"Recreation Travel Modeling: A Day-Use Park Visitation Model for the Alberta Northern Region."

by

James Vaillancourt, B.Sc.

A Master's Degree Project
submitted to
The Faculty of Environmental Design
The University of Calgary
in partial fulfillment of the requirements for the degree of
Master of Environmental Design
(Urban and Regional Planning)

Calgary, Alberta

December, 1991.

The University of Calgary
Faculty of Environmental Design

The undersigned certify that they have read, and recommend to the Faculty of Environmental Design for acceptance, a Master's Degree Project entitled:

"Recreation Travel Modeling: A Day-Use Park Visitation Model for the Alberta Northern Region"

submitted by James Vaillancourt in partial fulfilment of the requirements for the degree of Master of Environmental Design

Prof. Tom L. Harper
Faculty of Environmental Design

Dianne Draper

Dr. Dianne Draper
Faculty of Social Sciences
(Department of Geography)

Dr. David Henry
Faculty of Environmental Design

December, 1991

ABSTRACT

Recreation Travel Modeling:

A Day-Use Park Visitation Model for the Alberta Northern Region

The making of sound decisions about the allocation of resources for provincial parks requires a significant amount of information, in particular attendance data. The purpose of this project is to develop a statistical model to be used for policy analysis and decision making.

In this project, the model building process is reviewed and then applied to the development of a logistic regression model to predict day-use park attendance and the effect of policy changes on attendance. The data used consists of characteristics of both the visitor origins and facilities or features of the parks. This spatial interaction model establishes a relationship between the number of day-use visits to parks, the attractiveness of the parks, and the distance between origins and parks. The attractiveness of a park is based on the number of recreation opportunities and facilities offered. The "distance" parameter is most significant as it bears the greatest weight on park attendance forecasting. The model is evaluated in terms of its accuracy, ease of use, and data requirements.

An analysis is made of the spatial re-allocation of a fixed total number of day-use park visits resulting from four policy changes: increasing the attractiveness of a park, decreasing the attractiveness of a park, adding a park, and closing a park. Following the interpretation of the results of the policy analysis, model limitations are assessed and potential model refinements are recommended. Model refinements could include changes to: data sampling, model structure, and market segmentation.

Keywords: Logistic regression, model, park planning, spatial interaction, policy analysis.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

INTRODUCTION

In order for government departments to make sound decisions about the allocation of resources for provincial parks, a significant amount of information is required, and in particular attendance data.

The purpose of this project is to discuss the model building process and to apply this process to a particular park planning situation in Alberta. Specifically, the project aims at developing a statistical model to predict day-use attendance to selected provincial parks in northern Alberta (Figure 1).

The model structure is designed around the type of data readily available. The data pertain to both the characteristics of the origins of visitors and facilities or features of the parks. A relationship is established between the number of day-use visits to the parks, the attractiveness of the parks, and the distance between origins and the parks.

The model is evaluated particularly with respect to accuracy and ease of use. Recommendations are made about improvements to the model. The improvements pertain to data collection, model structure, and modeling procedure. With such improvements, it is expected that the model would become more accurate but also more costly in terms of data and maintenance.

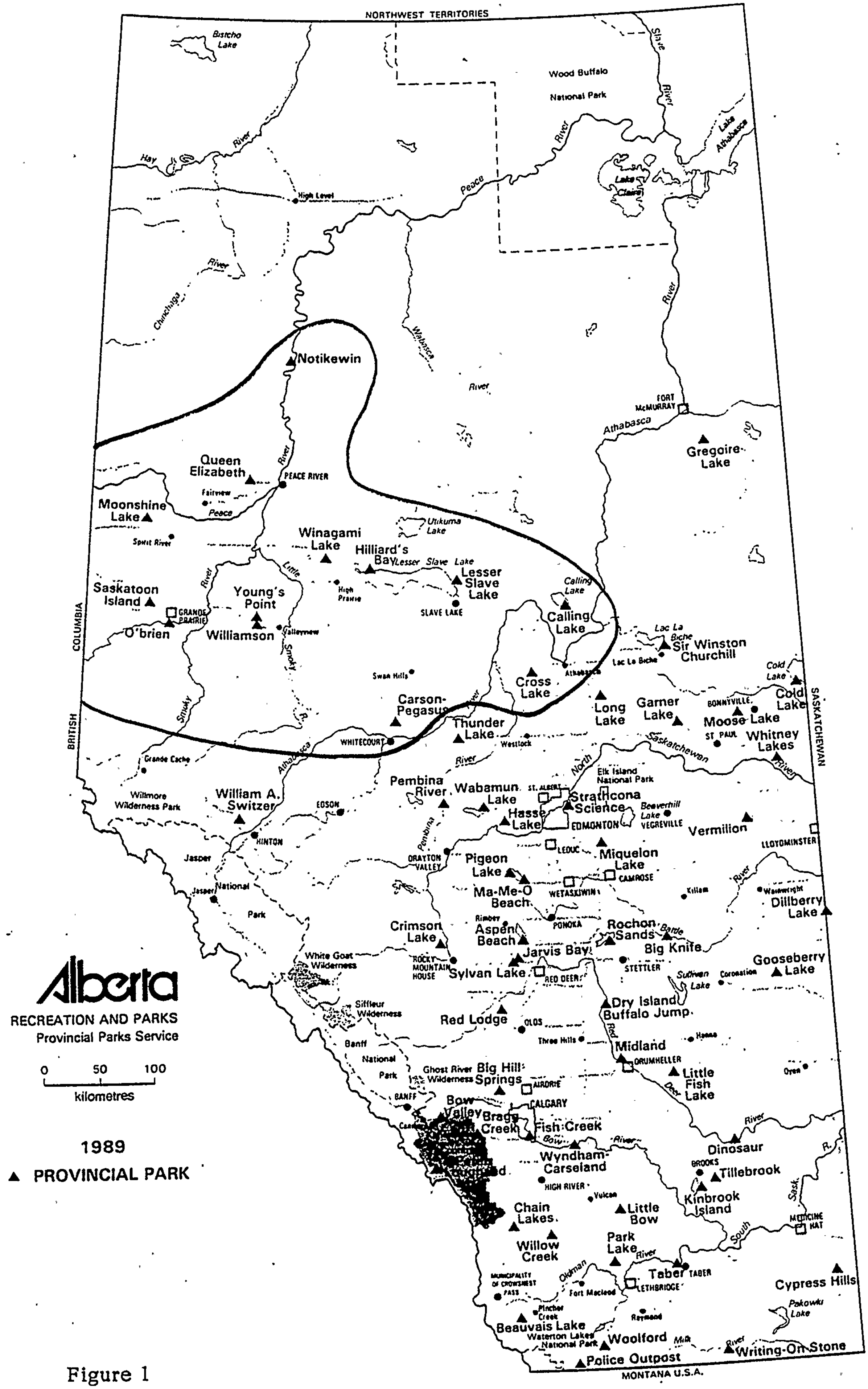


Figure 1

Produced by Support Section, Project Management Branch

Base constructed by Land Information Services Division, Alberta Forestry, Lands and Wildlife

CHAPTER 1

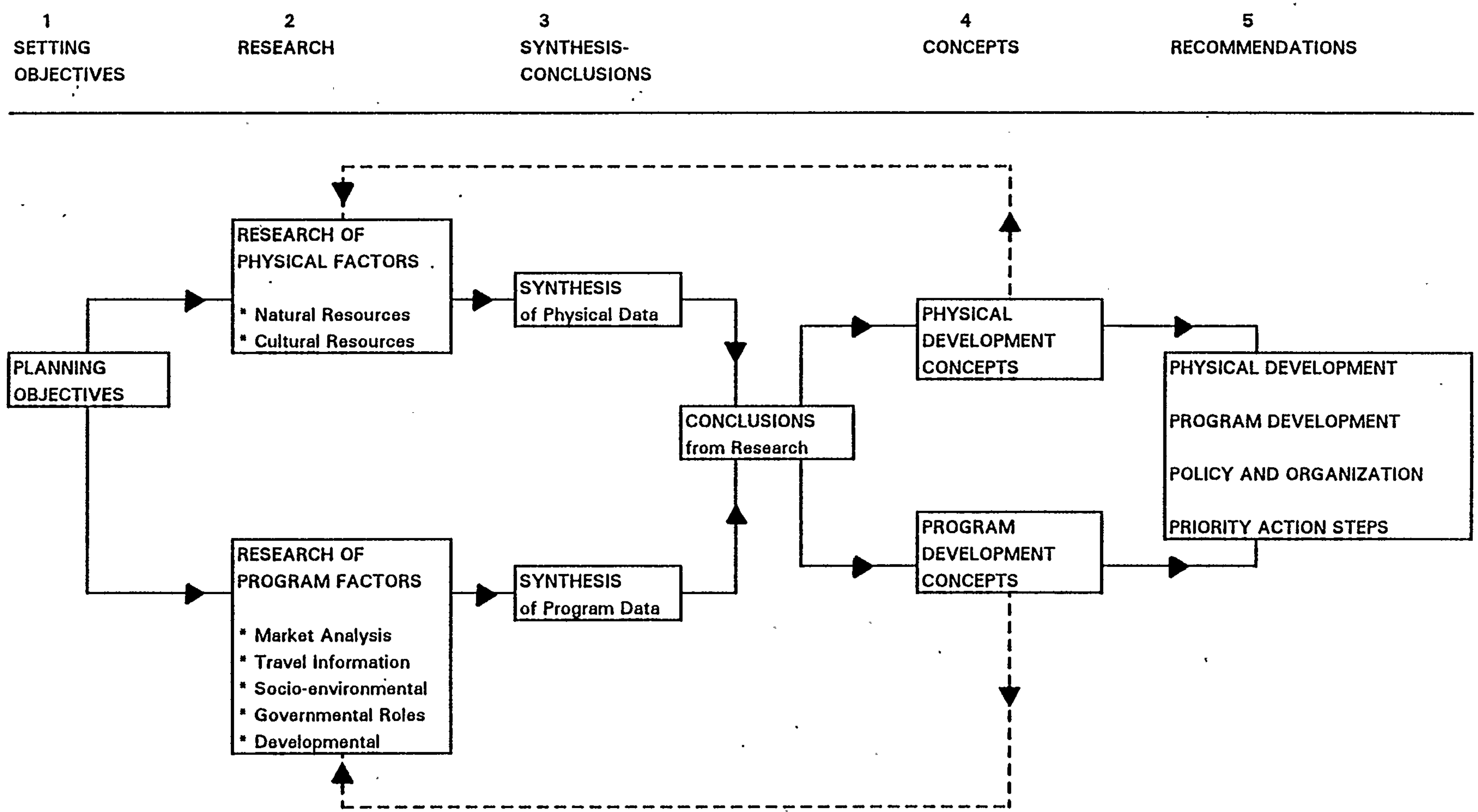
RECREATION TRAVEL FORECASTING: STATE OF THE ART AND FUTURE DIRECTIONS

An important phase of the regional and the community tourism (recreation) planning processes (Figure 2) is resource inventory and market analysis (Alberta 1988a, Gunn 1988, Perks 1986). Several types of information, internal and external, are required. Information is obtained through the research process. Some information is internal to the location (region, municipality, park) and pertains to factors that are, to some degree, controllable or measurable while other information pertains to external factors like socio-economic trends or policies that are beyond local control. For example, there is a need to inventory existing recreational facilities, and to know the type and number of various visitors who use or might want to use such facilities. There is a need to anticipate trend changes in economic conditions, population characteristics, consumer attitudes, and competing recreation opportunities. But other evaluation criteria (external factors) should be considered, for example, political, legal, technical, administrative and equity. A synthesis of the information leads to conclusions and finally decisions.

Although park visitation data are not the sole criterion of decision making, much effort has recently been applied to their accurate estimation (Fesenmaier 1988). This information is used in the next phase: the situational analysis, that is, the activity through which planners and decision makers assess the adequacy of the facilities in satisfying user needs and determine what are the opportunities for recreation facilities expansion and what are the potential constraints to development.

Obviously important to planning decisions is the total potential market, that is, the number of current visitors to each site plus the number of potential visitors.

FIGURE 2
REGIONAL STRATEGIC PLANNING PROCESS



(Adapted from Gunn 1988)

The main objective of recreation travel forecasting is to isolate the specific causal factors which generate interest in recreation travel and to estimate their quantitative impacts on travel behaviour. Recreation planners identify causal factors that can be controlled (to some degree) by government departments (policy variables) and then make recommendations about the provision and location of facilities with the general purpose of increasing park visitations and hence participation in recreation activities.

In this chapter the essential principles of models and the model building process will be outlined, quantitative and qualitative forecasting (modeling) approaches will be described with their most recent refinements, and future directions for research will be suggested.

Recreation forecasting techniques are usually divided into quantitative and qualitative approaches. Each technique will be evaluated in terms of its overall usefulness based upon five key evaluation criteria: time horizon, complexity, costs, data availability and accuracy. These evaluation criteria are defined and justified in the section entitled "Evaluation of Forecasting Techniques". This evaluation will guide the selection of the modelling (forecasting) technique to be used in this project.

MODELS AND THE MODEL BUILDING PROCESS

Simply stated, a model is a simplified representation (usually mathematical) of a complex real-life situation. Indeed, only the most relevant processes and factors are included in the model. By using a model, the planner can eliminate many of the complexities that obscure a problem to emphasize only the structural relationships between the important variables, so as to predict the outcome of a chosen policy.

It is possible to perform "what-if" analyses, that is to test the effect of policy changes on the dependent variable. For example, the planner could design a model to test the effect of an increase of park entrance fees on the total number of park visits. It must be determined, however, if the gains in simplicity and data manageability outweigh the loss of realism caused by the omission of descriptive details.

Planners are frequent users of models. Common model applications are: population projection (Chapin and Kaiser 1979), the attraction of shopping centres (Timmermans 1982); retail employment, residential population, and land use (a Lowry-type model, Batty 1972); recreation destination choice (Fesenmaier 1988). Planners prefer making recommendations based on predictions involving a certain degree of uncertainty to making decisions arbitrarily.

Types of models

A classification of model types may have different levels of refinements. In this discussion, only the most common model types will be listed, more or less in order of increasing complexity or abstractness (Stokey and Zeckhauser 1978).

Most people are familiar with *physical models* because they are often seen in public meetings. Indeed, planners frequently exhibit physical models of housing, transit system or park projects. These physical models are small scale versions of real size buildings, parks or other.

Somewhat more abstract are *diagrammatic models*: familiar examples are the traditional road map used to guide a traveller to a destination and a colour-coded land use map or plan for a lake area. Diagrammatic models normally make use of symbols like lines, shadings or colours to represent real properties like elevations, densities and land uses. In order to help the reader interpret the symbols, a legend is included on the map or plan. Graphs, flow charts and tables are also diagrammatic models.

Another type of model frequently used is the *conceptual model.* Conceptual models are often used to make qualitative predictions about the behaviour of groups or individuals. A good example is the functional tourism system model in Figure 3 (Gunn 1988).

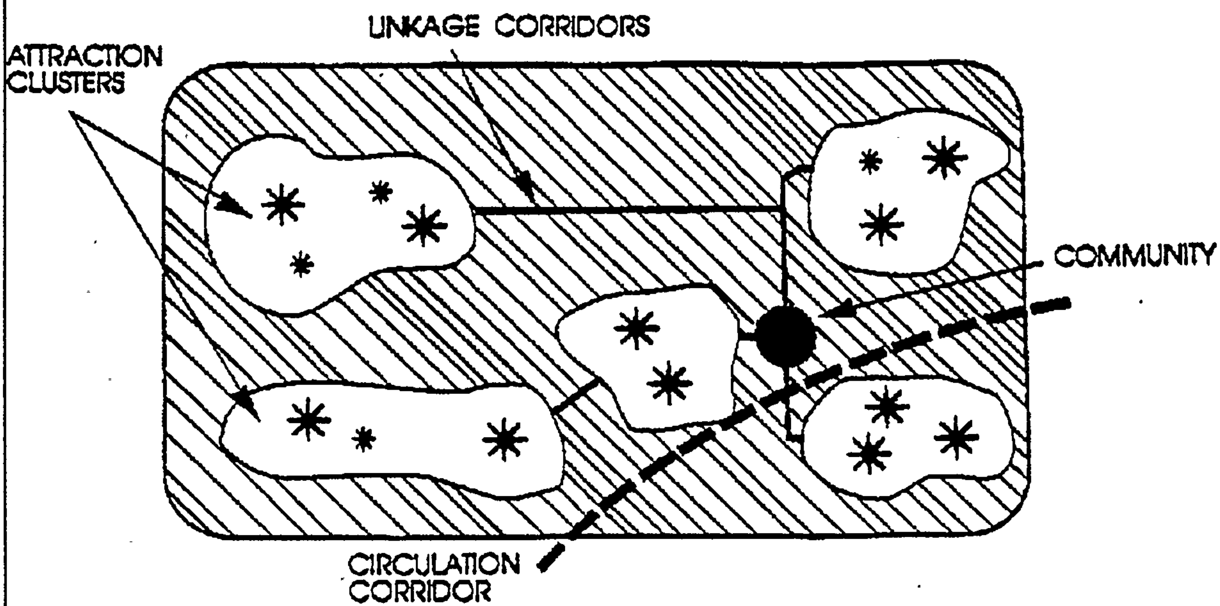


FIGURE 3 -DESTINATION ZONE CONCEPT

Source: Clare A. Gunn, Tourism Planning (2nd Edition), 1988.

The most complex and abstract model type is the *mathematical model.* Mathematical models usually consist of sets of equations representing the relationships among the variables of a particular problem situation. A computer is normally used because of the large number of equations and to allow the planner to simulate behaviour, for example recreation site choice. Although mathematical models are easier to manipulate than the other model types, they are generally more difficult to structure. It is because of their structural difficulty that more attention will be given to mathematical models in the next few pages.

Descriptive vs Prescriptive Models

A descriptive model attempts to describe or predict something like changes in a system or a component of a system; it focuses on the relationships between action and consequence. For example, a planner may use different net migration rates in a cohort-survival model to predict the effects on population numbers.

A prescriptive (or normative) model consists of two parts: one descriptive, the other procedural. The descriptive part predicts the outcome of a number of alternatives open to the decision maker. The procedural part gives and applies (a) rule(s) for choosing among the various alternatives. For example, a planner may use a residential growth model which predicts the attractiveness of different peripheral zones to residential land use based on spatial factors like accessibility to work areas, nearness to school, nearness to a major street etc. The decision rule would be to give most planning consideration to the zone(s) with the highest attractiveness rating(s).

Deterministic vs Probabilistic Models

A deterministic model is one in which the outcome is assumed to be certain. For example, in the population model mentioned above, for given initial conditions (birth rate, death rate, net migration rate etc.), the outcome is unique.

If the model predicts a range of possible outcomes for which the probabilities may be estimated, the model is called probabilistic.

The General Structure of Models

The general structure of models consists of variables and rules that determine how the variables relate to each other and how they predict the outcome. Chapin and Kaiser (1979) list the variable types as: input variables, status variables and output variables. They include as rules or operating characteristics: functional relationships, identities, constraints, and algorithms. These components of a model structure are briefly defined here.

The values of *input variables* are determined outside the model; they are often called "exogenous" variables. Common examples of input variables are the population of a town, the recreation participation rate or the origin-destination distance.

Variables that are internal to the model are called "endogenous" or *status variables* because they describe the status of a component of the model while the model is running. Status variables may obtain values from outside the model i.e. exogenously, but such values would normally change during the model run, and hence the model is said to generate values for these variables. Status variables are often used as iteration counters; they act as intermediaries between input variables and output variables.

The *output variables* are the dependent variables in the model i.e. the variables that express the outcome of the model. The values of the output variables depend on the values of the input and status variables also called the independent variables.

Possibly the most important rule or operating characteristic is the *functional relationship* because it shows how the outputs respond to and depend upon input. Mathematical forms that express the functional relationship can be linear, power or exponential. *Identities* are accounting or tautological statements, for example the holding capacity of a zone being the product of the land available for development and the number of dwelling units allowed per acre. *Constraints* are statements about allowable limits on values of variable. Upper or lower limits may be set, that is, x > = y or x < = y. *Algorithms* are computational rules, that is, a set of logical and mathematical operations performed in a specific sequence. The algorithm of the model may force the operation of the model to branch from one component to another. For example, the operation of the model may branch from the trip generation component to the trip distribution component.

The Model Building Process

A relatively detailed discussion of the model-building process is given in Chapin and Kaiser (1979). For the purpose of this project, however, only a discussion of the basic principles and procedures of model-building is needed. This discussion will pertain to the main activities of the model-building process which are:

1) Problem definition
2) Model specification
3) Model fitting
4) Model evaluation

Problem definition consists of defining the purpose of the model, that is, deciding what problem(s) the model attempts to resolve. For example, the model in this project is used to determine the generation and distribution of recreation trips.

The model specification procedure consists of formulating basic hypotheses about the variables and their relationships, determining how space and time should be treated and deciding on a solution method.

Model fitting, involves first choosing proxy variables, if applicable, to replace variables that are not directly observable or which would involve a complex and expensive data collection. Model fitting also involves the processes of verification and calibration of functional relationships. Verification and calibration are usually performed through statistical analysis, most frequently regression analysis. Verification consists of determining whether a variable is important (significant) in affecting the model's results. A variable is included in the model and the variable is verified. Then another variable is added and verified and so on. The modeler, through the verification process, can decide whether the model can be simplified by deleting a variable without a significant loss in predictive capability.

Calibration (often called estimation) means to find the values of the parameters (constants) that will result in the "best fit" between the model's output data and the observed data. The model fitting process may result in the creation of sub-models with different predictive capabilities and levels of complexity.

The *evaluation* process consists of deciding which model is better for the purpose at hand. The model will be evaluated in terms of its predictive capability; the costs of building, maintaining and using the data; and its overall usefulness relative to all costs.

The Advantages and Limitations of Models

From the description of the model building process given above, one can conclude that building a model involves a certain amount of discipline which forces us to consider fundamental principles. The model building process requires the modeler to think clearly about the problem at hand; the modeler is required to make many decisions about how to resolve the problem. The model forces the modeler to identify the variables that most influence results. Such important variables are often called policy variables because they represent factors that can be changed through government intervention.

Particularly valuable to planners is the possibility of experimenting with the model rather than the system itself. Indeed, through "what if" analysis, the planner can change the assumptions about the values of variables and predict the implications of such changes without incurring prohibitive costs of time and money. Through "what if" analysis, the critical tradeoffs become more evident.

One could agree that the limitations of a model result primarily from the constraints of time and data imposed on the model builder, but also from our understanding of the given model's ability to represent human decisions. The greater the resources decision makers are willing to allocate to refine a model, the less limited will the model be particularly in terms of its ability to predict. In assessing a model's limitations, the model builder may ask several questions. For example, how valid is the dependent variable, that is, is it really measuring demand for a site, transportation, or whatever other purpose the model is serving? Are the model results fairly stable from one run to another given stable inputs? What is the amount of error in predicted values?

QUANTITATIVE AND QUALITATIVE APPROACHES

Quantitative forecasting techniques that have been applied in the recreation field may be grouped into two broad categories: time series, and causal models which include regression models, and gravity and trip distribution models. Each of the above categories contains many categories. Time and space limitations make it impossible to discuss all of the types of models. One type of model, the simulation model (which generally combines both time series and causal components in a dynamic structure) is highly complex and not frequently used in recreation forecasting; it will not be described in this report.

Time Series Model

A time series is a sequence of observations obtained at successive points in time. Simply defined, a time series model is a technique which makes use of historical data to forecast the future . The underlying assumption is that what happens in the future is a function of what has happened in the past. Time series models are not concerned with explaining the causal factors of a forecast. All causal factors are considered in the aggregate. These models make the assumption that the net result of these causal factors (variables) is what has caused whatever trends that may exist in the data and that an extrapolation of a trend will yield an accurate forecast (Swart, Var and Gearing 1978).

Simple Trend Projection

Simple trend projection is one of the simplest time series models. Simple trend projection is a technique that fits a trend line to a series of historical data points and projects the line into the future. If a straight edge is used on graph paper, the function is assumed to be linear. An improvement over the "straight edge" method is the development of a linear trend function by a precise statistical technique called ordinary least squares (OLS). This technique is well explained by Wheelwright and Makridakis (1980). When used for projecting park visits, linear projections assume a constant growth rate of visits over time. For most applications over short time intervals, linear projection is the easiest technique to use and it produces useful results. Unrealistic results will be obtained, however, if the projections are made too far in the future (Burton 1981).

Alternative Growth Functions

There are alternatives to the linear growth function namely: exponential, logistic and product life cycle (Figure 4). Clawson and Knetsch (1966) illustrate the use of an *exponential* growth function in forecasting National Park System (NPS) visits. This function had the form of:

(1) $V = 10^{a+bt}$, and thus, $\log V = a+bt$

where:

- V is NPS visits in a given year t
- a, b are constants
- log is logarithm base 10
- t is time (in years)

In the use of this exponential function with Pre-World War II data, NPS obtained unrealistic projections to 1980 and 2000.

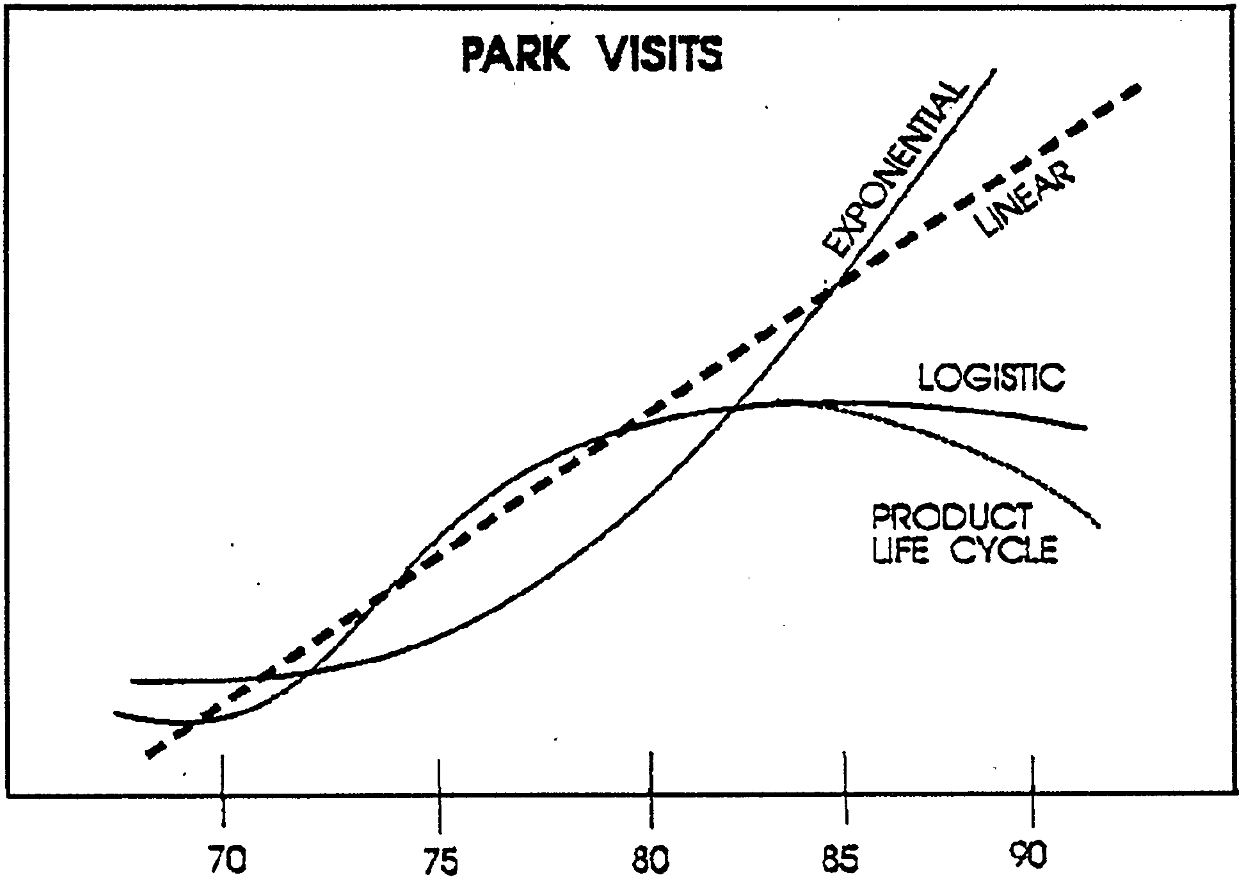


FIGURE 4 – GROWTH FUNCTIONS

The *logistic* function has been applied to a variety of population growth and other problems. Since the work of Thiel (1969), the logistic function has been recommended as a good candidate for capturing certain relationships in trip distribution models (Ewing 1980). The logistic function has been described by Stynes and Peterson (1984) as

> ".... well suited to processes which have start-up impediments and saturation effects, as the curve grows slowly at first, reaches a maximum rate of growth, and then proceeds to increase at a decreasing rate, approaching the saturation point as a limit."

The *product life cycle curve*, borrowed from marketing, follows the logistic trend with an eventual decline at the end of the cycle (Howard and Crompton 1980). Stynes and Spotts (1980) have observed that for long-range projections many recreation activities or products follow trends similar to the product life cycle.

Using only three data points of the dependent variable, the four functions (curves) fit the trend similarly but produce quite different long-range forecasts. It is obvious from Figure 4 that linear and exponential trends, being unbounded (without lower or upper limits to values), should not be projected too far in the future otherwise results will be unrealistic. It is also essential that as many data points as possible be used with simple trend projections.

More complex time series models, widely used in business forecasting, are seldom used in recreation forecasting primarily because of the cost of the refined data such models require (Wheelwright and Makridakis 1980). Although more complex, these techniques are frequently more appropriate or more powerful than simple trend projection.

Other Time Series Models

Time series models like simple trend projection, moving averages, and exponential smoothing, all make use of historical data.

Moving Averages

The moving averages model estimates the next period's value $\overline{V}_{t+1}$ as a linear function of n previous values. Mathematically, this model is expressed as:

(2) $$\overline{V}_{t+1} = (V_t + V_{t-1} + ... + V_{t-n+1})/n$$

where n is the number of periods in the moving average; for example, three, four or five years. The moving averages model is effective in smoothing out random fluctuations in a data series and thus providing stable estimates. However, this model is not without its problems. Increasing the number of periods averaged (the size of n) does smooth out random fluctuations better, but it makes the model less sensitive to real changes in the data. Also, moving averages cannot detect trends very well. Since they are averages, they will always remain within past levels and will not be very sensitive to a change to either a higher or lower level (Render and Stair 1988).

Exponential Smoothing

The exponential smoothing model is a type of moving average technique. The basic exponential smoothing formula can be expressed as:

$$(3) \qquad \overline{V}_{t+1} = \alpha V_t + \alpha(1-\alpha)V_{t-1} + \alpha(1-\alpha)^2 V_{t-2} + \ldots$$

$$(4) \qquad = \alpha V_t + (1-\alpha)\overline{V}_t$$

where α is a weight (or smoothing constant) that can assume values between zero and one inclusive. The concept is relatively simple. The smoothing constant, α, can be changed so that it is very responsive to recent data, when it is high, or to past data, when it is low.

The overall objective is to obtain the most accurate forecast. Exponential smoothing is more accurate than simply adjusting last year's figure by five or ten percent, particularly if last year was atypical. It should be noted that the choice of α is subjective and that as with any moving average technique, exponential smoothing fails to respond to seasonal or cyclical variations.

Seasonal Decomposition Techniques

Unlike traditional manufacturing/processing industries, the tourism and recreation industries are seasonal in North America. This seasonality is more a function of climate than any other factor. When a time series of data contains a seasonal pattern, it may be useful to separate the annual trend from the seasonal pattern to have a better idea of the direction of the annual trend. Wheelwright and Makridakis (1987) describe simple algebraic methods for decomposing seasonal data series. Such techniques can be applied to recreation participation or park visitation modelling.

There are a number of more complex forecasting techniques that are used in tourism forecasting but not in the recreation field. Techniques like the Box-Jenkins (ARIMA) model and others are discussed in Wheelwright and Makridakis (1987).

CAUSAL MODELS

It is assumed with time series models, that relationships do not change over time. Forecasts can generally be more accurate and more useful if planners understand the forces (causal factors) which underlie them (Stynes 1983). Both the multiple regression models and the trip distribution models reviewed in this section are causal models because they explicitly attempt to quantify the relationship between the dependent variable and a set of causal variables. In a causal model, park attendance in a given year is related to a set of independent variables that might explain or be correlated with attendance. Although causal models often describe a relationship rather than a cause-effect linkage, the literature on statistical modeling refers to them predominantly as causal models.

The general form of these models is:

(5) $$V_t = f(P_n, S_n, D_n) \text{ for } n = 1, 2, 3, \ldots$$

where V_t is an estimate of the number of visits to a specific park in period t, and P, S, D are sets of independent variables of n elements..

Socioeconomic variables such as population, average annual income, employment, etc. are represented by the Ps. Supply variables or measures of the quantity and quality of recreation opportunities are represented by the Ss. Most models use physical inventory data of site/facility attributes (for example, acres of parks, water surface area, number of trails). The Ds can be thought of as barriers or constraints to travel such as distance, costs or time.

The application of causal models to forecast recreation travel behaviour follows the model building process outlined above.

Multiple Regression Models

The most commonly used causal models are multiple regression models. Indeed, multiple regression models have been applied to both tourism demand modelling (Little 1980; Jujii and Mak 1980, Quayson and Var 1982; Uysal and Crompton 1985) and recreation travel modelling (Cheung 1972; Johnson and Suits 1983).

Unlike simple linear regression analysis which uses only one explanatory (independent) variable, multiple regression analysis hypothesizes that the dependent variable is a function of many explanatory variables. The modeler formulates a regression equation with the desired combination of variables. The fitting of the model to the observed data set involves estimating the values of the coefficients.

The most widespread estimation method is the ordinary least square method (OLS) but other methods exist; a discussion of the assumptions made with the various estimation methods is found in Smith and Munley (1978). The popularity of the OLS method can be attributed to the large number of computer programs which emphasize it (Iman and Conover 1983). In cases where the number of independent variables is large, the complexity of the model becomes unwieldly. To resolve this problem a procedure called stepwise regression is used to test the significance (importance) of each independent variable. Independent variables are added in order of significance (i.e. their degree of influence on dependent variables), taking into account the influence of the variables already added. The unimportant or non-significant variables are left out of the model. This procedure is explained in most intermediate statistics textbooks (Draper and Smith 1981).

In recreation travel forecasting, a typical multiple regression model takes the general form of equation 5 above.

For example, Cheung (1972) developed a recreation travel model for Saskatchewan which took the following form:

(6) $$V_{ij} = C_o + (C_1P_i + C_2P_iA_i + C_3T_j + C_4)/\, g(D_{ij})$$

Where:

V_{ij} is the number of vehicles in hundreds estimated to be travelling to park j from observation unit i per season,

P_i is population of observation unit i (in thousands),

A_i is alternative factor for observation unit i,

T_j is attractiveness of park j,

D_{ij} is road distance in miles from the largest population centre in observation unit i to park j

C_o, ...C_4 are parameters to be estimated, and $g(D_{ij})$ is a distance function. Cheung defined $g(D_{ij})$in a separate equation as either equivalent to D_{ij}, $D_{ij}/2$ or $D_{ij}^{3/2}$ depending on the origin-park distance.

A_i represents alternative choices for origin i. Cheung used distance as the most important factor. Thus the alternative factor A_i defined as:

$A_i = \Sigma\ 1/D_{ij}^{1/2}$

T_j in the attractiveness function is defined as:

$T_j = \Sigma\ S_e\ \Sigma\ R_m\ Q_m$

where

T_j = attractiveness of park j

S_e = relative popularity rating of activity e

R_m = relative importance rating of facility m, and

Q_m = score of facility m, according to its quantity or quality

The stepwise regression technique was applied to estimate the coefficients. The regression equation obtained was as follows:

(7) $V_{ij} = 1.33 + (120.31\ P_i - 36.60\ P_iA_i + 12.25\ T_j - 104.56) /\ g(D_{ij})$

The model explained a reasonably large percentage (91%) of the variation in the dependent variable, and the regression coefficients were all significant at the one percent probability level.

Multiple regression models have been particularly popular among tourism and outdoor recreation planners. This popularity is due to the ease and low cost of development and execution and the case of interpretation offered by these models (Makridakis and Wheelwright 1979). Their most obvious drawback, however, is the need to estimate the future values of the causal factors.

Gravity and Spatial Interaction Models

The recreation forecasting field has been dominated by the use of linear regression and gravity models in some cases without a clear understanding of the assumptions which underlie the mathematical structures of these models or their suitability in a given application. It seems that modellers have given precedence to the most readily available computer routines to estimate models at the expense of the validity of the assumptions relative to the problem at hand (Stynes and Peterson 1984).

As mentioned in the previous section, multiple regression models (with linear functions) have dominated the recreation field. Gravity models, however, have been a close second. Gravity models are similar in form to regression models except that greater emphasis is put on the effects of distance or travel time as a constraint to recreation travel. These models are called "gravity" because of the similarity of their structure to Newton's law of gravitation. Gravity models have a conceptually and technically different formulation than regression models. In contrast to regression models which are always estimated statistically, gravity models are usually "calibrated" by trial and error procedures (Cesario 1969). Most simple gravity models take the form of equation 8 below and are unconstrained power functions.

$$T_{ij} = P_i \, A_j \, f(C_{ij}) \tag{8}$$

where:

T_{ij} is the number of trips from origin i to destination j

P_i is the population of origin i

A_j is the attractiveness of destination j. A_j is usually a surrogate measure like lake area, the number of recreation activities possible at the site, etc.

C_{ij} is a measure of travel cost from i to j, or some surrogate of cost such as distance.

The recreation literature contains many examples of applications of gravity models (Wilson 1971). Saunders, Senter and Jarvis (1981) developed a gravity model to predict day-use recreation participation in the Upper Savannah River Basin, using regional participation rates, population projections and distance to recreation sites.

Their forecasting procedure consisted of two stages: first, the prediction of the total number of potential participants for an activity from each population source, and second, the allocation of the potential participants to the various recreation sites according to travel time and site capacity. The total number of potential participants to any activity was based on the projected population for each centre, the participation rate for each activity, and a factor accounting for the higher participation in urban centres as opposed to rural centres. They concluded that relatively simple gravity models can provide a "useful level of detailed information to planners in a short period of time, using available data sources, at very low cost to the department".

Although quite useful in some cases, simple gravity models like the ones above have a number of weaknesses. Ewing (1980) criticizes the simple gravity model in equation 8 as being too simplistic: without modification, it could not account for changes in the number of trips to existing destinations caused by the addition of new, competing destinations i.e. the "supply-generated-participation effect". Some British transportation planners (Batty and Mackie, 1972) have incorporated this "effect" in very sophisticated models. Ewing suggests (1) to the need to look at this "effect" in developing gravity models; (2) performing a sensitivity analysis to determine how substantially parameters change under changing conditions; and (3) the need for greater understanding of tourist perceptions (for example, rather than using the researcher's estimates on the importance of site physical attributes).

Ewing's criticism of simple gravity models has inspired many researchers in using the logistic curve and other refinements in the development of "improved gravity models" (Peterson, Anderson and Lime 1982; Peterson, Dwyer and Darragh 1983; Ewing 1983; Fesenmaier 1988). The properties of the logistic function were described previously in this chapter (page 14 and Figure 4).

Indeed, in addition to being well suited to processes which have start-up impediments and saturation effects, the logistic function has two properties which make it particularly well-suited to modelling discrete choices. The first property is its interval being restricted to values between 0 and 1. This makes the logistic function suitable as a probability function. The second property is that the logistic function can be transformed to a convenient linear form by using a natural logarithm.

Two primary examples of discrete choice are the choice of a recreation activity and choice of a recreation site (Fesenmaier 1988). In both instances, individuals select from a finite set of mutually exclusive alternatives.

A **spatial interaction model** of recreation travel usually consists of two components: a trip generation component and a trip distribution component. The trip generation component estimates the number of trips generated by a province, region or city as a function of its population and the quality, quantity and accessibility of the destinations to visit. The trip distribution component allocates those trips among alternative destinations. While the trip generation component estimates a total number of travellers from an origin going to a number of destinations, the trip distribution component estimates the probability of an individual or household at an origin choosing a specific destination (Ewing 1980; Peterson et al. 1983).

Such a spatial interaction model (sometimes called a complete model) can be expressed either as a logistic regression model or a multinominal logit (MNL) model:

a) Logistic regression model

(9) $T_{ij} = (T_{i\cdot})(P_{ij})$

(10) $= (T_{i\cdot})(e^{u} / 1+e^{u})$

= Trip generation coefficient X Trip distribution coefficient

b) Multinominal logit model

(11) $T_{ij} = (T_{i\cdot})(P_{ij})$

(12) $= (Ti.)(e^{Vij} / \Sigma\ e^{Vij)}$

= Trip generation component X Trip distribution component where:

T_{ij} is the number of trips from origin i to destination j,
$T_{i\cdot}$ is the total number of trips generated at origin i,
P_{ij} is the probability that an individual at origin i will choose destination j,
u is a linear function of park attributes in the form of

$$u = a + b_1x_1 + b_2x_2 + ... + b_nx_n, \tag{13}$$

V_{ij} is the utility derived from a trip from origin i to destination j,
e is the neperian number equal to 2.71828 ...

Logistic regression and multinominal logit (MNL) models, like linear or gravity (power) models, involve certain assumptions which may be more or less realistic to a particular problem. Fesenmaier (1985) lists the assumptions made by most recreation modellers who use the MNL and states that a number of those assumptions are often violated, particularly the assumption that a typical recreationist will repeatedly choose the same destination if faced with the same set of opportunities. It is difficult for any model, however, to simulate a choice based on the need for obtaining variety in someone's recreational experiences, that is, the diversification behaviour .

Linear and gravity (power models) make different assumptions about lower and upper limits, the rates of change of the dependent variable (slope) and elasticities. It is clear, however, that the functional form of the logistic regression and MNL models which use the logistic function have theoretical merits over the linear or power functions. This can be observed from Figure 4 and was elucidated on page 20.

Logistic regression and MNL models are nonlinear and the maximum likelihood method frequently used to estimate them leads to measures of goodness-of-fit that are more complex to interpret than those of linear models. It is argued, however, by Stynes and Peterson

(1984) that for many recreation travel applications, the properties of the functional form are more important than measures of goodness-of-fit.

Logistic regression and MNL models have a number of shortcomings. Some of these shortcomings are, however, shared with other quantitative models. For example, in a MNL recreation site choice model, a significant amount of error is inevitable if forecasts are made beyond 1 or 2 years, the weight or importance assigned to attributes (attractiveness, distance, etc.) may not be independent of the alternatives in the choice set (Stynes and Peterson 1984). Another shortcoming is the inability of the MNL model to account for intrapersonal variation in destination choice (diversification behaviour) i.e. not accounting for the possibility that a household (or day-use party) may not make the same choice of a park on repeated occasions (Fesenmaier 1985). Unlike the MNL model, which is normally used with data representing individual choices, the logistic regression model is most frequently used with aggregated data and thus models a distribution rather than individual choices.

Despite these shortcomings, logistic regression and MNL models provide an adequate structure for dealing with recreation travel problems such as the substitution of alternative sites and the segmenting of recreation markets.

Qualitative Approaches

Qualitative approaches are used to collect the pooled opinions of experts on the likely outcome of future events. Qualitative approaches are particularly appropriate where past data are unavailable or insufficient and where non-quantitative variables are expected to have an impact on tourism or recreation travel, for example, an increase in leisure time, changes in recreation activities etc.

The qualitative approaches that have been proposed since the 1960s include the Delphi technique (Dalkey and Helmer 1963), cross-impact analysis (Helmer 1979), the nominal group technique (Van de Ven 1974), the Gearing-Swart-Var (GSV) technique (Gearing et al 1976) and scenario writing (Bar On 1979). In this report only the most widely used qualitative techniques in the tourism and recreation fields will be described namely the Delphi technique and scenario writing.

The Delphi Technique

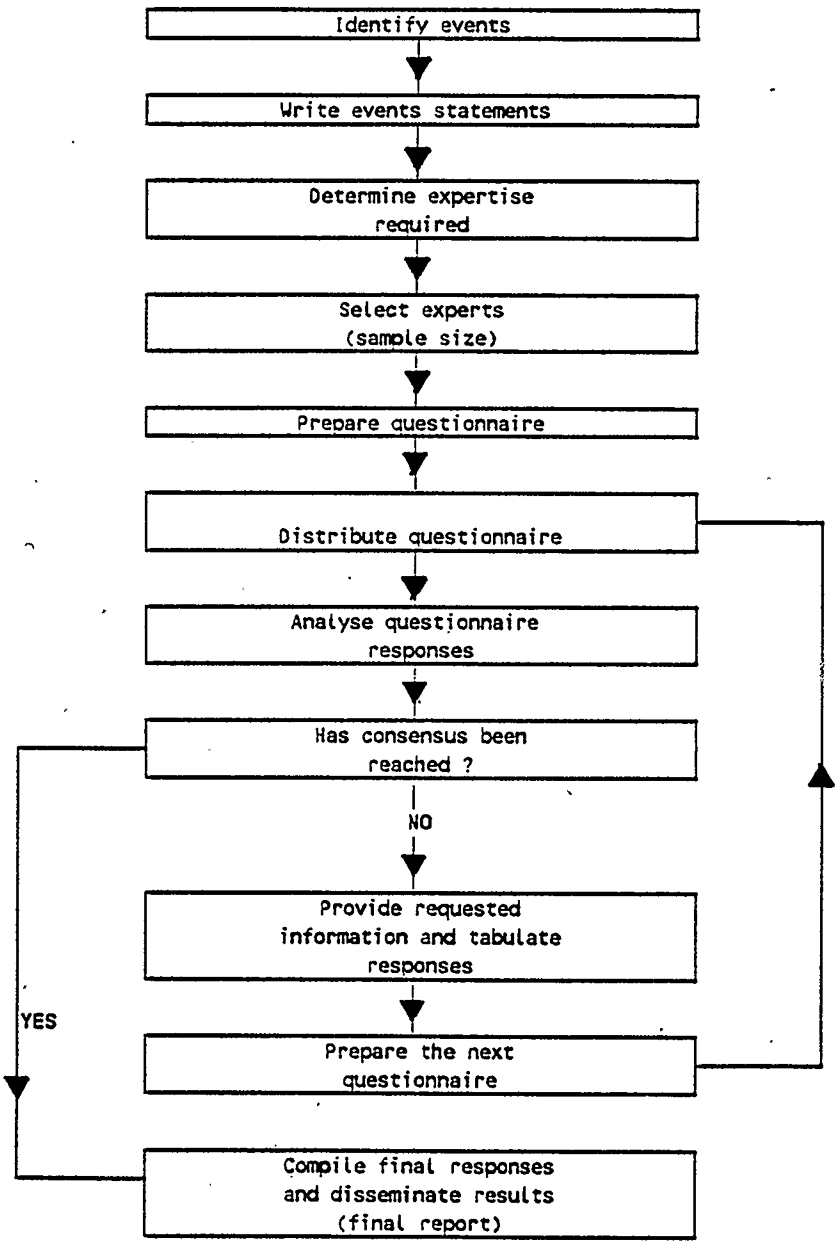
The purpose of the Delphi technique is to combine expert knowledge and opinion to arrive at a group consensus about the probability of occurrence of future events. The forecast time horizon for the Delphi technique is usually long (Stynes 1983, Green et al 1990). Instead of the traditional approach of brainstorming, the Delphi technique aims at developing consensus opinions among experts without direct confrontation or group pressure (Dalkey and Helmer 1963). Experts who take part in the Delphi process remain anonymous.

A questionnaire is administered to the panel of experts asking them for the probability of occurrence of certain events, to edit/comment on events that are ambiguous and to add events that should be included on the list. Their responses are collected and summarized. Each of the respondents is provided feedback from the other experts and then a second questionnaire is given out. The procedure is repeated until a consensus is reached (Figure 5). "Consensus" simply means the point at which the distribution of responses begins to stabilize. A good indicator of panel consensus is the spread of the interquartile range: a wide range indicates panel disagreement, a narrow one indicates panel consensus (Moeller and Shafer 1983).

The Delphi technique was used to gather expert opinions on the future impacts of tourism in Nova Scotia by the year 2000. The information generated was considered as an effective policy-making tool in solving tourism planning and management problems (Kaynak and Macauley 1984). The Delphi technique has, however, been more successful in predicting technological change than human behaviour.

FIGURE 5 - THE DELPHI TECHNIQUE

(Adapted from Teraine and Riggs, 1976)



The main advantage of the Delphi technique is its simplicity and its ability to provide useful future perspectives or estimates when there is simply no other technique available. The Delphi technique, however, has its weaknesses. It is difficult to identify panel experts and evaluate their expertise. They may all share misconceptions or miss trends. Experts, being busy people, may not wish to participate on a Delphi panel for an extended period of time. Some panel members may drop out before the results are fully evaluated. The director of the panel may have a strong influence on the results by editing the panel responses. Although quite simple, the Delphi technique may require more time, effort, and can be more costly than other forecasting techniques (Moeller and Shafer 1983).

Scenario Writing

Scenarios are realisable future images or states. Often part of strategic planning, scenario writing attempts to show how a particular future state or a set of alternative future states could eventually be achieved, given the current situation as the starting point. For example, an optimistic scenario could be written as: "Continued strong economic growth to 1992". Such a scenario could be achieved through a stable unemployment rate of 4.5% and a stable inflation rate of 5% per year. The modeler would then project the effects of this optimistic scenario on park visitation.

Van Doorn (1984) defines a thoroughly written scenario as one that contains at least three components: (1) a description of the current situation (baseline analysis); (2) at least one future image; i.e., a description of a future state; and (3) for each future image, at least one future path which indicates how the current situation could develop into the eventual future image.

One of the best examples of scenario writing using the required three components mentioned above is that of Bar On (1983) who used the approach to forecast short- to medium- term tourist arrivals to Israel. Three scenarios (optimistic, intermediate and pessimistic) were prepared for a time horizon of twelve to eighteen months. A local example of scenario writing can be found in Jamieson et al (1988) where four scenarios were developed as part of the Strategic Options Workshop for the Crowsnest Pass community of Alberta.

Scenario writing could become a very useful tool in tourism and recreation travel forecasting particularly through the use of mini-scenarios for small regions.

FUTURE DIRECTIONS FOR RESEARCH IN RECREATION FORECASTING

In any tourism or recreation organization, improvements to the forecasting process require improvements in the related components: namely, data collection and management, model development, and decision making. The main concern of model developers is the sparsity and often poor quality of data bases for forecasting. The first step toward improving data bases is the recognition of the importance and role of forecasting within the organization. It is only through this realization that appropriate investments in resources, time, and personnel will be made to improve data bases.

No dramatic breakthroughs have occurred in new technique development in the past several years. Most research efforts have been centred on refinements to existing models and on improving forecast accuracy through combined forecasts. A few researchers have written on the process of selecting a technique using as evaluation criteria: costs, expertise required and other factors (Georgoff and Murdick 1986).

Both time series and causal models specifications are continually being improved. Researchers have reported that traditional socioeconomic variables explain only a small portion of variation in recreation participation (Seneca and Davis 1976; Stynes 1983). It is essential then to include variables that better explain recreation behaviour. Recent refinements in time series and causal models (particularly trip distribution models) are leading to comprehensive and complex models which take into account more factors of change in recreation behaviour (Oliveira et al 1983; Lin et al 1988). Researchers are exploring how to include different market segments, family structures and stages of the life-cycle in recreation forecasting models. The refined models will have a broader variety of applications.

Recreation modellers can improve their forecasts in the following ways: combining forecasts, performing a sensitivity analysis and selectively applying judgment. According to Georgoff and Murdick (1986), the research on combining forecasts to improve accuracy is "extensive, persuasive, and consistent". Many researchers agree that the results of combined forecasts greatly surpass most individual projections, techniques, and expert analysis (Mahmoud 1984; Zarnowitz 1986; Calantone et al 1988). The greater accuracy of combined forecasts will be emphasized in the next section on the evaluation of the various techniques.

With the proliferation of personal computers and the advent of the electronic spreadsheet, it is easy for the forecaster to perform a sensitivity analysis on a model. Sensitivity analysis consists of determining the range of probable outcomes (forecasts) by changing the combination and values of input variables to the model. Through sensitivity analysis the modeler can determine what variables are most critical and thus deserve more accurate measurement, the range and distribution of expected outcomes, and the effects of different assumptions on the outcomes.

Forecasting should not be judged strictly on accuracy but also on its ability to improve decision making within the organization. When it is expected that the values of independent (explanatory) variables may change significantly from established patterns, it is advisable that the forecaster apply a judgmental adjustment to improve the forecast. What is also needed are clearer guidelines about when and under what circumstances one technique is to be preferred over the others.

EVALUATION OF FORECASTING TECHNIQUES

Tourism and recreation departments must make a variety of forecasts as part of planning and management. The selection of a particular forecasting approach is a difficult decision. Each forecasting approach has strengths and weaknesses. Every forecasting situation is limited by constraints like expertise, time, funds, and data. The greater the forecaster's familiarity with the strength and weaknesses of both quantitative and qualitative approaches, the better his/her ability to select the "best" technique for a given forecasting situation.

The evaluation of the accuracy of simple or combined techniques constitute a major research project of its own. Research of this type has recently been completed and reported (Wheelwright and Makridakis 1987; Mahmoud 1984; Zarnowitz 1983; Makridakis et al 1982). Obviously extensive accuracy evaluations and comparisons by the author are beyond the scope of this project.

The objective of this section is to comment on the meaning and/or significance of common evaluation criteria, to summarize the results of the evaluation of the main approaches/techniques, and to discuss the recent emphasis given in practice to combined forecasts. The information from this section will serve as the basis for the selection of the most appropriate technique or combination of techniques consistent with the objectives of this project.

Georgoff and Murdick (1986) identified sixteen evaluation criteria for the selection of a forecasting approach/technique. Only the five most important evaluation criteria at the top of their list will be considered here: time horizon, complexity, cost, data availability, and accuracy.

Time horizon. Forecast results can be extended as far into the future as needed. Different approaches have varying abilities to accommodate different time horizons. Normally, the longer the time horizon the greater the complexity of the forecasting technique, the greater the cost and time required to produce the final results, and the less accurate are the results.

Complexity. Complex forecasting techniques require a greater level of technical sophistication on the part of the forecaster. Although training in the use of microcomputers and statistical packages is becoming widespread in universities, not all planners or managers have sharpened their quantitative skills enough to be comfortable with some of the forecast results generated by a computer.

Cost. There are three main components of cost: data acquisition costs, development costs and running costs. These costs are: time, resources, and personnel. Data acquisition costs may involve field inventories, social surveys/questionnaires or literature reviews. Development costs are more important at the beginning when the technique is adapted and installed. Running costs are those required for the operation and maintenance of the forecasting process.

Data availability. Previously mentioned research studies on forecasting accuracy have revealed that more data tend to improve accuracy, and disaggregate (more detailed) data are more valuable than aggregate data for accuracy purposes. In choosing a forecasting technique, the forecaster must consider the "extensiveness, currency, accuracy, and representativeness of the available data" (Georgoff and Murdick 1986).

Accuracy. The maximum accuracy one can expect from a technique must be within a range of values bounded by the average percentage error of the random component of the data series. Accuracy is an important criterion but the forecaster may wish to forgo some accuracy in favour of other advantages like a more meaningful time horizon, the cost of the sampling method or the ease of use. Importance given to accuracy must therefore bear in mind the control the forecaster has over the predicted outcome, the constraints of time and resources imposed on him/her, and the tradeoff between accuracy and other advantages.

Many of the strengths and weaknesses of the various forecasting approaches have been mentioned in the previous section. It would seem appropriate at this stage to summarize the results of the evaluation of these approaches using the above mentioned criteria (except data availability). Table 1 summarizes the evaluation results of forecasting approaches/techniques of two major studies: Makridakis and Wheelwright (1979), and Mahmoud (1984).

A perusal of Table 1 leads to the following key observations:

1) Overall, quantitative techniques outperform qualitative techniques, particularly when data is readily available.

2) Simple forecasting techniques perform more accurately than, or at least as accurately as, complex techniques. They are also less costly, easier to learn and use. For example, exponential smoothing usually outperforms in accuracy the more complex Box Jenkins techniques.

3) Time series and simple linear regression are more accurate when used to forecast short time horizons (less than a year). Researchers found that exponential smoothing had the lowest errors when forecasting up to four periods, and that the simple moving averages technique had the lowest cumulative errors for all twelve forecast periods (Makridakis and Hibon 1979).

4) Multiple regression models are accurate when used to forecast from one to three years, that is, the medium term horizon. The main advantage of multiple regression models is that the modeler can simulate the effects of policy changes on the outcome. These models, however, are more difficult to develop

and require more data than time series models. There is also the persistent problem of forecasting the changes of independent (explanatory) variables first in order for the model to be useful.

5) Forecasting accuracy can be improved by combining techniques. Indeed by carefully combining two or more complementary techniques, the forecaster can offset the limitations of one technique with the advantages of another all the while retaining the strengths of the first.

The majority of tourism and recreation forecasting studies mentioned in the literature employ time series or regression techniques. It is only recently that greater interest has been given to combining the results of two or more simple forecasting models. For example, Calantone et al (1988) combined econometric (regression) models with time series (Box-Jenkins ARIMA) models. Their combined model produced "more accurate and more (managerially) useful forecasts than any one single method both in predictive power and accuracy as well as usefulness as a diagnostic (explanatory) tool".

Van Doorn (1984) suggests very practical combinations, for example: short-term time series or medium-term regression with longer-term forecasting techniques such as the Delphi technique or scenario writing. An example of combined statistical time series with scenario writing is that of Edgell and Seely (1980). There is a growing interest in the use of combined forecasts.

TABLE 1
A COMPARISON OF FORECASTING METHODS

METHOD	TIME HORIZON	COSTS	COMPLEXITY	ACCURACY
Straight line extrapolation	Short	Very low	Minimal quantitative capabilities required.	Limited practical level of accuracy
Time series extrapolation	Short to long	Minimal if data are available	Minimal quantitative capabilities required.	Normally accurate for trends
Moving averages	Short to long	Minimal if data are available	Minimal quantitative capabilities required.	Similar to regression models.
Exponential smoothing	Short to medium	Minimal if data are available	Minimal quantitative capabilities required.	Generally high accuracy for short term forecasts.
Regression models	Short to medium	Moderate	Fundamental quantitative skills required.	Can be accurate if the variable relationships are stable and the proportion of explained variance is high.
Delphi technique	Usually long	Moderately high	Minimal quantitative capabilities required, but it is difficult to learn the technique and interpret the results.	Not particularly accurate, but usually most accurate in the long term and when conditions are dynamic.

(Adapted from Georgoff and Murdick, 1986; Makridakis and Wheelwright, 1979)

GENERAL OBSERVATIONS

In this chapter general principles of models and the modelling process were discussed, both quantitative and qualitative approaches were described, the assumptions underlying the most common techniques were given, and the techniques were evaluated based upon five key criteria.

In all quantitative techniques, a certain amount of error is inevitable, particularly for long-term forecasts. There are two types of errors: the error in estimating the relationships among variables and errors in forecasting the independent variable(s). It was argued that simple quantitative techniques usually produce forecasts as accurate as, and at times more accurate than, the more complex and more expensive techniques. If the explanatory and predictive abilities of two models are about the same, the simpler technique should be chosen because data collection, parameter estimation and model interpretation will be easier.

Whatever quantitative model is selected, the model should be validated using data series other than the one used to estimate its parameters. Validation and sensitivity analysis are important to ensure that the parameter estimates are stable and that the model is not biased.

Qualitative techniques like Delphi or scenario writing do not produce accurate forecasts of long-term future events but only a set of subjective probabilities of occurrence of such events. Certainly qualitative approaches could complement the results produced by quantitative ones particularly for quantitative projects beyond a two-year time period. This view is shared by an increasing number of forecasters as evidenced by the more frequent use of multimethod approaches.

CHAPTER 2

THE SELECTION OF A MEDIUM TERM FORECASTING TECHNIQUE

The author agrees with Calantone et al (1988), that a multimethod approach produces "more accurate and more (managerially) useful forecasts than any one single method both in predictive power and accuracy as well as usefulness as a diagnostic (explanatory) tool". The author agrees with these researchers and Van Doorn (1984) and therefore suggests the use of a multimethod (combined) approach for forecasting recreation travel demand for the Northern Region of Alberta. This combined forecasts approach includes medium term (1 year) forecasts of regional attendance to parks (time series), medium term forecasts of specific park attendance (logistic regression), and long term forecasts of future states of recreation travel trends and park attendance (scenario writing).

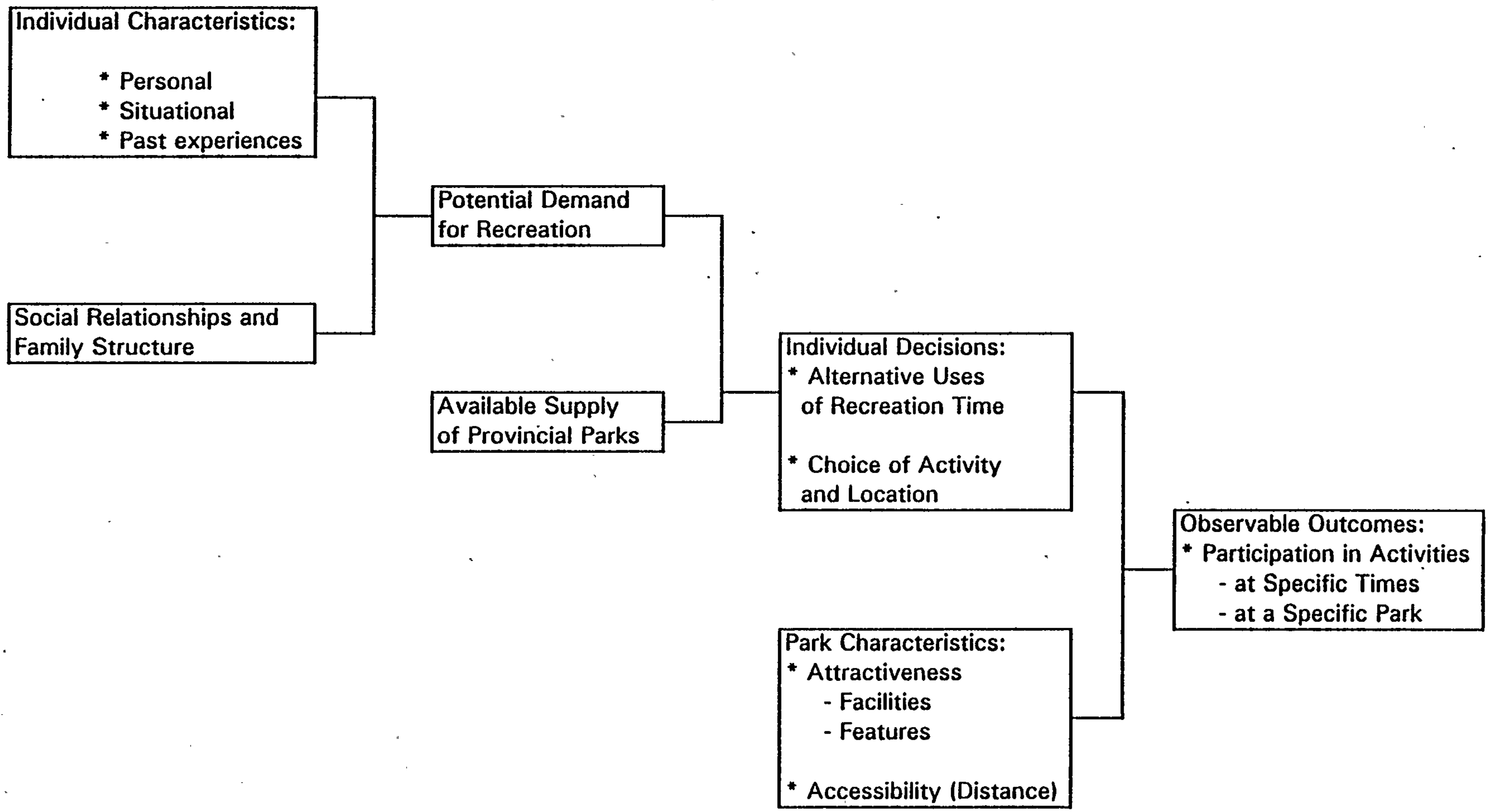
In this project, however, the emphasis will be put only on **medium term forecasting techniques** because resource constraints and the agreed upon scope of this project exclude the demonstration of long term scenario writing (or delphi). Long term scenario writing or the delphi require much more data, time and the involvement of many people.

Day-Use Park Attendance Forecasting Model

Conceptual Model of Recreation Choice

A person's choice of a recreational activity and the park where to participate in it depends on the person's preferences and his/her social relationships. This complex decision process is represented in an idealized (simplified) manner in the conceptual model of Figure 6. Since a model is a simplification of reality, only the quantitative components "attractiveness" and "accessibility" of a park will be considered in this project. In this model, the choice of a destination (i.e. a provincial park) is the sole dependent variable. It is assumed that the

FIGURE 6
ALBERTA PARKS RECREATION CHOICE MODEL



recreationist simultaneously selects an activity (swimming, picnicking, hiking etc.) and the park at which the activity will take place. Recreationists usually perform more than one activity during a park visit. (Fesenmaier 1988)

Model Assumptions

1) The household possesses adequate knowledge in order to make an "informed" judgement about alternative parks.

2) An increase in the attractiveness of a park or its accessibility (or decrease in distance) will result in an increase in the number of party visits generated from an origin.

3) A typical household will repeatedly choose the same park if faced with the same set of alternatives. At any time, the probability distribution of the number of visitors choosing the same park will be the same given the same choice of alternatives.

4) The greater the number and variety of facilities in a park, the higher the attractiveness of the park.

Model description

This is a **"complete" spatial allocation model** because it contains both a trip generation coefficient and a trip distribution coefficient. Figure 7 illustrates both these coefficients and the push and pull factors attached to them. The model is expressed as:

(14) $$T_{ij} = T_i P_{ij}$$

Where T_{ij} is the number of day-use party visits from origin i to park j, T_i and P_{ij} are the trip generation and trip distribution coefficients respectively.

The trip generation coefficient T_i is the number of **potential** visits from an origin i to all parks. T_i is a function of push factors, that is, factors that draw potential visitors to leave their origins to visit parks.

Mathematically, the trip generation coefficient T_i (for each origin i) is expressed as:

(15) $T_i = (\text{Total})\ P_i$

and

(16) $P_i = \text{Prob1}_i \ / \ \Sigma\ \text{Prob1}_i$

and

(17) $\text{Prob1}_i = f(\text{Pop, Urban, Att/Dist, Att/Dist}^2\text{, 1/Dist, 1/Dist}^2)$

The variable Total is the total number of day-use park visits for all the parks in the region. The value of Total is obtained through the time series technique called **exponential smoothing** as explained in Chapter 1 (page 16). One could define P_i as the "market share" of park visits ascribed to an origin (i). As will become evident later, Prob1_i is obtained through another technique (logistic regression), hence the forecasting is said to be a **combined approach.**

In Prob1_i, the independent variable Att/Dist (or its variations) is the **sum** of the ratio of attractiveness/distance of each park to an origin. Att/Dist^2, which is the sum of the ratios of attractiveness/(distance squared) from an origin to a park, is used to account for the possibility that a doubling of the distance may cause the pull of the park to drop by a quarter. There are, therefore, 28 values for this variable or one for each origin. Att/Dist shows the "regional" pull of the parks on an origin. The calculation of the "attractiveness" attribute of each park is explained further in this section (see page 50). The distance between an origin and a park is expressed in kilometres.

The independent variable Pop represents the population of the town of origin. It is assumed that the larger the population of the town, the greater the number of day-use visits originating from it. The dummy variable Urban is given a value of 1 if the population of the origin is greater than 100,000. Only Edmonton qualified for Urban = 1. The variable Urban was used to reflect the relationship between relatively large populations and park visits. The coefficients of the independent variables in $Prob1_i$ are estimated using the data for the whole region.

The trip distribution coefficient P_{ij} allocates the number of potential visits (T_i) generated from an origin to a **specific** park based on the probability that such visits can take place. P_{ij} is, therefore, the probability of visits from an origin i being made to park j. P_{ij} is based on pull factors (Figure 7), that is, factors that show (figuratively) how parks attract visits to them. In P_{ij}, however, there are some push factors: distance and competition. This probability is a function expressed as:

(17) $P_{ij} = Prob2_{ij} \; / \; \Sigma_j \, Prob2_{ij}$

and

(18) $Prob2_{ij} = f(Dist, Dist^2, Att, Att/Dist,$

$\quad\quad ...\; Att/Dist^2, Att^2/Dist, Compet)$.

Unlike the trip generation coefficient, the independent variables Attract and Dist are park specific, and hence, are not summations but the actual attractiveness of each park and the actual distance from an origin to a park.

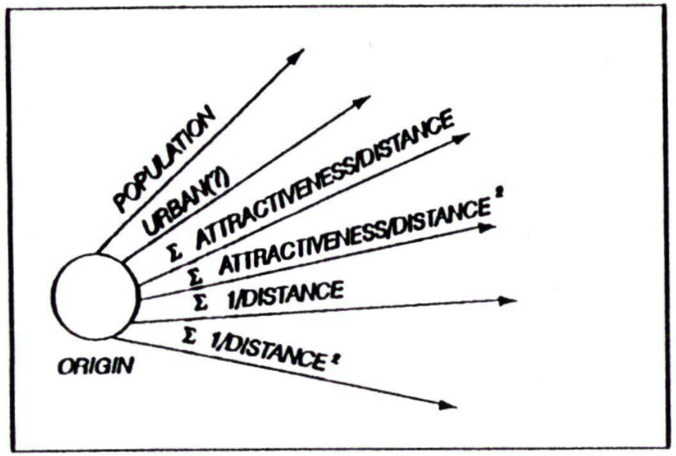
The independent variable Compet is the competition factor. This factor is calculated for each origin-park pair to account for the competition of other parks relative to that same origin. The competition factor is the sum of the ratios of Attract/Dist of the remaining 12 parks. The higher the competition factor for an origin-park pair, the greater the competition of the other parks relative to that origin.

FIGURE 7 - LOGISTIC REGRESSION MODEL

$$T_{ij} = T_i \, P_{ij}$$

DRAWN AWAY FROM AN ORIGIN

FACTORS IN THE TRIP GENERATION COEFFICIENT



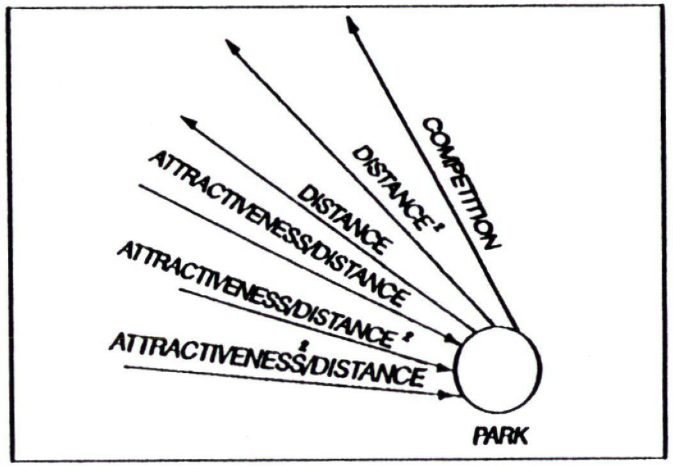
U_1

$$T_i = TOTAL \; P_i$$

$$T_i = TOTAL \; \frac{\left(\frac{e^{U_1}}{1 + e^{U_1}}\right)}{\Sigma \left(\frac{e^{U_1}}{1 + e^{U_1}}\right)}$$

DRAWN TO A PARK

FACTORS IN THE TRIP DISTRIBUTION COEFFICIENT



U_2

$$P_{ij} = \frac{\frac{e^{U_2}}{1 + e^{U_2}}}{\Sigma \left(\frac{e^{U_2}}{1 + e^{U_2}}\right)}$$

This model is standardized, that is, the sum of the probabilities equals one (1.0). Using the unit as the base reference, it is thus possible to compare the probabilities as if they were percentages. The continuous probability distribution of the logistic regression (LR) normally resembles the s-curve shown in Figure 4.

After "sketching" the design of the model above, it was necessary to choose the most "appropriate" **functional structure** (curve) for the probability coefficients P_i and P_{ij}. Among the alternatives were the linear, power, and logistic structures.

The advantages of choosing a logistic structure over the linear or power structures were discussed in Chapter 1 (page 14) and illustrated in Figure 3. Having chosen the logistic structure, there remained the choice of which **procedure** between the logistic regression (LR) and the multinomial logit (MNL). Many of the recently published recreation research papers involved the MNL procedure (Fesenmaier 1988). Although similar because of their use of the logistic curve, each of these procedures (LR and MNL) has a clear advantage depending on the application considered and the evaluation criteria chosen.

The MNL model is more complex than the LR model (see equation 12) and introduces a new assumption that may not be readily apparent (Stynes and Peterson 1984). This assumption, the Independence from Irrelevant Alternatives property (IIA) of the MNL constitutes both its major strength and weakness. The IIA property implies that the ratios of the probabilities for any two alternatives is independent of any other alternatives. In simpler terms, the relative probability (the log-odds) of selecting park A or park B is independent of the existence of parks C, D and E. It would seem that the IIA property would permit new parks (alternatives) to be easily added or changes in the physical characteristics of existing parks to be made without re-estimating the model. The weakness of the MNL is that the IIA property frequently yields counter-intuitive results.

A compromise solution that would reduce the problem of the breakdown of the IIA property is the use of a nested logit model. Nested choice models are based on a sequential set of choices. For example, an individual would first choose the park type (wilderness, primitive, developed) and then would choose the specific park. The provincial parks of Alberta are not differentiated in this manner and surveys of day-use visits are not sufficiently refined to include visitor perceptions about park types.

Stynes and Peterson (1984) advise that the MNL model be applied to recreation choices that are quite distinct. Depending on the degree of similarity (here the attractiveness attribute) between the alternative parks, the MNL may accurately predict the market share of the parks but may behave poorly in predicting the response to the addition, deletion or modification of parks.

As given in the Introduction, the objectives of this project include interpreting the implications of park policies on attendance. The policies to be modeled consist of the opening or closure of a park(s), and/or changes to the attractiveness attribute (Attract) of one or more parks.

Bearing in mind the above limitations (perceived or real) for policy analysis of the standard MNL model mentioned earlier, the limited refinement (if not roughness) of the survey data, the inherent advantages of a simpler model, and the desire to experiment with the design of a new model, the **logistic regression model** (LR) was chosen as the structure for forecasting park attendance.

The logistic regression model involves a binary choice represented by 1 and 0 (yes and no). One may state as an example, whether or not to visit a particular park. If we assume the probability P of visiting a park is a logistic function of the independent variables, the mathematical form is:

(18) $$P = e^u / 1+e^u$$

and the probability of <u>not</u> visiting is:

(19) $$1\text{-}p = 1 / 1+e^u$$

and

(20) $$p / 1\text{-}p = e^u$$

Taking the natural logarithms of both sides yields the logit L.

(21) $$L = \ln (p / 1\text{-}p) = u$$

where

(22) $$u = a + b_1x_1 + ... + b_nx_n$$

The logit or log-odds (L) as given in equation 21 is the natural logarithm of the ratio of the probability of visiting (P) and not visiting (1-P) a park. The independent variables $x_1 ... x_n$, in this project, consist of environmental and demographic factors which were described above with P_i and P_{ij}. The coefficients a, b_1 to b_n were estimated by the Stepwise Logistic Regression (LR) procedure of the BMDP statistical package (BMDP 1989). The logits u_1 and u_2 can be expressed as:

(23) $$u_1 = .000095673 \text{ Pop} - 51.009 \text{ Urban} + 5.7991 \text{ Att/Dist} - 44.431 \text{ Att/Dist}^2 - 166.070 \text{ (1/Dist)} + 1340.2 \text{ (1/Dist}^2\text{)} - 9.3042$$

(24) $$u_2 = - .032438 \text{ Dist} + .000033564 \text{ Dist}^2 - 1.1512000 \text{ Compet} + .054810 \text{ Attract} + .92572 \text{ Att/Dist} - 1.2876 \text{ Att/Dist}^2 - .0224260 \text{ Att}^2\text{/Dist} + 2.7927$$

The complete model in the **logistic regression structure** can be expressed as:

(25) $$T_{ij} = T_i \; P_{ij}$$

(26) $$= (\text{Total } P_i) \; P_{ij}$$

(27) $$= [\text{Total } (e^{u1}/1+e^{u1}) / \Sigma(e^{u1}/1+e^{u1})] \; (e^{u2}/1+e^{u2})/ \Sigma(e^{u2}/1+e^{u2})$$

Data Characteristics

The data collected for this medium term forecasting model pertain to 13 provincial parks and 28 towns of the Northern Region of Alberta as defined by Alberta Recreation and Parks (Figure 1 and Tables 2 and 10).

The data used for this medium term forecasting model include:

a) Monthly and fiscal year park attendance expressed as "day-use party visits"
b) Origins of visitors
c) Participation rates of major summer and winter recreation activities, that is, those for which facilities are provided by some parks.
d) Facilities and/or natural features associated with the main recreation activities.
e) Miscellaneous facilities and/or natural features associated with secondary summer/winter recreation activities.

The day-use attendance data reported by Alberta Recreation and Parks (1989) were collected through Automatic Counter Readings. Adjustments made to the vehicle counts include determining the ratio of day-use vehicles to all other vehicles entering and leaving the park. The adjusted vehicle counts and a mean party size are used to estimate the number of individual day-users. The adjustment factors are obtained from a day-use calibration survey conducted by Alberta Recreation and Parks every four years.

To obtain information on the origin of visitors, 13 random daily surveys are conducted throughout the May - September (peak season) period of each year. From these surveys, it is possible to determine the percentage of visitors from various origins and apply these percentages to the yearly attendence data and thus obtain an estimate of the allocation of total yearly visits to these origins (Tables 3 and 4).

TABLE 2

NORTHERN REGION PARK ATTENDANCE
(Day-use party visits)

			84-85	85-86	86-87	87-88	88-89
1	CA	Calling Lake	n/a	1887	2825	3325	3350
2	CP	Carson-Pegasus	14000	20487	18224	15825	17125
3	CL	Cross Lake	3900	5104	4653	4450	5400
4	HB	Hilliard's Bay	n/a	9350	12418	13625	14075
5	LS	Lesser Slave Lake	n/a	n/a	17240	23225	19875
6	ML	Moonshine Lake	9297	9273	10744	9950	8250
7	NO	Notikewin	n/a	2422	412	1225	1055
8	OB	O'Brien	13747	20006	18106	17850	14925
9	QE	Queen Elizabeth	7532	11134	10336	3375	3825
10	SI	Saskatoon Island	20092	1844	11622	13550	15500
11	WI	Williamson	2638	6341	2938	1150	5900
12	WN	Winagami	n/a	n/a	6129	7450	5250
13	YP	Young's Point	6251	6573	6748	5925	3450
		TOTALS	77457	94421	122395	120925	117980

TABLE 3

ESTIMATED DAY-USE ATTENDANCE PERCENTAGES (PARTY VISITS)
FISCAL YEAR 88-89

Origins/Parks	CA	CP	CL	HB	LS	ML	NO	OB	QE	SI	WI	WN	YP
Athabasca	39.5				0.3								0.3
Barrhead	0.5	1.4		0.2	0.4					0.2			
Beaverlodge				0.1		0.6		0.1		4.7	2.1	0.8	1.1
Edmonton	32.4	19.4	26.1	4.2	9.5	0.6	0.8	0.1	0.8	1.6	2.9	2.9	2.0
Fairview			4.3	0.2	0.1	1.4	0.8		4.3	0.6	0.4	0.8	0.3
Falher				1.8						0.1	0.8	9.5	1.4
Fort Saskatchewan					0.2	0.2						0.3	
Fox Creek					0.1					0.1	0.4		0.3
Gibbons		0.2											
Grande Prairie		0.2		0.7	0.6	10.9		92.7	0.4	63.8	18.1	0.8	46.8
Grimshaw		0.2		0.5	0.1	0.4	0.8	0.1	38.4	0.2		1.3	
High Prairie				52.0	0.8					0.3	0.8	24.6	0.3
Hines Creek				0.2						0.1			
Manning		0.2		0.4			42.1		1.2			0.8	
Mayerthorpe		0.9											
Mc Lennan				0.5							0.4	23.3	
Morinville	0.5				0.3							0.3	
Peace River				2.2	0.7	0.6	5.3		32.2		0.4	8.5	0.6
Sherwood Park	1.0	0.7		0.2	0.2								
Slave Lake	0.5			0.9	69.6					0.1		0.3	0.3
Spirit River				0.4		32.6			0.4	0.1		1.6	0.8
Spruce Grove	0.5	0.9			0.3								
St Albert	1.9	0.7	13.0	0.1	0.6							0.3	
Stony Plain		0.2			0.2					0.1	0.4		0.3
Swan Hills		3.0		0.2	0.5								
Westlock	0.5		8.7		0.7					0.1		0.3	
Whitecourt		54.2		0.1	0.2					0.1			0.3
Valleyview		1.2		0.2			0.8	0.1		1.0	34.9	0.8	12.6

TABLE 4

ESTIMATED DAY-USE ATTENDANCE (PARTY VISITS)
FISCAL YEAR 88/89

Origin / Total Party Visits	CA 3350	CP 17125	CL 5400	HB 14075	LS 19875	ML 8250	NO 1055	OB 14925	QE 3825	SI 15500	WI 5900	WN 5250	YP 3450
Athabasca	1323	n/a	n/a	n/a	60	n/a	n/a	n/a	n/a	n/a	n/a	n/a	13
Barrhead	17	240	n/a	28	80	n/a	n/a	n/a	n/a	31	n/a	n/a	n/a
Beaverlodge	n/a	n/a	n/a	14	n/a	50	n/a	15	n/a	729	124	42	48
Edmonton	1085	3322	1409	591	1888	50	8	15	31	248	171	152	87
Fairview	n/a	n/a	232	28	20	116	8	n/a	164	93	24	42	13
Falher	n/a	n/a	n/a	253	n/a	n/a	n/a	n/a	n/a	16	47	499	61
Fort Saskatchewan	n/a	n/a	n/a	n/a	40	17	n/a	n/a	n/a	n/a	n/a	16	n/a
Fox Creek	n/a	n/a	n/a	n/a	20	n/a	n/a	n/a	n/a	16	24	n/a	13
Gibbons	n/a	34	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Grande Prairie	n/a	34	n/a	99	119	899	n/a	13835	15	9889	1068	42	2036
Grimshaw	n/a	34	n/a	70	20	33	8	15	1469	31	n/a	68	n/a
High Prairie	n/a	n/a	n/a	7319	159	n/a	n/a	n/a	n/a	47	47	1292	13
Hines Creek	n/a	n/a	n/a	28	n/a	n/a	n/a	n/a	n/a	16	n/a	n/a	n/a
Manning	n/a	34	n/a	56	n/a	n/a	444	n/a	46	n/a	n/a	42	n/a
Mayerthorpe	n/a	154	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Mc Lennan	n/a	n/a	n/a	70	n/a	n/a	n/a	n/a	n/a	n/a	24	1223	n/a
Morrinville	17	n/a	n/a	n/a	60	n/a	n/a	n/a	n/a	n/a	n/a	16	n/a
Peace River	n/a	n/a	n/a	310	139	50	56	n/a	1232	n/a	24	446	26
Sherwood Park	34	120	n/a	28	40	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Slave Lake	17	n/a	n/a	127	13833	n/a	n/a	n/a	n/a	16	n/a	16	13
Spirit River	n/a	n/a	n/a	56	n/a	2690	n/a	n/a	15	16	n/a	84	35
Spruce Grove	17	154	n/a	n/a	60	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
St Albert	64	120	702	14	119	n/a	n/a	n/a	n/a	n/a	n/a	16	n/a
Stony Plain	n/a	34	n/a	n/a	40	n/a	n/a	n/a	n/a	16	24	n/a	13
Swan Hills	n/a	514	n/a	28	99	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Westlock	17	n/a	470	n/a	139	n/a	n/a	n/a	n/a	16	n/a	16	n/a
Whitecourt	n/a	9282	n/a	14	40	n/a	n/a	n/a	n/a	16	n/a	n/a	13
Valleyview	n/a	206	n/a	28	n/a	n/a	8	15	n/a	155	2059	42	548
	2591	14282	2813	9161	16975	3905	532	13895	2972	11351	3636	4054	2932
% day-use visits for which the origin was estimated	77%	83%	52%	65%	85%	47%	50%	93%	78%	73%	62%	77%	85%

Alberta household participation rates in various recreational activities were obtained from the Alberta Recreation and Parks (Alberta 1988b, Table 5).

Data Processing

The methodology chosen for processing the data for the attractiveness variable "Attract" is an adaptation of the methodology used by Cheung (1972). Although Cheung used the variable Attract in a multiple linear regression model, this variable was used here in a logistic regression (LR) model. The variable "Attract" used in this model differs from Cheung's by the inclusion of a miscellaneous features factor M_f. The mathematical expression of "Attract" is:

$$\text{Attract} = [\Sigma R_p \ \Sigma (F_m \ Q_m)] M_f \tag{28}$$

The attractiveness attribute **"Attract"** of a park is a function of:

R_p	-	the relative popularity rating of each recreational activity that can be practised in the park,
F_m	-	the relative importance rating of the facility(ies) available for the recreational activities,
Q_m	-	the rank scores of the facilities,
M_f	-	the miscellaneous features index.

The relative popularity rating $\mathbf{R_p}$ of each recreational activity was calculated relative to the activity "picnicking" using Alberta household participation rates $\mathbf{H_p}$ in various recreational activities (Alberta 1988b, Table 5).

Since the attractiveness of a park is based on its physical characteristics, it was necessary to examine the inventory data of facilities and features for each park (Tables 6 and 7). It was realized that not all outdoor facilities are equally important in attracting attendance to a park. In order to account for this, the Spearman's rank correlation coefficient procedure was used to determine the rank correlation between total day-use attendance of the 13 parks and each of the facilities at each park. The rank correlation coefficients for each park facility were then labeled as the **importance rating** of the facilities $\mathbf{I_m}$ (Tables 5 and 8).

The relative importance rating of a facility ($\mathbf{F_m}$) was calculated relative to picnicking facilities.

For example, the attractiveness measure Attract for the Calling Lake Provincial Park CA (Equation 28, Tables 5 and 9) was calculated as follows:

1) the rank scores Q_m of the facilities are determined,
2) for each recreational activity, the relative popularity rating of the activity R_p is multiplied by the sum of the products of rank scores Q_m and relative importance ratings F_m:

    for swimming only:

    0.80 * (0.69*7 + 0.73*4 + 0.69*6.5)

3) a sum is taken of the results in 2) for all activities
4) Attract is the product of 3) and the miscellaneous factor M_f

The **distance attribute** Dist is the distance in kilometres between the origin and the park visited (Table 10). Some modelers have redefined "distance" as travel time rather than kilometres, being convinced that travel time is the major deterrent in medium-distance travel (Ewing 1983). It can be argued that since distance and travel time are highly correlated, replacing one by the other would have no effect on the fit of the model.

The towns and cities of origin from which distances to parks were measured were selected using the criteria of a population of 1000 or more and a location within/near the Northern Region (Alberta Municipal Affairs, Official Population Lists 1989 and Table 10).

The next chapter contains the detailed results of the model runs and the interpretation of these results.

TABLE 5

POPULARITY RATING OF SOME DAY-USE ACTIVITIES
AND
IMPORTANCE RATINGS OF THEIR ASSOCIATED FACILITIES

ACTIVITY	ALBERTA PARTICIPATION RATE (Hp %)	RELATIVE POPULARITY RATING Rp = Hp/51 *	ASSOCIATED FACILITIES	IMPORTANCE RATING OF FACITITY, Im Im ≡ r_s	RELATIVE IMPORTANCE RATING OF FACILITY Fm = Im/0.45 *
Swimming	41	0.80	Bathing beach	0.31	0.69
			Change rooms	0.33	0.73
			Showers on the beach	0.31	0.69
Boating	28	0.55	Piers/Docks/Beaching area	0.40	0.89
			Boat rentals	0.23	0.51
Hiking	31	0.61	Trails	0.19	0.42
Picnicking	51	1.00	Picnic tables	0.45	1.00
			Picnic shelters	0.02	0.04
Golfing	40	0.78	Golf course	0.46	1.02
X-Country skiing	21	0.41	Winter trails	0.31	0.69
			Winter shelter	0.25	0.56

* The Choice of 51 and 0.45 as reference points for Rp and Im, respectively, is arbitrary.

TABLE 6

PHYSICAL INVENTORY *
(Fiscal 1988-89)

FACILITIES	CA	CP	CL	HB	LS	ML	NO	OB	QE	SI	WI	WN	YP
Bathing beach (length in m.)	180	200	150	225	5500	200	50	0	134	30	350	200	90
Change rooms	0	0	1	0	2	2	0	0	0	1	1	0	1
Showers on the beach	0	0	0	0	0	0	0	0	0	1	0	0	0
Piers/Docks/Beaching area (rating)	2.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Boat rentals	N	Y	N	N	N	N	N	N	N	N	N	N	N
Trails	0	1	4	3	2	0	0	4	3	3	0	0	6
Picnic tables	35	196	206	212	145	194	40	40	93	160	123	114	185
Picnic shelters	0	0	2	0	1	2	0	0	2	1	1	3	0
Golf course	N	N	N	N	Y	N	N	N	N	N	N	N	N
Winter trails	0	2	0	2	2	1	0	4	3	4	0	3	4
Winter shelter	0	0	0	1	4	5	0	0	3	1	0	7	0

* This table shows numbers unless otherwise stated.
Y = YES, N = NO, NA = NOT APPLICABLE

TABLE 7

MISCELLANEOUS FEATURES INDEX

Mf = 1 + 0.01n

PARKS

FEATURES	CA	CP	CL	HB	LS	ML	NO	OB	QE	SI	WI	WN	YP
Parking lot stalls (100s)	80	121	75	114	211	171	174	152	169	207	203	335	240
Washrooms buildings (10s)	4	12	1	1	36	19	3	1	7	9	15	7	6
Playground (2s)	0	2	2	4	4	4	0	2	3	4	1	2	2
Summer fishing	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y
Access to ice fishing	N	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y
Cleared skating area	N	Y	N	N	N	Y	N	N	N	N	N	N	Y
Downhill ski hill	N	N	N	N	N	N	N	N	N	N	N	N	N
Tobogganing hill	N	Y	N	N	N	N	Y	N	N	N	N	N	N
Winter day use	N	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y
Scenic viewpoints	N	Y	N	N	Y	N	Y	N	Y	Y	N	N	Y
Cultural/historical sites	N	N	N	Y	N	N	Y	N	Y	N	Y	N	N
Amphitheatre	N	Y	N	Y	Y	N	N	N	N	N	N	N	N
Park concessions	N	Y	N	N	N	Y	N	N	N	Y	Y	N	N
TOTALS:													
n =	2.20	11.41	4.85	8.24	12.71	10.61	7.04	4.62	5.89	7.97	9.03	8.05	9.00
Mf =	1.02	1.11	1.05	1.08	1.13	1.11	1.07	1.05	1.06	1.08	1.09	1.08	1.09

Y = YES = 1
N = NO = 0

TABLE 8

IMPORTANCE RATING OF THE "BATHING BEACH" FACILITY (1988-1989)

PARKS (Abbr.)	FACILITY (X) (Length in m.)	ATTENDANCE (Y) (Party visits)	RANK OF X (Qm)	RANK OF Y	R_x^2	R_y^2	R_xR_y
CA	180	3350	7	2	49	4	14
CP	200	17125	9	12	81	144	108
CL	150	5400	6	6	36	36	36
HB	225	14075	11	9	121	81	99
LS	5500	19875	13	13	169	169	169
ML	200	8250	9	8	81	64	72
NO	50	1055	3	1	9	1	3
OB	0	14925	1	10	1	100	10
QE	134	3825	5	4	25	16	20
SI	30	15500	2	11	4	121	22
WI	350	5900	12	7	144	49	84
WN	200	5250	9	5	81	25	45
YP	90	3450	4	3	16	9	12
TOTALS			91	91	817	819	694

Spearman's rho = r_s = 0.314921933

TABLE 9
PARK ATTRACTIVENESS (Attract)

PARKS	CA	CP	CL	HB	LS	ML	NO	OB	QE	SI	WI	WN	YP
						(Rank scores Qm)							
FACILITIES													
Bathing beach (length in m.)	7	9	6	11	13	9	3	1	5	2	12	9	4
Change rooms	4	4	9.5	4	12.5	12.5	4	4	4	9.5	9.5	4	9.5
Showers on the beach	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	13	6.5	6.5	6.5
Piers/Docks/Beaching area (rating)	3.5	8.5	5.5	10	12	13	3.5	1	2	5.5	8.5	11	7
Boat rentals	6.5	13	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5
Trails	3	6	11.5	9	7	3	3	11.5	9	9	3	3	13
Picnic tables	1	11	12	13	7	10	2.5	2.5	4	8	6	5	9
Picnic shelters	3.5	3.5	11	3.5	8	11	3.5	3.5	11	8	8	13	3.5
Golf course	6.5	6.5	6.5	6.5	13	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5
Winter trails	2.5	7	2.5	7	7	5	2.5	12	9.5	12	2.5	9.5	12
Winter shelter	4	4	4	8.5	11	12	4	4	10	8.5	4	13	4
TOTALS:													
Rp * (Fm * Qm)	22.03	39.45	39.15	43.26	49.72	44.59	21.32	23.86	28.39	39.49	35.63	35.23	38.55
Mf	1.02	1.11	1.05	1.08	1.13	1.11	1.07	1.05	1.06	1.08	1.09	1.08	1.09
Attract	22.47	43.79	41.10	46.72	56.19	49.50	22.81	25.05	30.09	42.65	38.84	38.05	42.02

TABLE 10

ORIGIN-PARK DISTANCES (km)

	CA	CP	CL	HB	LS	ML	NO	OB	QE	SI	WI	WN	YP
Athabasca	55	281	60	261	168	454	591	469	394	474	356	247	375
Barrhead	183	169	119	259	241	449	589	449	390	364	338	270	357
Beaverlodge	545	321	482	288	393	151	416	62	210	27	135	258	139
Edmonton	201	201	169	396	286	575	701	476	503	481	364	368	383
Fairview	518	390	453	268	356	86	261	143	63	140	203	188	200
Falher	378	252	314	124	230	131	285	191	88	219	102	45	121
Fort Saskatchewan	208	192	167	333	289	600	715	467	507	486	370	471	382
Fox Creek	443	90	309	223	299	295	443	217	245	222	105	201	124
Gibbons	193	197	154	339	275	557	690	473	496	486	369	351	383
Grande Prairie	503	287	438	248	353	109	376	20	178	25	92	226	97
Grimshaw	460	340	395	205	303	144	203	201	5	198	183	125	202
High Prairie	302	266	237	47	152	207	344	221	146	226	109	25	128
Hines Creek	547	418	482	297	385	100	290	159	92	156	225	217	216
Manning	534	424	469	279	377	228	119	285	79	282	257	199	286
Mayerthorpe	291	61	182	273	306	422	570	336	371	349	232	292	251
Mc Lennan	352	286	287	97	202	157	294	215	96	227	110	26	129
Morrinville	193	173	129	356	251	537	675	448	476	466	349	331	363
Peace River	436	310	371	181	279	168	217	225	19	222	159	101	178
Sherwood Park	221	216	192	416	306	601	719	494	521	499	382	388	403
Slave Lake	185	173	120	131	35	324	461	338	263	343	226	117	245
Spirit River	478	358	419	223	342	32	315	97	117	102	157	150	162
Spruce Grove	210	164	169	359	270	521	666	434	475	447	335	353	358
St Albert	188	178	149	386	263	557	683	447	488	466	434	348	363
Stony Plain	220	157	169	366	277	515	660	427	482	454	328	358	350
Swan Hills	300	60	211	159	149	352	489	352	291	365	256	170	267
Westlock	143	210	78	295	199	488	625	491	429	403	379	281	396
Whitecourt	369	20	227	228	225	377	525	291	326	304	187	247	206
Valleyview	392	170	327	137	242	209	357	131	159	136	19	115	38

CHAPTER 3

INTERPRETATION OF MODEL RESULTS

General Model Evaluation

Following model calibration, the next step in the model building process is the evaluation of the model. First, it is necessary to run the model, perform a sensitivity analysis, then interpret the results. The model in its original form (eqs. # 25, 26, 27) is called the General Model.

Without the benefit of years of testing, it is difficult to evaluate the model by criteria other than those of skills requirement and accuracy. It can be debated that whether a model is said to be complex or not is a matter of subjective interpretation. For this project, however, complexity means the level of quantitative skills required by the user to understand the model. The General Model, to be adequately understood by the user, requires the knowledge of quantitative skills which can be acquired from an intermediate statistics course.

The model is said to be accurate if the variable relationship is stable and the proportion of explained variance is high (Georgoff and Murdick, 1986). The stability of the variable relationship will be evaluated primarily in the policy analysis section below but a simple sensitivity analysis was conducted to determine the stability of the model when changes to the most important independent variable (distance) occur.

A sensitivity analysis is performed to determine in a general manner, if there are threshold values of the independent variable (like distance) where the dependent variable "day-use visits" either does not respond to changes of the independent variable or responds excessively. For this purpose, the "attractiveness" of one park (Lesser Slave Lake) was kept constant while the distance was allowed to change. At short distances (0-30 km) from a municipality, very little change in the number of visits occurred but as the distance increased , change in the number of visits responded as expected in a semi-linear manner.

Accuracy will be evaluated here using the forecasting results at both the individual park level and at the regional level (aggregation of the forecasting results of the 13 parks). To determine the model's accuracy, the model is run to obtain the expected day-use visits which are compared to the observed day-use visits thus giving an estimation of the model's forecasting error.

Table 11 shows for each park the number of observed day-use visits (O), the predicted (expected) number of day-use visits (E), the ratio of expected visits (E) to observed visits (O) and the error measurement $DIFF^2$. This error measurement is a form of the standard error of multiple estimate which is an estimate of the standard deviation. $DIFF^2$ is expressed mathematically as :

(29) $\sqrt{(O-E)^2} / n$

where n is the number of origin-park pairs, that is, 28.

Table 11 reveals that the General Model underestimates the number of day-use visits for 6 parks and overestimates the number of day-use visits for 7 parks. The averages of under- and over-estimation are 0.67 and 1.88 respectively. The model is obviously **not very accurate** at predicting day-use visits **for individual parks**; it either under- or over-estimates quite significantly. The parks for which the number of day-use visits are over/under estimated (i.e. E/O >1 and E/O <1) do not seem to have a distinct spatial distribution, that is, the estimation error occurs for parks throughout the study region.

TABLE 11

GENERAL MODEL AND POLICY ANALYSIS

POLICIES/PARKS	#1(CA)	#2(CP)	#3(CL)	#4(HB)	#5(LS)	#6(ML)	#7(NO)	#8(OB)	#9(QE)	#10(SI)	#11(WI)	#12(WN)	#13(YP)	#14NEW
General Model:														
Observed (O)	2591	14282	2813	9161	16975	3905	532	13895	2972	11351	3636	4054	2932	
Expected (E)	1509	8844	6981	5985	10031	5285	503	8916	3599	12631	8940	8935	6940	
Error (DIFF2) *	249.10	1145.21	429.65	794.75	1290.12	222.87	97.19	1086.54	143.59	145.16	650.45	622.91	595.61	
E/O	0.58	0.62	2.48	0.65	0.59	1.35	0.95	0.65	1.21	1.11	2.46	2.2	2.37	
Increase attraction of park #7	1406	8476	6547	6063	9797	5408	865	8779	4037	12482	9025	9240	6975	
% change from gen. model	-6.8%	-4.2%	-6.2%	1.3%	-2.3%	2.3%	72.0%	-1.5%	12.2%	-1.2%	1.0%	3.4%	0.5%	
Decrease attraction of park #5	1685	9436	7284	5265	3346	6097	589	10759	3947	15153	9745	8144	7649	
% change from gen. model	11.7%	6.7%	4.3%	-12.0%	-66.6%	15.4%	17.1%	20.7%	9.7%	20.0%	9.0%	-8.9%	10.2%	
Add park #14	1249	8325	6131	4940	8066	4312	382	8734	3566	14991	9420	9168	4948	4868
% change from gen. model	-17.2%	-5.9%	-12.2%	-17.5%	-19.6%	-18.4%	-24.1%	-2.0%	-0.9%	18.7%	5.4%	2.6%	-28.7%	
Remove park #12	1798	10768	8716	4149	10941	4871	546	10743	3611	17741	9486	N/A	5728	N/A
% change from gen. model	19.2%	21.8%	24.9%	-30.7%	9.1%	-7.8%	8.5%	20.5%	0.3%	40.5%	6.1%		-17.5%	

* DIFF2 is a form of the standard error of the multiple estimate and is calculated by SPSSX as:

The inaccuracy of the General Model at the individual park level can be explained mainly by the lack of representative sampling data for each origin. Indeed, as was explained in the previous chapter, the estimated number of day-use visits to a park in a year which serves as input data to the model is approximated from a proportion (or percentage) based upon a survey with a usually small sample ($n<40$). Due to this small sample size, it is clear that some origins may not be represented at the time the survey is taken despite a possibly large number of day-users from those origins. For example, the park survey showed no day-use visits to Notikewin (NO) from Grande Prairie. Although this park has a relatively low attractiveness (22.81) and is located 376 km from Grande Prairie, it is doubtful that no visits took place even if only as short stops on the way to further destinations. Other similar examples could be cited to emphasize the **poor representation** of the input data as presently collected.

It is important to bear in mind, however, the main purpose of this model. This model is primarily a spatial allocation model for policy analysis, that is, its main purpose is to allocate a total number of day-use visits (the variable TOTAL obtained from exponential smoothing) to the various parks of a region relative to a change in a park planning policy. Although important, forecasting the number of day-use visits to a specific park can be considered secondary to the model's policy analysis function. Model refinements are contemplated, however, to improve the accuracy in forecasting day-use visits for any specific origin-park pair.

As a measure of accuracy, it is very useful to determine for the entire region, what proportion of the variation between the observed and expected visits is explained by the model and conversely what proportion is attributable to chance. For that purpose the following variance equation was used.

$ST^2 = SA^2 + SE^2$ where:

ST^2 is the square of the total variation,

SA^2 is the square of the variation attributable to the model, and
SE^2 is the square of the variation attributable to error.

SE^2 is the square of the error $DIFF^2$ for the entire region using n = 28 x 13 = 364 origin-park pairs. ST^2 is the square of the standard deviation of visits (SDVISIT) for the entire region. $DIFF^2$ and SDVISIT were calculated by SPSSX and are respectively 696.22 and 1334.06. SA^2 is the difference between ST^2 and SE^2 and was calculated as follows:

$$SA^2 = (1334.06)^2 - (696.22)^2 = 1779556 - 424722 = 1294833.71$$

To determine the proportion of variation that is attributable to the model, one need only find the ratio of SA^2 to ST^2 as follows:

$$SA^2/ST^2 = 1294834/1779556 = 0.7276$$

This ratio shows that the model explains 73% of the total variation, on a regional level, between observed and expected day-use visits to the parks.

Although the model is not very accurate at predicting the number of visits to individual parks, it provides **a good estimate** of day-use park visits for the northern Alberta region taken as a whole.

Policy Analysis

The next step in the interpretation of the model results is the analysis of four recreation planning policies. This will be an analysis of the spatial allocation of the total number of day-use park visitors (i.e. the variable TOTAL) to the Provincial parks in the region in a given year resulting from policy changes.

As mentioned, this is a spatial allocation analysis and therefore for any analysis the variable TOTAL is held constant. Since TOTAL is assumed to be constant, any significant change in TOTAL would need to be factored in during the exponential smoothing process. For example, improving a park (or adding a new one) should increase TOTAL, and decreasing the attractiveness (or closing a park) should decrease TOTAL. The policy simulations that follow simply allocate TOTAL spatially and do not change its value. In this model, TOTAL cannot be changed through exponential smoothing because the data series are too short, that is, there is not a sufficient number of years of data. It becomes readily obvious that this is a combined approach to forecasting park attendance.

The policies of interest are: increasing the attractiveness of a park, decreasing the attractiveness of a park, adding a new park, and closing a park.

Increase the Attractiveness of Park #7 Notikewin (NO) by 50%.

The purpose of this policy simulation is to observe the effects on day-use park visits of increasing the attractiveness of Notikewin Provincial Park. This park has the second lowest attractiveness of the group with ATTRACT = 22.81 (Table 9). With this simulation its attractiveness would be increased by 50% to 34.23. Adopting this policy implies increasing the number of facilities and/or improving the existing services so as to make the park more attractive to potential visitors. The General Model results will serve as the basis for comparison (Table 11).

It is necessary at the outset, to determine if the model responds as expected to an increase in the attractiveness of NO. That is, whether $Prob1_i$ and $Prob2_{ij}$ for any origin increase with a higher ATTRACT value for NO. As a reminder, $Prob1_i$ is the probability of visiting any park from an origin based primarily on the origin's distance to all parks. $Prob2_{ij}$ is the probability of visiting a specific park based primarily on the competition between parks. Comparing the behavior of $Prob1_i$ and $Prob2_{ij}$ of this policy run with that of the General Model (using computer run printouts for the 364 origin-park pairs), it was observed that for most origins $Prob1_i$ and $Prob2_{ij}$ behaved as expected. A 50% increase in the attractiveness of NO caused the probabilities of a number of origins (and particularly the larger populated ones nearby) to increase. The overall effect of this policy relative to the base model (the General Model) is a 72% increase in the number of visits from all origins to Notikewin i.e from the 503 expected visits in the base case to 865 visits with this policy (Table 11).

Looking at examples of the behaviour of the probabilities for some origins, some ideal behaviours and some peculiarities were observed. Using Edmonton as an example (origin #4 located 701 km away from NO), $Prob1_i$ did not change but $Prob2_{ij}$ increased from .03 to .05 thus representing an increase in day-use visits from that city from 272 to 458 or a 68% increase. This significant increase in day-use visits from Edmonton can be explained by the preponderance of visitors to NO that normally originate from that city (Table 11).

Looking at the effect of this policy on the day-use attendance of the other parks (Table 11), one could speculate that the increased attractiveness of Notikewin and the resulting increase in demand for this park also entails an increase in the demand for other parks in the northern half of the region. In this case, this expectation does materialize as can be seen from Table 11, the number of visits to the northern parks (QE, WI, WN, YP, and HB) does increase if only slightly. The demand for more distant parks in the southern half of the region like CA, CP, and CL, however, decreased.

As mentioned previously, the TOTAL number of day-use visitors for the region was held constant; only the spatial allocation was changed.

Generally speaking, this policy simulation produced results that were expected intuitively. The effects on day-use visits to any park from a specific origin or from all origins may be exaggerated because of the lack of refinement of the input data. For example, the input data did not show any visits to Notikewin from Grande Prairie which is located 376 km away. Since Grande Prairie is the closest city to Notikewin, one would expect a number of visits from this origin but such visits were not represented during the 1989 origin surveys clearly attesting to the importance of accurate data.

Decrease the Attractiveness of Park #5 Lesser Slave Lake (LS) by 25%.

The purpose of this policy simulation is to observe the effects on day-use park visits of decreasing the attractiveness of LS by 25% i.e. from 56.19 to 42.14. The General Model results for LS will serve as the basis for comparison (Table 11). The Lesser Slave Lake Provincial Park has the highest attractiveness of the group at 56.19.

It is necessary at the outset, to determine if the model responds as expected to a decrease in the attractiveness of LS. That is, whether $Prob1_i$ and $Prob2_{ij}$ for any origin decline with a lower ATTRACT value for LS. Comparing the behavior of $Prob1_i$ and $Prob2_{ij}$ of this policy run to that of the General model (using the computer run printouts of probabilities for each of the 364 origin-park pairs) , it was observed that for most origins $Prob1_i$ and $Prob2_{ij}$ did behave as expected. A 25% decrease in the attractiveness of LS caused both probabilities to decline. This behavior of $Prob1_i$ and $Prob2_{ij}$ was more pronounced for origins near LS park and particularly for the Town of Slave Lake. The overall effect of this policy relative to the base model (the General Model) is a 66.6% decrease in the number of day-use visits to LS.

Looking at examples of the behaviour of the probabilities for some origins, some ideal behaviours and some peculiarities were observed. For example, it is difficult to explain why for Hilliard's Bay Provincial Park, $Prob1_i$ from the Town of Slave Lake declined significantly from .12 to .03 while $Prob2_{ij}$ increases from .09 to .11 resulting in the predicted number of day-use visits to HB from this Town to decline from 921 to 290 (or - 68.5%). One would have expected that a decrease in the attractiveness of LS would have resulted in visitors from the Town of Slave Lake to LS to go to HB instead. Also, why from the Town of Slave Lake does the propensity ($Prob1_i$) to visit parks in general decline? While the probability of visiting HB ($Prob2_{ij}$) remains relatively the same. Note however that the allocation of visits from any origin is the product of the probabilties, therefore even if $Prob2_{ij}$ increases, a decline in $Prob1_i$ may still result in a decline in the expected attendance.

In summary, comparing the results (i.e. day-use park visits) of this policy run with those of the General Model (Table 11), one can observe the following:

a) Decreasing the attractiveness of LS which is the most frequented park in the region results in a 67% decline in day-use visits to LS for that year.

b) The parks near LS, that is, HB and WN also experience a decline in day-use visits respectively of 12% and 9%.

c) Most other parks experience an increase in the number of day-use visits. For example, OB, SI and NO with respectively 21%, 20% and 17% increases.

Generally speaking, this policy simulation produces results that can be expected intuitively. This run can be very useful in simulating the effects of lake, playground, and/or concessions closures on day-use park attendance.

Add Park #14

The purpose of this policy simulation is to observe the effects on day-use park visits of adding a new park within the Northern Region. This new park is given the number 14 and has the characteristics of park #5 (Lesser Slave Lake - LS) with respect to attractiveness, distance, and the number of observed visits to it (VISIT) thus indirectly simulating the addition of a new park at the same location as LS.

It was assumed that adding a new park would not increase the total number of visits to parks in the region. It was also assumed that the new park #14 would have the same "drawing power" as park #5 and that some visits would be taken away from some parks and allocated to other parks. These assumptions may not be realistic. One could argue in favour of the supply-generated-participation effect which involves an increase in visits when the number of parks or facilities increase.

Had longer time series of TOTAL visits been available, it would have been possible to increase TOTAL through the exponential smoothing process and thus assume an increase in the TOTAL number of visits to parks in the region, a more realistic assumption. Normally, a time series of at least 10 years is required for exponential smoothing to produce accurate results.

If the model user wishes to show an increase in day-use visits from the addition of another park, he/she can do so by simply increasing the smoothing factor (α) in the exponential smoothing process from which the variable TOTAL would be predicted.

Except for the probabilities of the newly created park #14 which have no history, the probabilities of visiting parks from most origins would normally decline.

To determine if the model responds as expected, it is necessary to observe how $Prob1_i$ and $Prob2_{ij}$ behave for the various origins. For most origins $Prob1_i$ barely responds but $Prob2_{ij}$ declines. This indicates that the propensity to visit parks from any origin does not increase significantly with the addition of a new park and that from many origins the number of visits to any specific park may decline. Therefore, model results indicate that when TOTAL is not changed through exponential smoothing, the addition of a new park does not increase the total number of visits but simply allocates them to more parks. The sparsity of data which negates the use of exponential smoothing precludes the model from showing the supply-generated effect.

For a few origins, the decline of $Prob2_{ij}$ is quite significant. For example, for origin #20 (the Town of Slave Lake), the probability of visits to park #5 (Lesser Slave Lake Provincial Park) i.e. $Prob2_{ij}$ declines from .67 to .53 or from an expected 7071 to 5584 day-use visits or a 21% decline. This could be explained by the presence of the newly created park #14 for which, it was assumed, having the same characteristics of attractiveness as park #5 (LS) and would be located at the same distance from the Town of Slave Lake. The new park would be in direct competition with park #5 for the park visiting population of Slave Lake.

The number of visitors from Slave Lake would remain the same but the distribution to parks #5 and #14 would change i.e. a decline in day-use visits at park #5 and an increase in day-use visits to park #14.

A great increase of $Prob2_{ij}$ occurs, however, for park #10 (Saskatoon Island - SI) from origin #10 (Grande Prairie). $Prob2_{ij}$ increases from .44 to .55 representing an increase in day-use visits from 10509 to 13167 or 25%. It is interesting to note that $Prob1_i$ for Grande Prairie remains unchanged at .27. It is difficult to explain the above mentioned increase of $Prob2_{ij}$ because $Prob2_{ij}$ for this origin-park pair did not change with the policy of increasing the attractiveness of Park #7 (NO) reviewed in a previous section. This could be explained by the lack of representative input data for this park-pair and the great distance from Grande Prairie to NO.

Generally speaking, in this policy simulation, the model behaves as expected assuming no supply-generated participation effect. The addition of a new park adds competition (through the competition factor COMPET defined in the first chapter) against existing parks and allocates day-use visitors away from these parks by reducing $Prob2_{ij}$.

Closing Park #12.

The purpose of this policy simulation is to observe the effects on day-use park visits of closing park #12 (Winagami - WN). Winagami Provincial Park has an attractiveness of 38.05 and is located near the Town of High Prairie.

As with the previous analysis, it is necessary to determine if the model responds as expected to the policy in question. Bearing in mind that in this model the total number of visits remains the same and that only the allocation changes, one would expect intuitively that the closure of Park #12 would cause the number of visits to increase in other parks and particularly so in the nearby parks. First, $Prob1_i$ and $Prob2_{ij}$ are observed to determine if they increase as expected for the majority of origin-park pairs. The expected number of day-use visits for the specific parks will be compared to those of the General Model (Table 11) and anomalies will be noted.

For most origin-park pairs, $Prob1_i$ and $Prob2_{ij}$ increase as a result of the closure of Park #12 (WN). Table 11 shows that the total number of day-use visits to most parks increases, on average, 7%, with increases experienced by 8 of the 12 parks (67%) while only 3 parks of 12 experienced a decrease (25%). There were no changes in total day-use visits for Park #9 (QE). The greatest increase is for Park #10 (SI) at 40%.

In this simulation run, the model behaved unexpectedly for the nearby parks, namely Hilliard's Bay (HB) and Lesser Slave Lake (LS). Park #4 (HB) experienced the greatest decline in day-use visits of the group i.e. 31%. These simulation results are rather problematic; one could have expected the total number of day-use visits to HB to increase rather than decrease and particularly so since the attractiveness values (ATTRACT) of both Parks #12 (WN) and #4 (HB) are relatively close i.e 38.05 and 46.72 respectively. The total day-use visits for Park #5 (LS) with an ATTRACT value of 56.19 experienced a 9% increase which seems less than one might have expected since LS is a nearly park and should have captured visits from park #12. As mentioned above, park #10 (SI) experienced a 40% increase in visits. Most of them originated from Grande Prairie for which $Prob1_i$ changed from .27 to .33 and $Prob2_{ij}$ changed from .44 to .53 resulting in a change in the number of day-use visits from 10509 to 15540 or a 48% increase.

Generally speaking, the model behaves as expected; the closure of park #12 (WN) results in increases in the total number of day-use visits to most other parks. $Prob1_i$ and $Prob2_{ij}$ for specific origin-park pairs increase as expected. There are, however, a few anomalies. One would have expected the allocation to favour the nearby parks; park #4 (HB) and park #5 (LS). That is, most of the visitors who would have gone to Park #12 (WN) would have been expected, after the closure of Park #12, to go to the nearby parks HB and LS. The simulation shows, however, a significant decline of day-use visits for HB and only a small increase of day-use visits for LS. On the other hand, park #10 much further west and near

Grande Prairie received a substantial 40% increase in day-use visits mostly from Grande Prairie.

The analysis of the four recreation policies reveals that the model, in some cases, behaved differently than expected. The unexpected model results have been labelled idiosyncrasies. Normally, a researcher would investigate these idiosyncrasies since they may provide insight into new recreation choice behaviors. In this model, however, most of the idiosyncrasies contradict simple logic and, therefore, do not warrant further testing of hypothesis but rather the determination of possible causes for the anomalous results. The author is convinced that the idiosyncrasies of this model are caused by incomplete, and in many cases, inaccurate input data.

It is evident from the foregoing that the quality of the input data and possibly the model structure could be improved to make the forecasting of day-use visits to specific parks more accurate and the policy simulations more intuitively correct.

The next chapter will contain a discussion on potential model refinements.

CHAPTER 4

POTENTIAL MODEL REFINEMENTS

Any forecasting model is subject to improvements. Ideally, potential improvements are suggested after a few years of use. Improvements may pertain not only to the model structure but also to all aspects of model use, ranging from data collection to data interpretation. Refining a model can be an on-going process.

It was observed in the previous chapter that there are limitations to the General Model and to the results of the Policy Analysis runs. The General Model is not very accurate at forecasting the number of day-use visits to specific parks but reasonably accurate (73%) at forecasting the total attendance at the regional level. The Policy Analysis runs produce results that are intuitively correct, although there are a few idiosyncrasies.

Although improvements will be suggested to all aspects of this model including data processing and structure modification, the effects of such improvements can only be evaluated through actual model use over the period of a few years. Obviously, for practical reasons, the implementation and evaluation of these model refinements are beyond the scope of this project. Model refinements will be suggested to improve the representation of the input data and thus the forecasting accuracy. Also suggested will be an improvement to the attractiveness variable (ATTRACT) to reflect the visitor's perception of what site characteristics are most important or attractive in the selection of a park destination. The interpretation of model results and particularly the effects of the various policy simulations will be improved through market segmentation.

Data Sampling Refinements

Essential to sound park/recreation planning is knowing the amount and kind of use the various parks receive. Accurate estimates of attendance aid in cost-benefit analysis and budget justification. Also, knowledge of visitor distributions both among the parks and within each park helps identify overuse and underuse, provides a base for attendance forecasting, and helps in planning for the construction of future parks and facilities. The importance of attendance data sampling cannot be overemphasized.

Attendance data sampling, although important for all aspects of park planning, will be discussed here primarily with respect to the project forecasting model. If the model is to forecast accurately the number of day-use visits for each origin-park pair, it is essential that the sampling data be representative of the population of day-use park visitors for the region. It is desirable to obtain the most accurate count of day-use visitors/parties and their place of origin. As with all sampling techniques, the main difficulty arises in obtaining representative and therefore accurate data at a low cost. The greatest challenge comes from large parks with many access and egress points for both vehicular and pedestrian traffic.

Provincial parks are controlled or limited -access parks. Each park is equipped with Automatic Counter devices (usually at the main entrance) to count the number of vehicles entering or leaving the area. The automatic vehicle count is then adjusted by a statistician for non day-use visits. While providing a count of day-use visits, the Automatic Counter readings even if adjusted to account for non day-use traffic give no information about the "true" user distribution within the park. Which sections of the park are used and for what?

Information about the origin of visitors is surveyed during 13 random days throughout the May-September peak season. These surveys are rudimentary and do not cover the fall and winter seasons. It would be desirable to have origin surveys all year-round even though the winter season sees few visitors. The data on winter visits can be useful for making decisions about expanding winter facilities like cross-country skiing shelters. Another advantage of having year-round origin surveys rather than peak season surveys is the ability to better determine the origin of visitors. Also, since there are different user turnover rates, more frequent surveys of who are day-users would allow planners to better determine within any time span which individuals are representative of day-users.

It is well understood that in any park, a particular recreation activity is practiced more effectively and pleasantly in one particular section, that is, an activity is normally concurrent with one location. It would be worthwhile to find which section (or sectoral attendance count) has the highest correlation with total attendance. The section (and its concurrent activity) with the highest correlation will tell us which activity is most popular i.e. for which there is a strong demand. Then it becomes necessary to investigate why a section/activity is not well used/practiced. This information, in addition to helping in better predicting the park attendance, could help the modeler in giving a greater weight of attractiveness to the facility of the most frequented section.

The subject of refinements to the attractiveness variable, a structure modification, will be pursued in the next section. It is important to note that if all that is needed is seasonal attendance then sectionalization is not required. To summarize, sectionalization will reveal the best location for Automatic Counter devices, it will indicate which is the most popular activity, and thus which activity/facility contributes most (a weighting factor) to the attractiveness of the park.

As mentioned in the previous chapter, a weakness of the input data is that it is not representative of the various origins of visitors. This weakness may be due to an insufficient number of surveys about visitor origins. If the input data is not adequately representative of the origins, that is, if there is an insufficient number of surveys or an insufficient number of visitors from many origins who happened to be sampled, there will be an anomaly in the logistic distribution and the model's forecasting accuracy will be reduced. Therefore, its is essential to conduct many origin surveys and these surveys should be more comprehensive. Random origin surveys should be conducted in each season, that is, six randomly-selected week-day surveys and six randomly selected week-end or holiday surveys for a total of at least 24 origin surveys. These surveys, being conducted on-site, can be criticized for being biased but this is acceptable bearing in mind the cost of conducting larger surveys of the public at large.

Surveys about the origin of visitors are important for three reasons. First, these surveys can include other questions like travel costs/deterrents and preferred recreation activities. Second, information on origins is needed to calculate origin-park distances, a key element in the model structure. Third, the information is useful for marketing purposes.

The survey questionnaire would be given to day-use visitors by part-time workers who would have been briefed about how the questionnaire should be completed. The questionnaire should be kept to a minimum length, that is, it should contain no more than five questions or so. the questionnaire should include one question about the respondent's stage of his/her life-cycle, a priority ranking of recreation activities available at Provincial Parks (ranked first, second and third), the names of the parks that were visited twice or more during the past twelve months, and a priority ranking of the five most important visual (site) characteristics.

As will become more obvious in the next section, the information from the comprehensive visitors-origins survey, will be used to improve the model structure, particularly the attractiveness variable, and it will help in policy analysis by creating the opportunity for market segmentation.

Structure Modification

The method for calculating the attractiveness variable (ATTRACT) of a park is based on an inventory of facilities at each park and on the presence of recreational opportunities offered by the landscape. All of these factors, including the presence of on-site amenities like concessions, are used to arrive at a park attractiveness value without any input data about "perceived attractiveness" from the user's perspective.

With the visitors-origins survey, the park visitor would rank the five most important/attractive visual characteristics (facilities and features). The rationale for attaching importance to perceived attractiveness is based on research which demonstrates that a small change in visual resources may have a major effect on the perceived quality of the resource (Buhyoff and Wellman, 1980). This ranking of visual characteristics would serve as the basis for weighting the characteristics, that is, the characteristic given the highest rank or greatest importance would be given the largest weight in the calculation of the park attractiveness. The variable ATTRACT then becomes determined (to some extent) by the recreationists rather than planners.

Equation # 28 would then include a weighting factor W_f replacing the importance ratings I_m and Q_m . Instead of using I_m and Q_m which are based on Spearman's rank correlation between total day-use attendance and the facilities at each park, a weighting factor W_f would be used for each facility. It is believed that the use of a weighting factor would simplify the calculation of the attractiveness variable but it remains undetermined if forecasting accuracy would be improved. An inventory of facilities would still be needed.

Market Segmentation

Although park attractiveness is an important factor in recreation activity choice, it is believed that differences in individual tastes, motivations, and perceptions are the greatest influences on activity choice. The General Model hypotheses described in Chapter 2 were made with the assumption that both characteristics of the individual and attributes of the alternative parks affect the choice process (Figure 6A).

Several mechanisms may influence choices. One possibility, (that is , the utility theory) consists of using these characteristics as linear, additive terms in a utility function (in this model u_1 and u_2). In this case, the effect of the characteristics is marginally to add to or subtract from the utility of activities. Some researchers argue that the best way to enter the variables in a recreation choice model is to use them as a basis for population (market) segmentation (Watson and Stopher, 1974).

If one assumes that all possible market segments have a homogeneous perceptual space, that is, that park visitors operate with the same factor structures in their choice of a recreation activity or site, then it becomes necessary to find the best model structure to test for different weights on a given factor. The same model structure would be used in all segmentation tests; each segmentation would constitute a unique model.

Statistical tests, using the Student's t-distribution may be conducted on the coefficients of the models corresponding to the various (market) segments to determine if the models are based on independent samples. If the models originate from independent samples (segments), there will be significant differences between the coefficients of the various models. In addition, the likelihood-ratio test can be performed between the pooled results of the (market) segments and the unsegmented model. The likelihood-ratio test establishes whether or not the segmented models succeed in explaining more of the recreationists behavior than does the single unsegmented model.

Research on recreation choice models by Stopher and Ergun (1979) revealed that the most promising segmentation bases were life-cycle stages and geographic location. Of these market segmentation schemes, probably the most promising in terms of usefulness is the life-cycle stages. Figure 8 is a list of reasonable groupings of stages used to create the life-cycle variables (LC1 to LC5).

How the life-cycle variable expresses differences in recreation behavior is a matter of conjecture. It is reasonable to suggest, for example, that people in stages 1 and 2 (LC1) are likely to be more active in recreation because of a lack of various responsibilities and some independence from other people; whereas people in stages 3 and 4 (LC2) which constitutes the home-making stages would tend to be less active in recreation because greater responsibilities particularly to children and because of their need to save money. A number of similar arguments can be advanced to explain the behaviour of recreationists in other life-cycle stages.

FIGURE 8

LIFE-CYCLE VARIABLES

LC1	Young (<35 years), unmarried, living alone or with others.
LC2	Young, married, no children or oldest child <5 years.
LC3	Married, oldest child >5 years.
LC4	Older (>35 years), married, no children at home.
LC5	Older, unmarried, living alone or with others.

Source: adapted from Stopher and Ergun, 1979.

A stage of the life-cycle segmentation was done by Stopher and Ergun (1979) using 812 cases from two Chicago suburbs. They observed that the segmented models performed significantly better than the unsegmented model based on a likelihood-ratio test. They concluded that the segmentation scheme was worthwhile because it identified a number of underlying differences in behaviour although the stage of the life-cycle variable possibly operated only as a proxy for a complex set of constraints to recreational activities.

Undoubtedly, the life-cycle stage may not be the optimum segmentation scheme. Further research could indicate that multiple segmentation, that is, segmentation schemes based on more than one variable, could produce optimal model structures.

It is believed that the refinements mentioned above would greatly improve the quality of the input data, the model structure, and hence the usefulness of the model for both attendance forecasting and policy analysis.

CONCLUSION

Government plays an important role in the provision of recreation opportunities. Most of these recreation opportunities require large investments of capital and human resources to insure that recreation facilities are available to a large number of individuals. This allocation of resources is the means of implementing most park/recreation policies. The recreation planner or policy analyst must obtain reasonably accurate data or information to support the policies that he/she advances.

This project aimed at developing a model to forecast day-use park attendance to the various Alberta parks using information about the origin of visitors, the distance to the parks, and the attractiveness of the parks. The model would be used primarily for policy analysis. Different park planning policies would be simulated to determine the effect of such policies on day-use park attendance.

It was deemed necessary to present the elements of the model building process to better introduce the project subject matter. the advantages and limitations of models were presented to situate the project in a realistic context. Different quantitative and qualitative forecasting techniques were described and evaluated to provide the reader with information about the status of recreation forecasting techniques. Future directions for research are the refinements of existing techniques and the improvement of forecasting accuracy through combined or multimethod forecasts. The proliferation of microcomputers and sophisticated software like spreadsheet and statistical packages facilitate the development of powerful yet user-friendly models.

The author favoured the use of a combined or multimethod approach to forecasting travel choice. The combined forecasts approach includes medium term (1 year) forecasts of regional attendance to parks (time series), medium term forecasts of specific park attendance (logistic regression) and long term forecasts of recreation travel trends (scenario writing). The latter long term forecasting technique was discussed only summarily.

In this project, the emphasis was put only on the selection and development of a medium term forecasting model. The chosen day-use park attendance forecasting model is a "complete model" with both a trip generation coefficient and a trip distribution coefficient. Although the multinomial logit (MNL) model has had a large number of followers in the past few years, it was deemed preferable to chose a simple model for which parameter estimation software would be readily available and easy to understand and use. For these reasons, and the desire to experiment with the design of a new model, the logistic regression (LR) model structure was chosen.

The input data for the logistic regression model pertain to 13 provincial parks and 28 municipalities of the Northern Region of Alberta. The data comprises monthly and fiscal year park attendance, origins of visitors, participation rates, number and quality of facilities and features, and origin-park distances. An attractiveness attribute is ascribed to each park based primarily on the physical characteristics of the park.

The model was run and the results were interpreted. Without the benefit of years of testing, it is difficult to evaluate the model by criteria other than those of complexity and accuracy. The model was judged to be intermediate in complexity since it requires that the user possess intermediate skills with statistics. The model's accuracy was evaluated by comparing expected day-use visits with observed day-use visits thus giving an estimation of the model's forecasting error.

The model was found to be not very accurate at forecasting day-use visits to individual parks; it either under- or over-estimated quite significantly. The inaccuracy of the model at the individual park level can be explained mainly by the lack of representative sampling input data. When used to forecast "regional" day-use attendance, the model explained 73%

of the variation between observed and expected day-use visits to parks. At the regional level the model provides a good estimate of day-use park visits.

The model's main use, however, is policy analysis. Policy analysis in this project means the analysis of the spatial allocation of the total number of day-use park visits in the region in a given year resulting from policy changes. The policies of interest are: increasing the attractiveness of a park, decreasing the attractiveness of a park, adding a new park, and closing a park.

Generally speaking, the simulation runs showed that the model behaved as could be expected intuitively. There were, however, a few anomalies justifying the need for some improvements. Although potential improvements to a model should be based on the experience from a few years of use, two types of improvements were recommended based only on the evaluation of model results. These potential improvements pertain to both data collection and model use. Specifically, improvements need to be made to the input data sampling technique, to the model structure, and to data processing by means of market segmentation.

If the model is to forecast accurately the number of day-use visits for each origin-park pair, it is essential that the sampling data be representative of the population of day-use park visitors for the region. Better data representation can be obtained through an improved data sampling technique. A notable sampling technique is sectionalization. This technique will reveal the best location for Automatic Counter devices and hence, which activity/facility most contributes to the attractiveness of the park, More frequent random sampling is recommended using a comprehensive survey questionnaire. The comprehensive survey questionnaire would provide information about the respondent's stage of the life cycle and priority ranking of facilities/features.

The information from the survey questionnaire could be used to improve the model structure, particularly the way the attractiveness variable (Attract) is determined. Indeed, by using the respondents' priority ranking of facilities and features as weights in the attractiveness equation, park attractiveness would better reflect the visitors perspective.

Research has showed that segmented models perform significantly better than unsegmented models. The stage of the life cycle information from the survey questionnaire would likely constitute an optimum segmentation scheme.

The goals of this project have been accomplished, that is, the model was formulated, the results were interpreted, and recommendations were made to improve the model.

REFERENCES CITED AND BIBLIOGRAPHY

Alberta Government. 1988a. Alberta Tourism. Community Tourism Action Plan Manual (Revised). Book 3.

Alberta Government. 1988b. Alberta Recreation and Parks, 1988 General Recreation Survey.

Bar On, R.R.V. 1979. "Forecasting Tourism - Theory and Practice. " TTRA Tenth Annual Conference Proceedings, College of Business, University of Utah.

Bar On, R.R.V. 1983. "Forecasting Tourism by Means of Travel Series Over Various Time Spans Under Specified Scenarios." Third International Symposium on Forecasting, June 1983.

Batty, Michael J. 1972. "Recent Developments in Land Use Modeling: A Review of British Research", Urban Studies, 9(2): 151-177.

Burton, T.L. 1981. "You Can't Get There from Here: A Perspective on Recreation Forecasting in Canada." Recreation Research Review 9(1981): pp. 38-43.

Calantone, R.J., A. diBenedetto and D.C. Bojanic. 1988. Multimethod Forecasts for Tourism Analysis." Annals of Tourism Research, Vol. 15, pp. 387-406.

Cesario, F.J. 1969. "Operations Research in Outdoor Recreation." Journal of Leisure Research 1 (Winter), pp. 35-51

Chapin, F. Stuart Jr. and Edward J. Kaiser. 1979. Urban Land Use Planning (Third Edition). University of Illinois Press. Chicago, Illinois.

Cheung, H.K. 1972. "A Day Use Park Visitation Model." Journal of Leisure Research, 4(2), pp. 139-156.

Clawson, M., and J.L. Knetsch. 1966. Economics of Outdoor Recreation. The John Hopkins University Press. Baltimore. Ma.

Dalkey, N. and O. Helmer. 1963. "An Experimental Application of the Delphi Method to the Use of Experts." Management Science 9, April, pp. 458-467.

Draper, N.R. and H. Smith. 1981.Applied Regression Analysis, Second Edition. John Wiley & Sons Inc., Toronto.

Edgell, D. and R. Seely. 1980. "A Multi-Stage Model for the Development of International Tourism Forecasts for States Regions." Tourism Planning and Development Issues. Hawkins et al. eds. Washington D.C.: George Washington University.

Ewing, G.O. 1980. "Progress and Problems in the Development of Recreational Trip Generation and Trip Distribution Models." Leisure Sciences, 3, pp. 1-24.

Ewing, G.O. 1983. "Forecasting Recreation Trip Distribution Behavior." Recreation Planning and Management, pp. 120-140. S.R. Lieber and D.R. Fesenmaier, editors. State College, Pa. Venture Publisher.

Fesenmaier, Daniel R. 1988. "Integrating Activity Patterns Into Destinastion Choice Models" in the Journal of Leisure Research, 1988, Vol. 20, No. 3, pp. 175-191

Fesenmaier, Daniel R. 1985. "Modeling Variation in Destination Patronage for Outdoor Recreation Activity" in the Journal of Travel Research, Fall 1985, pp. 17-22.

Fujii, E.T. and J. Mak. 1980. "Forecasting Travel Demand When the Explanatory Variables are Highly Correlated." Journal of Travel Research, Spring 1980, pp. 31-34.

Gearing, E.C., W.W. Swart and T. Var. 1976. Planning for Tourism Development: Quantitative Approaches. Praeger Publishers, New York, N.Y.

Georgoff, D.M. and R.G. Murdick. 1986. "Manager's Guide to Forecasting." Harvard Business Review, January-February, pp. 110-120.

Green et al 1990. Application of the Delphi Technique in Tourism. Annals of Tourism Research, Vol. 17, pp. 270-279.

Gunn, Clare A. 1988. Tourism Planning (Second Edition). Taylor & Francis, New York, N.Y.

Helmer, O. 1979. "The Utility of Long Term Forecasting." Forecasting. Studies in the Management Sciences, editors: S. Makridakis and S.C. Wheelwright, vol. 12, Amsterdam, North Holland.

Heywood, John T. 1991. Visitor Input to Recreation Opportunity Spectrum Allocation and Monitoring. Journal of Park and Recreation Administration, Vol. 9, No. 4, Winter 1991, pp. 18-30.

Howard, D.R. and J.L. Compton. 1980. Financing, Managing and Marketing Recreation and Park Resources. William C. Brown. Dubuque, Iowa.

Iman, R.L. and W.J. Conover. 1983. Modern Business Statistics. John Wiley & Sons, Toronto.

Jamieson, W., N.J. MacDonald and W.T. Perks. 1988. "The Crowsnest Pass: Directions for the Future." Alberta and Northwest Territories Journal of Planning Practice. Number 7, Winter, pp. 81-105.

Johnson, R.L. and D.B. Suits. 1983. "A Statistical Analysis of the Demand for Visits to U.S. National Parks: Travel Costs and Seasonality." Journal of Travel Research, Fall 1983, pp.21-24.

Kaynak, E. and J.A. Macaulay. 1984. "The Delphi technique in the measurement of tourism market potential - The case of Nova Scotia." Tourism Management, June, pp. 87-98.

Lin, Y, G.L. Peterson and P.A. Rogerson. 1988. "A Nested Recreation Site Choice Model." Leisure Sciences, Volume 10, pp. 1-15.

Little, J.S. 1980. "International Travel in the U.S. Balance of Payments." New England Economic Review. May/June, pp. 42-55.

McAvory, Leo H. et al. 1986. "The Importance of Visual Environmental Quality in Site Selection for Water-Based and Water-Enhanced Recreation Activities". Recreation Research Review, Volume 12, No. 3, pp. 41-48.

Mahmoud, E. 1984. "Accuracy in Forecasting: a Survey." Journal of Forecasting, Vol. 3, pp. 139-159.

Makridakis, S. and S.C. Wheelwright. 1979. "Forecasting Framework and Overview." in Forecasting, editors: S. Makridakis and S.C. Wheelwright. Studies in the Management Sciences, vol. 12. Amsterdam: North Holland.

Makridakis, S. and M. Hibon. 1979. "Accuracy of forecasting: an empirical investigation." Journal of the Royal Statistical Society, 142, part 2, pp. 97-145.

Makridakis, S. et al. 1982. "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition." Journal of Forecasting, Vol. 1, pp. 111-153.

Moeller, G.H. and E.L. Shafer. 1983. "The Use and Misuse of Delphi Forecasting." Recreation Planning and Management. S.R. Lieber and D. R. Fesenmaier, editors. State College, Pa. Venture Publisher.

Oliveira, R.A., L.M. Arthur and A.C. Papastavrou. 1983. "A Distributed Lag Approach to Forecasting Wilderness Use." Journal of Leisure Research, Volume 15, pp. 52-64.

Perks, William T. and Lydia Ind-Kawun. 1986. "Strategic Planning for Small-Town Community Development" in Alberta Journal of Planning and Practice, No. 5, Fall 1986.

Peterson, G.L., Anderson, D.H. and D.W. Lime. 1982. "Multiple Use Site Demand Analysis: An Application to the Boundary Waters Canoe Area Wilderness." Journal of Leisure Research, First Quarter, pp. 27-36.

Peterson, G.L., Dwyer, J.F., and A.J. Darragh. 1983. "A Behavioral Urban Recreation Site Choice Model." Leisure Sciences, Volume 6, Number 1, pp. 61-81.

Peterson, G.L. and D.J. Stynes. 1986. "Evaluating Goodness of Fit in Nonlinear Recreation Demand Models." Leisure Sciences, Volume 8, Number 2, pp. 131-147.

Quayson, J. and T. Var. 1982. "A tourism demand function for the Okanagan, B.C." Tourism Management, June 1982, pp. 108-115

Render, Barry and Ralph Stair Jr. 1988. Quantitative Analysis for Management, Third Edition. Allyn and Bacon Inc. Boston.

Saunders, P.R., Senter, H.F. and J.P. Jarvis. 1981. "Forecasting Recreation Demand in the Upper Savannah River Basin." Annals of Tourism Research VIII(2), pp. 236-256.

Scott, David and Geoffrey C. Godbey 1990. Reorienting Leisure Research - The Case for Qualitative Methods. Society and Leisure Vol. 13, No. 1 (Spring 1990) pp. 189-205.

Seneca, J.J. and R.K. Davis. 1976. "A Cross Section Analysis of State Recreation Activity." Journal of Leisure Research, Volume 8, Number 2, pp. 88-97.

Siderelis, Christos, & Joe Roise 1991. An Optimal Apportionment Strategy for Park Operations. Journal of Park and Recreation Administration Vol. 9 No. 2 Summer 1991, pp. 48-58.

Smith, V.K. and V.G. Munley. 1978."The Relative Performance of Various Estimators of Recreation Participation Equations." Journal of Leisure Research, Volume 10, Number 3, pp. 165-176.

Stanley, R. Lieber and Dan R. Fesenmayier (Editors), 1983. "Recreation Planning and Management". Article by D. J. Stynes. State College, Pennsylvania. Venture Publishing.

Stopher, P.R. and G. Ergun. 1979. "Population Segmentation in Urban Recreation Choices." Transportation Research Record, 728, pp. 59-65.

Stopher, P.R. and A.H. Meyburg. 1979. Survey Sampling and Multivariate Analysis for Social Scientists and Engineers. Lexington Books, Toronto.

Stokey, Edith and Richard Zeckhauser. 1978. A Primer for Policy Analysis. W.W. Norton & Company. New York.

Stynes, D.J. and G.L. Peterson. 1984. "A Review of Logit Models with Implications for Modeling Recreation Choices" in the Journal of Leisure Research, Vol. 16, No. 4, pp. 295-310.

Swart, W.W., T. Var and E.C. Gearing. 1978. "Operation Research Applications to Tourism," Annals of Tourism Research, (Oct.-Dec.), pp. 33-51.

Thiel, H. 1969. "A multinomial extension of the linear logit model." International Economic Review. 10, pp. 251-259.

Timmermans, Harry. 1982. "Consumer Choice of Shopping Centre: an Information Integration Approach" in Regional Studies, Vol. 16.3, pp. 171-182.

Uysal, M. and J.L. Crompton. 1985. "An Overview of Approaches Used to Forecast Tourism Demand". Journal of Travel Research, Spring 1985, pp. 7-15.

Van Doorn, J.W.M. 1984. "An Unexplored Forecasting Area in Tourism: Scenario Writing." Problems of Tourism, Editor, Van Doorn, vol. 3, pp. 63-73.

Van de Ven, A.H. 1974. "Group Decision Making and Effectiveness." Organization and Administrative Sciences, 5, pp. 100-110.

Watson, P.L. and P.R. Stopher. 1974. "The Effect of Income on the Usage and Valuation of Transport Modes." Transportation Research Forum Proceedings, Vol. 5, No. 1, pp. 460-469.

Wheelwright, S.C. and S. Makridakis. 1980. Forecasting Methods for Management, Third Edition, Wyley, New York. N.Y.

Wilson, A.G. 1971. "A Family of Spatial Interaction Models and Associated Developments." Environment and Planning A 3(1971), pp. 1-32.

Wrigley, N. 1982. "Quantitative methods: Developments in discrete choice modeling." Process in Human Geography, 6, pp. 547-562.

Zarnowitz, V. 1984. "The Accuracy of Individual and Group Forecasts from Business Outlook Surveys." Journal of Forecasting, Vol 3, pp. 11-26.