

World Bank & Government of The Netherlands funded

Training module # SWDP - 11

How to compile rainfall data

New Delhi, February 2002

CSMRS Building, 4th Floor, Olof Palme Marg, Hauz Khas, New Delhi – 11 00 16 India Tel: 68 61 681 / 84 Fax: (+ 91 11) 68 61 685 E-Mail: dhvdelft@del2.vsnl.net.in DHV Consultants BV & DELFT HYDRAULICS

with HALCROW, TAHAL, CES, ORG & JPS

Table of contents

<u>Page</u>

1.	Module context	2
2.	Module profile	3
3.	Session plan	4
4.	Overhead/flipchart master	6
5.	Handout	7
6.	Additional handout	9
7.	Main text	10

While designing a training course, the relationship between this module and the others, would be maintained by keeping them close together in the syllabus and place them in a logical sequence. The actual selection of the topics and the depth of training would, of course, depend on the training needs of the participants, i.e. their knowledge level and skills performance upon the start of the course.

2. Module profile

Title	:	How to compile rainfall data					
Target group	:	Assistant Hydrologists, Hydrologists, Data Processing Centre Managers					
Duration	:	Five sessions of 60 minutes each					
Objectives	:	 After the training the participants will be able to: Compile rainfall data for different durations Estimate areal rainfall by different methods Drawing isohyets 					
Key concepts	:	 Aggregation of data to longer durations Areal rainfall Arithmetic average Weighted average Thiessen method Kriging method Inverse distance method 					
Training methods	:	Lecture, exercises, softwares					
Training tools required	:	: Board, OHS, computers					
Handouts	:	As provided in this module					
Further reading and references	:						

No	Activities	Time	Tools
1	GeneralImportant points	5 min	OHS 1
2	 Aggregatiuon of data to longer duration Objectives Plot of hourly data Plot of compiled daily data Plot of weekly data Plot of ten-daily data Plot of monthly data Plot of yearly data Multiple plots for various intervals (a) Multiple plots for various intervals (b) Working with HYMOS 	5 min	OHS 2 OHS 3 OHS 4 OHS 5 OHS 6 OHS 7 OHS 8 OHS 9 OHS 10
3	 Estimation of areal rainfall (1) Objective & definition Various methods Arithmetic & weighted average Example 3.1 – Arithmetic average Thiessen polygon method Example 3.2 (a) – Thiessen polygons Example 3.2 (b) – Thiessen weights & plot of areal average series Comparison of results from two methods Working with HYMOS 	15 min	OHS 11 OHS 12 OHS 13 OHS 14 OHS 15 OHS 16 OHS 17 OHS 18
4	 Estimation of areal rainfall (2) Procedure in Isohyetal Method, flat terrain Example Isohyetal Method Procedure Isohyetal Method, mountainous terrain Procedure Isopercental method Combining isopercentals with normals Drawing isohyets with additional data from normals Procedure Hypsometric Method Hypsometric Method application 	30 min	OHS 23 OHS 24 OHS 25 OHS 26 OHS 27 OHS 28 OHS 29 OHS 30

5	Estimation of areal rainfall (3)	90 min	[]
Ŭ	Rainfall interpolation by Kriging and Inverse Distance Meth.	55 1111	OHS 31
	 Estimation of values on a grid 		OHS 32
	•		OHS 33
	Rainfall interpolation by kriging (1)		OHS 33 OHS 34
	Assumption for ordinary kriging		OHS 35
	Kriging: unbiasedness and variance minimisation		OHS 35 OHS 36
	Kriging equations		OHS 30 OHS 37
	 Exponential spatial correlation function 		OHS 37 OHS 38
	Exponential co-variance function		
	Exponential semi-variogram		OHS 39
	Possible semi-variogram models in HYMOS		OHS 40
	• Sensitivity analysis on variogram parameters (1)		OHS 41
	Sensitivity analysis on variogram parameters (2)		OHS 42
	 Sensitivity analysis on variogram parameters (2) 		OHS 43
			OHS 44
			OHS 45
	Sensitivity analysis on variogram parameters (5)		OHS 46
	Application of kriging and inverse distance method		OHS 47
	Example Bilodra: spatial correlation		OHS 48
	• Example Bilodra: fit of semi-variance to spherical model (1)		OHS 49
	• Example Bilodra: fit of semi-variance to spherical model (2)		OHS 50
	Example Bilodra: fit of semi-variance to exponential model		OHS 51
	Example Bilodra: isohyets June 1984 using kriging		OHS 52
	Example Bilodra: estimation variance June 1984		OHS 53
	• Example Bilodra: isohyets June 1984 using inverse distance		OHS 54
6	Transformation of non-equidistant to equidistant series	3 min	
	General		OHS 19
_			
7	Compilation of maximum and minimum series	2 min	
	Statistical inferences		OHS 20
	• Example 5.1 (a) – Min., max., mean etc.		OHS 21
	Example 5.1 (b) – Tabular results		OHS 22
8	Exercise		
1	• Compilation of hourly rainfall data to daily interval and	30 min	
	observed daily interval to ten-daily, monthly and yearly		
	intervals		
	 Estimation of areal average using arithmetic and Thiessen 	30 min	
	polygon method		
	 Compilation of extremes for ten-daily rainfall data series for 	30 min	
	part of year (July 1 – Sept. 30)		
	• Fitting of semi-variogram for monthly rainfall in Bilodra	30 min	
	catchment (1960-2000), based on aggregated daily rainfall		
	series. Fit different semi-variance models and compare the		
	results in a spreadsheet		
	 Application of semi-variance models to selected months 	20 min	
	and compare the results of different models (interpolations		
1			
	and variances)	10 min	
	and variances)Estimate monthly isohyets by Inverse Distance Method	10 min	
	and variances)	10 min	

Add copy of Main text in chapter 7, for all participants.

6. Additional handout

These handouts are distributed during delivery and contain test questions, answers to questions, special worksheets, optional information, and other matters you would not like to be seen in the regular handouts.

It is a good practice to pre-punch these additional handouts, so the participants can easily insert them in the main handout folder.

Contents

1.	General	1
2.	Aggregation of data to longer durations	1
3.	Estimation of areal rainfall	5
4.	Transformation of non-equidistant to equidistant series	25
5.	Compilation of minimum, maximum and mean series	25

How to compile rainfall data

1. General

- Rainfall compilation is the process by which observed rainfall is transformed:
 - from one time interval to another
 - from one unit of measurement to another
 - from point to areal values
 - from non-equidistant to equidistant series
- Compilation is required for validation, reporting and analysis
- Compilation is carried out at the State Data Processing Centre; it is done prior to validation if required, but final compilation is carried out after correction and 'completion'.

2. Aggregation of data to longer durations

Rainfall from different sources is observed at different time intervals, but these are generally one day or less. For the standard raingauge, rainfall is measured once or twice daily. For autographic records, a continuous trace is produced from which hourly rainfall is extracted. For digital rainfall recorders rainfall is recorded at variable interval with each tip of the tipping bucket. Hourly data are typically aggregated to daily; daily data are typically aggregated to weekly, ten daily, 15 daily, monthly, seasonally or yearly time intervals

Aggregation to longer time intervals is required for validation and analysis. For validation small persistent errors may not be detected at the small time interval of observation but may more readily be detected at longer time intervals.

2.1 Aggregation of daily to weekly

Aggregation of daily to weekly time interval is usually done by considering the first 51 weeks of equal length (i.e. 7 days) and the last 52nd week of either 8 or 9 days according to whether the year is non-leap year or a leap year respectively. The rainfall for such weekly time periods is obtained by simple summation of consecutive sets of seven days rainfalls. The last week's rainfall is obtained by summing up the last 8 or 9 days daily rainfall values.

For some application it may be required to get the weekly compilation done for the exact calendar weeks (from Monday to Sunday). In such a case the first week in any year will start from the first Monday in that year and thus there will be 51 or 52 full weeks in the year and one or more days left in the beginning and/or end of the year. The days left out at the end of a year or beginning of the next year could be considered for the 52nd of the year under consideration. There will also be cases of a 53rd week when the 1st day of the year is also the first day of the week (for non-leap years) and 1st or 2nd day of the year is also first day of the week (for leap years).

2.2 Aggregation of daily to ten daily

Aggregation of daily to ten daily time interval is usually done by considering each month of three ten daily periods. Hence, every month will have first two ten daily periods of ten days each and last ten daily period of either 8, 9, 10 or 11 days according to the month and the

year. Rainfall data for such ten daily periods is obtained by summing the corresponding daily rainfall data. Rainfall data for 15 daily periods is also be obtained in a similar manner for each of the two parts of every month.

2.3 Aggregation from daily to monthly

Monthly data are obtained from daily data by summing the daily rainfall data for the calendar months. Thus, the number of daily data to be summed up will be 28, 29, 30 or 31 according to the month and year under consideration. Similarly, yearly rainfall data are obtained by either summing the corresponding daily data or monthly data, if available.

2.4 Hourly to other intervals

From rainfall data at hourly or lesser time intervals, it may be desired to obtain rainfall data for every 2 hours, 3 hours, 6 hours, 12 hours etc. for any specific requirement. Such compilations are carried out by simply adding up the corresponding rainfall data at available smaller time interval.

Example 2.1:

Daily rainfall at ANIOR station (KHEDA catchment) is observed with Standard Raingauge (SRG). An Autographic Raingauge (ARG) is also available at the same station for recording rainfall continuously and hourly rainfall data is obtained by tabulating information from the chart records.

It is required that the hourly data is compiled to the daily interval corresponding to the observations synoptic observations at 0830 hrs. This compilation is done using the aggregation option and choosing to convert from hourly to daily interval. The observed hourly data and compiled daily data is shown if Fig. 2.1 and Fig. 2.2 respectively.

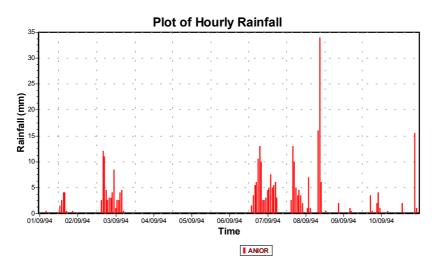


Fig. 2.1: Plot of observed hourly rainfall data

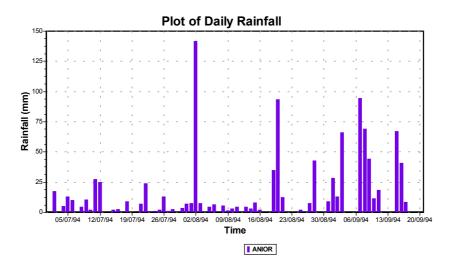


Fig. 2.2: Compiled daily rainfall from hourly data tabulated from ARG charts

Similarly, daily data observed using SRG is required to be compiled at weekly, ten-daily, monthly and/or yearly interval for various application and for the purpose of data validation. For this, the daily data obtained using SRG is taken as the basic data and compilation is done to weekly, ten-daily, monthly and yearly intervals. These are illustrated in Fig. 2.3, Fig. 2.4, Fig. 2.5 and Fig. 2.6 respectively.

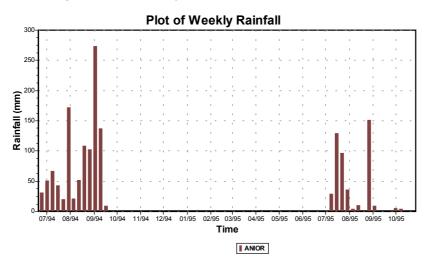


Fig. 2.3: Compiled weekly rainfall from hourly data tabulated from ARG charts

Page 3

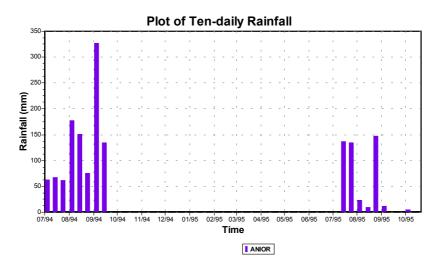


Fig. 2.4: Compiled ten-daily data from daily data obtained from SRG records

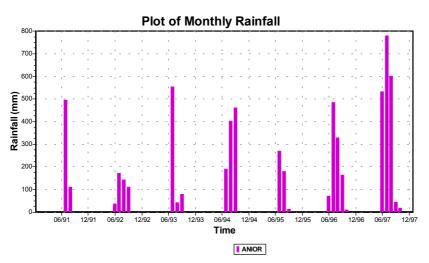


Fig. 2.5: Compiled monthly data from daily data obtained from SRG records

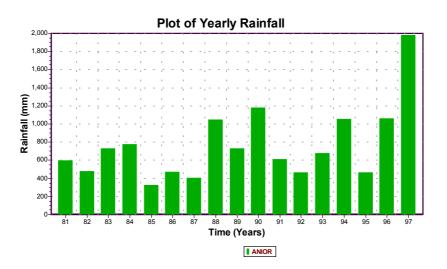


Fig. 2.6: Compiled yearly data from daily data obtained from SRG records

Page 4

3. Estimation of areal rainfall

3.1 General description

Raingauges generally measure rainfall at individual points. However, many hydrological applications require the average depth of rainfall occurring over an area which can then be compared directly with runoff from that area. The area under consideration can be a principal river basin or a component sub-basin. Occasionally, average areal rainfall is required for country, state or other administrative unit, and the areal average is obtained within the appropriate political or administrative boundary.

Since rainfall is spatially variable and the spatial distribution varies between events, point rainfall does not provide a precise estimate or representation of the areal rainfall. The areal rainfall will always be an estimate and not the true rainfall depth irrespective of the method.

There are number of methods which can be employed for estimation of the areal rainfall including:

- The Arithmetic average method,
- Weighted average method
- Thiessen polygon method.
- Kriging techniques

All these methods for estimation of areal average rainfall compute the weighted average of the point rainfall values; the difference between various methods is only in assigning the weights to these individual point rainfall values, the weights being primarily based on the proportional area represented by a point gauge. Methods are outlined below:

3.2 Arithmetic average

This is the simplest of all the methods and as the name suggests the areal average rainfall depth is estimated by simple averaging of all selected point rainfall values for the area under consideration. That is:

$$P_{at} = \frac{1}{N} (P_{1t} + P_{2t} + P_{3t} + \cdots + P_{Nt}) = \frac{1}{N} \sum_{i=1}^{N} P_{it}$$

Where:

P_{at}= estimated average areal rainfall depth at time t

 P_{it} = individual point rainfall values considered for an area, at station i (for i = 1,N) and time t,

N = total number of point rainfall stations considered

In this case, all point rainfall stations are allocated weights of equal magnitude, equal to the reciprocal of the total number of stations considered. Generally, stations located within the area under consideration are taken into account. However, it is good practice also to include such stations which are outside but close to the areal boundary and thus to represent some part of the areal rainfall within the boundary. This method is also sometimes called as unweighted average method since all the stations are given the same weights irrespective of their locations.

This method gives satisfactory estimates and is recommended where the area under consideration is flat, the spatial distribution of rainfall is fairly uniform, and the variation of individual gauge records from the mean is not great.

3.3 Weighted average using user defined weights

In the arithmetic averaging method, all rainfall stations are assigned equal weights. To account for orographic effects and especially where raingauges are predominantly located in the lower rainfall valleys, it is sometimes required to weight the stations differently. In this case, instead of equal weights, user defined weights can be assigned to the stations under consideration. The estimation of areal average rainfall depth can be made as follows:

$$P_{wt} = \frac{1}{N} (c_1 P_{1t} + c_2 P_{2t} + c_3 P_{3t} + \dots + c_N P_{Nt}) = \frac{1}{N} \sum_{i=1}^{N} c_i P_{it}$$

Where:

 c_i = weight assigned to individual raingauge station i (i = 1,N).

To account for under-representation by gauges located in valleys the weights do not necessarily need to add up to 1.

3.4 Thiessen polygon method

This widely-used method was proposed by A.M. Thiessen in 1911. The Thiessen polygon method accounts for the variability in spatial distribution of gauges and the consequent variable area

which each gauge represents. The areas representing each gauge are defined by drawing lines between adjacent stations on a map. The perpendicular bisectors of these lines form a pattern of polygons (the Thiessen polygons) with one station in each polygon (see Fig. 3.1). Stations outside the basin boundary should be included in the analysis as they may have polygons which extend into the basin area. The area of a polygon for an individual station as a proportion of the total basin area represents the Thiessen weight for that station. Areal rainfall is thus estimated by first multiplying individual station totals by their Thiessen weights and then summing the weighted totals as follows:

$$P_{at} = \frac{A_1}{A}P_{1t} + \frac{A_2}{A}P_{2t} + \frac{A_3}{A}P_{3t} + \cdots + \frac{A_N}{A}P_{Nt} = \sum_{i=1}^{N} (\frac{A_i}{A})P_{it}$$

where: A_i = the area of Thiessen polygon for station i A = total area under consideration

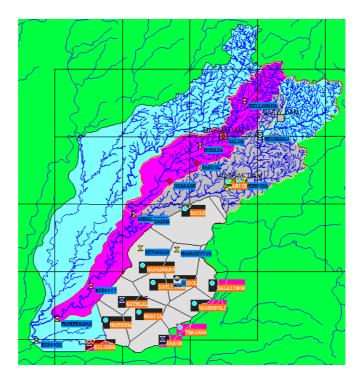


Fig. 3.1: Small basin upto BILODRA gauging site (portion shown with Thiessen polygons)

The Thiessen method is objective and readily computerised but is not ideal for mountainous areas where orographic effects are significant or where raingauges are predominantly located at lower elevations of the basin. Altitude weighted polygons (including altitude as well as areal effects) have been devised but are not widely used.

Example 3.1

Areal average rainfall for a small basin upto BILODRA gauging site (shown highlighted in Fig. 3.1) in KHEDA catchment is required to be compiled on the basis of daily rainfall data observed at a number of raingauges in and around the region. Areal average is worked out using two methods: (a) Arithmetic average and (b) Thiessen method.

(a) Arithmetic Average

For the arithmetic average method rainfall stations located inside and very nearby to the catchment boundary are considered and equal weights are assigned to all of them. Since there are 11 stations considered the individual station weights work out as 0.0909 and is given in Table 3.1 below. On the basis of these equal station weights daily areal average is computed. The compiled areal daily rainfall worked out using arithmetic average method is shown for the year 1994 in Fig. 3.2.

Table 3.1: List of stations and corresponding weights for arithmetic average method

Areal computation - Arithmetic Average

Areal series: BILODRA MA1

Station wei	ghts	
BALASINOR	=	0.0909
DAKOR	=	0.0909
KAPADWANJ	=	0.0909
BAYAD	=	0.0909
MAHISA	=	0.0909
MAHUDHA	=	0.0909
SAVLITANK	=	0.0909
THASARA	=	0.0909
VAGHAROLI	=	0.0909
VADOL	=	0.0909
KATHLAL	=	0.0909
Sum	=	0.999

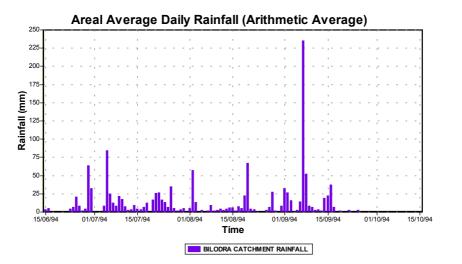


Fig. 3.2: Plot of areal daily rainfall for BILODRA catchment using arithmetic average method

(b) Thiessen polygon method

Computation of areal average using Thiessen method is accomplished by first getting the Thiessen polygon layer (defining the boundary of Thiessen polygon for each contributing point rainfall station). The station weights are automatically worked out on the basis of areas of these polygons with respect to the total area of the catchment. The layout of the Thiessen polygons as worked out by the system is graphically shown in Fig. 3.1 and the corresponding station weights are as given in Table 3.2. On the basis of these Thiessen polygon weights the areal average of the basin is computed and this is shown in Fig. 3.3 for the year 1994. In this case it may be noticed that there is no significant change in the values of the areal rainfall obtained by the two methods primarily on account of lesser variation in rainfall from station to station.

Table 3.2: List of stations and corresponding weights as per Thiessen polygon method

Areal computation - Thiessen Polygon Method

Areal series: BILODRA MA3

Station weights

ANIOR	=	0.0127
BALASINOR	=	0.0556
BAYAD	=	0.1785
DAKOR	=	0.0659
KAPADWANJ	=	0.1369
KATHLAL	=	0.0763
MAHISA	=	0.0969
MAHUDHA	=	0.0755
SAVLITANK	=	0.0724
THASARA	=	0.0348
VADOL	=	0.1329
VAGHAROLI	=	0.0610



= 1.00

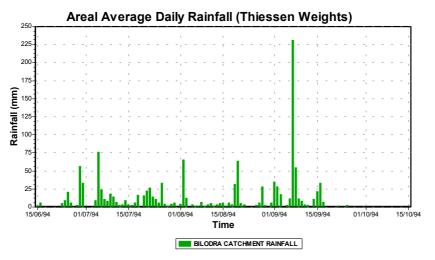


Fig. 3.4: Plot of areal daily rainfall for BILODRA catchment using Thiessen polygon method

3.5 Isohyetal and related methods

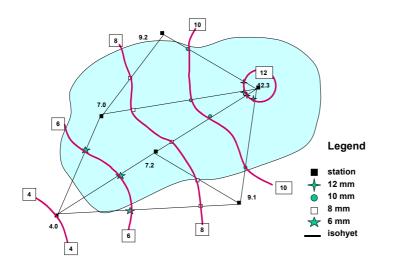
The main difficulty with the Thiessen method is its inability to deal with orographical effects on rainfall. A method, which can incorporate such effects, is the isohyetal method, where lines of equal rainfall (= isohyets) are being drawn by interpolation between point rainfall stations taking into account orographic effects.

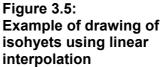
In flat areas where no orographic effects are present the method simply interpolates linearly between the point rainfall stations. Manually the procedure is as follows. On a basin map first the locations of the rainfall stations within the basin and outside near the basin boundary are plotted. Next, the stations are connected with their neighbouring stations by straight lines. Dependent on the rain depths for which isohyets are to be shown by linear interpolation between two neighbouring stations the position of the isohyet(s) on these connecting lines are indicated. After having completed this for all connected stations, smooth curves are

drawn through the points marked on the straight lines between the stations connecting the concurrent rainfall values for which isohyets are to be shown, see Figure 3.5. In drawing the isohyets personal experience with local conditions and information on storm orientation can be taken into account. Subsequently, the area between two adjacent isohyets and the catchment boundary is planimetered. The average rainfall obtained from the two adjacent isohyets is assumed to have occurred over the entire inter-isohyet area. Hence, if the isohyets are indicated by P_1 , P_2 , ..., P_n with inter-isohyet areas a_1 , a_2 , ..., a_{n-1} the mean precipitation over the catchment is computed from:

$$\overline{P} = \frac{a_1(\frac{P_1 + P_2}{2}) + a_2(\frac{P_2 + P_3}{2}) + \dots + a_{n-1}(\frac{P_{n-1} + P_n}{2})}{A}$$
(3.4)

It is noted that if the maximum and/or minimum point rainfall value(s) are within the catchment boundaries then P_1 and/or P_n is to be replaced by the highest and/or lowest point rainfall values. A slightly biased result will be obtained if e.g. the lowest (highest) isohyet is located outside the catchment area as the averaging over two successive isohyets will underestimate (overestimate) the average rainfall in the area bounded by the catchment boundary and the first inside isohyet.





For flat areas the isohyetal method is superior to the Thiessen method if individual storms are considered as it allows for incorporation of storm features like orientation; for monthly, seasonal or annual values such preference is not available. But its added value is particularly generated when special topographically induced meteorological features like orographic effects are present in the catchment rainfall. In such cases the above procedure is executed with a catchment map overlaying a topographical map to be able to draw the isohyets parallel to the contour lines. Also the extent of rain shadow areas at the leeward side of mountain chains can easily be identified from topographical maps. The computations are again carried out with the aid of equation 3.4. In such situations the isohyetal method is likely to be superior to the Thiessen method.

The **isopercental method** is very well suited to incorporate long term seasonal orographical patterns in drawing isohyets for individual storms or seasons. The assumption is that the long term seasonal orographical effect as displayed in the isohyets of season normals applies for individual storms and seasons as well. The procedure involves the following steps, and is worked out in Example 3.2:

1. compute point rainfall as percentage of seasonal normal for all point rainfall stations

- 2. draw isopercentals (= lines of equal actual point rainfall to station normal rainfall) on a transparent overlay
- 3. superimpose the overlay on the seasonal isohyetal map
- 4. mark each crossing of seasonal isohyets with isopercentals
- 5. multiply for each crossing the isohyet with the isopercental value and add the value to the crossing on the map with the observed rainfall values; hence, the data set is extended with the rainfall estimated derived in step 4
- 6. draw isohyets using linear interpolation while making use of all data points, i.e. observed and estimated data (see step 5).

Special attention is to be paid to situations where at the higher elevations raingauge stations are non-existing. Then the orographic effect has to be extrapolated from the lower reaches of the mountains by estimating a relation between rainfall and elevation which is assumed to be valid for the higher elevations as well. Using this rainfall-elevation curve a number of points in the ungauged upper reaches are added to the point rainfall data to guide the interpolation process.

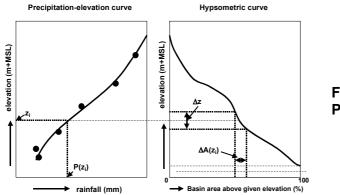


Figure 3.6: Principle of hypsometric method

(3.5)

A simple technique to deal with such situations is the **hypsometric method**, see e.g. Dingman, 1994, where a precipitation-elevation curve is combined with an area-elevation curve (called hypsometric curve) to determine the areal rainfall. The latter method avoids recurrent planimetering of inter-isohyet areas, whereas the results will be similar to the isohyetal method. The precipitation-elevation curve has to be prepared for each storm, month, season or year, but its development will be guided by the rainfall normal-elevation curve also called the orographic equation. Often the orographic equation can be approximated by a simple linear relation of the form:

$$P(z) = a + bz$$

This relation may vary systematically in a region (e.g. the windward side of a mountain range may have a more rapid increase of precipitation with elevation than the leeward side). In such cases separate hypsometric curves and orographic equations are established for the distinguished sub-regions. The areal rainfall is estimated by:

$$\overline{\mathsf{P}} = \sum_{i=1}^{n} \mathsf{P}(\mathsf{z}_{i}) \Delta \mathsf{A}(\mathsf{z}_{i})$$
(3.6)

where: \overline{P} = areal rainfall

 $P(z_i)$ = rainfall read from precipitation-elevation curve at elevation z_i

 $\Delta A(z_i)$ = percentage of basin area contained within elevation $z_i \pm 1/2\Delta z_i$

n = number of elevation interval in the hypsometric curve has been divided.

Example 3.2

In this example the application of the isopercental method is demonstrated (NIH, 1988). The areal rainfall for the storm of 30 August 1982 has to be determined for the catchment shown in Figure 3.7a. The total catchment area amounts 5,600 km². The observed and normal annual rainfall amounts for the point rainfall stations in the area are given in Table 3.3.

Sta	tion	30 August 1982 storm	Normal annual rainfall	Storm rainfall as percentage of annual normal
		(mm)	(mm)	(%)
1.	Paikmal	338.0	1728	19.6
2.	Padampur	177.0	1302	13.6
3.	Bijepur	521.0	1237	42.1
4.	Sohela	262.0	1247	21.0
5.	Binka	158.0	1493	10.6
6.	Bolangir	401.6	1440	27.9

Table 3.3: Storm rainfall and annual normals

For each station the point rainfall as percentage of seasonal normal is displayed in the last column of Table 3.3. Based on this information isopercetals are drawn on a transparent overlay, which is subsequently superimposed on the annual normal isohyetal map. The intersections of the isopercetals and isohyets are identified and for each intersection the isopercental is multiplied with the isohyet to get an estimate of the storm rainfall for that point. These estimates are then added to the point rainfall observations to draw the isohyets, see Figure 3.7b. The inter-isohyet area is planimetered and the areal rainfall is subsequently computed with the aid of equation 3.4 as shown in Table 3.4.

Isohyetal range	Mean rainfall	Area	Volume
(mm)	(mm)	(km²)	(km ² xmm)
110-150	130	80	10400
150-200	175	600	105000
177-200	188.5	600	113100
200-250	225	3370	758250
250-300	275	620	170500
300-400	350	230	80500
400-500	450	90	40500
500-521	510.5	10	5105
	Total	5600	1283355
		Average	1283355/5600=
			229.2 mm

Table 3.4: Computation of areal rainfall by isohyetal/isopercental method

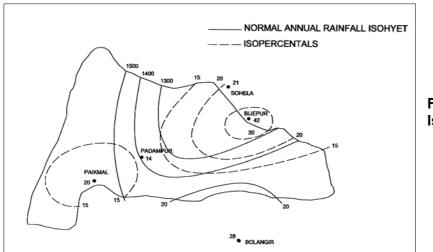


Figure 3.7a: Isopercental map

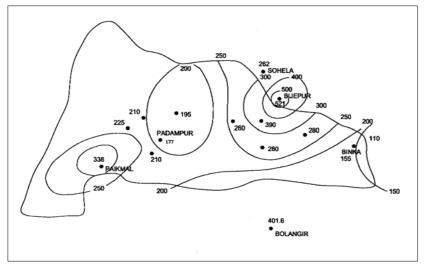


Figure 3.7b: Isohyetal map drawn by isopercental method

3.6 Kriging method

General

The Kriging Method is an interpolation method. It provides rainfall estimates (or estimates of any other variable) at points (point-kriging) or blocks (block-kriging) based on a weighted average of observations made at surrounding stations. In this section point-kriging will be discussed. In the application of the kriging method for areal rainfall estimation and drawing of isohyets a dense grid is put over the catchment. By estimating the rainfall for the gridpoints the areal rainfall is simply determined as the average rainfall of all grid points within the catchment. In addition, in view of the dense grid, it is very easy to draw isohyets based on the rainfall values at the grid points.

At each gridpoint the rainfall is estimated from:

At each gridpoint the rainfall is estimated from:

$$Pe_{0} = \sum_{k=1}^{N} w_{0,k} \cdot P_{k}$$
(3.7)

where: Pe_0 = rainfall estimate at some gridpoint "0"

 $w_{0,k}$ = weight of station k in the estimate of the rainfall at point "0"

- P_k = rainfall observed at station k
- N = number of stations considered in the estimation of Pe₀

The weights are different for each grid point and observation station. The weight given to a particular observation station k in estimating the rainfall at gridpoint "0" depends on the gridpoint-station distance and the spatial correlation structure of the rainfall field. The kriging method provides weights, which have the following properties:

- the weights are linear, i.e. the estimates are weighted linear combinations of the available observations
- the weights lead to unbiased estimates of the rainfall at the grid points, i.e. the expected estimation error at all grid points is zero
- the weights minimise the error variance at all grid points.

Particularly the error variance minimisation distinguishes the kriging method from other methods like e.g. inverse distance weighting. The advantage of the kriging method above other methods is that it provides besides the best linear estimate of rainfall for a point on the grid also the uncertainty in the estimate. The latter property makes the method useful if locations for additional stations have to be selected when the network is to be upgraded, because then the new locations can be chosen such that overall error variance is reduced most.

Bias elimination and error variance minimisation

The claims of unbiasedness and minimum error variance require further explanation. Let the true rainfall at location 0 be indicated by P_0 then the estimation error at "0" becomes:

$$e_0 = Pe_0 - P_0$$

(3.8)

with Pe_0 estimated by (3.7). It is clear from (3.8) that any statement about the mean and variance of the estimation error requires knowledge about the true behaviour of the rainfall at unmeasured locations, which is not known. This problem is solved by hypothesising:

- that the rainfall in the catchment is statistically homogeneous so that the rainfall at all observation stations is governed by the same probability distribution
- consequently, under the above assumption also the rainfall at the unmeasured locations in the catchment follows the same probability distribution as applicable to the observation sites.

Hence, any pair of locations within the catchment (measured or unmeasured) has a joint probability distribution that depends only on the **distance** between the locations and not on their locations. So:

- at all locations E[P] is the same and hence E[P(x₁)] E[P(x₁-d)] = 0, where d refers to distance
- the covariance between any pair of locations is only a function of the distance d between the locations and not dependent of the location itself: C(d).

The unbiasedness implies:

$$E[e_0] = 0 \quad \text{so:} \quad E[\sum_{k=1}^{N} w_{0,k} \cdot P_k] - E[P_0] = 0 \quad \text{or:} \quad E[P_0] \left(\sum_{k=1}^{N} w_{0,k} - 1\right) = 0$$

Hence for each and every grid point the sum of the weights should be 1 to ensure unbiasedness:

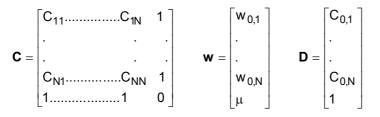
$$\sum_{k=1}^{N} w_{0,k} = 1$$
(3.9)

The error variance can be shown to be (see e.g. Isaaks and Srivastava, 1989):

$$\sigma_{e}^{2} = E[(Pe_{0} - P_{0})^{2}] = \sigma_{P}^{2} + \sum_{i=1}^{N} \sum_{j=1}^{N} w_{0,i} w_{0,j} C_{i,j} - 2\sum_{i=1}^{N} w_{0,i} C_{0,i}$$
(3.10)

where 0 refers to the site with unknown rainfall and i,j to the observation station locations. Minimising the error variance implies equating the N first partial derivatives of σ_e^2 to zero to solve the $w_{0,i}$. In doing so the weights $w_{0,i}$ will not necessarily sum up to 1 as it should to ensure unbiasedness. Therefore, in the computational process one more equation is added to the set of equations to solve $w_{0,i}$, which includes a Lagrangian multiplyer μ . The set of equations to solve the stations weights, also called **ordinary kriging system**, then reads:

where:



Note that the last column and row in **C** are added because of the introduction of the Lagrangian multiplyer μ in the set of N+1 equations. By inverting the covariance matrix the station weights to estimate the rainfall at location 0 follow from (3.11) as:

$$w = C^{-1} \cdot D$$
 (3.12)

The error variance is then determined from:

$$\sigma_e^2 = \sigma_P^2 - \mathbf{w}^T \cdot \mathbf{D}$$
(3.13)

From the above equations it is observed that C^{-1} is to be determined only once as it is solely determined by the covariances between the observation stations being a function of the distance between the stations only. Matrix **D** differs for every grid point as the distances between location "0" and the gauging stations vary from grid point to grid point.

Covariance and variogram models

To actually solve above equations a function is required which describes the covariance of the rainfall field as a function of distance. For this we recall the correlation structure between the rainfall stations discussed in module 9. The spatial correlation structure is usually well described by an exponential relation of the following type:

$$r(d) = r_0 \exp(-d/d_0)$$
 (3.14)

where: r(d) = correlation coefficient as a function of distance

- r_0 = correlation coefficient at small distance, with $r_0 \le 1$
- d_0 = characteristic correlation distance.

Two features of this function are of importance:

- $r_0 \leq$ 1, where values < 1 are usually found in practice due to measurement errors or micro-climatic variations
- the characteristic correlation distance d₀, i.e the distance at which r(d) reduces to 0.37r₀.
 It is a measure for the spatial extent of the correlation, e.g. the daily rainfall d₀ is much smaller than the monthly rainfall d₀. Note that for d = 3 d₀ the correlation has effectively vanished (only 5% of the correlation at d = 0 is left).

The exponential correlation function is shown in Figure (3.8).

The covariance function of the exponential model is generally expressed as:

$$\begin{array}{lll} C(d) = C_0 + C_1 & \mbox{for} & d = 0 \\ C(d) = C_1 \, \exp(-\frac{3d}{a}) & \mbox{for} & d > 0 \end{array} \tag{3.15}$$

Since according to the definition $C(d) = r(d)\sigma_P^2$, the coefficients C_0 and C_1 in (3.15) can be related to those of the exponential correlation model in (3.14) as follows:

$$C_0 = \sigma_P^2(1-r_0)$$
; $C_1 = \sigma_P^2 r_0$ and $a = 3d_0$ (3.16)

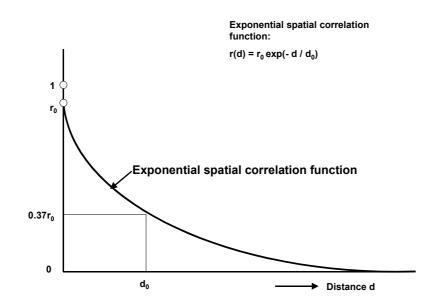


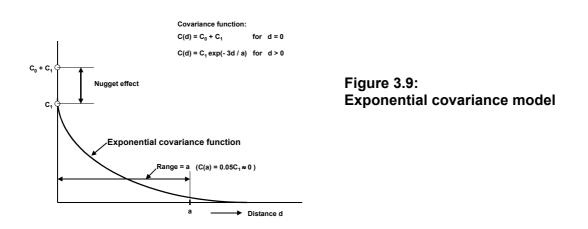
Figure 3.8 Spatial correlation structure of rainfall field

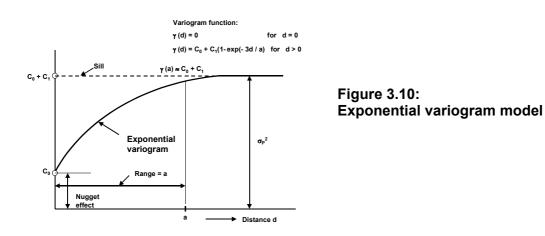
In kriging literature instead of using the covariance function C(d) often the semi-variogram $\gamma(d)$ is used, which is halve of the expected squared difference between the rainfall at locations distanced d apart; $\gamma(d)$ is easily shown to be related to C(d) as:

$$\gamma(d) = \frac{1}{2} E[\{P(x_1) - P(x_1 - d)\}^2] = \sigma_P^2 - C(d)$$
(3.17)

Hence the (semi-)variogram of the exponential model reads:

$$\gamma(d) = 0$$
 for : $d = 0$
 $\gamma(d) = C_0 + C_1 \left(1 - \exp(-\frac{3d}{a}) \right)$ for : $d > 0$ (3.18)





Features of the exponential model are following:

- C₀, which is called the **nugget effect**, provides a discontinuity at the origin; according to (3.16): C₀ = σ_P²(1-r₀), hence in most applications of this model to rainfall data a small nugget effect will always be present
- The distance 'a' in the covariance function and variogram is called the **range** and refers to the distance above which the functions are essentially constant; for the exponential model a = 3d₀ can be applied
- $C_0 + C_1$ is called the **sill** of the variogram and provides the limiting value for large distance and becomes equal to σ_P^2 ; it also gives the covariance for d = 0.

Other Covariance and semi-variogram models

Beside the exponential model other models are in use for ordinary kriging, viz:

- Spherical model, and
- Gaussian model

These models have the following forms:

Spherical:

$$\gamma(d) = C_0 + C_1 \left(\left(\frac{3}{2} \frac{d}{a} - \frac{1}{2} \left(\frac{d}{a} \right)^3 \right) \quad \text{if} \quad d \le a$$

$$\gamma(d) = 1 \quad \text{otherwise}$$
(3.19)

Gaussian:

$$\gamma(d) = C_0 + C_1 \left(1 - \exp\left(-\frac{3d^2}{a^2}\right) \right)$$
(3.20)

The Spherical and Gaussian models are shown with the Exponential Model in Figure 3.11.

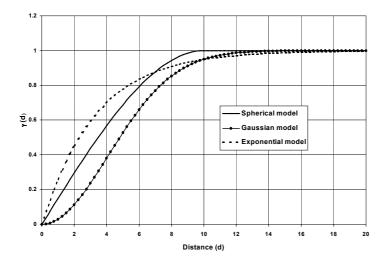
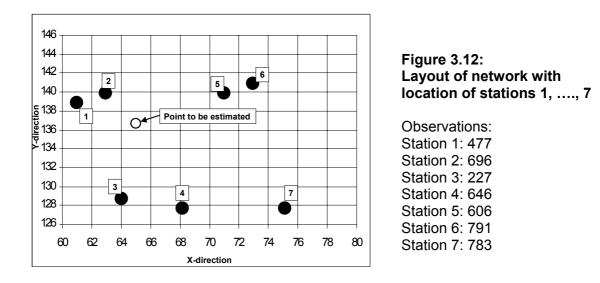


Figure 3.11: Example of Spherical, Gaussian and Exponential type of variogram models, with $C_0=0, C_1=1$ and a = 10

The spherical model has a linear behaviour at small separation distances near the origin, with the tangent at the origin intersecting the sill at about 2/3 of the range "a". The model reaches the sill at the range. The gaussian model is fit for extremely continuous phenomena, with only gradually diminishing correlation near the origin, much smoother than the other two models. The range "a" is at a distance the variogram value is 95% of the sill. The exponential model rises sharper than the other two but flattens out more gradually at larger distances; the tangent at the origin reaches the sill at about 1/5 of the range.

Sensitivity analysis of variogram model parameters

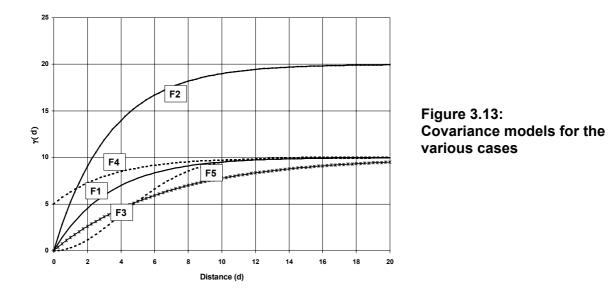
To show the effect of variation in the covariance or variogram models on the weights attributed to the observation stations to estimate the value at a grid point an example presented by Isaaks and Srivastava (1989) is presented. Observations made at the stations as shown in Figure 3.12 are used. Some 7 stations are available to estimate the value at point '0' (65,137).



The following models (cases) have been applied to estimate the value for "0":

Case 1:	$\gamma_1(d) = 10 \left(1 - \exp(-\frac{3d}{10}) \right)$	$C_0 = 0$	C ₁ = 10	a = 10	
Case 2 :	$\gamma_2(d) = 20 \left(1 - \exp(-\frac{3d}{10}) \right) = 2\gamma_1$	1(d) $C_0 = 0$	C ₁ = 20	a = 10	
Case 3 :	$\gamma_{3}(d) = 10 \left(1 - \exp(-3(\frac{d}{10})^{2})\right)$	$C_{0} = 0$	C ₁ = 10	a = 10	(Gaussian)
Case 4 :	$\gamma_4(d) = 0$ for	or : d = 0			
	$\gamma_4(d) = 5 + 5 \left(1 - exp(-\frac{3d}{10})\right)$ for	or: $d > 0$ $C_0 = 5$	6 C ₁ = 5	a = 10	
Case 5 :	$\gamma_5(d) = 10 \left(1 - \exp(-\frac{3d}{20})\right)$	$C_0 = 0$	C ₁ = 10	a = 20	

Figure The covariance and variograms for the cases are shown in Figures 3.13 and 3.14.



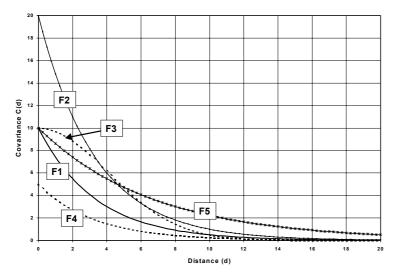


Figure 3.14: Semi-variograms for the various cases

Page 20

The results of the estimate and variance at point "0" as well as the weights of the stations computed with the models in estimating point "0" are presented in Table 3.5.

Case	Estimate	Error	Stations	1	2	3	4	5	6	7
	at "0"	variance	Distance to "0"	4.47	3.61	8.06	9.49	6.71	8.94	13.45
	(mm)	(mm ²)					weights			
1	593	8.86		0.17	0.32	0.13	0.09	0.15	0.06	0.09
2	593	17.91		0.17	0.32	0.13	0.09	0.15	0.06	0.09
3	559	4.78		-0.02	0.68	0.17	-0.01	0.44	-0.29	0.04
4	603	11.23		0.15	0.18	0.14	0.14	0.13	0.13	0.14
5	572	5.76		0.18	0.38	0.14	0.07	0.20	0.00	0.03
ID	590	-		0.44	0.49	0.02	0.01	0.02	0.01	0.01

Table 3.5:Results of computations for Cases 1 to 5 and ID
(=Inverse Distance Method) with p = 2

From the results the following can be concluded:

- Effect of scale: compare Case 1 with Case 2
 In Case 2 the process variance, i.e. the sill is twice as large as in Case 1. The only effect this has on the result is a doubled error variance at "0". The weights and therefore also the estimate remains unchanged. The result is easily confirmed from equations (2.12) and (3.13) as both C, D and σ_P² are multiplied with a factor 2 in the second case.
- Effect of shape: compare Case 1 with Case 3 In Case 3 the spatial continuity near the origin is much larger than in Case 1, but the sill is the same in both cases. It is observed that in Case 3 the estimate for "0" is almost entirely determined by the three nearest stations. Note that kriging does cope with clustered stations; even negative weights are generated by stations in the clusters of stations (5, 6) and (1, 2) to reduce the effect of a particular cluster. Note also that the estimate has changed and that the error variance has reduced as more weight is given to stations at small distance. It shows that due attention is to be given to the correlation structure at small distances as it affects the outcome significantly.
- The nugget effect: compare Case 1 with Case 4 In Case 4, which shows a strong nugget effect, the spatial correlation has substantially been reduced near the origin compared to Case 1. As a result the model discriminates less among the stations. This is reflected in the weights given to the stations. It is observed that almost equal weight is given to the stations in Case 4. In case correlation would have been zero the weights would have been exactly equal.

• Effect of range: compare Case 1 with Case 5

The range in Case 5 is twice as large as in Case 1. It means that the spatial correlation is more pronounced than in Case 1. Hence one would expect more weight to the nearest stations and a reduced error variance, which is indeed the case as can be observed from Table 3.5. Cases1 and 5 basically are representative for rainfall at a low and high aggregation level, respectively (e.g. daily data and monthly data).

There are more effects to be concerned about like effects of anisotropy (spatial covariance being direction dependent) and spatial inhomogeneity (like trends due to orographic effects). The latter can be dealt with by normalising or detrending the data prior to the application of kriging and denormalise or re-invoke the trend after the computations. In case of anisotropy the contour map of the covariance surface will be elliptic rather than circular. Anisotropy will require variograms to be developed for the two main axis of the ellips separately.

Estimation of the spatial covariance function or variogram.

Generally the spatial correlation (and hence the spatial covariance) as a function of distance will show a huge scatter as shown in Figures 1.1 to 1.4 of Module 9. To reduce the scatter the variogram is being estimated from average values per distance interval. The distance intervals are equal and should be selected such that sufficient data points are present in an interval but also that the correct nature of the spatial correlation is reflected in the estimated variogram.

Alternative to kriging

HYMOS offers an alternative to station weight determination by kriging through the inverse distance method. In this method the station weights and estimate are determined by:

$$Pe_{0} = \sum_{k=1}^{N} w_{0,k} \cdot P_{k} \qquad w_{0,k} = \frac{1/d_{k}^{p}}{\sum_{j=1}^{N} 1/d_{j}^{p}}$$
(3.21)

It is observed that the weights are proportional to the distance between "0" and station j to some power p. For rainfall estimation often p = 2 is applied.

Different from kriging the Inverse Distance Method does not take account of station **clusters**, which is convincingly shown in Table 3.5, last row; the estimate for "0" is seen to be almost entirely determined by the cluster (1, 2) which is nearest to "0". Hence, this method is to be applied only when the stations are more or less evenly distributed and clusters are not existing.

Example 3.3: Application of Kriging Method

The kriging method has been applied to monthly rainfall in the BILODRA catchment, i.e. the south-eastern part of the KHEDA basin in Gujarat. Daily rainfall for the years 1960-2000 have been aggregated to monthly totals. The spatial correlation structure of the monthly rainfall for values > 10 mm (to eliminate the dry periods) is shown in Figure 3.15.

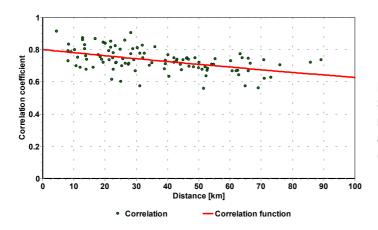


Figure 3.15: Spatial correlation structure of monthly rainfall data in and around Bilodra catchment (values > 10 mm)

From Figure 3.15 it is observed that the correlation only slowly decays. Fitting an exponential correlation model to the monthly data gives: $r_0 \approx 0.8$ and $d_0 = 410$ km. The average variance of the monthly point rainfall data (>10 mm) amounts approx. 27,000 mm². It implies that the sill of the semi-variogram will be 27,000 mm² and the range is approximately 1200 km ($\approx 3 d_0$). The nugget is theoretically $\sigma_P^2(1-r_0)$, but is practically obtained by fitting the semi-variogram model to the semi-variance versus distance plot. In making this plot a lag-distance is to be applied, i.e. a distance interval for averaging the semi-variances to reduce the spread in the plot. In the example a lag-distance of 10 km has been applied. The results of the fit overall and in detail to a spherical semi-variogram model is shown in Figure 3.16. A nugget effect (C_0) of 2000 mm² is observed.

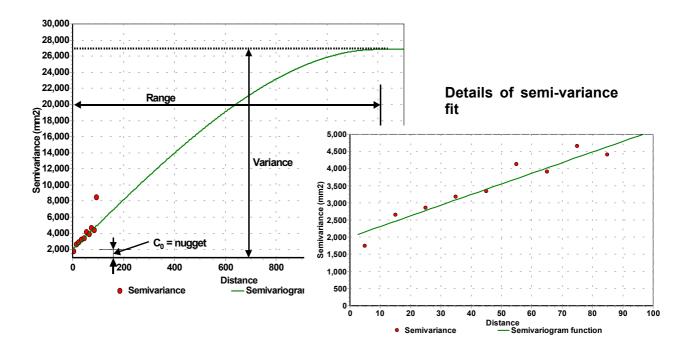


Figure 3.16: Fit of spherical model to semi-variance, monthly rainfall Bilodra

Similarly, the semi-variance was modelled by the exponential model, which in Figure 3.17 is seen to fit in this case equally well, with parameters $C_0 = 2,000 \text{ mm}^2$, Sill = 27,000 mm² and Range = 800 km. Note that C_0 is considerably smaller than one would expect based on the spatial correlation function, shown in Figure 3.15. To arrive at the nugget value of 2,000 mm² an r_0 value of 0.93 would be needed. Important for fitting the semi-variogram model is to apply an appropriate value for the lag-distance, such that the noise in the semi-variance is substantially reduced.

The results with the spherical model applied to the rainfall of June 1984 in the BILODRA catchment is shown in Figure 3.18. A grid-size of 500 m has been applied. The variance of the estimates is shown in Figure 3.19. It is observed that the estimation variance at the observation points is zero. Further away from the observation stations the variance is seen to increase considerably. Reference is made to Table 3.6 for a tabular output.

For comparisons reasons also the isohyets derived by the inverse distance method is shown, see Figure 3.20. The pattern deviates from the kriging results in the sense that the isohyets are more pulled towards the observation stations. As was shown in the sensitivity analysis, the nearest station(s) weigh heavier than in the kriging method.

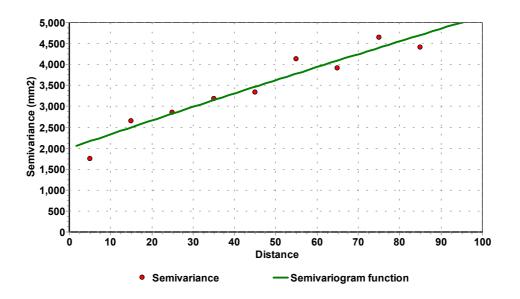


Figure 3.17: Fit of exponential model to semi-variogram, monthly data Bilodra

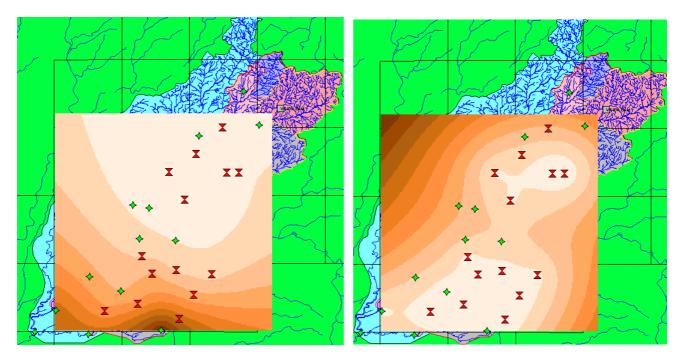


Figure 3.18 Figure 3.19 Isohyets of June 1984 rainfall in Bilodra catchment using by spherical semi-variogram model (Figure 3.18) and the variance of the estimates at the grid-points (Figure 3.19).

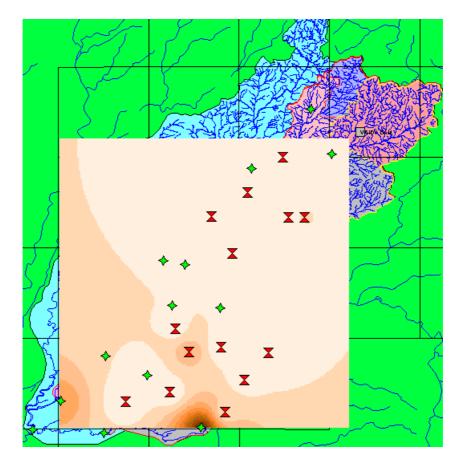


Figure 3.20: Isohyets derived for June 1984 rainfall in Bilodra catchment using inverse distance weighting (compare with Figure 3.18)

Example output of interpolation by kriging Table 3.6:

Variogram parameters			
Nugget (CO): Sill (C1) Range (a):	2000.000000 25000.000000 1200.000000		
Grid characteristics: Number of cells in X, Y Origin of X and Y Block Size of X and Y Blocks:	s: 0.00000E+00		
Search Radius: Minimum number of samples Maximum number of samples)	
Data: ANIOR MP2 1 at Data: BALASINOR MP2 1 at Data: BAYAD MP2 1 at Data: BHEMPODA MP2 1 at Data: DAKOR MP2 1 at Data: KAPADWANJ MP2 1 at Data: KATHLAL MP2 1 at Data: MAHEMDABADMP2 1 at Data: MAHISA MP2 1 at Data: MAHISA MP2 1 at	59.317 49.430 70.017 39.945 32.921 24.756	63.440 value: 21.981 value: 52.552 value: 63.390 value: 552 value: 29.687 value: 15.644 value: 7.998 value: 10.327 value: 7.421 value:	.00000 1.00000 18.70000 176.00000 11.00000 .00000

Data: Data: Data:	SAVLITANK THASARA VADAGAM VADOL VAGHAROLI	MP2 MP2 MP2	l at l at l at	36.817 46.848 43.604 45.951 52.382		value: value: value:	54.00000 22.00000 .00000 .00000 5.00000
Estimated average variance		400 17.58 101.39	3128				
Column		Row		Estimate	Variance		
1		1		45.806480	2685.862000		
1		2		45.719250	2680.906000		
1		3		45.625660	2676.289000		
1		4		45.525860	2672.001000		
1		5		45.420020	2668.018000		
etc.							

4. Transformation of non-equidistant to equidistant series

Data obtained from digital raingauges based on the tipping bucket principle may sometime be recording information as the time of each tip of the tipping bucket, i.e. a non-equidistant series.

HYMOS provides a means of transforming such non-equidistant series to equidistant series by accumulating each unit tip measurement to the corresponding time interval. All those time interval for which no tip has been recorded are filled with zero values.

5. Compilation of minimum, maximum and mean series

The annual, seasonal or monthly maximum series of rainfall is frequently required for flood analysis, whilst minimum series may be required for drought analysis. Options are available in HYMOS for the extraction of minimum, maximum, mean, median and any two user-defined percentile values (at a time) for any defined period within the year or for the complete year.

For example if the selected time period is 'monsoon months' (say July to October) and the time interval of the series to be analysed is 'ten daily', then the above statistics are extracted for every monsoon period between a specified start and end date.

Example 5.1

From daily rainfall records available for MEGHARAJ station (KHEDA catchment), ten-daily data series is compiled. For this ten-daily data series for the period 1961 to 1997, a few statistics like minimum, maximum, mean, median and 25 & 90 %ile values are compiled specifically for the period between 1st July and 30th Sept. every year.

These statistics are shown graphically in Fig. 5.1 and are listed in tabular form in Table 5.1. Data of one of the year (1975) is not available and is thus missing. Lot of inferences may be derived from plot of such statistics. Different pattern of variation between 25 % ile and 90 % ile values for similar ranges of values in a year may be noticed. Median value is always lower than the mean value suggesting higher positive skew in the ten daily data (which is obvious owing to many zero or low values). A few extreme values have been highlighted in the table for general observation.

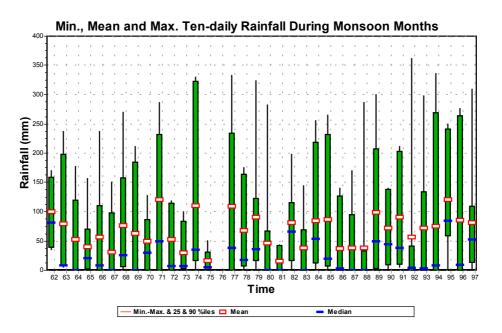


Fig 5.1: Plot of statistics of ten-daily rainfall series at MEGHARAJ station

Year	Min.	Max.	Mean	Median	25 %ile	90 %ile
1961	34.54	170.39	99.6	81.03	39.36	158.47
1962	5.6	237.6	78.9	8.6	8.4	197.5
1963	0	177.44	53.0	0	0	119.1
1964	0	157.2	39.7	20.7	1.7	69.6
1965	0	237	56.3	8	0	110.6
1966	0	151	31.4	0	0	98
1967	0	270	75.9	26	6	158
1968	0	211	63.0	0	0	185
1969	0	128	49.2	30	0	87
1970	0	287	120.7	50	0	232
1971	0	118.5	53.1	7	0	114
1972	0	99.6	29.9	7	2.6	83.3
1973	0	330.4	110.8	34.8	17	322.6
1974	0	51	16.5	5	1.5	31.2
1976	0	333.4	108.8	38.2	0	234.2
1977	0	175.4	67.6	18	7	164
1978	0	324	90.3	36	16	123
1979	0	282	46.0	0	0	67
1980	0	43	15.3	0	0	42
1981	0	198	81.0	65.5	16	115.5
1982	0	144	38.5	0	0	69
1983	0	256	84.7	54	12	219
1984	0	265	87.0	19.5	7.5	231.5
1985	0	140.5	36.9	3	0	127
1986	0	170	38.4	0	0	94.5
1987	0	287	38.5	0	0	33
1988	0	300	99.0	50	3	207
1989	0	140	72.3	44.5	9	138.5
1990	5	211.5	91.1	38.5	10	203.5
1991	0	361.5	56.7	4	0	41.5
1992	0	298	72.2	3	0	134
1993	0	336.5	75.7	8	0	269
1994	0	249	121.1	85	58.5	241.5
1995	0	276.5	85.9	9.5	0	264
1996	0	309	81.9	52.5	13.5	109
1997	0	391	105.7	23	10	242.5
Full Period	0	391	68.7			

Table 5.1: Ten-daily statistics for MEGHARAJ station between 1st July and 30st Sept