Journal of Hydrology **11** (1970) 69-78; © *North-Holland Publishing Co., Amsterdam*

89

ON THE COMPUTER EVALUATION OF THIESSEN WEIGHTS*

M. H. DISKIN**

Abstract: An improved method is presented for evaluating the Thiessen weights or coefficients by a digital computer. The new method is based on covering the area of the watershed with uniform, equally spaced test points on a rectangular grid and assigning each such point to the nearest rain-gaging station. The number of points assigned to any station relative to the total number of points that fall inside the boundaries of the watershed gives the weight of the station. The new procedure is faster and gives more accurate results than a previous program based on random test points.

Introduction

In a recent article[1]) a method was proposed for computing the Thiessen coefficients or weights by a digital computer. The method was based on a Monte Carlo procedure for evaluating areas of known boundaries. A detailed description of the method is given in the article mentioned above, as well as an example of its use.

Practical experience with the proposed procedure indicated that the procedure is slow. The time taken to complete a set of computations is relatively long, and the change in weights from one set of computations to the next tends to become rather small after the initial larger differences in the first two or three sets. It follows that if accuracies of better than about 0.5% are needed, the computer time used to obtain the Thiessen coefficient becomes excessive.

An analysis of the causes for the above disadvantages of the existing procedure led to the development of a new, more efficient procedure for computation of the Thiessen weights. The new program is based on the systematic and uniform coverage of the rectangle that circumscribes the

* Contribution of the Southwest Watershed Research Center, USDA, Agricultural Research Service, Soil and Water Conservation Research Division, Tucson, Arizona.

** Research Hydraulic Engineer, Southwest Watershed Research Center, USDA, Agricultural Research Service, SWC, 442 East Seventh Street, Tucson, Arizona 85705. The author is a visiting scientist on leave from Technion-Israel Institute of Technology, Haifa, Israel.

watershed boundaries instead of the random process used in the old program. The purpose of this note is to discuss the reasoning that led to the new procedure, describe the new computer program developed, and present some comparative results obtained with the old and the new programs. The new computer program is written in Fortran IV, and the old program was also translated into the same language for comparison.

The new procedure

Figure 1 is a schematic drawing used to define the terms used in the paper. It shows the boundaries of a watershed, the location of the rain-gaging stations, and a rectangle that encloses, or circumscribes, the watershed. With reference to this drawing, the original procedure consisted of the following three steps:

(a) Selection of a random point (XT, YT) inside the rectangle.

(b) Test if the point is inside the watershed boundaries by first computing

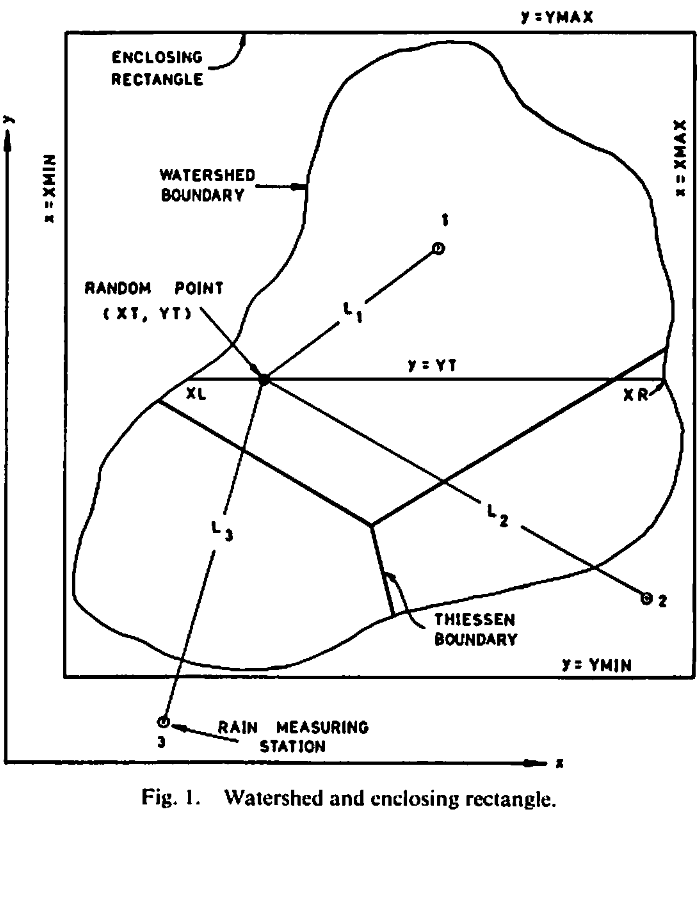

Fig. 1. Watershed and enclosing rectangle.

the points of intersection (XL, YT) (XR, YT) of the line y=YT with the boundaries and then by checking if XL≤XT<XR.

(c) Assignment of the point, if it is inside the watershed, to the nearest rain-gaging station.

The computation of the points of intersection in the first part of Step (b) involves scanning successive pairs of points for all the points that define the boundaries of the watershed. If the line y=YT intersects the boundary in more than one pair of points, there is a subsequent sorting of the points of intersection in an increasing order. The various computations and comparisons needed to accomplish this first part of Step (b) are time consuming, and in reviewing the procedure it was obvious that considerable computer time could be saved if this part could be carried out not for every point generated but only once for every group of such points. This could be obtained, for example, by grouping the random points into sets of points having the same value of YT or points falling within a narrow strip centered about the line y=YT. The scheme of computation would then be to generate a value of YT, compute the corresponding values of XL and XR, and then generate a string of values of XT for the number of points belonging to the given set.

While the above method of grouping the test points reduces the time needed to complete one set of computations, it does not affect the slow convergence of the results. This was traced to the vagaries of the random number generator used to produce the coordinates (XT, YT) of the random point. These coordinates were generated with uniform probability along the axes of the circumscribing rectangle, but the number of points generated in each set of computations was apparently not large enough to ensure the uniformity needed. Consequently, the weights obtained at the end of each set of computations differed by more than the specified amount from those produced by the previous set, and convergence was slow.

At this stage, it was decided to abandon the random generation of test points and substitute a uniform grid of points generated in a systematic way to cover the entire area of the watershed. In addition to uniformity, this method also gives some additional saving in time since it is unnecessary to call in the random number generating procedure. Also, the number of points falling outside the boundaries of the watershed was greatly reduced.

The flow chart of the new procedure is very similar to that of the old procedure (Fig. 2 of Ref. 1) except that the computations and comparisons are made with reference to a point that moves systematically over the area of the rectangle instead of a random point used previously. This is achieved by defining an incremental length (DLL) equal to a known fraction (usually

$\frac{1}{50}$ to $\frac{1}{200}$) of the range of x values (RNGX) or of y values (RNGY), whichever is smaller.

$$\text{RNGX} = \text{XMAX} - \text{XMIN}$$
$$\text{RNGY} = \text{YMAX} - \text{YMIN}$$
$$\text{DLL} = \min(\text{RNGX/NDIV}, \text{RNGY/NDIV})$$

where NDIV is the number of subdivisions or of increments specified for completely covering the shorter side of the enclosing rectangle. Values of YT are then chosen at the middle of the strips of width DLL.

$$\text{YT} = \text{DLL} * \text{K} - 0.5 * \text{DLL}$$

where K is the sequential number of the strip. Values of K are chosen to cover uniformly the range of values of the variable y

$$(\text{YMIN/DLL}) \leq \text{K} \leq (\text{YMAX/DLL}).$$

For each value of YT thus generated, a test is run to determine the points of intersection of the line y = YT with the boundaries. If at least one pair of points of intersection is obtained, a string of values of XT are chosen, at the center of square elements of side DLL, by the following equation:

$$\text{XT} = \text{DLL} * \text{JX} - 0.5 * \text{DLL}$$

where JX is the sequential number of the square in the strip. Values of JX are chosen to cover the interval

$$\text{XL} \leq \text{XT} \leq \text{XR}$$

defined by successive pairs of points of intersection for each line y = YT. The number of points generated and included in each computation set is thus less than the total number required to cover the rectangle

$$\text{NPTS} < (\text{RNGY/DLL}) * (\text{RNGX/DLL}).$$

The number of lines for which the points of intersection with the boundaries are determined is (RNGY/DLL), which is of the order of the square root of the total number of points (NPTS) included in the set

$$(\text{RNGY/DLL}) \approx (\text{NPTS})^{\frac{1}{2}}.$$

Some computer time can be saved by choosing the coordinate system in such a way that

$$\text{RNGY} < \text{RNGX}.$$

Except for the above changes, the new program is similar to the previous program. Experience with the new program indicated that for most purposes only one set of computations is needed. It appears that because of the uniform coverage of points, the errors in the weights, compared to the weights obtained with a very large density of points, are quite small. However, provisions were made in the new program for carrying out a number of sets of computations in which the number of strips or subdivisions (NDIV) is increased by a factor larger than unity (FCT), and the number of points is increased by a factor of $(\mathrm{FCT})^2$, from one set of computations to the next set. The computations will terminate when the change in the values of the weights from one set to the previous set is less than a specified value (LIMW), and the change in the value of the area of the watershed relative to that of the enclosing rectangle is less than a specified value (LIMA). If the value specified for the factor (FCT) is less than 1.05 or if no value is specified, the program will produce only one set of computations. In this case it is not necessary to specify values of LIMW and LIMA.

A new feature of the revised program is that it computes the coordinates of the centroid of the area represented by each rain gage. This is done simply by adding the x and y coordinates of all the points assigned to a particular rain gage and dividing the sum by the number of points assigned to that station.

$$\mathrm{CX} = \Sigma \mathrm{XT}/\mathrm{NS}$$
$$\mathrm{CY} = \Sigma \mathrm{YT}/\mathrm{NS}.$$

The program also computes the distance from the centroid to the location of the rain gage.

The complete listing of the new computer program in Fortran IV could be obtained from the author or from the Southwest Watershed Research Center, Tucson, Arizona.

Applications

The new program was tried on a number of watersheds for which values of the Thiessen weights were derived both by the conventional-graphical method and by the old computer procedure. The results obtained were comparable, and the time of computation for the new program was smaller by a factor of between 5 and 10 in comparison to the computer time used by the old program. An example of the results of one such comparison, for one of the subwatersheds of the Walnut Gulch watershed described below, is given in Table 1. The area of the watershed was 3.18 square miles, and the available map of the watershed was at a scale of 1:24000. Three independent Thiessen polygons were drawn for the manual-graphical determination of weights, and the areas were measured by a planimeter. For the computer evaluation the boundaries were defined by 31 boundary points, and their coordinates were punched directly on cards by means of a chart reader. The number of points specified per set (NSET) for the old procedure was 2000. The number of divisions (NDIV) specified in the new procedure

TABLE 1

Thiessen weights in percent by various methods

Rain gage station No.	Manual-graphical method				Old computer program				New computer program	
	A	B	C	Mean	1 set	5 sets	10 sets	17 sets	100 Div.	200 Div.
1	15.0	14.4	14.5	14.6	13.1	14.4	14.3	14.3	14.8	14.8
2	14.0	14.7	14.1	14.3	15.5	14.6	14.5	14.3	14.3	14.3
3	6.9	6.9	6.9	6.9	6.0	6.3	6.5	6.6	6.8	6.8
4	35.0	34.5	34.1	34.5	34.7	36.1	36.0	35.6	34.9	34.9
5	11.5	11.8	13.0	12.1	12.5	11.5	11.4	11.6	11.5	11.6
6	17.6	17.7	17.4	17.6	18.1	17.0	17.2	17.5	17.7	17.6

was 100 and 200. The central processing unit time taken for the old program was 6.7 sec per set of computations, but some 5 to 7 sets of computations were required for an accuracy of 0.5%. The corresponding time for the new program was 5.3 sec for a set based on 100 divisions and 15.8 sec for a set using 200 divisions. The difference in weights for the last two sets was not more than 0.1%.

Another example based on a watershed with somewhat more complicated boundaries and a larger number of rain gages is given below. The comparisons in this case are only for results obtained with the old and new procedures, without the manual-graphical procedure included in the previous example. The computations of weights are for a network of rain-gaging stations in the Walnut Gulch watershed, which is an experimental watershed in southeastern Arizona operated by the Southwest Watershed Research Center, Soil and Water Conservation Research Division, Agricultural Research Service, USDA[2]). The watershed is 57.7 square miles in area and is equipped with some 90 recording rain gages distributed over the watershed and its immediate neighborhood. Eighteen of the rain gages were selected to comprise the network for which the Thiessen weights were computed. The boundaries of the watershed were represented by straight line segments specified by the locations of 35 boundary points. Figure 2 shows the simplified boundaries of the watershed, the location of the 18 rain-gaging stations and their identification numbers.

The Fortran IV versions of the old program and the new program were run on the CDC* 6400 computer at the University of Arizona. The com-

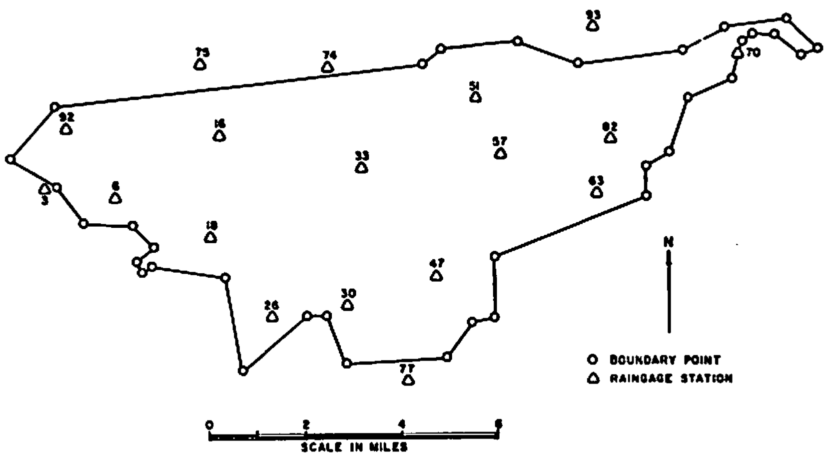

Fig. 2. Walnut Gulch Experimental Watershed.

* Use of a trade name does not imply endorsement by the U.S. Government but is used for the reader's benefit.

piling time for both programs was about the same, taking approximately 1.3 sec of the central processing unit and 9.6 sec of the peripheral equipment. Another 9.6 sec of peripheral equipment time was usually taken in printing out results. The time taken per set of computations of the old program was about 4.1 sec of the central processing unit when the set comprised 1 000 random points. The time was practically doubled when the number of points in the set was increased to 2 000. The maximum and mean errors in the weights obtained for some sets of computations, relative to the true values of the weights, are given in Table 2.

Using the same data, the new program was run with the number of subdivisions of the shorter side of the rectangle between 20 and 300. The number of points included in each set, the central processing unit time, and the maximum and mean errors obtained for each set of computations are given in Table 2. The errors given in the table were computed with reference to the results obtained with the set of computations containing the largest density of points, which were assumed to represent the true values. The same values were also used as a basis for comparison of the results obtained with the old procedure as given above.

Conclusions

The use of a uniform grid of test points in the computation of the Thiessen weight instead of the random points used previously has resulted in a reliable and more efficient computer program for the evaluation of the weights. The time taken for one set of computations is somewhat larger in the new program, but because of the uniformity of coverage only one set of computations is needed in comparison to 5 to 7 sets of computations needed in the previous program. The net result is a saving in computer time so that the machine time used by the new program is of the order of 20% of the time needed in the old program for a comparable degree of accuracy.

The number of points included in a set of computations can be specified to produce any accuracy needed. This is done by specifying the number of divisions or increments to be used on the shorter side of the circumscribing rectangle. Provision is made in the program for carrying out multiple sets of computations with increasing numbers of points per set of computations until the change in weights between one set and the previous set is less than a specified amount.

The number of subdivisions of the shorter side of the rectangle recommended for general work is 100, but good results may be obtained with 50 subdivisions. The time consumed by the central processing unit of the computer is roughly proportional to the square of the number of subdivisions specified for the set of computations. Some computer time can be saved if the

TABLE 2

Comparison of new and old computer programs

Old computer program					New computer program				
Number points per set	Number of sets	CPU time in seconds	Error in weights Mean	Error in weights Max.	Number of subdivisions	Number of points	CPU time in seconds	Error in weights Mean	Error in weights Max.
1000	1	4.0	0.79	2.37	20	530	0.8	0.28	0.67
	5	20.2	0.32	0.68	30	1100	1.4	0.09	0.21
	10	40.5	0.22	0.65	40	1800	2.4	0.07	0.18
	15	60.8	0.19	0.54	60	3900	4.6	0.03	0.07
	20	81.1	0.18	0.48	90	8500	10.0	0.02	0.05
2000	1	8.0	0.59	1.91	100	10500	10.6	0.02	0.05
	4	32.2	0.42	0.87	150	23100	23.6	0.015	0.04
	8	64.4	0.27	0.70	200	40700	41.2	0.005	0.02
	12	96.6	0.21	0.70	300	90100	90.2	-	-
	15	120.6	0.20	0.55					

Note: Errors (in percent) are computed as difference between weights (in percent) and those obtained with 300 subdivisions.

coordinate system is chosen so that the y-axis is along the short side of the circumscribing rectangle.

References

1) M. H. Diskin, Thiessen coefficients by a Monte Carlo procedure. J. Hydrol. 8 (1969) 323–335.
2) K. G. Renard, The hydrology of semiarid rangeland watersheds. U.S. Department of Agriculture Publication ARS-41 162 (1970)