Validating a Bus Operations Simulation Model

Kevin T. Ryan

Department of Computer Science

University of Kansas

Lawrence, Ks. 66045

Abstract

An existing but unvalidated digital simulation model of a system of urban bus routes is reviewed. A data collection project is described which was to yield sufficient field data to allow at least partial validation of the model. Data collection and analysis were completed, although with some difficulty. Previous field results were reviewed and found to be reasonably consistent. High levels of variance were observed so that model inputs derived by averaging, were very approximate. A theoretical fault in the model was also found. Partial validation was attempted, but with limited success. The project provided insights on the model, the problem of bus scheduling, and on the methodology of simulation validation.

1. INTRODUCTION

In the developed countries of the world, traffic chaos and energy shortages have recently caused a revival of interest in urban transport systems. At the same time, in underdeveloped countries the vast majority of urban dwellers will be dependent on public transport for the foreseeable future. In developed countries research on urban bus systems has attempted to increase the efficiency and reliability of the service, since recent studies (e.g. Markham (1)) have shown that car owners seem more sensitive to quality of service than to cost. One particular difficulty is in maintaining a reliable and sufficient service during peak periods ("rush hours") when demand is at its highest but traffic congestion at its worst. While this is by no means the only problem area facing a bus manager, it would appear to be well suited to the application of Operations Research methods. Analytic studies of simple bus systems are numerous. For example, Potts and Tamlin (2) explained the observed phenomenon of buses pairing along routes as being due to "self-generated instabilities." Bullock (3) extended analysis to include bus capacities and some variation of bus journey times and passenger demand. Any further extension of an analytical model to include a limited fleet operating on a group of overlapping but distinct routes seemed to be unpromising, especially as many of the variables involved might be represented stochastically and be highly correlated, so research moved towards digital simulation of bus systems using observed data as input. Furthermore, a reliable, cheap and easy to use simulation model would facilitate the testing not alone of alternative schedules but also of different bus types and capacities, as well as giving estimates of the sensitivity of the quality of service to variations in passenger demand or in travel time along any or all of the routes. Before placing any confidence in such a model we must measure its validity, that is the accuracy with which it replicates the real world. This paper outlines an existing model and describes an attempt to validate it using field data.

2. The Model

2.1 Overview

The Bus Operations Simulation Model (BUSOPS) was developed at the IBM (United Kingdom) Scientific Centre in Peterlee, England, by J.P.D. Gerrard (IBM) and D. Brook (University of Newcastle-Upon-Tyne). It is fully described by Gerrard and Brook (4) and by Ryan (5). While it drew upon aspects of previous models, notably those of Dekindt and Griffe (6) and Oliver (7), the model was the first to tackle the problem of the congested "urban corridor," along which passengers are carried by a number of routes. These routes share a common stem before branching to separate outlying termini. Buses may overtake one another and limited stop services may be operated.

The model is a discrete event stochastic digital simulation model in which the main entities are buses, passengers and bus queues. In operation the model is event-driven with the arrival of a bus at a stop as the main event. Buses are traced through the system from stop to stop

1979 Winter Simulation Conference

while queues are created and serviced with bus arrival. The primary output statistic is the average time spent waiting by the passengers who board.

The operating environment of a bus route or routes includes factors which can be varied by the manager and external factors which are outside his control. The former can be called:

Internal or controlled factors:
(a) number and type of buses
(b) schedule of services for each route
(c) fare-collection method (e.g. one man operation)

while the latter are:

External or uncontrolled factors:
(d) bus journey times between any two stops
(e) number of passengers arriving at any stop and the routes they wish to board
(f) number of passengers alighting at any stop
(g) time required to board and alight.

Controlled factors are represented directly in BUSOPS. They are deterministic in the real world and are modelled deterministically. Specification of the bus management strategy is part of the model input.

Modelling of the uncontrolled factors involves the use of hypothesized patterns which may be supported by observed data. To each of these patterns BUSOPS adds a random variation (positive or negative) to represent the unexplained (and hence "random") variation which is observed in real life.

2.2 Bus Journey Times

Since the model is primarily designed for rush-hour simulation, the variation in journey time between each pair of stops is taken to have a particular peaked form:

$$f(x) = c + \kappa(\rho x - \tau)e^{(\rho x - \tau)} \quad \text{for } x \geq \frac{\tau}{\rho}$$
$$f(x) = c \quad \text{for } x < \frac{\tau}{\rho} \qquad (1)$$

where
$f(x)$ = journey time between two stops (i.e. one link) at time x.
c = constant (off-peak) journey time for this link.
κ = scaling factor for this link.
$\frac{\tau}{\rho}$ = time peak period begins.
$\frac{1}{\rho}$ = time to reach maximum.

The parameters τ and ρ do not vary from stop to stop and govern the shape of the curve, which has a sharp rise followed by a slow decay, while c and κ are determined by observation for each pair of stops. A stochastic component is added to the result given by equation (1). This "excess" component also has the gamma form and is adjusted to have mean of zero with a limit of $-\frac{1}{3}f(x)$ and a long positive tail giving occasional large positive deviations. Furthermore, the excess times (positive or negative) of successive bus trips between the same pair of stops are serially correlated with an effect which is exponentially damped out over time. Thus if X_{i-1} is the excess time of the previous bus which covered a given link at a time t_{i-1}, then the correlated excess time (X_i) of the bus traversing the same link at time t is given by:

$$X_i = D\,X_{i-1} + \sqrt{(1 - D^2)}\;T_x \qquad (2)$$

where
T_x = stochastic (excess) component of this bus (uncorrelated)
$D = e^{-T_d|t_i - t_{i-1}|}$
T_d = "serial correlation discount factor"

The values given by equations 1 and 2 are added to give the total journey time over a given link.

2.3 Passenger Arrivals and Classification

Individual passengers are not traced in the BUSOPS model.

Instead, passengers are divided into types depending on the choice of routes available to them to reach their destination. Clearly, if there are n distinct routes in a system, there can be at most 2^n-1 different passenger types. The passengers on buses or in queues are grouped by passenger type. The pattern of arrivals over the period to be simulated is input for each stop and is used to calculate the number of simulated passengers added to a bus queue between one bus arrival and the next. Inter-arrival times are generated with a Poisson distribution whose parameter is the average arrival rate over the time period (t_2-t_1) between two bus arrivals. A passenger is counted for each inter-arrival time generated and the process ceases when the total of inter-arrival times exceeds t_2-t_1. Distribution of these arrivals by type depends on the "attractiveness" of potential destinations for each passenger type. A measure of each stop's "attractiveness" as a destination is calculated as shown below.

Since the model is oriented towards measuring the waiting time of passengers, rather than their "trip-time" or travel time, origin-destination data is not required. The model aims to replicate the actual number boarding and alighting at each stop, but assumes independence of origin and destination. Instead, the origin-destination matrix for the stops in the system is taken to consist of entries of the form $\lambda_i\mu_j$ where λ_i is the loading index of the origin stop i, and μ_j is the unloading index of the destination stop j. Each μ is an overall measure of a stop's relative "attractiveness" as a destination. In the case of outbound passenger flow the origin-destination matrix for a system of n stops, numbered in ascending order in the outward direction, will yield a set of non-linear equations of the form:

$$R_i = \lambda_i \sum_{j=i+1}^{n} \mu_j \qquad \text{for } i = 1,2,3,\ldots n$$
$$C_j = \mu_j \sum_{i=1}^{j-1} \lambda_i \qquad \text{for } j = 1,2,3,\ldots n \qquad (3)$$

where

R_i = the sum of the ith row of the origin-destination matrix (and hence the total number boarding at stop i)

C_j = the sum of the jth column (and hence the total alighting at stop j)

If the values of R_i and C_j are known, for all values of i and j, the set of equations in equation (3) yields solutions of the form:

$$\mu_j = a_j t \text{ and } \lambda_j = \frac{b_i}{t}$$

where

a,b = constant

t = parameter.

In practice an arbitrary value is assigned to μ_n and the system solved explicitly by back substitution. The unloading indices are calculated independently and are input to the model. They are used as follows to estimate the proportion of each passenger type in any group of arrivals. The probability of passenger arrivals at stop i being of type k is given by:

$$P_{ik} = \frac{\Sigma\mu_j \; \forall_j \varepsilon D_{ik}}{\Sigma\mu_j \; \forall_j \varepsilon D_i} \qquad (4)$$

where

D_i = set of all destination stops from stop i

D_{ik} = set of all destination stops, for passengers of type K, from stop i.

2.4 Passenger Alighting

The probability of passengers of type k alighting at a given stop i is given by an expression similar to equation (4):

$$O_{ik} = \frac{\mu_i}{\mu_i + \Sigma\mu_j \forall_j \varepsilon D_{ik}} \qquad (5)$$

where

D_{ik} = set of all possible destinations beyond stop i, for passengers of type k.

When a bus arrival is being processed, the probabilities given by equation (5) are applied to each passenger class on board the bus and the sum of this calculation gives the total number alighting.

In considering alighting and boarding, only the outbound flow of passengers has been mentioned. Inbound flow would require similar calculations, which would be simplified by the smaller number of passenger types (only one, unless there is a limited stop service).

2.5 Boarding and Alighting Times

The time required for passengers to board and/or alight at a stop is taken to be:

$$T_{bap} = \text{Max.}(n_b t_b, n_a t_a) \quad \text{for parallel board/alight} \qquad (6a)$$

and

$$T_{bas} = n_b t_b + n_a t_a \quad \text{for sequential board/alight} \qquad (6b)$$

where

n_b, n_a = number of passengers boarding, alighting

t_b, t_a = time required for one passenger to board/alight.

An overhead for each board/alight operation is the acceleration and deceleration time penalty which is effectively an additional constant term to equations (6a) and (6b) above.

3. The Project

3.1 Objectives

A model which has been implemented as a computer program must be made credible before any reliance can be placed on it. The term "validation" has been used to describe this entire process, for example by Teorey (8), but others, including Schatzoff and Tillman (9) and Mihram (10, 11) distinguish between verification and validation. Following Mihram (11) we can define verification as:

"The determination of the rectitude of the completed model vis-a-vis its intended algorithmic structure"

Verification therefore includes not only the normal debugging of the computer programs used to implement the model, but also tests, whether quantitative or qualitative, which confirm the desired behavior of all or part of the model.

Once a model has been verified, its adequacy as a representation of the real system must be established. This validation Mihram (11) defines as:

"The comparison of responses emanating from the verified model with available information regarding the corresponding behavior of the simulated system"

Validation involves comparing the model's behavior with that of the real system over as wide a range as possible. This ideal may not always be attainable, especially where variation of the real world system is costly or difficult.

In the case of the BUSOPS model it was possible to attempt first to validate some of the underlying assumptions before proceeding to overall model validation. Field data, which was needed as input to the simulation model, could first be analyzed with respect to the BUSOPS modelling hypotheses.

In particular it was necessary to test the following assertions:

(a) Profiles of passenger arrival rates are repeatable from day to day.

(b) Passenger boarding and alighting times are accurately represented by equation (6).

(c) Totals boarding and alighting at each stop are reasonably consistent from day to day.

(d) Journey times between pairs of stops are repeatable from day to day and follow the Gamma form of equation (1).

Once these hypotheses had been tested, validation by comparing system and real world outputs, could proceed.

Ideally, validation would require the greatest possible variation in both the controllable inputs (bus types, frequencies and capacities) and the uncontrollable inputs (passenger arrivals, traffic congestion, total numbers alighting at each stop) so that the accuracy of the model in predicting the real system's behavior can be tested over as wide a range as possible. This is clearly impracticable in the case of bus route simulation because of the difficulty and cost of changing even the controllable inputs. The alternative method, which is that of collecting sufficient data to give a wide range of inputs, is severely restricted by the limited time and money available. However, if the model's primary outputs (in this case the average passenger waiting time at each stop) are derived during simulation, from all of the model's inputs and hypotheses, and if for a given setting of the inputs the observed and simulated values of the outputs do not differ significantly, the model can be said to be partly validated.

As the definition given above implies, verification does not require the use of field data. Instead, one-sample statistical tests are applied to test the distribution of model output. However as the verification of a large model will inevitably involve error detection (or debugging) it is advisable to have realistic inputs which roughly correspond to those of the real world. For example, the adequacy of a fixed length queue (as used in BUSOPS) depends on the service and arrival rates. A subsidiary objective of data collection was therefore to provide a realistic set of testing data which could be used during verification of the model program. This approximate data was updated as necessary before being used in the attempt to validate the model.

3.2 Data Collection

Once preliminary contact had been established in June 1973 with Coras Iompair Eireann (C.I.E.), who operate the Dublin City bus services, it was necessary to specify exactly what data was required. It was decided that over a two to three week period the following data on the evening rush hours were required:

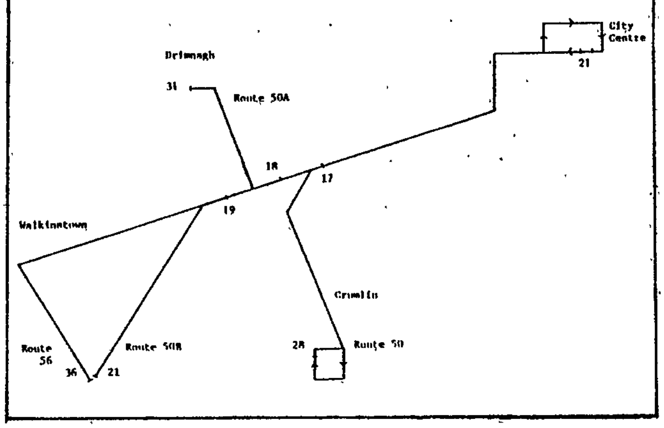
(a) Passenger arrival rates at stops.
(b) Total numbers boarding/alighting at stops.
(c) Journey times between each pair of stops.
(d) Information of bus types, route structure and schedules.
(e) Passenger boarding and alighting rates.
(f) Passenger waiting times at stops.

Items (a) to (e) are model inputs, while items (a), (b), (c) and (e) are necessary to test the validity of model assumptions. Item (f) is required for model validation as described in 3.1 above.

Permission was obtained from C.I.E. to collect data over the period 18 July, 1973 to 3 August, 1973, inclusive, but in the time available it was not possible to obtain union consent for observers to travel on board the buses while collecting data. While this was not an insurmountable difficulty, it caused considerable complications in the analysis of the journey time data.

The group of routes numbered 50/50A/50B/56, which run from Dame St. in Dublin City Centre to the Crumlin/Walkinstown/Drimnagh housing areas in the southwest of the city, were chosen for the data collection project. A sketch of the route structure, with the termini and some stop numbers marked, is shown in Figure 1. From the model viewpoint these routes had the advantage of a long common "corridor" of 17 stops, little "interference" from other routes and a high level of congestion in the city centre terminus area. Furthermore, unlike many of Dublin's bus routes, they were not yet radio-controlled.

Figure 1 DUBLIN BUS ROUTES USED IN DATA COLLECTION



Twenty-six students were employed on a part-time basis to collect data during the evening peak period. They were briefed individually before beginning data collection. Three different data forms were used:

Form A: for passenger arrival data only
Form B: for bus timings (arrival and departure) and for passenger alighting data.
Form C: for a sample of passenger waiting times.

The following can be noted about the forms:

(a) No boarding data was collected, as it is only the total number boarding that is required and this can be deduced from the other data.
(b) Since both Form A and Form C require the full-time attention of the observer, it was necessary to obtain the remainder of the data from Form B, if a maximum of three observers were to be used at any one stop at the same time. Form B may therefore be over-demanding on the observer.
(c) In Forms B and C the times were recorded in minutes and seconds after the hour. It was appreciated that one second accuracy was unlikely, but one minute accuracy would have been insufficient, particularly on the shorter links between stops.
(d) Blank spaces on Form C were used to describe the passenger who began waiting

so that he or she would be recognized when boarding later.

(e) In addition to the route number, Form B observers also recorded the right-most three digits of the bus registration number (vehicle identification).

Since the set of routes chosen had about 40 stops, it would have required 120 observers to monitor all types of data, at all stops, each day. Instead, it was decided to "cover" stops with a frequency relative to their importance, concentrating the A and C (arrivals and waiting time) data collection in the area of greatest passenger boarding. Form B data collection (bus timings) on the other hand, which was required for all stops in the system, involved having two observers with synchronised watches taking timings at consecutive stops. Consequently, when the data collection team assembled for briefing each day at 4 p.m., watches were synchronised, and when debriefing after 6:30 p.m., the accuracy of synchronisation was measured. Results from this test are described below. The observers who kept the most accurate times were assigned to Form B data collection, since it was only for journey-time estimation that a high degree of synchronisation was essential.

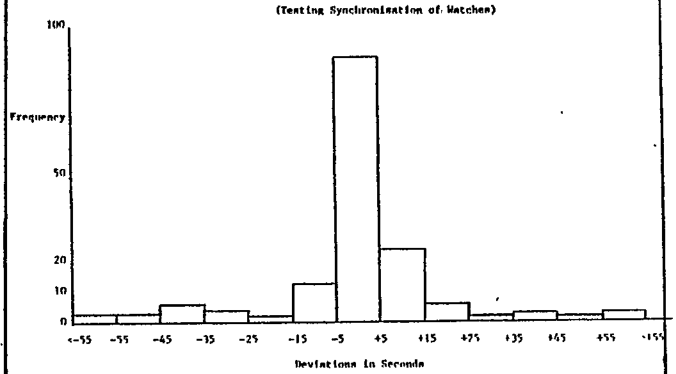
Throughout the data collection operation the students were kept informed of the project's purpose and progress, which helped to maintain a high level of interest and diligence. No antagonism was encountered from the bus crews, who seemed to have been reassured by their management and trade unions.

3.3 Difficulties and Errors

The synchronisation test at debriefing required each observer to record the time of a staged event, under classroom conditions. Deviations from the daily mean were accumulated for the days on which tests were made and were grouped, giving the histogram in Figure 2. It can be seen that 90% of the timings were within ±15 seconds of the mean. The deviations are attributable to a number of factors. The observer may have erred in synchronising his watch before data collection. The observer's watch may have "drifted" and lost accuracy during the data collection period, or the observer may have erred in reading the watch while timing the event. The first of these errors was minimised by giving repeated time checks until each observer was satisfied with his

Figure 2 FREQUENCY DISTRIBUTION OF DEVIATIONS FROM THE DAILY MEAN

(Testing Synchronisation of Watches)



watch's synchronisation. The effects of the second and third types of error on Form B data were minimised by assigning the more accurate observers to collect this data.

Nevertheless, because of the remaining errors in synchronisation and because of the circumstances of field data collection, the accuracy of bus timings cannot be taken to be better than ±10 seconds. Where the journey time between stops was of this order of magnitude, as it was on some outer suburban links, deductions about bus journey time variations cannot usefully be made.

A second major difficulty was specifying the base time for timings. Initially the field instructions were for times to be taken "from the time data collection began." As this complicated timing, especially when observers started a few minutes late, it led to errors in recording minutes. From the second day onwards, timings were taken from the hour.

There was less difficulty with Form A (arrivals) and Form C (waiting times). On a few occasions observers at the city centre terminus lost count of arrivals during a peak period and estimates were used for the numbers arriving. Observers collecting waiting time data were asked to record as high a proportion of the arrivals as they could accurately maintain. This varied from "almost all" at stops having few boarders to "about one in eight" at city centre termini.

Collectors of arrival (A) and waiting (C) data reported that lack of queue discipline made observations difficult, on a few occasions.

3.4 Data Validation

In all collection projects the adequacy of the advance planning, form design, field instructions, etc., can only be judged when data has been successfully validated. The BUSOPS project allowed insufficient time for deliberation and planning. As a result, data validation was difficult. Passenger arrival data was least complicated and consequently almost free of (logical) errors.

Timing proved to be the major problem. Waiting time data were difficult to check as the change of hour was not explicitly shown. Observers recorded only minutes and seconds past the hour. In the bus journey timings this was compounded by the sheer volume of data required, by the field conditions (eg simultaneous bus arrivals, rapidly passing non-stop bus) and by the difficulty of identifying buses. Errors of observation and/or timing led to spurious timings and/or fictitious buses. The ideal solution to most of these difficulties would have been to have observers on board buses. This proved impossible and the present method had to be adopted at short notice. In consequence the Form B data required an excessive amount of manual checking and correcting, so that the effort expended in obtaining marginal data was out of all proportion to its usefulness. Nevertheless all data was eventually verified so that the analysis could proceed.

4. The Results

4.1 Bus Journey Times

Previous studies of bus journey times are numerous (e.g. 3,12,13). One of the more recent is that of Charlesworth and Walmsley (14) who studied the variation between different periods of the day and between single days in July 1974 and November/ December 1974. As Charlesworth and Walmsley did not consider time of day, other than to distinguish between peak and off-peak periods, they did not hypothesize any particular form of peak in the journey time variation. Furthermore, they eliminated day of the week variation by sampling, as far as possible, on the same day. However, their conclusion that, where road conditions are unchanged, there is no evidence of significant seasonal variation between November and July, serves to mitigate one limitation of the present study, where all the data were collected in the same season (i.e. in July/August of 1973).

A more relevant study for the present purpose is that of Markham and O'Farrell (15), where not alone was the objective to study variations in bus journey times between routes, time of day and days of the week, but in addition, among the routes studied were the 50A and 50B both of which were included in the present project. Hence it was possible to compare directly the results obtained. The only major differences between the two studies were:

(a) Markham and O'Farrell had a four week survey period in June 1971, whereas the present study took place approximately two years later and was limited to two and a half weeks.

(b) Markham and O'Farrell had observers at four fixed locations on the routes, on every day of the study, whereas the present study monitored almost all the stops on the routes, but with varying frequency.

A preliminary comparison could be made between the results obtained in the two studies by taking a long link (stop 3 to stop 17) and analyzing its average timings within fifteen minute intervals. This was compared to the results of Markham and O'Farrell, who do not provide any raw data but base their analysis on fifteen minute interval averages (commencing at 16.45). The links being compared do not correspond exactly. They do however have a long section in common so that the form of any variation, if not the actual values, should be comparable. From the comparison given in Figure 3, it can be seen that the histograms are similar in shape. The overall average in Markham and O'Farrell, is higher, at 19.5 minutes, than the result for link 3 to 17 which is 16.725 minutes. Although the latter is the longer link, the interval measured by Markham and O'Farrell included a greater length in the more congested inner city area. The results were scaled to have equal means and the resulting histograms compared using the Chi-square test. A test statistic of 0.045 with 6 degrees of freedom was recorded. This shows very high correspondence between the two sets of data.

SPSS (16) routines were used to analyze the raw data. Correlations were computed between T, the journey time in seconds, and the following functions of D, the minute of departure,

D, D^2, D^3, D^4, e^D, De^D and $\log_e D$

Correlations were low, the highest (0.37) being with D^4.

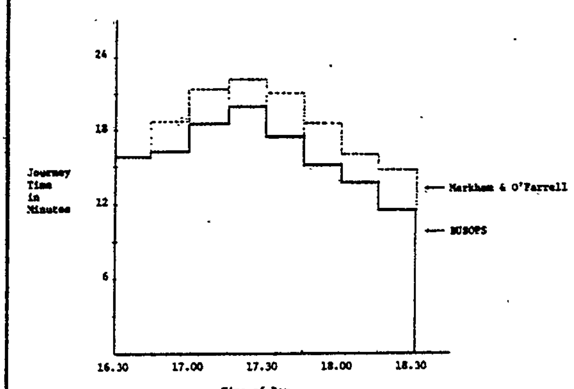


Figure 3 Comparison of Two Studies of Interval Average Journey Time

A stepwise regression was performed with T as the independent variable and using the above functions of D. The process stopped after four steps yielding the equation:

$$T = 0.29D^4 + 245.80D - 4.81D^3 - 179.80\log_e D + 687.35 \quad (7)$$

The dependent variables are shown in equation (7) in the order in which they entered the calculation. The fit of this equation to the data was very poor giving an R^2 statistic of 0.29.

It cannot be concluded from this analysis that any of the proposed forms of variation in journey time with time of day, is confirmed by the data. Neither the quadratic form deduced by Markham and O'Farrell nor the gamma form (te^t) hypothesized by Gerrard and Brook and used in BUSOPS provided an adequate fit to the field data.

The overall conclusion must be that although an overall daily trend was discernible, and although it did resemble that detected in an earlier study, the factors included in the analysis, which were the time of day the journey began and, to a lesser extent, the day of the week, were sufficient to explain only a small proportion of the variance. The high "random" factor in the observed journey times might be attributed to one or more of the following factors:-

(a) Observation errors.
(b) Differences between bus drivers' behavior.
(c) Variations in weather conditions.
(d) Traffic control either by police or traffic signals.
(e) Accidents, emergencies, bomb hoaxes, etc.
(f) Variations in individuals' travel habits.

If further analysis justified the inclusion of any of these factors in a model of the journey times, it would require a higher level of detail in the model.

In the meantime we must conclude that bus journey times show much greater variance than the BUSOPS model of equations (1) and (2) allows.

4.2 Passenger Arrival Rates

To establish the repeatability of the passenger arrival rates there are two aspects to be considered. The variation in total arrivals, at each stop, from day to day, and the variation at each stop in the pattern of arrivals during the time period being studied.

Figure 4 shows the mean, bracketted by the standard deviations, of the arrivals at all stops where data was collected. The number of days data upon which the calculation is based is shown in parentheses. It can be seen that the amount of variation was large at all stops. Only at the less used stops (numbers 5,7,8,9, and 17) does the pattern resemble a Poisson distribution with mean and variance (σ^2) approximately equal. As the busier stops have more variation than would be expected for a purely Poisson distribution, the day of the week might have a noticeable effect. The following table gives the frequency of maximum and minimum total arrivals, for all stops, by day of the week. e.g. There was only one stop for which Wednesday provided the maximum arrivals.

DAY	MON.	TUE.	WED.	THU.	FRI.
MAXIMA	0	1	1	4	5
MINIMA	5	0	4	1	1

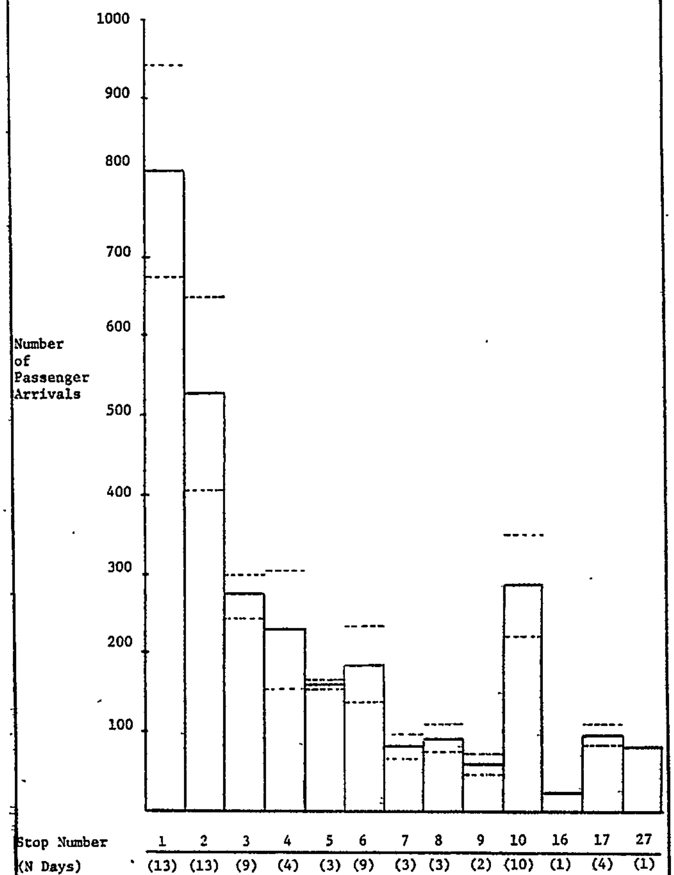
It is clear that Thursday and Friday have "more" arrivals, while Monday and Wednesday have "less". There is insufficient data to prove the significance of these variations.

The conclusion regarding the day to day variation of passenger arrival totals is that because of the high level of variation observed, the construction of a very accurate model would entail consideration of day of the week effects, and hence a considerably increased data collection effort. The BUSOPS model would require fundamental modification, to include day of the week effects, before it could be expected to model passenger arrival totals with great accuracy.

The second pattern in the arrival data to be studied was the variation with time of day. For each stop a histogram of ten minute interval totals was computed. Two typical results are shown in Figure 5. Stop 3 was in the city centre, stop 6 was in the inner suburbs. Each was observed on 9 days. It can be seen that arrival rates increase from 16.30 to a peak at or shortly after 17.00 hours. Then there is a slight drop followed by a second peak at 17.30 hours, after which the arrival rate declined until 18.30 when data collection ended. This accords well with the fact that most Dublin business premises close at 17.00 or 17.30. Demonstrating a repeatable pattern, for example by using X^2 test, was not attempted because of the high observed variation and the (relatively) small number of days for which data was available.

Figure 4 OBSERVED AVERAGE TOTAL ARRIVALS BY STOP

Standard Deviation Shown as -----



In spite of the limitations described above it was possible to derive the input data for the BUSOPS model. Given the relatively coarse level of detail in the model, and the consequent inherent error levels in simulation and outputs, passenger arrival rates were sufficiently regular to justify the BUSOPS assumptions.

4.3 Passenger Boarding and Alighting Rates

The final data to be used in checking the BUSOPS model's assumptions were the timings for passenger boarding and alighting operations. Three parameters are input to the model, and together they determine the total time needed for a bus to stop, to allow passengers to board or alight and for the bus to regain its normal driving speed. The three parameters are:-

(a) The average boarding time per passenger: t_b

(b) The average alighting time per passenger: t_a

(c) The additional time "penalty" incurred in decelerating and accelerating a bus, regardless of the number of passengers boarding or alighting.

In the case of buses having parallel boarding

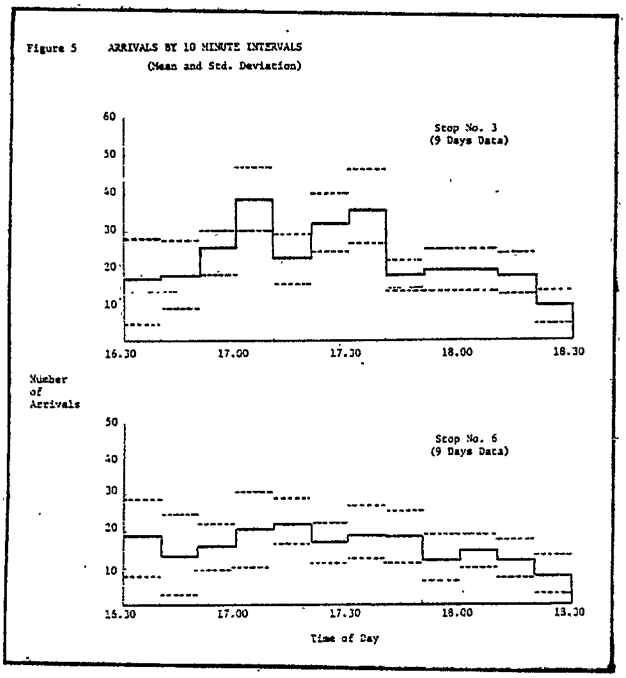


Figure 5 ARRIVALS BY 10 MINUTE INTERVALS (Mean and Std. Deviation)

and alighting, and all the buses in the present study were of this type, the boarding and alighting time (T_{ba}) is calculated in BUSOPS using the formula:

$$T_{ba} = \text{Max.}\ (n_b t_b,\ n_a t_a) \quad (8)$$

where n_a and n_b are the number of passengers boarding and alighting respectively. The model makes no explicit allowance for so-called "dead time" (i.e. there is no constant term in equation (8)), but if such an allowance was required it could be included in the acceleration/deceleration penalty.

The data collection was limited to approximately one hour of the evening peak period on two successive days. A straight and uncongested stretch of the route was selected and four observers were positioned as follows:-

```
direction of travel ——→     0 Bus Stop
 ↑——————————————————————————|——————————↑
 1                         2,3         4
```

Observers 1 and 4 recorded the registration number and time of passing of each bus. If the bus halted at the stop, observer 2 recorded the exact time the bus stopped (wheel stop) and the time the bus began to move off. Meanwhile observer 3 counted the number of passengers who boarded and alighted from the bus. Although the experiment was limited, partly by choice because of the wide range of previous studies available (see Chapman (17) for a comprehensive survey of boarding time studies), and partly by necessity due to the relatively high number of observers required, some useful results were obtained nevertheless.

The derivation of the acceleration/deceleration time penalty proved difficult as there were only 3 non-stop buses observed by observers 1 and 4. These gave an estimate of the non-stop journey time from observer 1 to observer 4. The journey times for buses that halted were adjusted to allow for the time spent stationary while passengers were boarding or alighting, and the average nett journey time from observer 1 to observer 4 was calculated. The result was approximately 4 seconds greater than the non-stop average and this value was taken as an estimate of the acceleration/deceleration time penalty.

The total number of observations of boarding and/or alighting was twenty-five. This is much less than the minimum (100) recommended by Chapman. An additional complication was that most of the buses observed had both boarding and alighting at the stop. A function of the form given in equation (8) was chosen, and by trial and error a reasonable fit was obtained for $t_b = 4.5$ and $t_a = 1.5$. The work of Ellis (18), which was conducted around the same time, gave:

$$\text{for boarding } T_b = 6.3 + 1.1 n_b \quad (9a)$$

$$\text{for alighting } T_a = 7.3 + 3.5 n_a \quad (9b)$$

These figures correspond reasonably well with equation (8), although for small values of n_a and n_b the absence of a constant (or intercept) term makes the T_{ba} value of equation (8) somewhat smaller than those of (9a) and (9b). When viewed in the light of the results surveyed by Chapman the marginal, and in this case average, boarding time of 4.5 seconds per passenger seems a little higher than average for buses with on-bus ticketing. Judging from Chapman most results lie in the range 2.5 to 4.5 seconds. To determine whether the

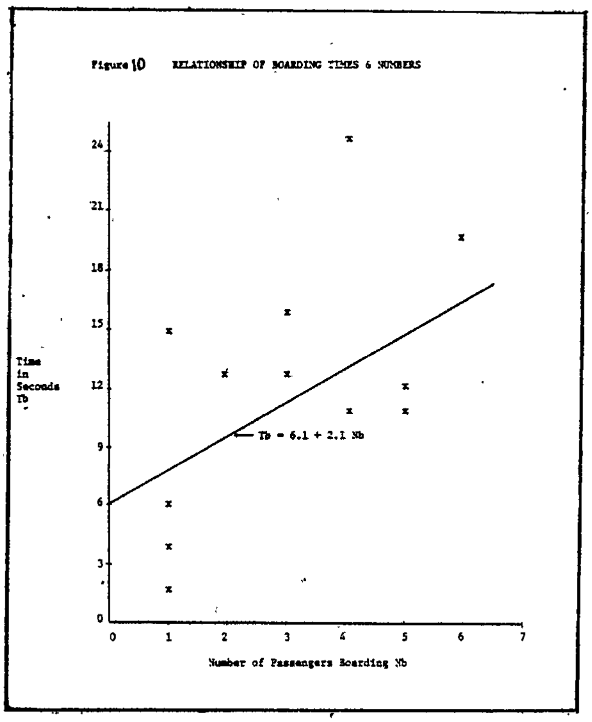


Figure 10 RELATIONSHIP OF BOARDING TIMES & NUMBERS

present study's results would be better fitted to the conventional model of a "dead-time" plus a marginal boarding time per passenger, it was first of all necessary to select only observations where the boarding time was dominant. A straight line was fitted to the twelve observations thus obtained, and yielded the equation:

$$T_b = 6.1 + 2.1n_b \qquad (10)$$

This is shown, with the data points, in Figure 10. It can be seen that all the observations were for six or less passengers boarding and the spread from the fitted line was quite large. The "dead-time" figure, given by the y-intercept, is 6.1 seconds, which is large by comparison with the results surveyed by Chapman but accords quite well with the result of Ellis, given from equation (9a) as 6.3 seconds. Although it is generally believed (Jenkins 19) that the marginal boarding time decreases for a high number of passengers boarding, the value of 2.1 seconds, which was obtained with relatively low numbers boarding, is quite low on the range of values surveyed by Chapman.

The relative importance of boarding and alighting times as factors affecting the reliability or punctuality of bus services has not been definitely established. Chapman (17) states that "as a rule of thumb one fifth of total journey time is spent at bus stops." More relevant for the present study is the sensitivity of the average passenger waiting time to changes in the average time taken to board or alight. It was found from experiment that within the range of realistic values the average waiting times in the BUSOPS model showed a lack of sensitivity to variations in board/alight times. Whether this is the case in reality could not be established from the present study where there was only one type of bus in operation. For the purpose of model input however, this insensitivity in BUSOPS would have made it wasteful to expend excessive effort in obtaining highly precise estimates of the actual boarding and alighting times.

Indeed, we can conclude that the BUSOPS assumptions regarding boarding and alighting rates, while crude and approximate, were sufficient for this relatively simple model.

4.4 A Theoretical Flaw

As described in Section 2.3 above, the observed boarding and alighting totals are used to compute a set of indices (μ_i, i = 1 to n), which are described as measures of the "attractiveness" of each stop as a possible destination. The method of solution of the set of equations (given as Equation 3 above) which should yield the values of μ_i for i = 1 to n is illustrated in the original BUSOPS report for a single route system. The procedure to be adopted for a branched outward flowing system is not given in detail. Instead it is asserted that "the system of equations can obviously be augmented to include one or more branches from the mainstream" (Gerrard and Brook, pp. 16). In fact, this proved impossible to implement.

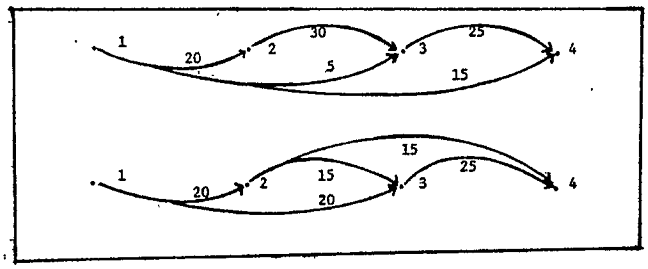
Within the BUSOPS model, passengers arriving at a stop are distributed by type, using the relative cumulative attractiveness of all stops which are reachable from this stop by a given route.

The derivation of the unloading or alighting indices (the μ's) in a branched network is not defined. The omission of impossible values ($\lambda_i \mu_j$ where stop j is not reachable from stop i) from the Origin-Destination matrix can yield an underdetermined set of non-linear equations. The difficulty is due to a conflict between the BUSOPS assumption of independence of passenger origin and destination and its use of an O-D matrix in calculating load/unload indices, and can be illustrated as follows.

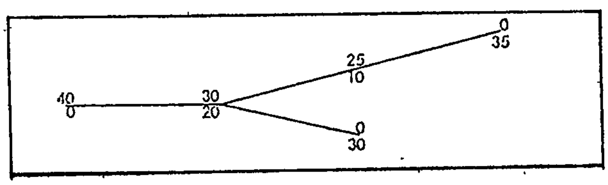
Consider a single route system (i.e. no branching), with passenger boarding and alighting totals as shown.

BOARD	40	30	25	0
ALIGHT	0	20	35	40

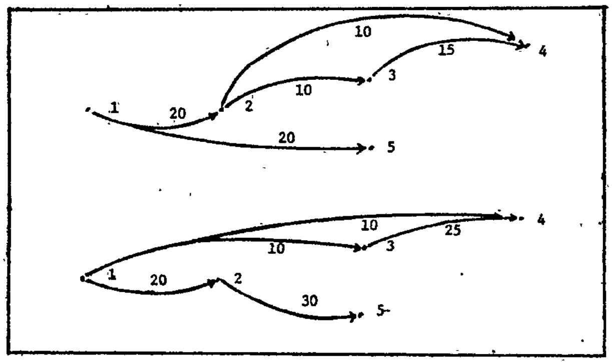
There are many origin-destination flows which would accord to this distribution, e.g. the following two:



Both of these conjectured flows, and all others that would yield the correct board/alight totals, give the same nett flow of passengers at any point along the route. This means that the loading factor (i.e. percentage occupancy) for buses is independent of passenger origin and destination and that therefore in this route system the level of service to passengers, as it is measured by BUSOPS, is also unaffected by the actual origin and destination pattern. This is only possible because BUSOPS considers only the waiting time to boarding in its measure of the level of service and takes no measure of, for instance, passenger in-vehicle time, or total journey time. However, if an example is taken where there are two routes, and hence three possible passenger types, the above argument fails. Consider the following branched system where boarding and alighting totals are again shown above and below the stop positions.



Among the possible origin-destination patterns that would conform to this boarding/alighting distribution are the following two.



It can easily be seen that in this system the level of service to passengers boarding different routes, as it is measured by the average waiting time to boarding, will depend on the loading factor (percentage occupancy) of buses on the desired route or routes. This is in turn determined by the choice of destination of passengers already boarded. In the above system the level of service to passengers boarding at stop 2 depends both on their ultimate destination and the destination of passengers arriving at stop 1 to board. It follows that even if the Origin-Destination matrix could be used to find values of "loading and unloading indices," the result would be an arbitrary choice from among many alternatives and would not necessarily reflect the actual journey patterns in the system.

As a method to yield a reasonably accurate set of values for the unloading indices (μ's) it was decided to solve four separate systems, one for each route being studied. Breakdown of totals alighting by stop and route was computed. Statistics for boarding by route were estimated using the passenger waiting time data and the overall boarding totals. Each route yielded an independent set of μ values and these were scaled, in proportion to the totals carried by each route, before being combined to give an overall index for each stop. These composite indices were then used as input to the BUSOPS model.

Subsequent results, given below, cast doubt on the validity of computing the μ values in this way. However it is difficult to see any other way of choosing a realistic solution to the set of equations given as equation (3) above.

4.5 Comparison of Observed and Simulated Results

The analysis of model output must always be an iterative process involving simulation, adjustment and further simulation. However a point can be reached, beyond which improvements in model validity, as measured by its accuracy in replicating the real world system, can only be schieved by a radical change in the model. When this point was reached with the BUSOPS model it was decided to evaluate the comparison of an average set of model outputs with those that had been observed in the field. An average of ten successive runs was taken to give an estimate of the average model output. The average numbers simulated as boarding or alighting, and the average passenger waiting times, were calculated manually, as the BUSOPS model provides no averaging facility. The average model output was then compared to the average observed statistics, for those stops for which a useful quantity of data had been collected. There were three distinct comparisons involved:

- Comparison A — Average total boarding (by stop)
- Comparison B — Average total alighting (by stop)
- Comparison C — Average passenger wait time (by stop).

It must be emphasized that Comparison A, using observed and simulated results, was largely a verification of the BUSOPS process for passenger generation, which had the observed passenger arrival rate as one of its inputs. Comparison B would test, not only the passenger generation process, but slso the realism of the BUSOPS alighting procedures, which in turn rely heavily on the indices of attractiveness (the μ's discussed above). This comparison would validate a part of the BUSOPS model. The principle measure of the model's validity, as had been intended from the outset of the project, was to be the comparison of passenger waiting times, Comparison C above. The average passenger wait time was not only a measure of the quality of service received, but was also a statistic which was a product of all the model's operations and hence could provide a useful validation criterion.

The comparison of observed and simulated boarding totals is given in Figure 6. The observed figures are averages, scaled to reflect losses to other routes. (Some of the lesser used stops were estimated from short samples.) It can be seen that overall, the fit is poor, and this is reflected in a χ^2 statistic of 237 for 14 degrees of freedom. The simulated results tend to be lower and are particularly so at the less used stops, numbers 11 to 18. This may be attributed in part to the fact that BUSOPS, as implemented, attempts to generate passengers before every bus arrival. If buses are frequent and arrivals are few, there will be a severe truncation effect. e.g. If .4 of a passenger is generated 5 times, there will be no arrival recorded. The deviation at stops 17 and 18 was even more noticeable. These stops lie at the end of the common stem. Passengers boarding here must require a specific route. Accordingly, their arrival rate may reflect the published (or observed) arrival times of buses on these routes. It is difficult to see how BUSOPS, or indeed any aggregate model, could accurately reflect these effects.

As the passenger generation method cannot be considered verified from the preceding comparison, the examination of passenger alighting total, observed and simulated, might be considered pointless. However, the actual total boarding, whether simulated or observed, was reasonably consistent and the alighting data was available for a larger number of stops. The observed and simulated alighting totals are shown in Figure 7. The goodness of fit is poor and is reflected in a χ^2 statistic of 249 with 24 degrees of freedom. Although the fit is quite good for stops 4 to 15, the difference is very large at some outlying stops. This may be caused by the shortfall in passengers generated for boarding (discussed above) or by the method of computing the alighting indices (μ's). Further study is needed to quantify these effects.

Figure 6 COMPARISON OF SIMULATED AND OBSERVED BOARDING TOTALS

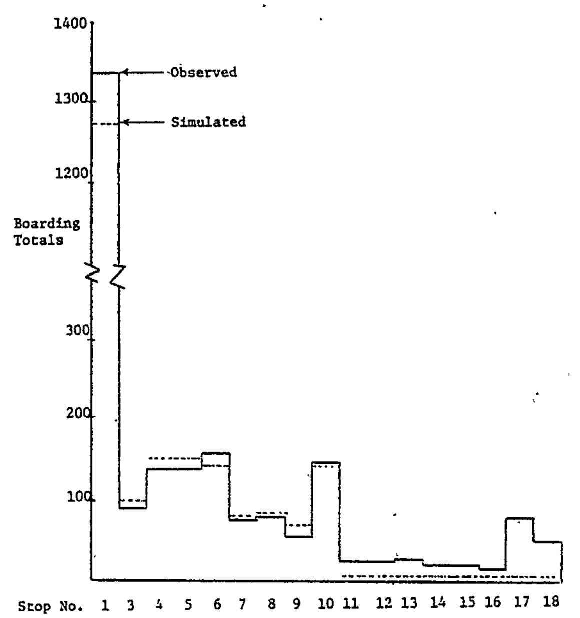
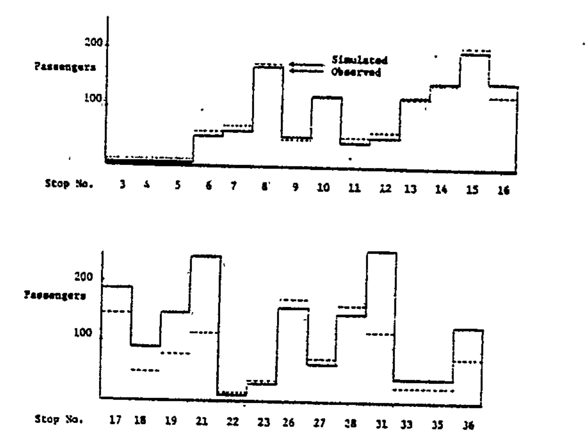


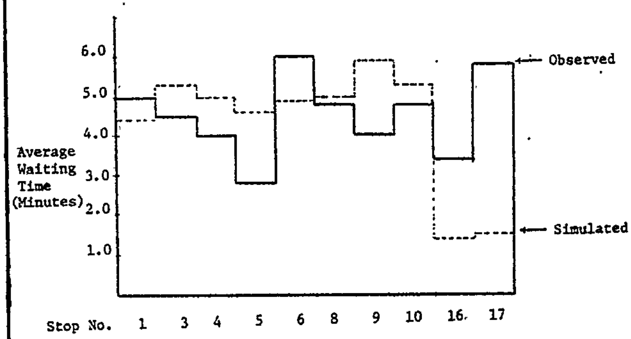
Figure 7 COMPARISON OF SIMULATED AND OBSERVED ALIGHTING TOTALS



Average passenger waiting times had been observed for only 10 effective stops (stops 1 and 2 had been combined). The observed results for these stops were compared to the corresponding simulated results, as shown in Figure 8. The results conform quite well for the first eight stops, although the simulated figures are slightly higher, but there is a wide divergence at stops 16 and 17. The χ^2 statistic of 50, with 9 degrees of freedom, shows a significant ($\rho < 0.01$) difference. However if the two furthest stops are omitted from the calculation a χ^2 of 2.53 is obtained and this shows a high level of conformity. ($\rho < 0.05$) The divergence at stops 16 and 17 may be due, in part, to the relatively low arrival rates at these stops. As explained already, it then takes a long inter-bus interval to allow the generation of a passenger. At the end of such an interval, a "bunch" of buses can be expected to arrive, one of which is likely to suit the intending passenger. Offsetting this consideration however is the suggestion that passengers time their arrivals to coincide with scheduled (or observed) buses. If this is the case, the BUSOPS would assert that they do considerably worse (by a factor of 2 or 3) than if they arrived at random!

The BUSOPS model was found to have a serious theoretical flaw when used with branched networks. The modified model's output was compared to some observed results. The results did not significantly increase the model's credibility, as although the waiting times simulated were (mainly) within the correct range of values, the model proved very poor at reproducing the passenger boarding/alighting behaviour of the real world system. The extent of attempted validation was very restricted, however, being confined to a single set of inputs, and did not accurately test many aspects of the model, e.g. its sensitivities to variation of inputs.

Figure 8 COMPARISON OF SIMULATED AND OBSERVED WAITING TIMES



5. Findings

5.1 The BUSOPS Model

An extensive data collection effort had provided a substantial test-bed of realistic data so that model validation could be attempted. Those model inputs which directly derived from field data were obtained, although not without difficulty, caused mainly by the organization of the data collection project. The hypothesized regularity in the observed data was not generally confirmed. Passenger arrivals were reasonably consistent but bus journey times showed extremely wide variation. A significant day of the week effect was also noted. Results of previous studies were reviewed and did not provide any marked contrast in trends or approximate values. The limited validation intended was further restricted by the failure of an underlying BUSOPS hypothesis on the effects of independence of passenger origins and destinations. Some validation tests were made, although these were of necessity on single sets of average values. The results did not add much to the model's credibility and tended to reinforce the belief that some of the fundamental premises of the model were inadequate. However, the range of validation was too small to allow a definite conclusion. It had been hoped to observe the model's response over the (small) observed range of inputs. However the level of variation was so high that the amount of data available could not provide distinct and realistic subsets. It seems unlikely, in view of the model's demonstrated shortcomings, that the expense of further validation would be justified, involving as it would, either the controlled modification of the real-world system or extensive data collection.

5.2 The Bus Scheduling Problem

However great may be the merits of simulation, its usefulness can never be universal. Some systems are naturally suited to being modelled, in particular those which have a clear boundary and include random factors which make analytical solutions tedious or impossible. Queueing systems, chemical reactions or motorway intersections are all amenable to simulation to a degree of detail which is limited only by the cost of field data and the availability of computing power. More complex systems, having somewhat arbitrary boundaries, can be modelled simply, provided sufficiently accurate approximations can be made, or they can be simulated by equally complex models. Available evidence indicated that the assumptions made in BUSOPS were oversimplified, especially in their representation of bus journey times. The basic question here is whether any satisfactory approximation can be found to model the observed variations in journey times. If none can be found, within the rather arbitrarily demarcated system which is being simulated, then one must either conclude that the bus scheduling problem is unsuitable for simulation, or one must expand the system being simulated to include such additional entities as will allow accurate approximations to be made. A possible addition would be to subdivide the inter-stop links to take account of intersections, traffic lights, bottlenecks, etc. Factorial analysis of observed data might provide an alternative method, but the factors that would explain the observed behavior are not immediately obvious.

Minor difficulties which arose while testing and evaluating the BUSOPS model reflect inherent difficulties in simulating bus systems. Factors were ignored whose influence may actually be large. Among these were the effect of weather on passenger demand and behavior, and the actual, as distinct from the theoretical, capacity of the buses. There is also the problem of "second-best" choices, when passengers choose an alternate route, when they've waited for some time for their first choice route. This clearly cannot be resolved, even partially, without major increases in model complexity and in data collection costs. On a wider plane there are fundamental questions about the purpose and usefulness of simulating, without question, an existing public transport system. The aim is generally stated as increasing efficiency and improving the reliability of services. In a static situation these limited goals would be worthy of the efforts of many scientists. But urban transport is far from static. In many cities, including Dublin, the public transport system has declined, due mainly to increased wealth and the consequent dominance of private cars. Under these circumstances, it can be argued that limited scientific resources are better employed in tackling the more fundamental problems of urban transport. How can the overall cost of transport be minimized? What are the effects of urban planning or zoning legislation? How can unnecessary travel be avoided?

These problems may be more difficult to simulate but their solution would have greater long term benefits than we can hope to achieve simply by tuning the current declining systems.

5.3 Simulation Methodology

Few scientists would regard simulation using digital computers as a precise and rigorously defined method. Due partly to its novelty but mainly to the problems it addresses, simulation is very much an art, wherin trial and error are unavoidable. Some skills and approaches are prerequisites if a simulation model is to be successfully developed and implemented on a computer. Other techniques and abilities, although not essential, will help the model builder to reach a successful conclusion with less expenditure of time and effort. The primary requirement for worthwhile modelling is a clear and comprehensive understanding of the system being modelled and especially of the use to be made of the final model. As with any exercise in systems analysis, costs and benefits must be carefully weighed, so that resources are directed with priorities that reflect overall objectives. The third major prerequisite is a critical and sceptical approach that never loses sight of the fact that all models are approximations of an imperfectly understood reality.

Simulation is by definition an interdisciplinary activity, so the range of desirable techniques or knowledge, is very broad. It is so broad, in fact, that one person can rarely provide all that is required and a team approach is generally preferable. Besides, a team allows for constant cross evaluation and is less likely to lose sight of the overall objectives. The work which has been described above illustrates the strengths and weaknesses of the personnel involved.

The BUSOPS model was developed on the basis of very limited field observation and made many unsubstantiated assumptions. Validation of the model was very much an afterthought. When validation was proposed a data collection project was successfully undertaken at short notice. The field data collected is believed to exceed previous samples in the Dublin area, not only in its volume, but also in its accuracy. In other areas there were noticeable shortcomings. The data collection experiment would have benefitted from more advanced planning and more sophisticated sampling methods. A pilot experiment might have highlighted problems with the data collection forms and with the data verification. Finally, a knowledge of commercial data processing would have helped both in verifying the data and in its subsequent analysis.

The overall conclusion is that the construction and evaluation of simulation models are best undertaken by a team which has within it a broad range of knowledge and abilities. Furthermore, validation should be a consideration from the outset of model design with the primary consideration being that an unvalidated complex model is never preferable to a simple and even partially validated alternative.

BIBLIOGRAPHY

1. Markham, J. "Choice of Mode for the Journey to Work in Dublin - some multivariate aspects." Royal Statistical Society Conference on Multivariate Analysis and its Applications. Hull University, April 17, 1973.

2. Potts, R. and Tamlin, E.A. "Pairing of Buses," Proceedings of the Second Conference of the Australian Road Research Board. 1964.

3. Bullock, D.A. "Scientific Methods in Bus Control Assessments." Proceedings of R.R.C. Symposium on Public Road Transport Analysis. London, September 1970.

4. Gerrard, J.P.D. and Brook, D. "A Computer Simulation Model for Bus Operations Planning." I.B.M. U.K. Scientific Centre Report No. 40. September 1972.

5. Ryan, K.T. The Development and Evaluation of Simulation Models for Bus Operations Planning, Ph.D. Thesis, University of Dublin, Trinity College, August 1977.

6. Dekindt, H. and Griffe, P. "Simulation d'une Ligne d'Autobus." Proceedings of Symposium of Traffic Control. Versailles 1970.

7. Oliver, A.M. :The Simulation of the E3 Bus Route - Intended Approach." O.R. and Special Systems Group, London Transport Executive, Technical Note TN23. 1971.

8. Teorey, Toby J. "Validation Criteria for Computer System Simulations." Simuletter. Vol. 6, No. 4, July 1975.

9. Schatzoff, M. and Tillman, C.C. "Design of Experiments in Simulation Validation." I.B.M. Journal of Research and Development. No. 19, May 1975.

10. Mihram, A.G. Simulation: Statistical Foundations and Methodology. Academic Press, 1972.

11. Mihram, A.G. "Some Practical Aspects of the Verification and Validation of Simulation Models." Operations Research Quarterly. Vol. 23, No. 1, March 1972.

12. Welding, P.J. et al. "Operations Research into Regularity of Bus Services and Results of Calculations with an Electronic Computer." O.R. Section, Management Services Office, London Transport Executive Research Report. No. 57, May 1956.

13. Benedetti, M. "Etude d'une Bande de Roulement Reservee aux Autobus Establissement d'un Modelle Previsionnel de Simulation." Revue Francaise d'Informatique et de Recherche Operationelle. Vol. 1, 1969.

14. Charlesworth, J.A. and Walmesley, J. "A Study of Bus Operations in Newcastle-upon-Tyne and Recommendations for Future Similar Studies." Universities Transport Studies Group, 8th Conference. University of Aston, January 1976 (unpublished).

15. Markham, J. and O'Farrell, P.N. "A Factorial Analysis of Peak Period Bus Journey Times in Dublin." Transportation Planning and Technology. Vol. 2, 1973.

16. Nie, D., Bent, S. and Hull, B. Statistical Package for the Social Sciences. McGraw Hill, 1970.

17. Chapman, R.A. "Bus Boarding Times - A Review of Studies and Suggestions for Interpretation." Transport Operations Research Group Working Paper No. 8. University of Newcastle-upon-Tyne, January 1975.

18. Ellis, A. Observations on Measures of Perceived Passenger Reliability for a City Bus Service. M.Sc. Dissertation, University of Newcastle-upon-Tyne, 1973.

19. Jenkins, I.A. "A Comparison of Several Techniques for Simulating Bus Routes." Transport Operations Research Group. Working Paper No. 14, University of Newcastle-upon-Tyne, January 1976.