

Thiessen—Software for Automatic Estimation of Weighted Average Rainfall

N. Panigrahy, S.K. Jain, Vijay Kumar and P.K. Bhunya

National Institute of Hydrology
Roorkee, INDIA

ABSTRACT: Estimation of average areal rainfall from point measurements is the first most important step in hydrological modelling. The Thiessen polygon method is commonly used for the purpose. This paper presents the development and usage of a MS Windows based software for estimation of weighted average rainfall over a basin using Thiessen Polygon Method. The software requires the vertex points of the basin boundary described by a polygon, location of raingauges as points and the monthly rainfall data at all the raingauges. The software uses a vector method, which is very fast and accurate, for the estimation of Thiessen weights depending on data status (availability of data at different gauges) of the basin for a specific month. Then it computes the average rainfall over the basin using the estimated weights. Along with the weighted average rainfall series, the software also gives other useful information as output such as data status at the gauges, Thiessen weights and Thiessen polygon in graphical form for different combination of gauges etc. There is provision for saving the text output in simple text as well as in popular spreadsheet (Excel) format, and the graphical output in commonly used graphical formats (bmp, tif, jpg, png, emf and wmf). In a case study the software is used to estimate the monthly weighted rainfall of a basin for 105 years, which it completed in 35 seconds on a Pentium IV PC.

INTRODUCTION

A detailed water balance study requires hydrological modeling of a basin that can be by empirical, conceptual or physically based distributed models. Precipitation data forms the basic data for most watershed models. The empirical and conceptual models require average areal rainfall over the watershed whereas the distributed models require the spatially varied rainfall to estimate the runoff. Hence, estimation of spatial variation of rainfall or average areal rainfall is the first and the foremost requirement of hydrological modeling.

The accuracy of average areal rainfall estimation depends on the network configuration. The estimation error of a raingauge network with a regular configuration is less than that of a random network (Seed and Austin, 1990). When the network density is high and raingauge stations are uniformly distributed simpler methods such as arithmetic average method may give reasonable accuracy in estimation of average areal rainfall and inverse (square) distance method may be useful to represent the spatial variability of rainfall. However, when the raingauge network is sparse, the accuracy of estimation depends on the method used. Many methods have been devised to estimate the average areal rainfall for sparse networks using different principles, such as, the Principle of Proximal Mapping (Thiessen, 1911), the Multivariate Estimation Theory (Bras and Rodriguez-Iturbe, 1976a,

1976b), Double Fourier Series (Thorpe *et al.*, 1979), Principal Component Analysis (Lane *et al.*, 1979), Kriging (Bastin *et al.*, 1984), Multiquadric Equations (Balascio, 2001). Tabios and Salas (1985) reported a comparative analysis of the techniques for spatial variation of rainfall and concluded that the optimal interpolation and Kriging techniques are the best for estimating annual precipitation.

Field engineers mostly rely on Thiessen Polygon method because of its simplicity and reasonable accuracy. This method can be used to quickly estimate the average rainfall over a basin. In this method, weights are assigned to each raingauge in the network and the weighted average rainfall is computed as the estimate of the areal average rainfall in the basin. Hence, to use this method, the weight of each raingauge in the network should be determined beforehand. To determine the Thiessen polygons and associated weights manually, a map containing the basin boundary, represented by a closed polygon, and the raingauge locations, represented by points, is prepared. All the points representing raingauge locations are joined by straight lines and perpendicular bisectors are drawn to these. The intersection of the bisector lines among themselves and with the basin boundary, results in the Thiessen polygons for the raingauges. The area of each polygon is determined by planimetry or any other suitable means and the weights are determined as the

ratio of area of influence of each gauge to the total area of the basin. When the network density is relatively high with few closely spaced raingauge stations, it becomes difficult to interpret the actual influence area for each gauge. Hence, Geographical Information System (GIS) software are increasingly used nowadays for determination of Thiessen weights.

Usage of GIS software for estimation of weights has reduced the human interpretation as well as the time requirement. However, when the configuration of the network changes due to addition or discontinuance of some gauges or unavailability of data at certain gauges, the whole GIS operations need to be repeated to recalculate the weights. This situation is frequent in real-life applications. Under this situation, one may consider estimating the weights for every possible combination of raingauges and select the weights depending on the prevailing combination. However, this requires enormous amount of data processing. For a network consisting of n raingauges, selection of any two raingauges with concurrent data can be done in nC_2 ways. Therefore, the total number of possible combinations will be,

$$\sum_{i=2}^n {}^nC_i = 2^n - n - 1$$

For instance 1013 possible combinations can be made out for a basin with 10 raingauges. Hence, it is desirable to estimate the weights in an automated method which is capable of accommodating the changes in the network dynamically.

For computerized estimation of Thiessen weights significant work has been done and methodologies have been developed (Diskin, 1969, 1970; Shih and Hamrick, 1975). Croley and Hartmann (1985) presented the merits and limitations of these methods and suggested a modified raster based method using edge detection technique to reduce the computational time. However, the accuracy of estimated weights was dependant on the resolution of rasterisation. The main objective of this study was to devise a methodology to estimate Thiessen weights quickly without compromising on accuracy.

ESTIMATION OF THIESSEN WEIGHTS

In this study, a vector approach was used to estimate the Thiessen weights. In this approach, the basin is divided (sliced) successively to determine the area of influence of each gauge. The boundary of the basin is defined by a polygon having a finite number of vertices numbered consecutively in clockwise (or anticlockwise) direction and the location of raingauges are defined by points (Figure 1).

Computations start with gauge no. 1. Assuming that only one gauge (i.e., gauge no. 1) is available in the basin, the total area of the basin represents the area of influence for gauge no. 1. When a second gauge (no. 2) is added to the network, depending on its position, the area of influence for gauge no. 1 will change. This change is reflected in the polygon representing area of influence of gauge no. 1 which results in reduction of number of vertices of the polygon. This is illustrated in Figure 1 where gauges 1 and 2 are joined by a straight line segment, $L(1, 2)$. The straight line $B(1, 2)$ is the perpendicular bisector of $L(1, 2)$. $B(1, 2)$ is then extended infinitely on both the sides of $L(1, 2)$ so that it cuts the area of influence of gauge 1 between two pair of vertices ($[i, i+1]$ and $[j, j+1]$) The points of intersection are designated as $ip1$ and $ip2$ respectively. Since the points i and $j+1$ are nearer to gauge 1, the boundary segment from vertex $i+1$ to j is replaced by only $ip1$ and $ip2$ (i.e., points of intersection) in the polygon. The new polygon, thus formed (the hatched portion) is treated as the area of influence for gauge no. 1 with the network having two gauges.

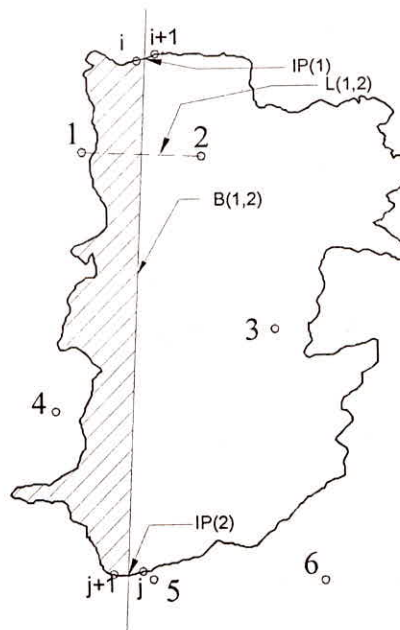


Fig. 1: Area for gauge 1 after gauge 2 is added

Addition of a third gauge (no. 3) further changes the area of influence for gauge no. 1 (Figure 2). In this figure, similar operations are performed as explained above, but in place of gauge 2, gauge 3 is considered as the addition to the network and in place of the whole basin, only the area of influence for gauge 1 (after consideration of effect of gauge 2) is taken. In this manner the remaining gauges in the network are added one by one and the area of influence for the

gauge no. 1 is modified for each addition. After addition of the final gauge, the area of influence of the first gauge represents the Thiessen polygon for that gauge (Figure 3).

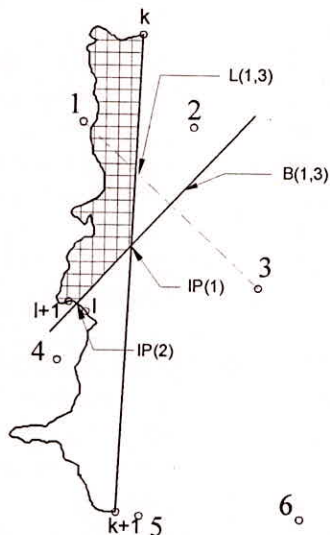


Fig. 2: Area for gauge 1 after gauge 3 is added

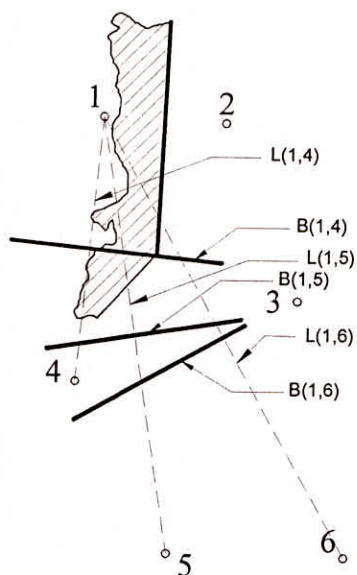


Fig. 3: Area for gauge 1 after all gauges are added

In some cases, the perpendicular bisector may not intersect the polygon of the area of influence, which means that the entire area is nearer to either the gauge under consideration or to the gauge added. The nearer gauge is determined by comparing the distance of any vertex point from both the gauges. If the area is nearer to the gauge added, the weight for the gauge under consideration is treated as zero and the polygon representing the area of influence will have no vertex. Under such situation, no further addition of gauge to the network is done. If the entire area is nearer to the

gauge under consideration, no correction in the polygon vertices is required. It can be observed in Figure 4 that after addition of gauge 4 to the network there is no effect of addition of gauges 5 and 6 to the network because the bisectors $B(1, 5)$ and $B(1, 6)$ do not intersect the area of influence boundary for gauge 1. Further, all the vertices of the polygon are nearer to gauge 1 (i.e., the gauge under consideration).

Repeating the same procedure for all the gauges results in the Thiessen polygons for all the gauges are determined and the sum of areas of these polygons gives the total area of the basin. The weight for each gauge is computed as the ratio of its area of influence to the basin area.

All these operations such as bisection of a line segment, determination of point of intersection of two lines, distance between two points, area of a polygon etc. can be done analytically using simple algebraic calculations by representing the points in a two dimensional Cartesian co-ordinate system and lines by linear equations.

DEVELOPMENT OF SOFTWARE

Data associated with this problem can be divided into three types viz. static, semi-static and time series data. The static data are the boundary points (BP), number of points describing the boundary (NB), the semi-static data are raingauge details i.e., its name and location (RG), and number of raingauges (NG). The time series data are the monthly rainfall data of all Raingauge Stations (RF). The details of the data used are presented in Table 1.

Table 1: Description of Data Types

Data	Details	Storage
BP	Structure type (point) x, y : real (double precision)	Array of NB points
NB	Basic type integer	Single integer
RG	Structure type Name: string of 30 characters z : structure type (point)	Array of NG raingauges
NG	Basic type integer	Single integer
RF	Basic type double precision	Two dimensional array, where, the 1 st dimension represents the year and the 2 nd dimension represents the month <i>Defined for each gauge</i>

Since the values of NB and NG are known only during run time, BP and RG are declared as pointer variables and memory for these are allocated

dynamically when the values of NB and NG are defined. The number of years (N_k) for which data are available for a specific gauge 'k' is also a variable quantity and known during run time; whereas, the number of months of data in a specific year is always known (12 months). Hence, RF is declared as a pointer to an array of 12 real numbers. Memory is allocated dynamically for storage of RF when the value of N is defined.

Using the principle described in the previous section, a subroutine (ThiessenWeights) was developed in C language for estimation of Thiessen weights. The inputs for this subroutine are the array of boundary points, number of points defining the boundary, array of rain-gauges and number of rain-gauges in the combination. An array containing the weights for all the rain-gauges forms the output.

Linking the main subroutine (ThiessenWeights) the software was developed to estimate of historic average areal rainfall. The Graphical User Interface (GUI) was developed using VisualStudio (VC++ 7). The software works on the following steps:

1. Read the static, semi-static and time series data (i.e., NB, BP, NG, RG and RF).
2. Determine the period for which estimation is to be done (start and end years & months) from the rainfall data of all the stations or as user input. Set the current time (month and year).
3. Determine the effective network depending on data availability.
4. If the current time is other than the time of start, check if the weights for the same network are already available in temporary storage. If weights are not available then call subroutine Thiessen Weights to estimate the weights for all the gauges in the network at the current time (W_i) and keep the weights for the network in temporary storage. Otherwise read the weights from the temporary storage.
5. Estimate the mean areal rainfall using formula $\bar{R}(y, m) = \sum_{i=1}^N R_i(y, m) W_i$ and store it in the output file.
6. Increment the time by one month and repeat steps (iii) to (v) till the final year and month for which estimation is required.

The flow chart of the software is presented in Figure 4. Estimation of Thiessen weights for a combination is the most time consuming step in the process. It can be observed from the flow chart that the program estimates Thiessen weights for selective combinations only, depending on the data availability as opposed to

estimating the weights for each time step. Hence, substantial amount of time is saved when the period of estimation is large.

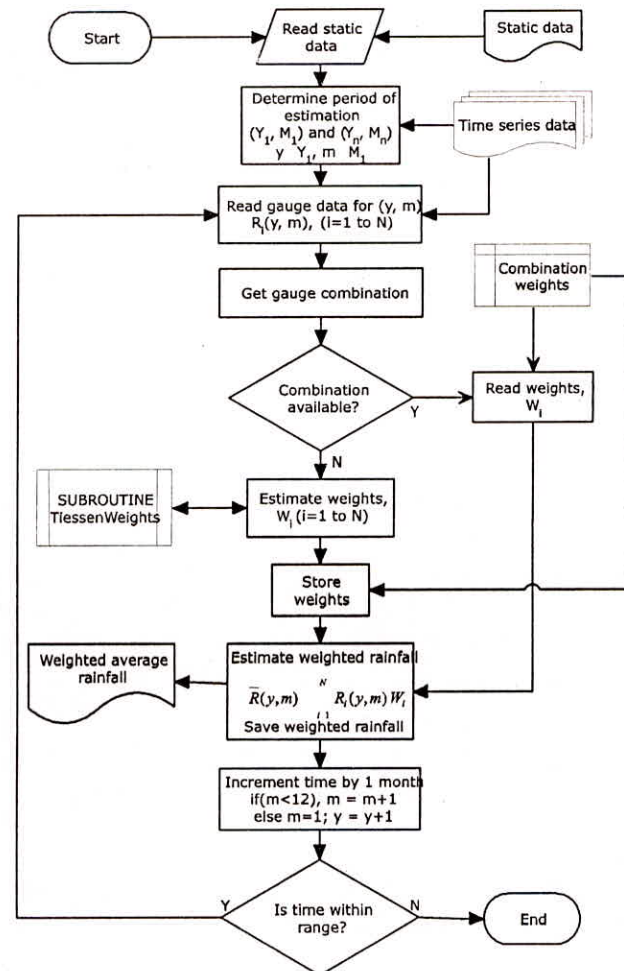


Fig. 4: Flow chart for Thiessen

SOFTWARE DESCRIPTION

The software developed in this work is a windows dialog based application. The initial dialog is shown in Figure 5. The filename fields (input and output) are filled when file(s) are selected through buttons [...] adjacent to these fields, which open the normal Windows file open dialog when clicked. In this dialog a previously created file can be selected or a name can be entered to create a new file. The output file name is automatically created, the name can be changed using [...] button against it.

Clicking [Create/Edit] button results in the basin description creation/editing dialog (Figure 6) in which the basin name is given, and list of rain-gauges with their coordinates and basin boundary coordinates are entered.

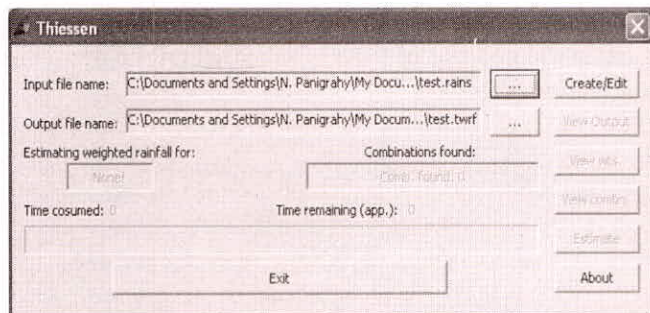


Fig. 5: Initial window of Thiessen

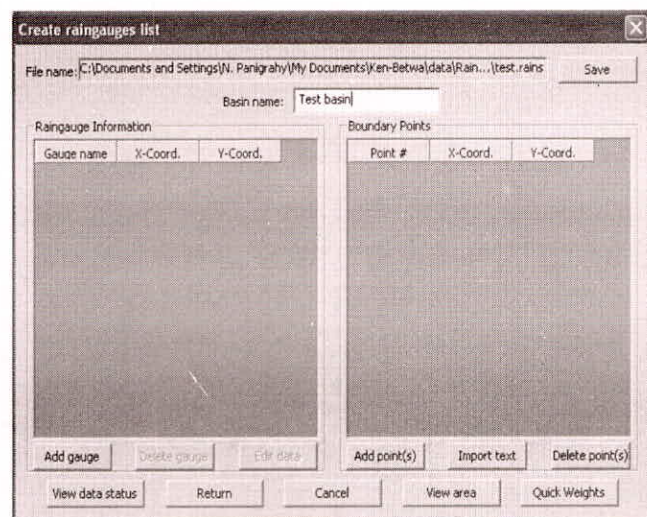


Fig. 6: Window for creating basin and raingauge description

The dialog for creation of raingauge name (Figure 7) is produced when [Add gauge] button is clicked. In this dialog:

1. A single raingauge can be created by entering a unique name.
2. A list of gauges along with their coordinates stored in a text file can also be imported using [Import file] button.
3. A raingauge file saved earlier can also be imported using [add an existing gauge] button. However, the file should be in the same folder as that of the main input file. Further, no duplication in gauge name is allowed.

The [Add points] button produces a dialog (Figure 8) in which the number of points and position at which these points are to be inserted are specified. This produces specified number of empty cells in the boundary points list. The boundary points stored in a text file can also be imported using [Import text] button.

Double clicking the gauge name or through [Edit data] button the rainfall data for the gauge can be entered/edited using a dialog box (Figure 10).

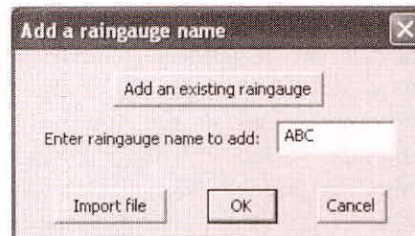


Fig. 7: Dialog for creating a raingauge

After creation of gauges and boundary points the co-ordinates can be entered as shown in Figure 9.

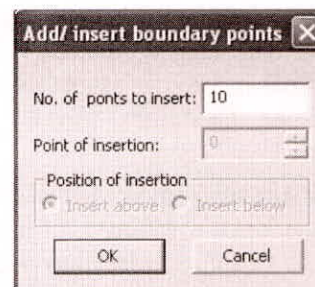


Fig. 8: Dialog for adding/inserting boundary point(s)

If some data are already available for the gauge, the data are shown in the data area. A blank data area means no data are available for the gauge. After the first and last years of data availability are entered, clicking the [Create/Modify Series] button creates or modifies the grid for data entry. The data area is colour coded (a yellow background refers no data is available for the specific cell). The months for which data are not available are to be left blank (no negative number or text is to be entered). After entering, the data should be saved.

For estimation of average areal rainfall for a period, the combinations of gauges are created dynamically depending on availability of data at different gauges for a specific month. However, the Thiessen weights for different combinations can be estimated without entering any data for these gauges. (The weights are dependant on the location of the gauges only). This can be accomplished by using the button [Quick weights] in the widow for creating basin and raingauge description (Figure 6). Clicking this button, a dialog containing all the raingauge names (Figure 11) is created in which the required combination can be selected by putting ticks (clicking on the box against each gauge). After deciding the combination, clicking the button [Estimate], the weights are estimated and shown against each gauge in the designated column. To view the graphical output [Show Figure] button may be used. This produces the result as shown in Figure 12.

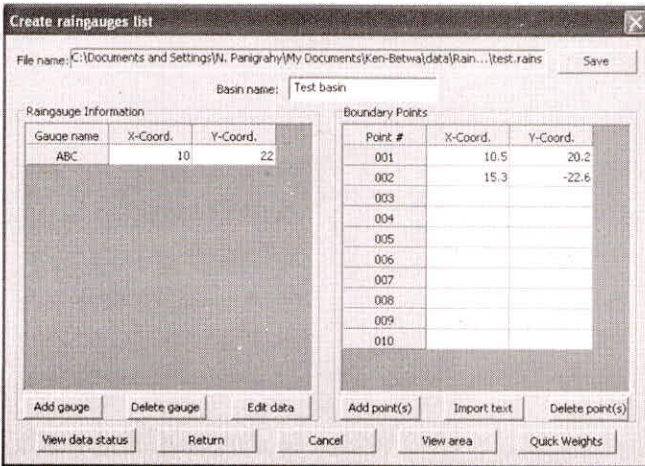


Fig. 9: Widow for entering basin and rainauge description

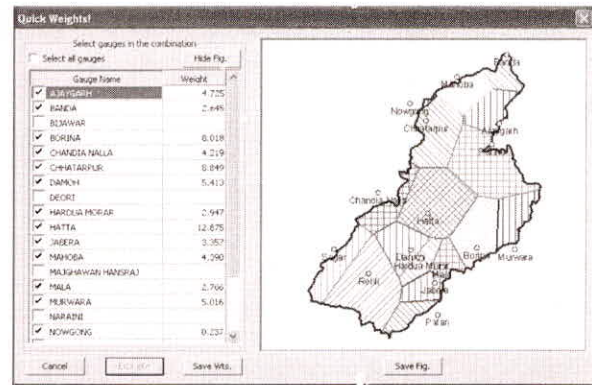


Fig. 12: Dialog showing estimated values of weights (in %age), optionally showing the Thiessen polygon figure

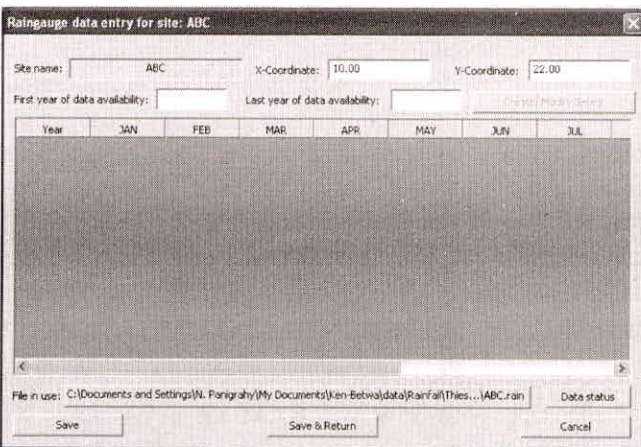


Fig. 10: The monthly rainfall for a gauge (ABC) is edited through this dialog

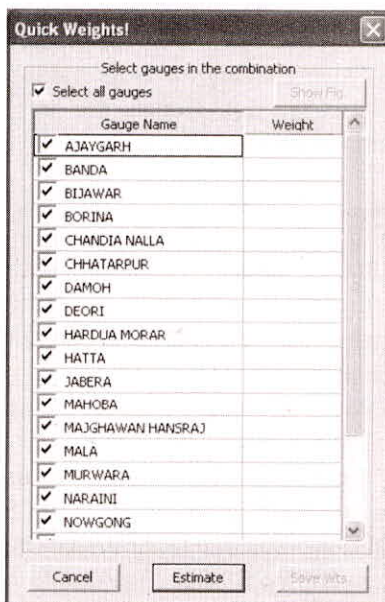


Fig. 11: Dialog to select stations for computing quick weights

The figure can be saved using [Save Figure] button which opens a normal Windows File Save As dialog. The figure can be saved in any of the commonly used file formats such as BMP, JPG., GIF, TIF, PNG, EMF, WMF selectable in the dialog. By default EMF is selected. The weights can also be saved to a text file clicking button [Save Wts.].

In the dialog for creating basin and rainauge description, option [View area] may be used to view the basin boundary along with the location of gauges. The generated figure can also be saved as above. In the same dialog clicking [Delete gauge] deletes the selected gauge name from the list. Where as using [Delete point] button, a number of points on the basin boundary can be deleted. Clicking [Save] in this dialog the input file is saved. [Return] button returns to the main (first) dialog.

After entering the data for the gauges, the average areal rainfall for the basin may be estimated for any period (within the data availability) clicking the button [Estimate] in the main (first) dialog.

If the output file is already available, a warning message is displayed. If the gauges combination is changed by addition or deletion of gauges after any previous estimations, this warning is to be ignored. The program then asks the range of years for which estimation is to be done.

While estimation is on, the progress is shown in various forms (Figure 13). The estimation can be stopped using [Stop!] button.

After the estimation is over or [No] is selected in the warning, three buttons [View output], [View wts.] and [View combn.] become active.

Clicking [View output], the weighted average rainfall for the basin can be viewed as shown in Figure 14. For easy navigation the Go to → Year, Month drop boxes can be used. Double clicking any cell opens the estimation details for that particular month (Figure 15).

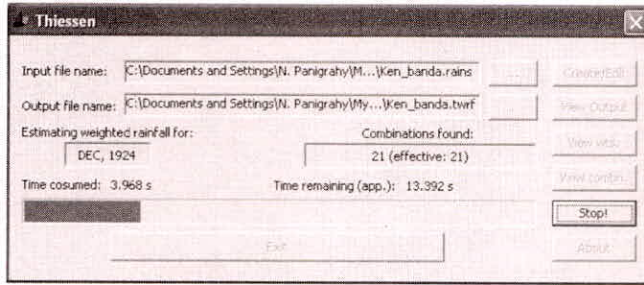


Fig. 13: Main dialog while estimating the weighted average rainfall for the range

YEAR	1901	1902	1903	1904	1905	1906	1907
JAN	55.445	11.380	2.939	2.096	6.465	0.936	6.130
FEB	18.342	10.415	1.315	13.327	2.997	19.759	82.399
MAR	11.784	0.069	0.098	54.719	7.333	13.442	1.930
APR	6.345	3.113	0.098	0.035	3.906	0.000	19.710
MAY	2.377	3.662	30.210	6.063	2.519	3.426	7.247
JUN	39.890	26.950	64.384	92.557	23.070	179.953	68.963
JUL	304.963	447.459	192.426	562.182	197.446	405.478	198.701
AUG	662.918	145.425	349.603	335.865	196.955	134.647	549.105
SEP	152.272	259.686	263.048	95.463	188.979	652.107	15.108
OCT	4.910	22.280	169.601	23.592	0.000	5.133	0.000
NOV	0.000	59.737	0.000	2.695	0.000	0.000	8.681
DEC	0.000	0.712	0.000	18.062	2.975	0.000	0.000

Fig. 14: Dialog showing the weighted average rainfall for the selected range of years

GAUGE	Rainfall	Weight (%)	Eff. rainfall
AYAYGARH	46.30	4.725	2.188
BANDA	67.60	2.645	1.728
BJAWAR	40.10	6.468	2.594
BORINA	---	---	---
HANDIA NALLA	---	---	---
CHHATARPUR	65.60	5.665	3.716
DAMOH	43.00	8.684	3.620
DEORI	---	---	---
ARDUA MORAI	---	---	---
HATTA	72.10	15.330	11.053
JABERA	28.00	9.335	2.614
MAHOBA	56.10	4.390	2.463
JHAWAN HANE	---	---	---

Fig. 15: Dialog showing the actual calculation performed along with weights and rainfall at each gauge for the month

In the dialog (Figure 15), using [Show Figure] button, the Thiessen polygon map for the month can be seen (Figure 16). The map can be saved as discussed earlier. The entire calculation can be saved to a text file clicking button [Save data].

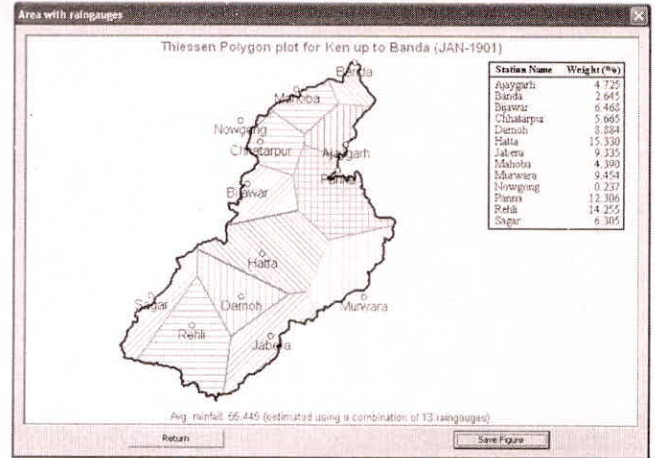


Fig. 16: Dialog showing the figure of Thiessen polygons for the estimation of average areal rainfall

While viewing the average rainfall (Figure 14), the data can be saved to a text or excel file by clicking on the button [Save to text/Excel file]. All the combinations found during estimation of average areal rainfall of the basin can also be seen on a tabular form or graphical form using buttons [View wts.] and [View combn.] respectively in the main dialog. It takes long time to create tabular data if the number of combinations is large.

MINIMUM REQUIREMENTS AND LIMITATIONS

The software has been developed and tested on a Pentium IV machine with 256 MB RAM on a Windows XP operating system. There are no specific hardware requirements for the software. Since it uses GDI+ functions, the Dynamic Link Library (DLL) for GDI+ must be available to run the software on other Windows (9x or 2000) OS. Although the main subroutine (ThiessenWeight) is portable, other GUI related functions are not. Hence, the software cannot be used for non-Windows OS.

CASE STUDY

Ken River basin located in the central India was selected as the study area. The basin lies between 23°12' N and 25°54' N latitudes, and 78°30' E and 80°36' E longitudes (Figure 17) having a catchment area of 28058 sq. km. The raingaugage network comprises of 21 gauges of which four are located outside the basin. Monthly rainfall data for these gauges are available from 1901 to 2005. Some of the raingaugage stations became operational after 1901, a few became operational afterwards and a few of these have been discontinued in between. Hence, the length of data is not same for all the stations.



Fig. 17: Ken River basin with its rain gauge network

The software was used to estimate the Thiessen weighted monthly average areal rainfall of the study area for the period from 1901 to 2005. There were 179 combinations of stations in 105 years of data. The program took only 35 second to estimate the average areal rainfall for the basin which includes the time required for dynamic selection of network.

CONCLUSION

In this study a computer-based method for automatic estimation of Thiessen weights was developed. Using the principles of the method a GUI based Windows application software was developed. The software was used to estimate the average areal monthly rainfall for 105 years for a basin. The software performed well in estimating the weighted rainfall with minimum manual labour.

REFERENCES

- Balascio, C.C. (2001). "Multiquadric equations and optimal areal rainfall estimation." *J. Hydrol., Engg., Am. Soc. Civ. Eng.*, 6(6), 498–505.
- Bastin, G., Lorent, B., Duque, C. and Grevers, M. (1984). "Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations." *Water Resour Res.*, 20(4), 463–470.
- Bras, R.L. and Rodriguez-Iturbe, I. (1976a). "Rainfall network design for runoff prediction." *Water Resour Res.*, 12(6), 1197–1208.
- Bras, R.L. and Rodriguez-Iturbe, I. (1976b). Network design for the estimation of areal mean rainfall events." *Water Resour Res.*, 12(6), 1185–1195.
- Croley II, T.E. and Hartmann, H.C. (1985). "Resolving Thiessen polygons." *J. Hydrol.*, 76, 363–379.
- Diskin, M.H. (1969). "Thiessen coefficients by a Monte Carlo procedure." *J. Hydrol.*, 8, 323–335.
- Diskin, M.H. (1970). "On the computer evaluation of Thiessen weights." *J. Hydrol.*, 11, 69–78.
- Lane, L.J., Davis, D.R. and Nnaji, S. (1979). "Termination of hydrologic data collection (A case study)." *Water Resour Res.*, 15(6), 1851–1858.
- Seed, A.W. and Austin, G.L. (1990). "Sampling errors for rain gauge derived mean areal daily and monthly rainfall." *J. Hydrol.*, 118, 163–173.
- Shih, S.F. and Hamrick, R.L. (1975). "A modified Monte Carlo technique to compute Thiessen coefficients." *J. Hydrol.*, 27, 339–356.
- Tabios-III, G.Q. and Salas, J.D. (1985). "A comparative analysis of techniques for spatial interpolation of precipitation." *Water Resour. Bull.* 21(3), 365–380.
- Thiessen, A.H. (1911). "Precipitation averages for large areas." *Monthly Weather Review*, 39(7), 1082–1084.
- Thorpe, W.R., Rose, C.W. and Simpson, R.W. (1979). "Areal interpolation of rainfall with Double Fourier Series method." *J. Hydrol.*, 42, 171–177.