COMPUTER RANDOM MODEL GENERATION FOR PREHISTORIC SETTLEMENT STUDIES

Dee F. Green and
Evan I. DeBloois

United States Forest
Service.

Introduction

Under the assumption that human behaviour is patterned rather than random, archaeologists seek to explain the patterning. Some archaeologists are satisfied with historical explanations, others look for explanation through ecological relationships and still others seek those patterns thought to represent cultural law. Regardless of the theoretical framework into which the archaeologist may cast his research, he is always faced with the problem of determining whether the data he is using are the results of patterned human behaviour or are results of some random activity. In many cases this determination is self evident. Pottery designs are obviously the result of human behaviour while artifacts recovered from an eroded context are random vis-a-vis their human deposited context.

In many cases it is not readily apparent whether a given body of archaeological data has a random or a non-random (patterned) distribution. With quantified data bodies, inductive statistical techniques can often be used to help solve this problem. Another technique available to the archaeologist is random model generation. One form of this latter technique based on a suggestion by Plog (1971) is discussed in this paper.

A random model offers several advantages to the archaeologist. In the first place, since he assumes that human behaviour is patterned any departure of the real data from the random model can be viewed as patterned behaviour. On the other hand data which are identical or very similar to the random model are probably random. Another and equally important advantage is that the random model is free from cultural bias. This quality enhances comparability by giving the archaeologist a known standard for comparison. Archaeologists often set up single sites or time periods as models for comparison. All such models contain built in cultural bias. Furthermore, the bias is usually uncontrollable since the parameters of population, sample, trajectory and other variables are unknown. Because a random model is free of cultural bias it can be used as a known standard against which two or more data bodies can be compared. Differences between data bodies can then be specified by measuring their distances from the random model.

Problem and Method

Archaeologists assume that site locations are not random, but occur as adaptations to social and ecological phenomena. A variety of variables may affect site location. In the arid southwestern United States water is thought to be a variable of prime importance in prehistoric site location. Thus real site locations should, on the average, be closer to water than a randomly generated settlement pattern. Data for testing the above notion was available to the writers based on an intensive survey of part of the Monticello Ranger District, Southeastern Utah. The survey included portions of two stream courses (Allen and South Cottonwood Canyons) within the National Forest boundary (Green 1971). Prehistoric sites in both canyons are primarily small Pueblo I and II agricultural villages along with some limited activity areas. Water would have been a necessary resource for both crops and culinary activity and sites should be located to take advantage of this resource.

The above can be cast in a mini-max, optimization, or "least-cost" model as proposed by Hill (1971). He formulates a general proposition which states, "Sites are located so as to minimize the amount of pursuit time in obtaining critical resources" (Hill 1971:58). In this paper we are considering only a single critical resource: water.

Our hypothesis states: Water is the variable of major importance in determining prehistoric village site location in the southwestern United States, i.e., the distance from the village to the nearest water will be minimized.

This hypothesis was set up with the expectation that it would eventually be disproven although not necessarily by this study. It was felt that sites in both Allen and South Cottonwood canyons would be located closer to the stream course than would random distributions. If this proved to be true then the hypothesis would neither be confirmed nor disproved since it specifies the entire southwest, and obviously other tests would have to be made. However, if either canyon showed a random distribution or locations farther from the stream courses than the random model then the hypothesis would be disproved since one case is sufficient to do so. It would not then be possible to generalize that village sites in the southwest were located primarily on the basis of water resources. Thus despite the general aridity water would not always constitute the prime resource determining site location.

The following procedures were used to test the hypothesis.

1. Site location data from the 1971 survey were available on aerial photographs. These locations were transferred to a USGS quadrangle map (scale 1:24000) using a Forest Service Kelsh photogrammetric plotter.

2. Two areas of equal size and about equal site density were then selected, one each for Allen and South Cottonwood canyons. The plot size was 3.05 square kilometers. Sixty-nine sites were located in the Allen Canyon plot and 71 in South Cottonwood.

3. For ease in working, the plot scale was increased in size and site locations, stream course, and contour intervals were transferred to clear polyester drafting film.

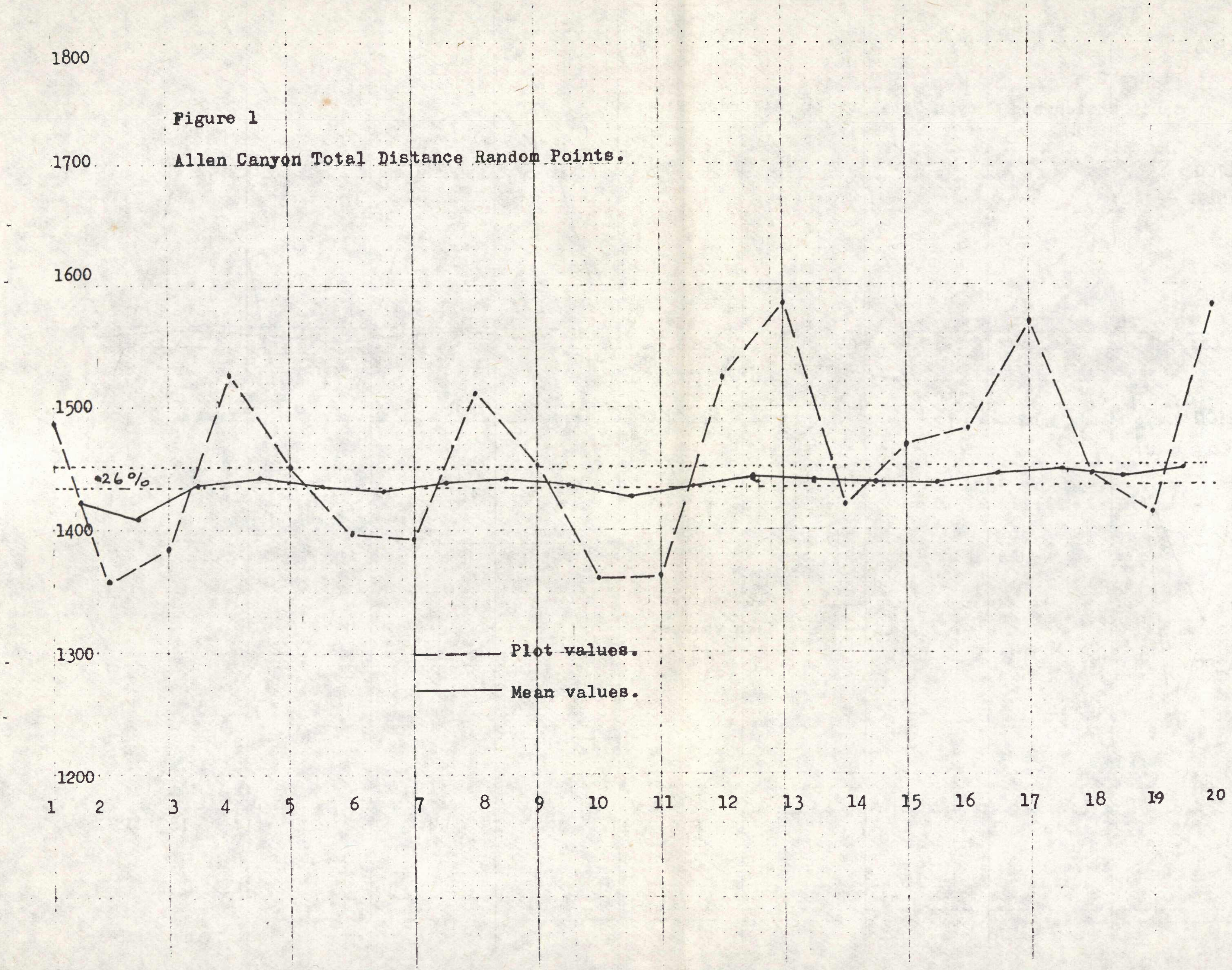


Figure 1
Allen Canyon Total Distance Random Points.
1800
1700
1600
1500
1400
1300
1200
26%
Plot values.
Mean values.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

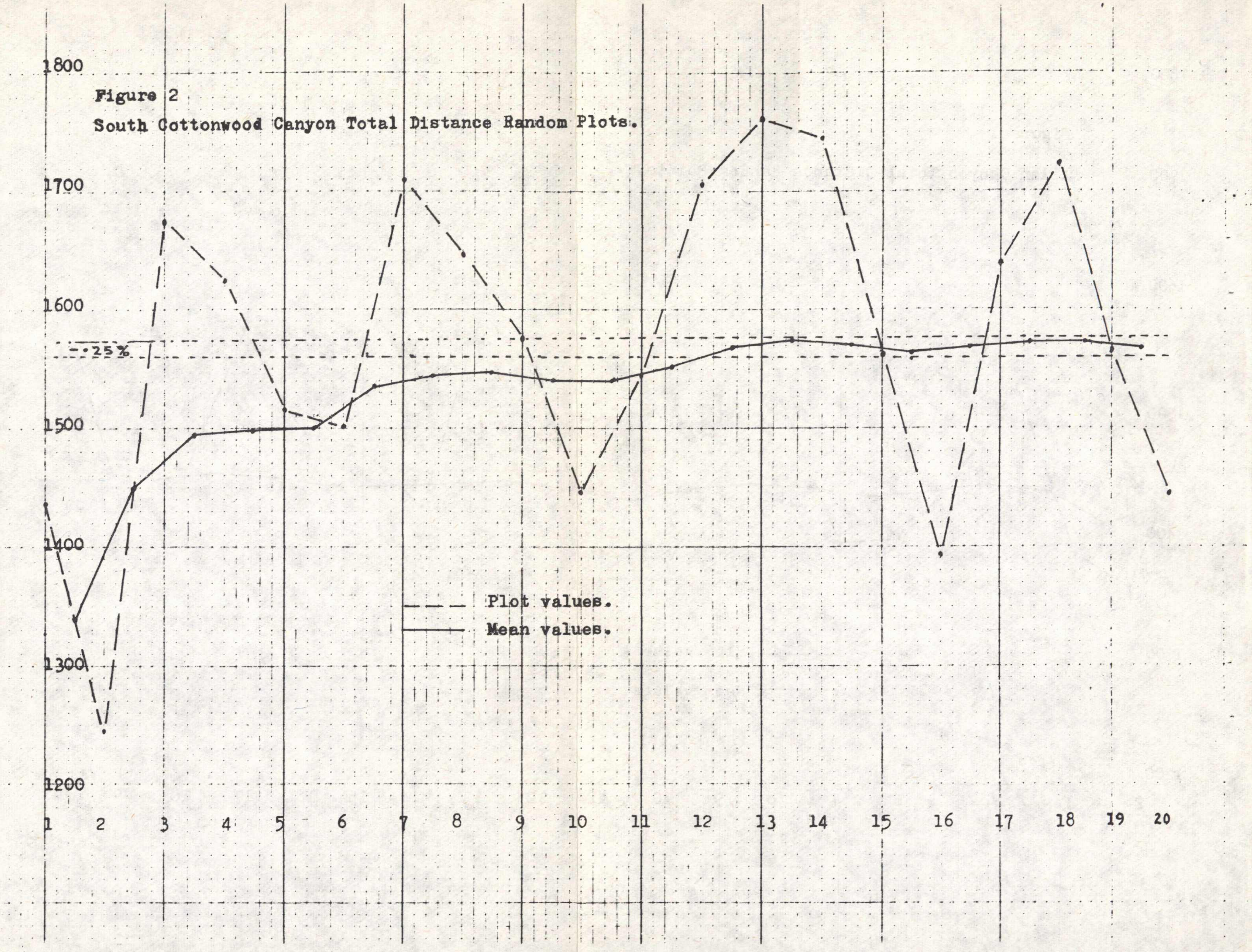


Figure 2
South Cottonwood Canyon Total Distance Random Plots.
1800
1700
1600
1500
1400
1300
1200
-.25%
Plot values.
Mean values.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

4. A WANG 720B programable calculator and X-Y plotter generate and plot random site distributions at the same scale as the real distributions in 3 above. Each of the 20 different plots contained 71 random site locations.

5. The materials from 3 and 4 above were then used to make the following sets of measurements for both streams courses.

(a) Real site horizontal distance to stream.
(b) Random site horizontal distance to stream.
(c) Real site vertical distance to stream.
(d) Random site vertical distance to stream.

Horizontal distance was measured by direct line in millimeters. Vertical distance was arrived at by counting the number of contour intervals between the site and the stream course and assigning a value.

6. Results were compared by using a modification of Plog's percentage of variability technique (Plog 1971, 53-54). The distance from all randomly located sites to the nearest point on the stream course is measured and summed. Then the distance from all real site locations to the nearest point on the stream course is measured and summed. These two values are subtracted and that figure divided by the random sum which gives a percentage point difference (Table 1).

Results

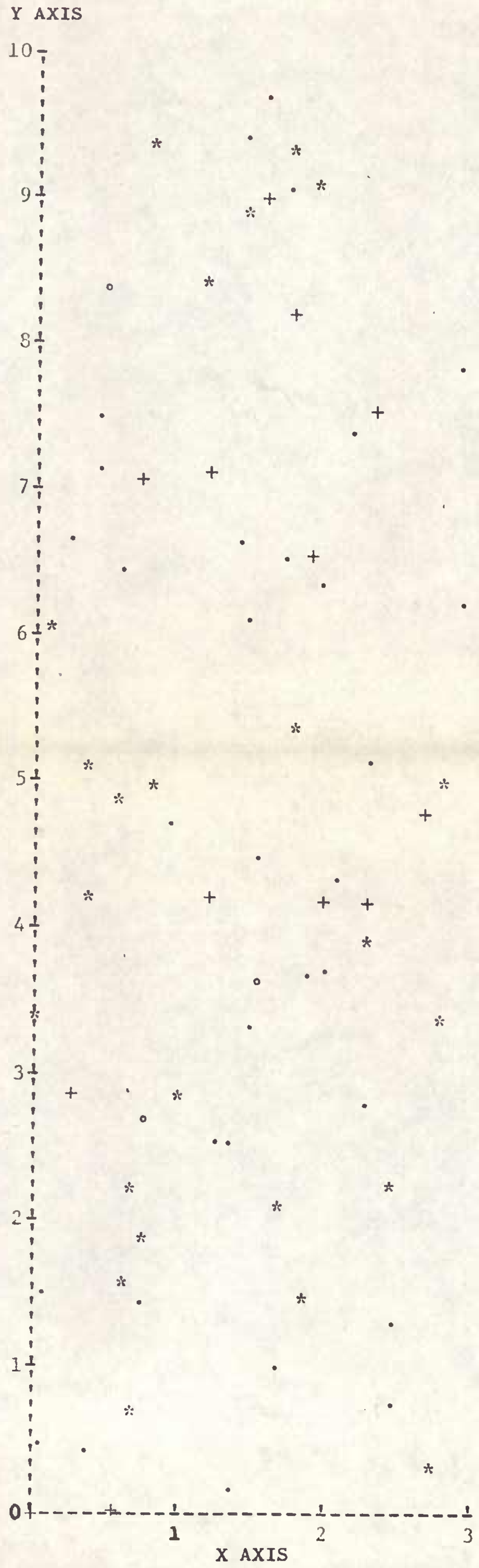
Since our purpose is to demonstrate the use of a random model as well as test the proposed hypothesis we will discuss some of the results of using this form of random modeling prior to a consideration of the test results themselves.

The generation of any single random model is a sufficient device against which data may be tested. However, by repeated generation a mean of random models can be arrived at which is more accurate. The number of repetitions necessary to achieve a mean which has little change as additional repetitions are added will vary (DeBloois N.D.). In this study 20 repetitions were used. That is, 20 random model site distributions of 71 sites each were generated, their distances measured to the stream courses and their means computed. Figures 1 and 2 show the above results for both stream courses. It is noted that in the case of Allen Canyon (Fig. 1) the mean of the total random distances does not fall outside the final percentage of variability after repetition 12. That is, after 12 repetitions the answer does not change so that in this case 12 repetitions would have been sufficient. In the case of South Cottonwood (Fig. 2) the mean of the total random distances still shows some variation through repetition 20. However, the curve has begun to straighten out and the final value after 20 repetitions is probably accurate. The same 20 random plots were used on both stream courses and all random plots for any one stream course totalled above or below the real site totals.

TABLE 1

Variable	Distance	Repetitions	N	%
		Allen Canyon		
Random Horizontal	1275	20	69	
Real Horizontal	942		69	.26
Random Vertical	175	20	69	
Real Vertical	120		69	.31
Random Total Distance	1450	20	69	
Real Total Distance	1062		69	.26
		South Cottonwood Canyon		
Random Horizontal	1375	20	71	
Real Horizontal	1714		71	-.25
Random Vertical	197	20	71	
Real Vertical	257		71	-.30
Random Total Distance	1572	20	71	
Real Total Distance	1971		71	-.25

Figure 3



We were also interested in whether or not horizontal or vertical distance was a better and/or sufficient measurement. These figures were, therefore, totalled separately as well as together and percentage variablity calculated for each case. (Table 1.).

It is evident from Table 1 that the horizontal distance is highly accurate and sufficient in this case. In areas with much greater vertical relief this would probably not be true.

Computer use in this study was confined to the generation and plotting of the 20 random site distributions. Had the software been available we could also have used a computer to take the measurements and, of course, do the calculations. Figure 3 gives one example of a computer generated random plot similar to those used in this study. The points (sites) are located within the defined space using a random number generator routine that forms random numbers from the lower order position of a number generated by the log function. Note that different symbols may be used at the located points. This enables the user to test a number of different data sets without generating a new group of random plots every time the number of sites changes. Size of the plots must remain constant, however, since points are plotted randomly within the defined space. Any test which uses a different scale would require a new set of plots at that scale. The WANG program was written to allow such changes with ease. Restrictions involve the physical limits of the paper size at the upper end and symbol overlaps beyond readability at the lower end.

While we have only used the random plot with settlement type data it is obvious that other applications in archaeology are possible especially where relationships in space are being considered, for example, artifact distributions within a site, room or room complex.

Turning to the test of the hypothesis we find from Table 1 that in Allen Canyon, on the average, sites are located closer to the stream than are those of the random model. However, in South Cottonwood canyon the sites are located farther from the stream than would be expected with a random distribution. Although the South Cottonwood test does not confirm the hypothesis as stated it does provide information about the relationship of sites to water resources and demonstrates the utility of random modeling.

Generalizing for the southwestern United States, water is <u>not</u> the variable of major importance in determining site location although it may be for any particular time period, site, or group of sites. Other variables obviously play a role and it is probably the case that one variable, seldom if ever determines site location. In the case of Allen and South Cottonwood canyons we can suggest that soils may have been a critical variable. South Cottonwood Canyon has a deep sandy alluvial bottom a few meters wide. Allen Canyon, on the other hand, has no alluvium and flows either on sandstone bedrock or on a bed of dioritic porphyry cobbles washed down from the Abajo Mountain laccoliths. In South Cottonwood Canyon corn fields could be grown on the alluvial bottom close to the stream bed with living sites located farther away off the alluvial soils. In Allen Canyon, however, farm plots would need to be placed on the same colluvial soils as the living areas and there may have been no advantage in placing the living areas farther from the stream. Of course, the above needs testing.

Allen Canyon is an adjacent tributary of South Cottonwood Canyon and the sites in both canyons are identified as participants in the Anasazi cultural tradition. Except for the geological differences noted above there is no obvious reason to expect a difference in settlement pattern vis-a-vis water resources between the two canyons. Through the use of a random modeling technique free from cultural bias it has been possible to demonstrate that water is not always the major variable in determining site location among the prehistoric Anasazi even for agricultural communities who need water resources in proximity.

REFERENCES

DeBloois, Evan I. (in press) The Elk Ridge Archaeological Project: A test of Random Sampling in Archaeological Surveying. Ph.D. thesis, University of Washington.

Green, Dee F. 1971 The Elk Ridge Archaeological Project Manti-LaSal National Forest Summary of the 1971 Field Season. United States Department of Agriculture, Forest Service. Intermountain Region, Ogden, Utah.

Hill, James N. 1971 Research Propositions for consideration Southwestern Anthropological Research Group. In, The Distribution of Prehistoric Population Aggregates, edited by George J. Gumerman. Proceedings of the Southwestern Anthropological Research Group. Prescott College Anthropological Reports No. 1.

Plog, Fred 1971 Some operational Consideration. In, The Distribution of Prehistoric Population Aggregates, edited by George J. Gumerman. Proceedings of the Southwestern Anthropological Research Group. Prescott College Anthropological Reports No. 1.