

TN 18: A METHOD FOR PREDICTING ENROUTE OVERNIGHT PARK USE

BY H.K. CHEUNG, S. SMITH & J. BEAMAN

ABSTRACT

In this paper a regression model is presented for predicting overnight use at a park where campers are of the enroute type. These are campers who are enroute to somewhere other than where they stop.

In the multiplicative model developed, major arterial highways that are close to parks being studied were considered to be "sources of visitors". The level of traffic flow was used to define the "origin population". Two other explanatory variables were employed in the model - the number of campsites and the road distance between highway and park. In planning for a new park or changing an existing one these latter variables are suggested to be ones which, in selecting and developing a park, can be manipulated by a park planner to influence the amount of use a park will receive.

When the model parameters were estimated using linear regression (after a logarithmic transformation of the multiplicative model) all regression coefficients were found to be significant at the five percent probability level according to the F-test and all had the expected signs. The root mean square percentage error in predictions for individual sites obtained was 22.59 percent, indicating that, at least according to this measure, the fit was good.

PURPOSE

The purpose of this note is to report on research that has resulted from attempts to derive a predictive equation for enroute camping use in Northern Ontario. The equation derived is especially relevant to park planning because some of the explanatory variables used in its derivation are planning variables in the sense that a park planner may be able to manipulate them to influence the volume of use that a park receives.

INTRODUCTION

Generally speaking, park users can be classified into four basic types: (1) main destination day-users, (2) enroute day-users, (3) main destination overnight users and (4) enroute overnight users (for further elaboration of these classes see TN 8 and 30). An enroute overnight user, which is the one of concern here, is a camper who stays at a park for one night and then moves on.

The usefulness of disaggregating park users has been pointed out by Pankey and Johnston (1962). For example, in a reservoir study they found that the root mean square (RMS) error of a linear equation for predicting total use (day and overnight) was 114. This was reduced when total use was broken up into day-use (RMS = 73) and overnight use (RMS = 80). They concluded that greater precision could be obtained when a researcher predicted day and overnight use separately and then summed them to obtain total use.

The point of pursuing the visitor classification issues raised above is that the model described in this paper (and applied herein to twenty provincial parks) is the outgrowth of one developed to forecast total use of Pukaskwa National Park. Pukaskwa is situated in a relatively isolated region in Northern Ontario (see Figure 1), about 192 road miles from Sault Ste. Marie and 220 miles from Thunder Bay. Each of these two distances includes a 12-mile link of paved road connecting the park to the Trans-Canada Highway. The park's location suggests that it would not generally be used as a main destination, as is the case with many other Canadian parks.

Indeed, an examination of the Ontario Camper Statistics (1971 Ontario Provincial Parks Statistical Report) revealed that most of the provincial parks in Northern Ontario had more enroute campers than main destination campers. Clearly, Pukaskwa's visitors would include many of those who travel the Trans-Canada Highway and who, while on their trip, would deviate, say 10 to 30 miles to the park as a convenient place to stop over.

By way of further introduction, the reader may wish to note briefly how the model presented here evolved. The original request by planners for assistance in developing a plan resulted in an early attempt to define a model. Planners were asked about their development ideas so that it would be possible to decide what types of users the researchers should consider. When it was recognized that the park would be established with the main stopping area approximately 25 miles from a major road, it became clear that it would be necessary, to concentrate on enroute camping use estimates. It was not considered likely that people would merely drive into the main stopping area of the park to say they had seen the park when it would take them 50 miles out of their way. The researchers and planners also noted that there were no large population centres within 200 miles of the park. The primary source of visitors would be the traffic on the Trans-Canada Highway.

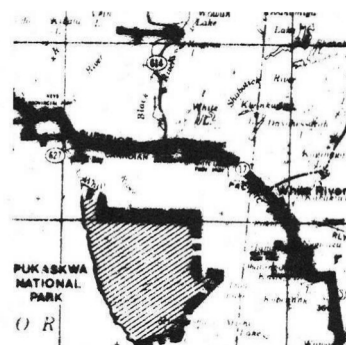
An important concern of the planners and park managers was the percentage of visitors who would be looking just for overnight accommodations as opposed to those who would be looking for a wilderness experience, which was the main purpose for establishing the park. These users who (in terms of TN 23) made only partly desirable use of the park were important for two reasons: (1) the park has an important but secondary objective of regional development, and (2) it is these people for whom there was a need to plan because, in line with not preserving the park for a few hardy wilderness users, it becomes necessary to provide the space and facilities that will encourage some enroute visitors to further investigate how they can make use of National Parks (the education and interpretation objective for National Parks).

There are some data available from other parks which are in similar situations, so it would have been possible to consider a more elaborate the model. However the researchers felt that it was adequate to pursue matters such as the number of interior wilderness use sites intuitively. The most important concern was to make sure that there would be sufficient capacity in the park so that the number of sites for the long-term visitors to the interior of the park could be expanded to meet demand as the use of the park increased. An initial capacity that seemed reasonably adequate to accommodate this type of user when the park was first open was already proposed in the parks draft master plan. So there seemed little merit in spending additional time and money to further define the need for wilderness sites.

Figure 1:

PUKASKWA NAT. PARK AND THE SURROUNDING AREA

▲ PROVINCIAL PARKS



THEORETICAL CONSIDERATIONS

Given a specific interest in developing an enroute camper model, a conceptual problem of model formulation is one of deciding the key variables affecting people's choice to stop over at a given park. A variable that is obviously important is the distance which people have to travel from their main route to reach the stopover park. How many people will deviate more than 15 miles from their main route if they do not intend to stop for any length of time at a given park but are, in fact, heading for some distant main destination? How many people will go even five miles out of their way on a gravel or poorly conditioned road? Clearly, both the distance and the condition of road should influence greatly one's decision to make an enroute visit to a park.

So it is reasonable to suggest that a major highway at a certain distance from a park serves as a generator of visitors in much the same way that a city at a given distance from a park serves as a generator of visitors (see TN 30 and Reference 28). The higher the weekday, weekend, (etc., depending on circumstances) annual average summer traffic volume, the more visitors one may expect to leave the highway and make an enroute visit to a given park. But there will be a certain volume of traffic that is commercial or business travel that will not stop at a park. Here it is assumed that the portion of potential park visitors in that flow of traffic is available from traffic partition counts. To avoid unnecessarily awkward descriptions of variables, one may refer to the traffic on a highway and mean it to be an estimate of the volume of the total traffic that is potential visitors.

Another variable considered to be important in influencing the decision to make an en route visit to a park is the number of campsites the park has. Almost certainly, the probability of people deviating from the main route of their trip will depend on the probability of finding accommodation at the stopover park. It is reasonable to suggest that the chance that a site is available is higher at a park with a large number of campsites than at a park with a small number.

Cheung (see TN 1) has suggested that the facilities at a park influence its attractiveness, and that an attractiveness index based on the quantity and quality of the facilities could be used as a variable to predict park use. Both he and Beaman (TN 9) have suggested that the existence of other parks around a given park affects use either by modifying attractiveness or in some other way (see TN 11 and 33). One may consider the features at a park and alternative recreation opportunities as important variables affecting park attendance.

Given all of the preceding, enroute camping use of a park is hypothesized to be a function of highway traffic counts, the number of campsites of a park under study, and the road distance between this park and one or more highways. Variables such as road condition, park attractiveness and alternative parks were excluded in this analysis because of a lack of data. However these variables are not expected to affect the R^2 values substantially because (1) most of the parks studied have gravel roads connecting them to a major highway; (2) the specific concern in developing an enroute camper model was with an isolated site so alternative parks have negligible influence on park use; and (3) enroute campers perhaps have little concern for anything beyond having an acceptable campsite relatively easily accessible, with certain amenities such as drinking water and flush toilets. Regarding (3) the same basic amenities are provided in all the sites considered because they conform to Ontario Provincial Park Standards.

DATA

Data used in defining the dependent variable, the number of enroute campers visiting a park during a season, were drawn from the 1971 Ontario Camper Statistics (OCS) compiled by the Ontario Ministry of Natural Resources. The OCS included information on type of campers,

average party size, and length of stay. Since most of the parks selected for this study had more enroute use than main destination use, these numbers of campers were regarded, for practical purposes, as if they were all enroute campers.

Traffic count data used as the homologs of origin population were the annual average daily traffic. Information on the number of campsites in a park was obtained from the 1971 Ontario Camper Survey. Road distances between major highways and parks were obtained from an Ontario highway map.

METHOD

Twenty provincial parks in northern Ontario were used in the analysis. They are located near major arterial highways and provide similar facilities such as drinking water, flush toilets and fireplaces. Fifteen sites were actually used in developing the model; the remaining five were used for model validation.

$$(1) V(j) = b(1)T(j)^{b(2)} S(j)^{b(3)} e^{-b(4)D(j,k)}$$

WHERE

$V(j)$ = the dependent variable, which is the observed number of "enroute" campers visiting park j during a season,

$T(j,k)$ = annual average daily traffic count on a section of the arterial highway, k , between two intersections leading to park j ,

$S(j)$ = the number of developed campsites in park j ,

$D(j,k)$ = the distance to the junction with highway segment k , and

$b(m)$, $m=1$ to 4 are parameters to be estimated.

The multiplicative model defined by Equation 1 was changed into a linear one by using logarithmic transformation that allowed the use of linear regression analysis to estimate the parameters of the equation. The estimated equation was:

$$(2) \ln(V(j)) = 3.22 + 0.37 \ln(T(j,k)) + 0.80 \ln(S(j)) - 0.03D(j)$$

RESULTS

The R^2 standard error of estimate and the F-value for Equation 2 are 0.89, 0.27, and 31.17 respectively. The equation derived is significant at the .01 level. The regression coefficients, their standard errors, and the increase in R^2 value at each step of the stepwise regression analysis are presented in Table 1. The coefficients of the model are significant at the five percent probability level according to the F-test, and all have the expected signs.

It is sometimes the case that the relative importance of the explanatory variables is suggested by the order in which they enter in a stepwise regression. Also there is some merit in discussing the increase in the amount of variance in $\ln V(J)$ each explained when each new variable is added. From the convention perspective taken in interpreting stepwise regression results the number of developed campsites, a variable reflecting the capacity of a park, is the most influential variable of the three explanatory variables. The coefficient of the variable is 0.8, implying that a 10 percent increase in the number of campsites of a given park will result in an eight percent increase in use, *ceteris paribus*. Essentially, what the coefficient means is that attendance increases as the number of campsites increases, but at a decreasing rate (the problem with this interpretation is commented on subsequently).

TABLE 1: SUMMARY OF THE VARIABLES ENTERED INTO THE REGRESSION EQUATION BY STEP

Regression Step	Variable	Regression Coefficient	Standard Error	R ²	Increase in R ²
1	Number of developed Campsites	0.80	0.14	0.70	0.70
2	Traffic Count on Arterial Highway	0.37	0.14	0.83	0.13
3	Distance from Arterial Highway to Park	—0.03	0.01	0.89	0.06

The fact that the coefficient of the variable, "traffic counts" (0.37) is much less than unity may be taken to imply that use does not increase indefinitely as traffic volume increases.

The distance variable entered the equation last and explained six percent of the variation in the transformed dependent variable. The relatively small explanatory power of distance as a predictor variable in this model was expected because the average distance from an arterial highway to the entrance of one of the 15 parks used to develop the equation is only about four miles.

DISCUSSION

Measuring how well a regression explains data involves a number of complex issues (see TN 19 and 35) but certainly a measure such as % RMSE that gives an idea of the average percent error in estimates is a useful guide. Here the percent root mean square error (% RMSE) is defined as:

$$\%RMSE = [1/n \sum (\% \text{ error in prediction at park}(j)^2)]^{1/2}$$

One may note that the percent root mean square error is also called the standard deviation of the percent error of prediction (see Reference 17.)

The 22.59 percent root mean square error (% RMSE) obtained indicates that the fit was relatively good. (On expected error and structural adequacy of models see TN 19 and 35.) The components of %RMSE, the percent error of prediction for the individual parks, varied from - 5.54 to 36.63 (see Table 2). Some of these large errors may be caused by the absence of variables accounting for the effects of alternative parks and park attractiveness. In particular, the over-prediction of total season camping at Aaron may have stemmed from the failure to weigh the importance of the three competing parks lying close to it, namely, Sandbar Lake to the east, Blue Lake to the west, and Ojibway to the north. On the other hand, the underprediction for Neys and Sandbar may have arisen from ignoring the fact that the former is the only park among those studies that has boats for hire, and that the latter contains an excellent fishing area. The situation in the latter case is further complicated by the failure to recognize that not all enroute campers are the same in their preferences for a campground. Some may be strongly oriented towards some

activity and thus perceive a different attractiveness for a park than those who are just looking for a place to sleep. A site with something "special" to offer can be expected to attract either some significant use as a main destination or use from activity-oriented enroute campers as well as use from the more ordinary enroute camper.

Thus far the model has been evaluated in terms of the plausibility and significance of the regression coefficients and the accuracy with which it can be used to make predictions for fifteen provincial parks. To further assess the predictive utility of the model, it was applied to the five parks that had originally been set aside for validation. The results presented in Table 3 are further evidence that the model can be used for predicting levels of enroute use at predicted park sites.

CONCLUSIONS

The conceptual framework proposed here was found to be substantiated by empirical results. In fact, validation shows that the model developed should give accurate enough predictions for planning and policy purposes. So, the results of the analysis have shown that by varying the number of campsites in a park, or the distance between the park and a major highway, or both, a park planner can reasonably and consciously manipulate the amount of enroute use a park will receive. This paper provides specific guidelines as to what the consequences of such manipulations will be.

TABLE 2: OBSERVED AND PREDICTED VALUES AND PERCENT ERRORS OF PREDICTION FOR THE PARKS USED IN DERIVING THE VISITATION MODEL

Park	Observed 1971 Total Campers	Predicted 1971 Total Campers	Error (%)
Lake of the Woods	5,695	5,700	0
Invanhoe Lake	12,688	12,780	0
Pancake Bay	38,621	35,579	- 7
Obatanga	13,468	17,080	26
White Lake	25,316	27,853	10
Aaron	-10,837	14,205	31
Greenwater	4,949	6,762	36
Klotz Lake	3,688	3,406	- 7
Blacksand	5,104	6,945	36
MacLeod	7,931	7,240	- 8
Neys	21,917	15,630	-28
Rainbow Falls	30,441	28,314	- 6
Sandbar	9,093	6,641	-26
Sioux Narrows	5,821	5,499	- 5
Inwood	18,072	11,310	-37

Root mean square error = 22.59%

The use of the number of campsites as an explanatory variable is particularly useful when the carrying capacity of a park, in terms of camping use, is known. If one assumes that the "physical capacity" of a site is measured in terms of campsites that can be developed (for further discussion of carrying capacity in this context see TN 16 and Reference 22), one can use Equation 1 to determine the user—days of use to be expected to see if they can be developed

without exceeding the social carrying or biological capacity of the site based on the level of management expenditure to be devoted to maintaining the site in a "sustained yield" condition (see TN 16).

Nevertheless, the use of the number of campsites as an explanatory variable in a regression equation causes concern for some outdoor recreation researchers. Their legitimate concern is related to what is called a "circularity problem". In essence, they argue that more campsites cause more use, which results in more campsites being developed (see Figure 6 in TN 40). It is impossible, they feel, to resolve this "chicken—egg" cycle in a way that justifies using the equation in the way described above. The author argues that correlations which when studied using regression analysis give the equation determined should not be read to imply a cause and effect relationship. It may be found that

for established campgrounds there is an equilibrium condition reached which puts an end to "circularity": here because of the need to reach some conclusion the equation is treated as defining some kind of an equilibrium condition that "would eventually be reached", not as indicating anything about the dynamics of how the equilibrium would be reached.

Other independent studies have also shown the importance of the number of campsites on the attractiveness of parks. Cesario (TN 4) found that the number of campsites in a park was the major component of attractiveness. Should a researcher be concerned about using Cesario attractivities because using them in some way implies that the number of campsites is used as an independent variable explaining camp use? Perhaps, but if the Cesario attractiveness do not seem to be useful for park use predictions what will be? The point is that there may be some problems with the model developed but, on the whole, there are many positive features. It is simple and requires relatively simple data to yield quite accurate estimates for other sites not used during the model's development. With appropriate projections of the highway traffic variable, and given accurate values of the other two variables, predictions can be made and there is good reason to think that they will be reasonably accurate. Clearly, the use of the model for prediction assumes that the parameters are accurately estimated and that they will not change appreciably during the prediction period and that the effects of alternative parks and park attractiveness have been correctly considered (see TN 37).

TABLE 3: OBSERVED AND PREDICTED VALUES AND PERCENT ERRORS OF PREDICTION FOR THE PARKS USED IN TESTING THE PREDICTIVE ABILITY OF THE VISITATION MODEL

Park	Observed 1971 Total Campers	Predicted 1971 Total Campers	Error (%)
Lake Superior	42,020	46,777	11.32
Wakami	4,183	3,872	- 7.43
Quetico	15,364	15,034	- 2.15
Kettle Lake	13,994	17,578	25.61
Nagagamisis	4,529	4,860	7.32

Root mean square error = 13.40%