MICRO-SIMULATION OF RESIDENTIAL LOCATION
CHOICE AND ITS VARIATION

K. A. Raju, P. K. Sikdar and S. L. Dhingra
Civil Engineering Department, Indian Institute of Technology, Bombay,
Powai, Mumbai 400 076, India

ABSTRACT. Traditionally land use models are designed to represent aggregate
behaviour even though land use development patterns are the manifestation of
micro-level individual location choices. This paper develops a micro-simulation
model for residential location choice in a medium-sized city in India. The model
consists of two major parts: (a) the simulation of the characteristics of the
individual decision units, the households of the city, and (b) the simulation of
their locational choice for residences. The first submodel simulates the
characteristics of households using the Monte Carlo method; the second is
developed using Artificial Neural Network (ANN) theory. The integrated model
simulates variations in the locational choices of the population and the residential
land use distribution by aggregating the decisions of all the individual decision
units. This model is developed and tested using data from the Indian city of
Guwahati.

INTRODUCTION

Housing, industries, transportation, and other activities related to the economy are the
main components of an urban system. Urban development models are commonly referred
to as land use models in which land use is described by the location of the activities.
Basically, a land use model is a simplified representation of the planner’s understanding of
the land development process in his/her particular planning jurisdiction over a particular
time period. Urban planners are concerned with the most appropriate arrangement of the
subsystems to yield an integrated urban structure in the form of master plan. A clear
knowledge of the spatial distribution of different land uses and their intensities is necessary
to estimate the required magnitude of the infrastructure facilities to be provided.

It is a general practice to characterise urban centres by their total population, their
location patterns, and their population densities. Therefore, among the different land uses
in the city, the residential land use pattern generally dictates several other policy measures
that follow these patterns. Identifying the important role of residential land use
distribution in the overall system of urban development and management, various models
have been developed for the residential location of a population which have different
theoretical backgrounds and data requirements.
Most of the traditional land use models deal with the behaviour of aggregate masses of the population. These aggregate models embody the critical assumption that households within a given zone are fairly homogeneous and that variations in the zonal average accurately reflect the variations among individuals. This assumption causes serious problems in understanding the behaviour of highly heterogeneous populations with a wide spectrum of socio-economic levels in developing countries. As a result, a model may fail to serve as an instrument for assessing the policy impacts on future growth patterns of residential land uses.

Moreover, any urban system is a complicated structure containing millions of interacting units such as individuals, households, and firms. It is these units which actually make decisions related to spatial location, trip making, and so on. As a result, it is reasonable to expect that a model’s predictions would be more successful if they were based on knowledge about these elemental decision-making units.

The term micro-simulation has appeared with increasing frequency in recent years and its implied definition varies with its context. The preferred broad definition which has been adopted in this paper is a modelling approach which is based on the representation of the individual decision maker. The basic assumption is that residential location choices can be defined in terms of the decision making of individual units, with all the possible outcomes linked to the characteristics of the relevant decision making unit. A single hypothesis or theory is rarely sufficient for this purpose. It is possible only by linking several relevant theories together in a meaningful way.

A micro-analytic model has been formulated as an integrated form of two submodels. The first submodel estimates the required attributes of the decision maker (or worker) under consideration (as any choice is influenced by the attributes of the decision maker). The second submodel simulates the decisions pertaining to the residential location of this decision maker. This ultimately results in the spatial distribution of population when the outcomes of the decisions of all the workers in the study area are combined.

MODELLING THE CHOICE PROCESS

The problem of estimating the decision of an individual unit making a particular choice when faced with a number of possible alternatives occurs in many fields when human behaviour is being modelled. The selection of a residential location zone depends on the socio-economic background of the decision maker such as income, vehicle ownership, and so on, and on the relative attributes of the zones such as travel time from work place, overall accessibility, and the availability of utility services. Therefore, in a generic form, the choice process can be represented as:

\[
\text{Choice} = f (\text{Characteristics of decision maker, Attributes of the alternatives}).
\]

Random utility models such as logit, structured logit, and probit are widely used in modelling the spatial location choices of individuals. However, the predictive ability of these models goes down drastically with an increase in the number of available alternatives (the choice set). For example, in the case of residential location choice, it may be necessary to consider all the zones in the study area as the choices available to a decision maker. Thus, the size of the choice set will be huge and the number of parameters to be estimated
will be larger, causing computational difficulties. In utility-based models, the size of the choice set is often reduced by excluding infeasible choices for a particular decision maker based on certain assumptions.

Recent developments attempt to develop models that represent a wide variety of locational choices and include a large number of policy variables. Therefore, successful behavioural modelling research which is closer to real world observations requires the availability of suitable analytic and computational techniques. In this regard, the applicability of Artificial Neural Network (ANN) theory, which has been successfully employed to classification problems, can be considered. This is true because, in a way, choice modelling is nothing but a classification problem. For example, a trip maker whose mode choice is car can be classified as a car user.

**ARTIFICIAL NEURAL NETWORKS FOR CHOICE ANALYSIS**

ANN can be viewed as massive parallel networks composed of many computational elements connected by links with variable weights. Neural network models can be found in the literature for numerous applications. All of these models attempt to achieve good performance through dense interconnection of simple computational elements (or neurons). The different network paradigms that have been developed so far vary greatly in the range of vector mappings that they can represent. The nature of the mapping relationship between the input and output vectors is determined by the values of free variables (often called weights) within the network.

Many paradigms have been developed in ANN theory for updating the synaptic weights. However, the back-propagation algorithm was used for the choice analysis in this research as it is the most popular and well-tested algorithm (Wasserman, 1993) of the available paradigms. The feed-forward network, trained by back-propagation, was a key development in the history of ANN (Rumelhart et al., 1986). The performance of ANN models has proven to be far better than multivariate regression techniques which are the most widely used analysis tools in the urban planning field.

Supervised network training for choice analysis was done with a back-propagation algorithm. The target vector represents the set of desired values from the network when the input vector \(X\) is applied. More formally, the back-propagation algorithm minimises the error between the output vector and target vector. The error is compared with a permissible error level to control the training phase as shown in Equation 1.

\[
\frac{1}{\text{NTP}} \sum_{c=1}^{\text{NTP}} \sum_{k=1}^{\text{NON}} (y_{kc} - \text{Out}_{kc})^2 \leq E, \tag{1}
\]

where \(y_{kc}\) is desired output at the \(k\)th output node for \(c\)th training pattern, \(\text{Out}_{kc}\) computed output at the \(k\)th output node for \(c\)th training pattern, NTP total number of training patterns, and NON number of output nodes.

**THE STUDY AREA**

Guwahati is the capital of Assam state with a population of 0.6 millions and 34 municipal wards. These wards are very large in size for analysis and were further divided.
into 101 internal analysis zones for the study. Household survey data were collected by Assam Engineering College as part of a sponsored research project. The survey questionnaire consists of two parts: (a) socio-economic characteristics of the household; and (b) travel details of household members. The survey contained a sample of 3,850 households, representing 3.5% of the households in the study area, and was collected in 1991–1992, and used to develop and validate the submodels. The sample size collected through the survey was not sufficient for choice analysis in some of the analysis zones. In order to adequately represent all of the zones in the sample and the choice frequencies, land use parameters, and geographical continuity of the zones, the 101 traffic zones in the study area were grouped to form 32 group zones.

The performance of a single residential location choice model that was developed for the whole study area, was found to be poor. This indicated the necessity of dividing the study area into a finite set of geographical sectors. Therefore, the study area was divided into five spatial sectors. These sectors were based on the residential location choice of the employees which were observed from the sample data. As a result, depending on the context, the study area has been viewed with any of the four spatial structures shown in Figure 1 consisting of hierarchical zones with: (a) 5 sectors, (b) 32 group zones, (c) 34 municipal wards, and (d) 101 traffic zones.

**SOCIO-ECONOMIC CHARACTERISATION SUBMODEL**

The basic idea is that the population distribution in the study area can be taken as the cumulative effect of the location decisions of all the households in the study area. Therefore, it is necessary to have complete information on the socio-economic attributes of all the decision-making units which influence the choice process. Collecting such an enormous data set is practically impossible. However, this problem can be solved by using a model which internally stimulates the required socio-economic attributes without deviating too much from reality. This was done by formulating a submodel which generates the characteristics of all the individual actors/decision-making units, based on the information obtained through the household survey. The distribution of the socio-economic characteristics and the pattern of residential location choices observed in the sample are assumed to be applicable to the study area. The Monte Carlo simulation approach was used to assign attributes to individual actors.

The socio-economic characterisation submodel is used to simulate the socio-economic characteristics of a worker who is randomly selected from the total set of workers in the study area. Cumulative probability distributions of the workers’ socio-economic characteristics are obtained conditionally on the basis of variables which are dependent on the problem specification. For this purpose, the household survey data have been cross-classified by the set of variables which are assumed to influence the response variable (residential location choice). The values from the cross-classification matrix are then converted into row-proportions to give conditional probabilities. Cumulative values for the row-proportions provide the cumulative distribution of the conditional probabilities. The socio-economic variables to be used in the simulation are initially selected on the basis of their relative importance in determining residential location choice.
FIGURE 1. Spatial levels in Guwahati for modelling.

(a) Sectors or Regions

(b) Group Zones

(c) Municipal Wards

(d) Traffic Zones
RESIDENTIAL LOCATION SUBMODEL

Factors Affecting Residential Location Choice

In an aggregate location model, activities are allocated to zones according to a function which measures the attractiveness of a zone, relative to all other zones in the study area. The measurement of locational attraction is one of the most important problems in model development and has never been solved satisfactorily (Batty, 1976). Attraction can be seen as a function of many factors such as residential density, accessibility to other activities, location relative to the Central Business District (CBD), the presence of public transport facilities, and so on. Travel times and travel distances which are dependent on the transport network characteristics are also major factors affecting decision making concerning residential location.

It is advisable to study the relative importance of all the possible variables which an analyst feels would affect the decision process and ignore variables which play a minor role. “Partitioning of weights of the ANN model” (Antony, 1994) has been used for this purpose, and the factors which have a significant influence on the residential location are identified based on their relative importance scores. Based on this investigation, it was observed that the most important variables affecting residential location choice are home ownership and the number and type of personalised vehicles.

Model Formulation

For the present model it was assumed that household-level decisions are dependent on the convenience of its working members. Thus, a worker, either a basic or service sector employee, is the key person in the household because of his active economic status. Decisions related to residential location are primarily dependent on the working member’s characteristics and their associated household-level characteristics. In the present model, the eldest male employee in a house is assumed to be the head of the household. If the household does not have a male employee, then the eldest female worker is treated as the head of the household.

As stated earlier, the 101 traffic zones were grouped to form 32 group zones which made up the location alternatives which are available to a decision maker. Thus, the locational choice is a function of the characteristics of the household and the attributes of the group zones.

The ANN model that was developed for the residential location choice is a feed-forward type, fully connected with three layers. The number of nodes in the input layer is decided by the factors that are assumed to affect residential location choices. The representation form and the corresponding number of nodes in the input layer for each variable are given in Table 1. The total number of nodes in the input layer is 116 and the number of output nodes is equal to the number of group zones in the choice set, i.e., 32. While training the network, the observed group zone choice is represented by “0.9”, indicating the higher probability of its selection; the others are represented by “0.1”. Therefore, the target vector for each individual decision unit consists of 32 components with 0.9 assigned to the selected group zone and 0.1 assigned to the rest.
Training and Testing

The performance of the ANN model is evaluated in terms of its ability to reproduce the trained vectors (reproducibility) and a set of patterns that were not used for training purpose (predictability). Separate ANN models were developed for each of the five spatial sectors based on the independent structure of their locational preferences. The total sample patterns available for each of the sectors are split into two parts for training and testing the neural net. 85% of the patterns in each of the sectors was used for training; the remaining patterns were used for testing purpose. The training process was terminated when the error level is reduced to 0.01 per pattern or the total number of iterations equals 50,000. The output node which yields the maximum output value is considered to represent the selected group zone. The synaptic weight matrices obtained at this stage are used to test the reproducibility and predictability of the neural net.

Table 2 gives the results of performance tests on the trained neural networks. Reproducibility evaluates the capability of the model to estimate the observed choices if the trained data is used as the input; this indicates the goodness-of-fit for the model. The data which were withheld from the training are used to estimate the model’s predictability which, in a way, indicates its transferability. As shown in Table 2, the neural network

<table>
<thead>
<tr>
<th>Sector</th>
<th>Total sample (Training + Testing patterns)</th>
<th>Reproducibility (%)</th>
<th>Predictability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>463 + 82</td>
<td>99.3</td>
<td>85.4</td>
</tr>
<tr>
<td>2</td>
<td>576 + 102</td>
<td>90.4</td>
<td>79.4</td>
</tr>
<tr>
<td>3</td>
<td>704 + 124</td>
<td>86.7</td>
<td>75.8</td>
</tr>
<tr>
<td>4</td>
<td>681 + 120</td>
<td>85.5</td>
<td>74.1</td>
</tr>
<tr>
<td>5</td>
<td>478 + 84</td>
<td>97.5</td>
<td>88.1</td>
</tr>
<tr>
<td>Study area</td>
<td>2640 + 512</td>
<td>90.1</td>
<td>79.7</td>
</tr>
</tbody>
</table>
models developed for the employees working in sectors 1, 2 and 5 show good levels of performance, while the models for sectors 3 and 4 failed to achieve the same level. However, the performance of the models are still at an acceptable level, with a predictability of 75% or more in each case.

VARIATIONS IN CHOICES

It must be recognised that individuals behave in different ways as a result of a whole range of intangible and tangible factors. Even if the tangible factors are measurable, responses to them will be a function of their perception, and they may not act rationally. As a result, people with same characteristics may behave differently when faced with a similar decision environment.

Similar conclusions can be drawn from the results of Table 2 which suggest that some people may not have selected the best alternative available to them (as ranked by the activation values of the output nodes of ANN model). It is possible that in some cases alternatives were selected which were inferior to the best one (e.g., second best, third best, etc.). The strict assumption that people always select the best alternative may therefore lead to a bias in estimating the actual demand. This demonstrates clearly that a variation exists in choice making. The model must be able to represent this with more accuracy. As a result, the Monte Carlo method was used again to simulate these variations in modelling the choice related decisions. For this purpose, the observed variations in choice making are estimated from the sample data and applied to the whole study area.

The direct outputs of the ANN models for the choice analysis are the activation values of the output nodes in the output layer. The activation value of an output node reflects the potential that an individual would select the option represented by that output node. This means that the higher the activation value, the greater the possibility that the option represented by that output node is selected. It may seem that the option should be selected which is represented by the output node with maximum activation value (which can be considered to be as equivalent to the utility in MNL), on the hypothesis that an individual will choose the alternative which yields the maximum utility. However, as already discussed, all of the decision makers may not act rationally and an element of variation exists in selecting the alternatives. In the present case, this variation means that the decision maker selects an alternative other than the one with maximum activation value.

The following methodology was used to estimate the probability of selecting an alternative on the basis of the rank obtained from the activation values of the output nodes.

1. Consider the data used for the calibration/training of the choice ANN model.
2. Find the corresponding activation values for the training pattern by applying it to the choice model.
3. Rank the alternatives (output nodes) in descending order according to their activation values.
4. If the alternative with rank ‘n’ is the observed choice, then the occurrence of nth ranked alternative is expected.
5. Compute the frequency of occurrences for all the ranks of the selected alternatives by repeating steps 2–4 using all the patterns of the training set.
6. Compute the cumulative probability values for the occurrences of all the ranks of the alternatives from the cumulative frequency distributions of their occurrences.
This algorithm was used to calibrate separately the residential location choices of the employees working in each spatial sector. It can be observed from Figure 2 that some of the households did not select the best alternative and the 15th best alternative (as ranked by the activation values of output nodes) was chosen. These cumulative distributions were used with the Monte Carlo method to simulate variations in choice making in the integrated model in order to better represent the actual demand.

**THE INTEGRATED MODEL AND SIMULATION SEQUENCE**

The integrated residential location micro-simulation model generated the population distribution for the entire study area. Unlike other micro-simulation models in which only a sample is simulated, this study attempted to simulate the total population distribution for the entire study area. However, the socio-economic characteristics, choice pattern, and variations in the choices observed in the household survey sample were used in a structured fashion to simulate the behaviour of the entire population.

While traditional land use models begin with only the basic employment in the zones, the present model uses the total employment as an exogenous input for its residential location process. It generates the household (and population) values internally and does not attempt to generate or locate services. The model has been formulated to simulate the residential location choices with zonal level constraints on population. A household is initially allocated to a group zone (G) and allocated to zone (R) based on its relative
attractiveness among the constituent zones in the group zone \((G)\). The overall model of micro-simulation is illustrated in Figure 3. The various inputs to the model and the simulation sequence are described in the following sections.

**Model Inputs**

Exogenous inputs to the model were obtained from the household survey data containing the sample population’s socio-economic characteristics and travel diaries. Travel time and distance matrices for different modes were obtained from network data. Employment levels in the zones were obtained from secondary sources and processed as required. Before beginning the simulation process, the model first estimates the accessibility levels of zones using the travel time matrix and the information on the service employment by applying Hansen’s approach. The controlling distributions for simulating the socio-economic characteristics of the households and household members, are prepared using the household survey data.

**Simulation Process**

The simulation process starts by generating a random number between 1 and 101 which represents the zone number for the employee to be simulated. This employee is assumed to be the head of the household and the socio-economic characteristics of the corresponding household are generated using the socio-economic characterisation submodel. These socio-economic characteristics along with the attributes of the group zones are the inputs to the residential location submodel.

The outputs from the residential Location submodel are the activation values for each node (the choice set) in the output layer. These activation values indicate the potential that the group zone represented by that output node will be selected. Therefore, the preference ordering of the group zones for the decision maker under consideration are ranked according to the descending order of activation values.

A second-level Monte Carlo method generates a random number and its comparison with the cumulative probability distribution of the ranked choices indicates the rank of the choice that the decision maker would select. This means that the group zone \((G)\) pointed to by this rank is treated as the choice of the decision maker. Thus, the choice, including its variation, is simulated to bring the results closer to reality.

A check is made in simulating the base year residential land use before allocating a household to a group zone \((G)\) to ensure that the population to be assigned does not exceed the observed population of the group zone. If the model is used for forecasting, a check is done against the holding capacity of the group zone. If the assigned population is less than the observed population or the holding capacity, the household is assigned to the group zone indicated by Monte Carlo method. If the assigned population exceeds the observed population or the holding capacity of the group zone, the Monte Carlo method is used once again to find the next possible group zone \((G)\) for residential location. The process continues until the model finds a group zone which satisfies the criteria on the population check.

A household that is assigned to a group zone \((G)\) must then be allocated to a specific traffic analysis zone \((R)\). This is done by comparing the attractiveness of the constituent
FIGURE 3. Structure of the micro-simulation model.
zones within a group zone, and allocating the household to the most attractive traffic analysis zone (R). The measure of attractiveness of zones is shown in Equation 2.

\[ A_{GR} = \frac{E_{GR}}{P_{GR}} \times \frac{1}{t_{WR}} \]  

(2)

where: \( A_{GR} \) is measure of attractiveness of zone ‘R’ in group zone ‘G’, \( E_{GR} \) service employment in zone ‘R’ of group zone ‘G’, \( P_{GR} \) population in zone ‘R’ of group zone ‘G’ at that stage of simulation, \( t_{WR} \) travel time between zone ‘R’ and zone ‘W’, where W indicates the work zone of the worker selected.

It can be seen that simulation process reduces the attractiveness of a given zone as the simulation process proceeds by reducing the service employment available per unit of population already assigned by the simulation. The population capacity check is conducted at the traffic zone level as it was for the group zones. A significant number of households have more than one worker. Therefore, the work zones for the second and third employees in the household, if any, must be established. This is required because the residential location process automatically provides work-to-home linkages, and, therefore, the total work travel in the city. While the traditional land use models interactively generates the service employment locations and their work-to-home linkages, the present model considered the total employment in simulating their residences.

In the absence of sufficient information on the work location of these employees, a heuristic measure called the opportunity to work measure has been used to locate the work places for the second and third employees, given their place of residence. The opportunity to work measure is given in Equation 3

\[ D_{WR} = \frac{(E_{W} - SE_{W})}{t_{WR}} \]  

(3)

where \( D_{WR} \) is opportunity to work in a zone W, when the residence is in zone R, \( E_{W} \) total employment in zone W, \( SE_{W} \) simulated employees in zone W at that stage of the simulation, and, \( t_{WR} \) travel time between zones W and R by private mode.

It is assumed that the zone which provides the maximum opportunity to work is the work zone for the second or third employees in the household. After simulating all the required socio-economic characteristics and residential location for a household and the employment locations of the workers other than the head of the household, the model generates another random number between 1 and 101 to simulate the choices of another employee from the chosen traffic zone.

Aggregating the simulated residential location choices of all the households yields the population and residential workers distributions for the entire study area.

**MODELLING THE BASE YEAR SCENARIO (1991)**

The various submodels and their computational processes have been discussed in the preceding sections. These submodels work together in an integrated fashion to determine the locational distribution as discussed in the previous section. Before the integrated model can be used for forecasting, its performance in simulating the base year residential land use distribution should be examined.
The submodels were developed with sample household data which captured the population’s locational behaviour and their predictability with unseen patterns. These are now used to simulate the base year location choice patterns for the study area. Thus, the model’s performance is evaluated with respect to its ability to reproduce the residential land use distribution in the base year. The model uses the Monte Carlo method and 15 simulation runs of the model were used with different seed values and the results were averaged. The averaged simulation results are presented in Table 3 along with the variation calculated at the 95% confidence level.

An important assessment of the model’s performance compares the simulated number of households in each zone with the observed number. Because only ward-level household information is available from the population census, the comparison for each ward is shown in Figure 4. It can be observed that the simulated number of households closely

![Figure 4. Simulated and actual number of households in wards (1991).](image)

<table>
<thead>
<tr>
<th>Table 3. Study Area Characteristics from Simulated Results for 1991</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study area characteristics</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Number of households</td>
</tr>
<tr>
<td>Mean household size</td>
</tr>
<tr>
<td>Number of vehicle in the study area</td>
</tr>
<tr>
<td>Cars</td>
</tr>
<tr>
<td>Two-wheelers</td>
</tr>
<tr>
<td>Bicycles</td>
</tr>
<tr>
<td>Vehicle ownership of households</td>
</tr>
<tr>
<td>No. of households with 1 car</td>
</tr>
<tr>
<td>No. of households with 2+ cars</td>
</tr>
<tr>
<td>No. of households with 1 two-wheeler</td>
</tr>
<tr>
<td>No. of households with 2+ two-wheelers</td>
</tr>
</tbody>
</table>

The submodels were developed with sample household data which captured the population’s locational behaviour and their predictability with unseen patterns. These are now used to simulate the base year location choice patterns for the study area. Thus, the model’s performance is evaluated with respect to its ability to reproduce the residential land use distribution in the base year. The model uses the Monte Carlo method and 15 simulation runs of the model were used with different seed values and the results were averaged. The averaged simulation results are presented in Table 3 along with the variation calculated at the 95% confidence level.

An important assessment of the model’s performance compares the simulated number of households in each zone with the observed number. Because only ward-level household information is available from the population census, the comparison for each ward is shown in Figure 4. It can be observed that the simulated number of households closely
matches the observed number of households in each ward; the $R^2$ value is 0.986. Thus, the model has simulated the existing situation quite well. The total number of households in the study area as simulated by the model is 1,19,280 which is comparable to the 1,25,906 households reported by census of India in 1991. The small variations between model runs as shown in Table 3 also show the robustness and reliability of the integrated model.

**FORECAST FOR THE YEAR 2001**

Changes in the transport network characteristics are likely to cause changes in the location of activities and travel behaviour of the population. Therefore, it is useful to simulate the location and travel behaviour of the population in the future with and without improvements to the transport system. This will also evaluate the model's capability for policy testing. The improvements in the transport network include the additional transport links which are expected to be constructed by the year 2001 and all proposed improvements to the existing network.

As indicated earlier, the model operates with the total employment for the work zones. The total employment level of the city was projected for the year 2001 on the basis of trend analysis and the development policies initiated by the planning authorities. The total employment was then distributed to zones on the basis of the base year employment distribution and other zonal employment development parameters.

Two different options are considered in forecasting the population distribution and their travel behaviour. The first has no improvements to the transport network. The second includes all of the planned improvements to the network.

**Option I with No Transport Network Improvements:**

This option assumed that there are no changes in the transport network from the base year situation. Therefore, the travel times and travel distances are the same as they were for the base year case. However, the accessibility values are not same due to changes in the service employment levels in the zones.

**Option II with the Improved Transport Network:**

This option takes into account the improvements in the transport network which are expected to be made by 2001. Therefore, the travel times, travel distances and accessibilities will be different from those in Option I.

To make the results of the two of simulations comparable, the same population values with their corresponding household and other characteristics must be used in the simulations. Thus, the same seed values were used to generate the random numbers for both options and there was no variation in the socio-economic characteristics of the population. In a way, it can be said that the two simulation runs handled the same population for different transport network conditions. The test evaluates the population’s locational behaviour for different transport network configurations.
The model was run for both policy options and it was found that there is no significant difference in the spatial distribution of the population for the two options. This suggests that the proposed improvements in the transport network for 2001 are insignificant and will only marginally influence the locational choices of the population. The reason for this is that the network improvements did not change the accessibility and travel time patterns in any significant manner. The model results show high growth in zones which are adjacent to the well-served public transport corridors. This is to be expected as most commuters are captive to transit due to low levels of vehicle ownership. The simulation results for the two network options are shown in Table 4.

The advantages of the micro-simulation strategy have been shown by a residential land use model. The individual submodels have been shown to perform very well for both calibration and prediction. The variation in choice making by individuals was considered innovatively by using the Monte Carlo method to make the simulation results more realistic. The socio-economic characteristics of the simulated population and their location pattern were found to closely match the observed distributions from the household survey.

Two alternative policy scenarios were formulated on basis of two network development options. These options were designed to determine the response of the same population’s location choices for two different transport networks. Changes to the accessibilities of the zones due to changes in the network characteristics and the employment levels in each zone did not generate any major shifts in the simulated locational patterns. The model forecasts for the horizon year 2001 showed a population growth of 41% in one decade and the zones with well-served public transport corridors were found to attract more population. The operational model developed here has an inherent advantage for considering different types of disaggregated policy evaluations and is free from any aggregation bias.

### Table 4. Study Area Characteristics from Forecast for 2001

<table>
<thead>
<tr>
<th>Study area characteristics</th>
<th>Year 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>8,23,381</td>
</tr>
<tr>
<td>Number of households</td>
<td>1,77,246 ± 3,080</td>
</tr>
<tr>
<td>Mean household size</td>
<td>4.7 ± 0.80</td>
</tr>
<tr>
<td>Number of vehicle in the study area</td>
<td></td>
</tr>
<tr>
<td>Cars</td>
<td>17,657 ± 1,243</td>
</tr>
<tr>
<td>Two-wheelers</td>
<td>38,816 ± 1,306</td>
</tr>
<tr>
<td>Bicycles</td>
<td>76,924 ± 1,721</td>
</tr>
<tr>
<td>Vehicle ownership of households</td>
<td></td>
</tr>
<tr>
<td>No. of households with 1 car</td>
<td>13,818 ± 1,204</td>
</tr>
<tr>
<td>No. of households with 2+ cars</td>
<td>1,787 ± 142</td>
</tr>
<tr>
<td>No. of households with 1 two-wheeler</td>
<td>31,722 ± 1,282</td>
</tr>
<tr>
<td>No. of households with 2+ two-wheelers</td>
<td>3,335 ± 196</td>
</tr>
</tbody>
</table>
REFERENCES


