TN 2: ATTRACTIVITY INDICES

BY J.H.C. ROSS

ABSTRACT

In this study, a methodology to estimate the attractivity of parks to a given type of user from data concerning the spatial interactions of individuals who patronize these sites is presented. The model proposed is utilized to produce an ordinal scale defining the relative attractivity of the alternative sites in the system being studied.

The 1969 CORD Study Park User Survey data were used for this analysis. Actually, only information collected at 12 Saskatchewan parks on 3,254 individual trips were used out of a larger data set. Documentation of the data collection and processing is not provided but other sources are cited where this information is available.

Using the data on origin destination flows just referred to, an attractivity ranking of the twelve parks was established. The consistency of the decisions which led to this ranking was found to be significantly non-random, although there is a degree of confusion evident in how different people appeared to rate the parks. Reasons for this confusion are introduced that suggest that much of the confusion is attributable to the nature of the data, and inhomogeneity of user purpose rather than being a result of problems with the methodology.

A postscript contains comments on what kind of research to which the original research has led.

INTRODUCTION

With all the difficulties involved, it seems entirety possible to develop specific, and rather objective, rating scales for different outdoor recreation areas and for major different uses of each. These scales would have great utility in planning, other research, and administration. The talents and knowledge of different kinds of specialists might well be used in devising and testing such rating scales. (Clawson and Knetsch 1963).

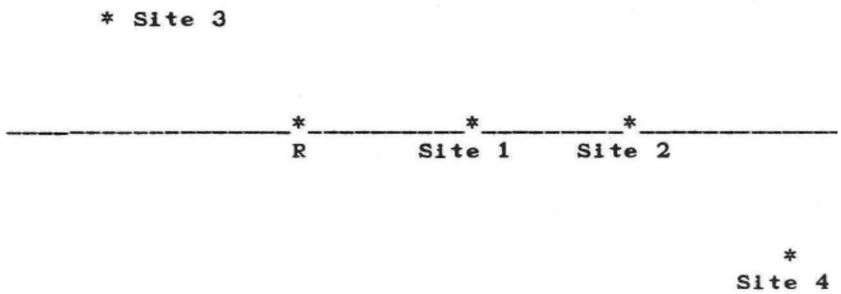
A general dissatisfaction with the more traditional methods of deriving such scales has led the author to develop an approach which avoids the strong metric assumptions necessary for the derivation of site attraction scales by the gravity model approach (Rogers 1966) yet, at the same time, avoids arbitrarily assigning attraction values solely on the basis of the number and capacity of site facilities. (See References in TN 9 and 16.)

METHODOLOGY

The model of human behavior proposed is concerned with attractiveness, A(j), which is a measure of an inherent property of site J, and distance D(r,j), a measure of the difficulty of travelling from the residence of visitor r to site j. It is assumed that the effects of increasing distance are such that the "degree of impedance" always increases as distance increases — although the function relating impedance and distance need not be defined (see TN 14.) This assumption really means that for any distance D(r,j), travelling further is seen as being more costly and less desirable unless something about the destination reached defrays (compensates for) the extra travel. Wolfe (1972; see also TN 14 and Beaman 1974) implies that this assumption is not always valid, but it is accepted here. Now, in one dimensional space, consider an individual residing at point r as shown in Figure 1. The assumptions stated imply that choosing to patronize a site at point 2 means that the site is more attractive than Site 1. If the individual had gone to Site

1 no similar statement could be made about the relative attraction of Site 1 compared to Site 2. For the visitor, distance rather than site attractivity may have been the "force" resulting in visit to 1 rather than some more attractive site.

FIGURE 1: TWO—DIMENSIONAL SPACE



The extension of this reasoning to two dimensional space involves the assumption that each individual perceives a given distance to be of the same magnitude regardless of the distance he must travel. If this assumption is made one may state that a person implicitly judges the site he selects to be more attractive than any alternative site which is closer to his origin than the selected site. One can make no such judgments with regard to the relative attraction of sites which are more distant than that which an individual selected. So, in Figure 1, if one accepts that the individual's residence is located at point r, while the points 1 - 4 represent alternative sites, a number of inferences can be made. For example, if the individual chooses to visit Site 4 it can be assumed that he judges that site to be more attractive than any other site closer to his residence.

THE COMPARISON MATRIX

Now, if for a number of parties we consider that one individual made the decision to visit a particular site, the number of times that any site I can be inferred to be more attractive than any other site j can be recorded to form a site-by-site comparison matrix C (Figure 2). In this matrix the i-jth entry is the number of times that site i was chosen over site j. In order to prepare such a matrix for further analysis, one must, taking each sample individual in turn, calculate the distance from his residence to each of the alternative sites. Then denoting the site visited as C(i,j) is incremented by one for every site j which is closer to the individual's residence than site i.

FIGURE 2: COMPARISON MATRIX: DISTANCE = 200 MILES

Site*	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	1	61	2	2	1	1	5	8	4	14
2	6	0	3	5	5	5	1	5	7	6	1	21
3	13	4	0	14	47	10	3	10	10	14	3	10
4	11	0	8	0	9	10	4	6	7	19	5	8
	7	0	9	14	0	2	1	9	1	7	1	6
6	5	0	5	5	8	0	2	4	1	8	2	4
7	42	3	4	10	7	13	0	1	51	44	40	43
8	20	1	2	20	4	1	1	0	2	19	0	18
q	56	2	5	53	50	7	1	43	0	55	0	55
10	203	3	0	227	1	1	1	2	8	0	1	9
11	6	6	1	4	2	5	5	0	146	99	0	5

12	191	1	4	30	6	7	3	7	4	33	4	0

* The names of sites 1 to 12 are given in Table 1.

THE PROPORTIONS MATRIX

From the comparison matrix C, a proportions matrix P may be determined. The entries in this matrix give the proportion of times that any site i was Judged to have a higher attraction than site J. An entry in the P matrix is defined as:

P(i,j) = C(i,j)(/C(i,j) C(j,i))

WHERE P(i,j)=the proportion of times site i was chosen over site j,
C(i,j)=the number of times site i was chosen over site j, and
C(j,i)=number of times site J was chosen over site i.

Because situations arise in which site i and site j were never compared, the fact that the comparison is missing must be noted. Throughout this paper, the value -1.00 is used to denote this.

In an ideal situation, when all individuals have similar perceptions as to which sites are most favorable, the elements of P, P(i,j) take only values of 1 and 0. In the real world, however, especially when the sites being compared are very similar in attraction, the value of P(i,j) will be intermediate between 1 and 0.

Psychophysicists, in attempting to derive measurement scales from similar paired comparison matrices, have developed a number of approaches. Many of these are based on the "Law of Comparative Judgment" developed by Thurstone (1972). If we were to use Thurstone's approach, we would have to assume that each entry in the table could be viewed as a normal deviation from the column mean. But, due to the fact that the individual subject's inferred decision that one site was more attractive than another was conditional on the spatial arrangement of the sites, one cannot regard the P(i,j) as normally distributed variables. Specifically, the large numbers of zeros and ones in the matrix present problems because the probability of getting either under Thurstone's assumptions is infinitely small.

As an alternative approach which avoids these problems, regard any P(i,j) greater than 1/2 as indicating a majority preferring i, and any value less than 1/2 as a majority preference for j. When this is done, one can construct a scale which is consistent with these majority judgments by finding the average value of each row of the P matrix. These values are found by summing the valid entries in each row and dividing by the number of valid entries in that row. The result is defined to be the index of attraction:

A(i) = e(i,j)P(i,j)/ Σ e(i,,j)

WHERE A(i) = the attraction index of site i
P(i,j) = the proportion of times site i was chosen over site j, and
e(i,j) = 1 if P(i,j) is not equal to -1.00, 0 otherwise. and the sum is on j.

Values of the index of attraction calculated using the matrix presented as Figure 3 are presented as Figure 4.

INDEX Q1 CONFUSION

A further measure, the index of confusion, may be calculated. It is defined as the proportion of valid entries in each row of the P matrix which indicate unanimous decisions. The preceding implies that it is calculated by dividing the number of entries in each row by the number of valid entries in the row, and subtracting the result from 1.0. Values of this index based on the P matrix presented as Figure 3 are shown in Figure 4.

CONSISTENCY

Consider three sites, j, and k. If A(i) is greater than A(j), and A(j) is greater than A(k), then A(i) must clearly be greater than A(k). Such a triplet is termed a transitive or non-circular triad. The weak transitivity of the P matrix is defined as the degree to which:

IF $P(i,j) > .50$ and $P(j,k) > .50$, then $P(i,k) > .50$ for all j, and k.

If the P matrix is incomplete the standard formulae for determining the consistency of the matrix cannot be applied directly. Instead, the n(n-1)(n-2)/6 possible triads must be examined separately. When the P matrix is incomplete, four different outcomes are possible when inspecting a triad. These are:

1. the triad may be non-circular,
2. the triad may be circular or intransitive,
3. the triad may be incomplete but of the form $P(i,j) > .5$ and $P(i,k) > .5$,

- it this is the case it must be transitive, and

4. the triads may be unknown because two elements are absent, (-1 occurs twice).

Kendall's Coefficient of Consistency (Kendal 1962) is defined as:

$$K = 1.0 - (NC/MC)$$

WHERE NC = the number of circular triads observed, and
MC = the maximum number of circular triads possible.

Reasoning that the maximum degree of inconsistency should be observed judgments between any sites are made at random, the following procedure has been adopted. Create a dummy P matrix's, by replacing all the valid entries below the diagonal with a uniformly distributed random number between (and including) 1.00 and 0.00. P(i,j)'s above the diagonal are refaced by the complement of the below-diagonal number. That is to say:

$$P'(i,j) = X, \; P'(j,i) = 1.00 - X$$

WHERE P'(i,j) = the i,jth entry in the dummy proportions matrix P', and
X = a uniformly distributed random number between 1.00 and 0.00.

FIGURE 3: PROPORTIONS MATRIX: DISTANCE 200 MILES

Site*	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0	0.0	0.07	0.85	0.22	0.29	0.02	0.05	0.08	0.04	0.40	0.07
2	1.00	0.00	0.43	1.00	1.00	1.00	0.25	0.83	0.78	0.67	0.14	0.95
3	0.93	0.57	0.00	0.64	0.84	0.67	0.43	0.83	0.67	1.00	0.75	0.71
4	0.15	0.0	0.36	0.00	0.39	0.67	0.29	0.23	0.12	0.08	0.56	0.21
5	0.78	0.0	0.16	0.61	0.00	0.20	0.13	0.69	0.02	0.88	0.33	0.50
6	0.71	0.0	0.33	0.33	0.80	0.00	0.13	0.80	0.36	0.89	0.29	0.36
7	0.98	0.75	0.57	0.71	0.88	0.87	0.00	0.50	0.98	0.98	0.89	0.93
8	0.95	0.17	0.17	0.77	0.31	0.20	0.50	0.00	0.04	0.90	-1.00	0.72
9	0.92	0.22	0.33	0.88	0.98	0.64	0.02	0.96	0.00	0.87	0.0	0.93
10	0.96	0.33	0.0	0.92	0.13	0.11	0.02	0.10	0.13	0.00	0.01	0.21
11	0.60	0.86	0.25	0.44	0.67	0.71	0.11	-1.00	1.00	0.99	0.00	0.56

12	0.93	0.05	0.29	0.79	0.50	0.64	0.07	0.28	0.07	0.79	0.44	0.00

*** -1.00 indicates a missing value. NOTE: The names of sites 1 to 12 are given in Table 1.**

FIGURE 4: REORDERED PROPORTIONS MATRIX

Site*	7	2	3	11	9	8	6	12	5	4	10	1
7	0.00	0.75	0.57	0.89	0.98	0.50	0.87	0.93	0.88	0.71	0.98	0.98
2	0.25	0.00	0.43	0.14	0.78	0.83	1.00	0.95	1.00	1.00	0.67	1.00
3	0.43	0.57	0.00	0.75	0.67	0.83	0.67	0.71	0.84	0.64	1.00	0.93
11	0.11	0.86	0.25	0.00	1.00	-1.00	0.71	0.56	0.67	0.44	0.99	0.60
9	0.02	0.22	0.33	0.0	0.00	0.96	0.64	0.93	0.98	0.88	0.87	0.92
8	0.50	0.i7	0.17	-1.00	0.04	0.00	0.20	0.72	0.31	0.77	0.90	0.95
6	0.13	0.0	0.33	0.29	0.36	0.80	0.00	0.36	0.80	0.33	0.89	0.71
12	0.07	0.05	0.29	0.44	0.07	0.28	0.64	0.00	0.50	0.79	0.79	0.93
5	0.13	0.0	0.16	0.33	0.02	0.69	0.20	0.50	0.00	0.61	0.88	0.78
4	0.29	0.0	0.36	0.56	0.12	0.23	0.67	0.21	0.39	0.00	0.08	0.15
10	0.02	0.33	0.0	0.01	0.13	0.10	0.11	0.21	0.13	0.92	0.00	0.96
1	0.02	0.0	0.07	0.40	0.08	0.05	0.29	0.07	0.22	0.85	0.04	0.0

*** (-1.00 indicates a missing value)**

The number of intransitive triads in this dummy matrix is counted and stored. The operation is then repeated, in our case twenty times, and a running sum and sum of squares of the number of circular triads is kept. This allows the mean and standard deviation of the number of circular triads to be calculated. The mean is taken as a guide to the number of intransitive triads which usually occur. The standard deviation allows one to place confidence limits on the difference between this number and the number of circular triads observed, thus permitting one to text the hypothesis that fewer triads have occurred because there are real site preferences. Verifying whether or not the observed number of triads is significantly less than the number expected by chance is the statistical test endorsed to show that the results of a given analysis show a real preference structure.

VISUAL INTERPRETATION OF TIE P MATRIX

Although the Proportions matrix may be interpreted visually it is helpful to reorder the rows and columns of the matrix in such a way as to place the highest scoring sites at the top and left of the matrix. Figure 4 presents the data of Figure 3 reordered according to high to low site attractiveness as given in Table 1. Ideally, if the matrix were perfectly transitive, and all entries unanimous, all values to the right and above the diagonal would be ones and all to the left and below zeros however, such a situation seldom arises.

Several useful observations can be made from an inspection of the recorded matrix. The inconsistent judgments can be identified as the values greater than .5 below the diagonal, or as values greater than .5 above it. Similarly, the degree of confusion about the attractivity of any site may be readily observed. So, although the display of the P matrix in this fashion adds no new information, it is useful in pinpointing sites which do not appear to fit into the general pattern of the system being studied.

DATA BASE AND PREPARATION

The study was based on Saskatchewan data collected as part of the CORD Study Park User Surveys (see the Data Documentation Volume). The data received by the author were in the form of lists of users at twelve parks in Saskatchewan. Each individual record contained the user's origin and the date he visited the park. These data were tabulated by hand to get origin destination flows. These flow data were then keypunched by punching the code number of each park, the coordinates of each locatable origin from which visits to that park were made and the number of such visits. In summary, the data analyzed consisted of one card for each origin, each card contained the identity number of the origin, the origin name, its coordinates, and the number of visitors from that origin to each of the twelve parks.

One of the flow patterns actually observed is reproduced as Figure 5. Similar material was prepared for one way travel with upper limits of 220 and 867 miles. One may find it interesting to note that Figure 5 is based on a total of 3,254 trips.

RESULTS OF THE STUDY

The attractivity indices program was run twice, the first run disallowing any trip exceeding 200 mites in length (one-way), and the second, which is not reported here, omitted those longer than 225 miles. The magnitude of the distance limit imposed is of course arbitrary, but about 200 miles one-way as an upper limit on day trips appears reasonable to the author. Regardless, the imposition of this limit invalidated less than three percent of all trips.

TABLE 1 :INDICES OF ATTRACTION AND CONFUSION: 200 MILES

(Maximum Distance = 200 Miles)

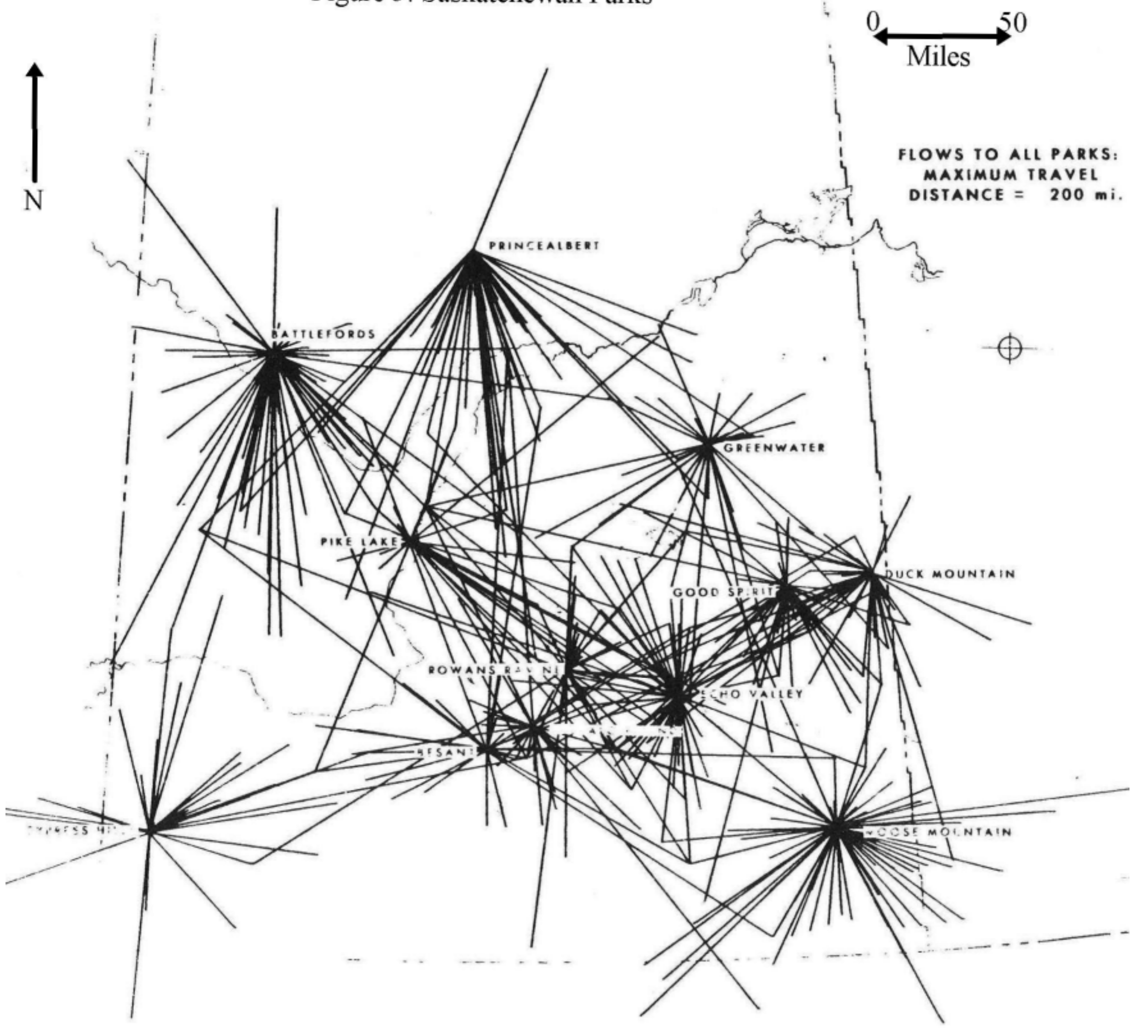
Site	Site Name	Attraction	Confusion
1	Buffalo Pond	.190	.909
2	Cypress Hilts	.732	.545
3	Duck Mountain	.730	.909
4	Echo Valley	.277	.909
5	Good Spirit	.390	.909
6	Greenwater	.456	.909
7	Prince Albert	.821	.727
8	Moose Mountain	.473	.900
9	Pike Lake	.6i4	.727
10	Rowan's Revine	.266	.727
11	Battlefords	.619	.900
12	Besant	.439	.727

Inspection of the comparison matrix C(introduced earlier in Figure 2) reveals that the number of visitors which may be inferred to have compared any two sites from 0 in the case of Sites 8 and 11 to 241 in the case of Sites 4 and i0. Due to the method of calculating the proportions matrix P (Figure 3) by dividing P(i,j) by P(i,j) + P(j,j) the varying numbers in the cells of the comparison matrix need not prompt undue concern.

It is evident from the proportions matrix P that almost one-third of the entries suggest that users deviate from unanimity by as much as 25% in showing their preferences. This indicates that there is no consensus as to whether one site is more attractive than another. Indeed, only 12 of the 66 comparisons are unanimous. Examination of the confusion indices (see Table 2) indicates that there are differing degrees of confusion for different sites. There is clearly no site about which there is no confusion.

The attraction indices (Table 1) ideally have a range of 1 (for a site which was never judged to be less attractive than any other site) to 0 (for a site which was never judged to be better than any other site). These indices, although regarded as only ordinal in nature, allow one to rank the sites. This was done in Figure 4, using the information from Figure 3. As described earlier, it is in terms of Figure 4 that one may understand a great deal about the preference structure discovered.

Figure 5: Saskatchewan Parks



It remains to assess the proportions matrix to determine the statistical significance of the results of the analysis. There were 17 circular triads counted, and after 20 simulations the average number of circular triads observed in the randomly generated P matrix was 53, with a standard deviation of 6.09 (this was true with both 200 and 225 mile travel limits). Thus it may be concluded that since the observed number of circular triads was 5.9 standard deviations from the mean, the results obtained were significantly non-random and there was a high degree of agreement between the sample subjects regarding the attraction of sites.

DISCUSSION

Although at present (1971) there is no way of assessing the correctness or accuracy of the results of this study, apart from reviewing the logic of the approach and the accuracy of the computer program used to analyze the data. (This is being written from a 1973 perspective, and a review of the methodology outlined above has been carried out. See TN 9 for details.

It is the author's opinion that the apparent difficulty of making judgments as to whether one site is more attractive than another (as revealed by the lack of unanimity in the P matrix) is rather disappointing, although if all twelve parks are considered to have similar levels of attractivity it is logical to find a relatively high degreeof confusion between them. In unpublished work on retail stores, the author found that eighty percent of the choices resulted in unanimous decisions. However, a great deal of the problem may be traced to the nature of the data set itself. The classification "day-user" may not result in a sufficiently homogeneous class of users as it includes fishermen, picnickers and swimmers, to name but a few, all of whom may perceive the attractiveness of a particular site differently. Also, in the case of this study, there are many provincial, local, regional and private recreation areas which provide alternate destinations for day trips. No comparable visitation data for these sites was available.

Furthermore, we have no way of knowing whether or not the individuals who visited any given site felt afterwards that the attraction of the site had justified the expense of the trip to it. In other words, the assumption of a well-informed user implicit in most models of recreation behaviour may not be valid. Ideally, this approach is best applied to situations in which it may be assumed that the individual has sufficient knowledge about all the alternative sites closer to his origin than the site he selects to make a rational decision as to which offers the greatest satisfaction-- presumably that which he patronizes most often. With the present data one does not have this information, but only knows that a trip was made to a specific site on at least one occasion.

Finally, the question of whether or not the entire benefit the individual receives from his day trip can be attributed to the site he chose to visit is one which has been addressed here.

CONCLUSION

The method employed above to analyze the Saskatchewan day use data has provided a ranking of twelve Saskatchewan parks. The consistency of the decisions which led to this ranking has been found to be significantly non-random, although there is a certain degree of confusion evident in the ordering of the pairwise comparisons. Still, having some behaviourally based attractivity assessment for parks is better than relying on intuition. This is particularly true when it is recognized that possible reasons for the apparent high confusion about various parks' attractivity have been pointed out as probably attributable to the nature of the data. If the conjectures about why the levels of confusion were so high are correct, then the method offers even more promise than is evident from the results achieved.

ATTRACTIVITY INDICES - POSTCRIPT

The main objectives of the attractivity indices model just presented were to determine a unique rank ordering of service sites with respect to their inherent attractions and to establish the degree of consistency present in the individual inferred judgments upon which the ranking is based. Following the establishment of consistent indices of site attraction, investigation of a data set may proceed by either modelling of the spatial choices inferred, or by the statistical explanation of the "pure" attraction indices which have been revealed. Each of these approaches is discussed below.

THE CONCEPT OF PREFERENCE SURFACES

The concept of preference surfaces is based on the assumptions that an individual in need of a service (in this case recreation) attempts to minimize some function of the attractiveness of each possible alternative and the distance to that alternative, and that each Individual faces what may be regarded as a unique array of spatial alternatives. The suggested thought process that each individual employs is a mental search routine whereby each alternative spatial opportunity is compared against all other available opportunities. He then chooses that alternative which will maximize his benefit. If, for example, we postulate a preference function of:

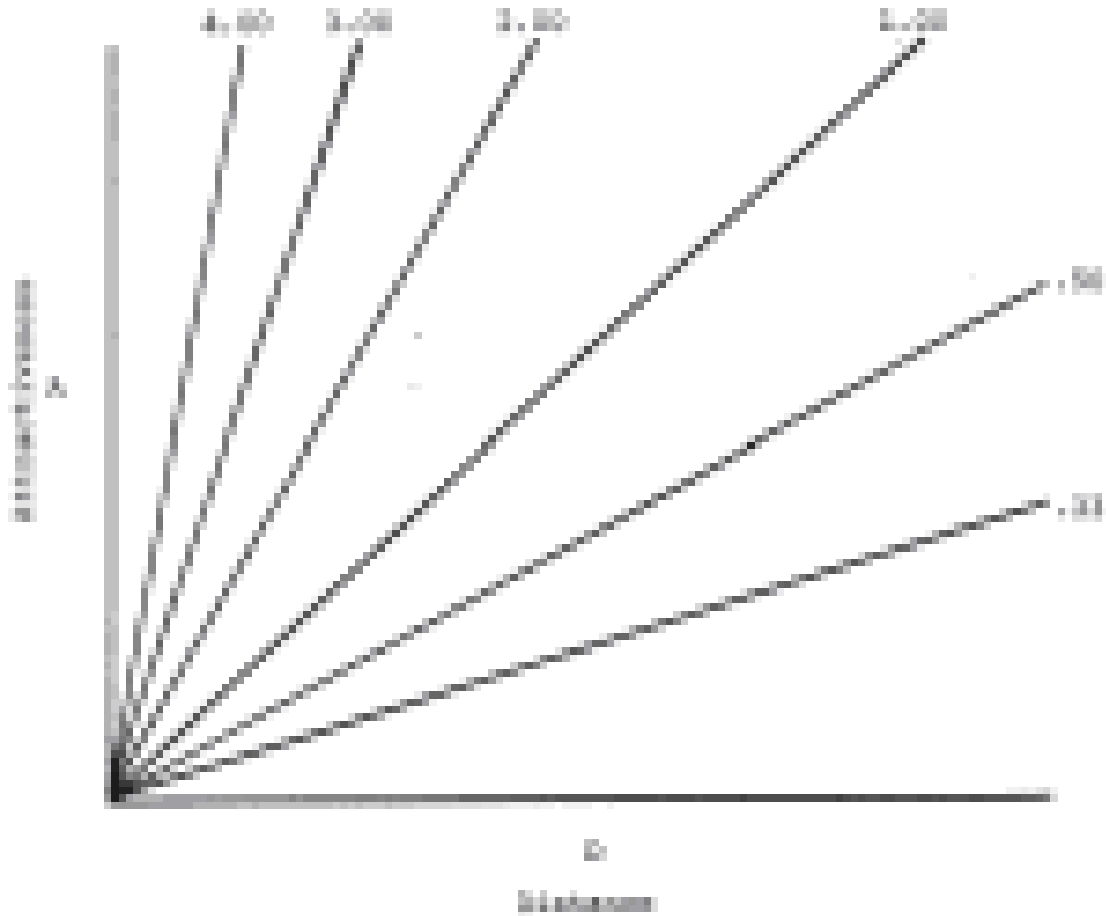
$$U(i,j) = A(j)/D(i,j)$$

WHERE $U(i,j)$ = net benefit to accrue to a visitor from i who chose to patronize site j, $A(j)$ = the attraction of site j, and $D(i,j)$ = the distance between i and j.

We define a preference surface. Such a surface would be similar to that shown in Figure 6. Given such a surface, one can proceed to inspect actual choices made to determine the extent to which the inferred choices were consistent with the postulated surface - that is to say, the proportion of times that any subject i chose to patronize that site j which yielded the highest value of U. One should note that in using a preference surface model it is assumed that a subject will be indifferent to the various spatial opportunities which lie on the same "contour" of the surface, e.g. those opportunities i,j, k,l, and men for which $U(i,j)= U(k,l) =U(m,n)$.

Initial investigation of the degree to which the observed recreation patterns in the Saskatchewan day use data used to develop the results presented earlier in the paper are explained by various surfaces has been carried out at the University of Western Ontario. Results obtained (based on only the 600 individuals) appear promising (unpublished).

FIGURE 6: HYPOTHETICAL PREFERENCE SURFACE



ESTIMATION OF INFERRED SITE ATTRACTION INDICES FROM OBSERVED SITE CHARACTERISTICS

Height at any point (i,j) = U(i,j) = A/D,

WHERE A = the attraction index of alternative sites;

D = distance to each alternative.

NOTE: The origin of the surface is always located at the individual's home.

An individual would be indifferent to alternatives x and y (because the utility of both is .33), due to the fact that the added attraction of y is compensated for by its greater distance. He would, however, favor z over either x or y.

It should also be noted that an individual's discriminatory powers are not perfect, and that each of the infinite number of contours which may be drawn on this surface really represents only the bisector of a zone of indifference.

Various attempts to quantitatively relate the number and quality of facilities to derived attraction indices have been reported in the literature (see the literature review in TN 9). Although several studies have reported a high degree of explanation, the combinatorial rules followed have generally been specific to each data set, thus rendering the formulation of general rules difficult. One of the major problems appears to be that there is little evidence that the supply of facilities and user response to them is a simple, relatively invariant one - although in certain cases, such as that treated by Wennergen and Neilson (1970) in their analysis of demand for fishing) it may very well be the case. In more complex situations it would seem that response to an increased number of any particular type of recreation facilities on one site should be a function of the relative availability of that facility on all alternative sites, and of what a person (party) really wants to do, rather than solely a function of what is at the site in question.

In view of difficulties involved in estimating the influence of increased diversity of facilities on consumers, it is believed that, at this time, the investigation of the relationship between site variables and attraction scores should be approached cautiously, possibly through the use of non-metric techniques. (See TN 4 on the use of metric techniques in recreation analysis problems.)