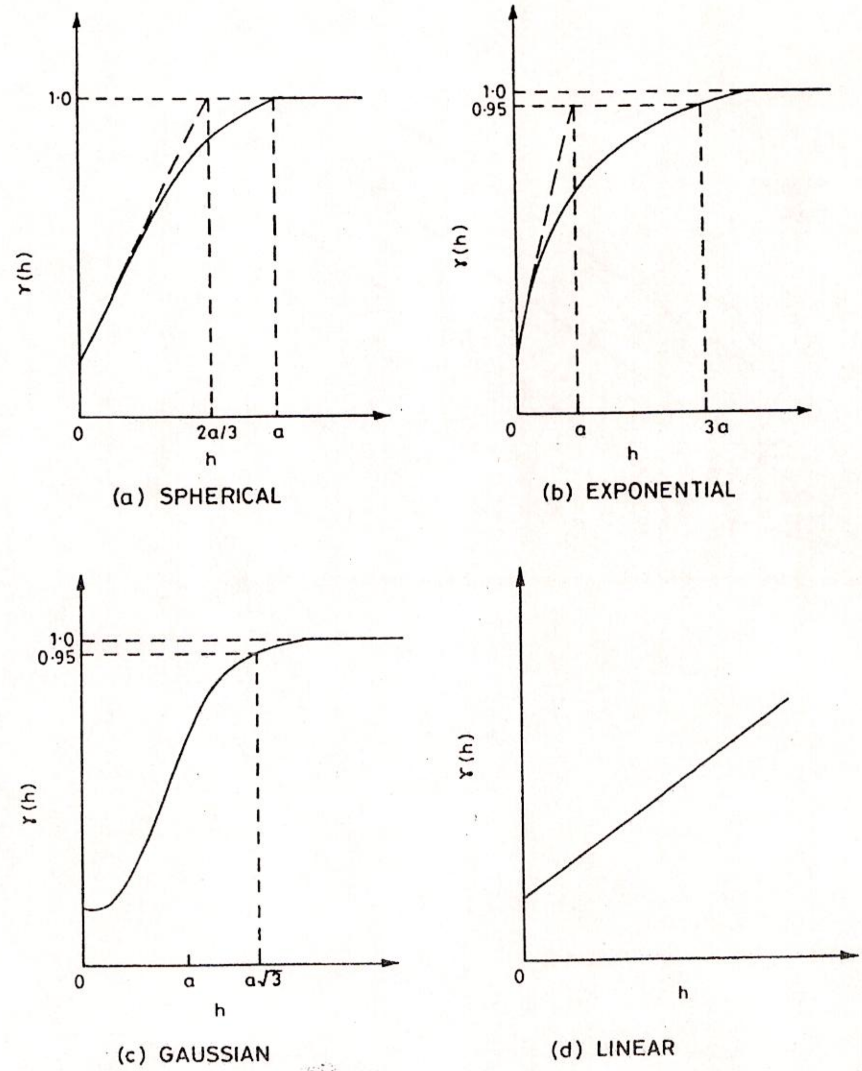

Fig. 3 Theoretical models of semivariogram

Kriging

Kriging is a geostatistical method for spatial interpolation. It is a technique of making optimal, unbiased estimates of regionalized variables at unsampled locations using the structural properties of the semivariogram and the initial set of data values. Consider a situation in which a property is measured at a number of points, x_i, within a region to give values of $z(x_i)$, i=1,2,3,...,N. (x_i is the coordinate of the observation point in 1, 2 or 3-dimensional space). From these observations, the value of the property at any place x_0 can be estimated as

$$z^*(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \qquad i=1,2,3,\ldots\ldots,N \qquad (6)$$

where,

$z^*(x_0)$ = estimated value at x_0

λ_i = weights chosen so as to satisfy suitable statistical conditions

$z(x_i)$ = observed values at points x_i, $i = 1,2,3,\ldots\ldots.,N$

N = sample size

In kriging, the weights λ_i are calculated so that $Z^*(x_0)$ is unbiased and optimal i.e.

$$E\{Z^*(x_0) - Z(x_0)\} = 0 \tag{7}$$

$$Var\{Z^*(x_0) - Z(x_0)\} = \text{minimum} \tag{8}$$

The best linear unbiased estimate of $Z(x_o)$ is obtained by using Lagrangian techniques to minimise Eq. (8) and then optimising the solution of the resulting system of equations when constrained by the nonbiased condition of Eq. (7). The following system of equations, known as kriging system, results from the optimization:

$$\begin{cases} \sum_{j=1}^{N} \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0) & i = 1,2,3,\ldots\ldots., N \\ \sum_{j=1}^{N} \lambda_j = 1 \end{cases} \tag{9}$$

where,

μ = Lagrange multiplier

$\gamma(x_i, x_j)$ = semivariogram between two points x_i and x_j

Solution of the above set provides the values of λ_i, which can be used with Eq. 6 for estimation. The minimum estimation variance, or kriging variance, is written as:

$$\sigma_k^2(x_0) = \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) + \mu \tag{10}$$

Network Optimization

The network optimization method uses the results of interpolation systems which provide in addition to the interpolated values an estimate of the error. Its reduction is the most used criterion for the optimization of observation network.

The kriging variance is a powerful tool for optimizing a network because in the expression

$$\sigma_k^2(x_0) = \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) + \mu \qquad (11)$$

the kriging variance depends only on the semi-variogram and the configuration of observation points in relation to the point to be estimated. It does not depend on the observation value. Therefore if the semi-variogram is already known, the kriging variance for any particular hydromet observation scheme can be determined before putting that scheme into effect. The kriging variance is therefore an indicator of precision to determine the location of any additional measuring point or to decide whether or not to strengthen the observation network. The kriging variance involves only the variogram model, the number of stations and their geographical positions. It is independent of the observed values and therefore the network can be planned. It is this characteristic that determines the optimal network.

To determine the accuracy of existing network, contour map can be prepared by kriging at pre-determined grid interval. This map is based on all available information. The estimation error contour map thus produced shows the uncertainty already present in the existing observation network. The reduction in number of observations without increasing the existing uncertainty is achieved by superimposing a square grid pattern over the area and selecting one observation site per square grid. The data of these sites is then kriged to prepare the contour map and estimation standard deviation contour map. This is repeated by varying the size of square grid. The size of square grid whose data set provide the contour maps for kriged parameter and estimation standard deviation closed to the existing network is selected. This gives the minimum number of observation wells required in estimating the hydromet.

In selecting the optimum number of observation sites required to reduce already existing average uncertainty, the location of addition sites can be suggested by observing the contour map of estimation standard deviation of reduced network. New fictitious measurement points are added in the area where the standard deviation is high. In this way, theoretical network of observational sites is constructed using the sites already existing in the area.

*****