Transpn Res.-B Vol. 15B, No. 4, pp. 249–266, 1981
Printed in Great Britain

0191-2615/81/040249-18$02.00/0


# A MATHEMATICAL MODEL OF AN URBAN BUS ROUTE

PER-ÅKE ANDERSSON

Operations Research Division, Department of Mathematics,
Linköping University, S-58183 Linköping, Sweden

and

GIAN-PAOLO SCALIA-TOMBA

Statistical Research Group, Department of Mathematical Statistics,
Stockholm University, Box 6701,
S-1385 Stockholm, Sweden

(*Received* 16 *April* 1979; *in revised form* 30 *April* 1980)

**Abstract**—The present paper is a mathematical description of an urban bus route in peak hour traffic. "Bus route" is here used in a collective sense to denote a set of more or less parallel-going sub-routes, here called "service variants". The deterministic and stochastic mechanisms of bus operation are analysed—with focus on time evolution—and general models for route structure, boarding and alighting events, link travel times and stop times are formulated. Various measures of goodness of fit are defined, for validation and model choice. The models are primarily intended for use in an interactive simulation program (see Andersson *et al.*, 1979).

## 1. INTRODUCTION

The present model of an urban bus route in peak hour traffic is primarily intended for use in an interactive simulation program (see Andersson *et al.*, 1979). The work has been sponsored by the Transport Research Delegation and supervised by Sven Erlander, Linköping Institute of Technology. The principal aim of the project has been the evaluation of control strategies that a bus company can apply in order to reduce irregularities and delays in the bus service. Input data for the model have been supplied by Stockholm Transport. These data have been obtained by a Bus Passenger Monitoring (BPM-)equipment. Parts of the model have been influenced by the nature of these data. In these cases the considerations involved are declared specifically. The actual implementation and evaluation of the model for the simulation of a bus route in central Stockholm is presented elsewhere (Andersson *et al.*, 1979).

This model has been developed almost simultaneously with its application. Therefore many of the examples given in this text are related to the actual conditions in Stockholm traffic. However, the model has also been designed for future use on more complex route networks, which motivates the generality of many formulae and related notational apparatus. To fit the simulation purpose of evaluating control actions taken on a specific bus at a specific point of time or position on the route, it must be possible to study each bus individually. It is thus important to describe and evaluate the changes which the specific bus, its passengers before and after the control action, and other buses are subjected to. Therefore the entire run of each individual bus along the route must be carefully modelled, while the description of the individual passengers may be less refined.

A typical urban bus route in central Stockholm—like in many other cities—is frequented by buses operating under different schedule conditions. Some of the buses run on a fixed schedule, others are injected into the traffic when travel demand is particularly high and usually service only a part of the route. The latter do not run on a fixed schedule and may overtake other buses. The different *service variants*, i.e. the sequences of bus stops frequented by the various types of buses, decompose the bus route into separate sections, each one specified by its set of variants. This is important when it comes to queueing at a stop which is common to two or more variants or interacting bus routes. It seems natural to let the model place the arriving passengers in different queues according to the destination section demanded, i.e. to the set of available variants, even in the absence of detailed data on such queue divisions. The adopted model of the route structure is quite general, although only a simple example is given in the text. It is well suited for simulation use, since the route and the related operations of loading and unloading at stops are described from the "point of view" of the buses. A more detailed description of the interplay between the route structure model and the representation of the route in the simulation can be found in Andersson *et al.* (1979).

Another characteristic is the *dynamics of rush hour traffic*: from "normal" levels the bus loads and travel times rise to their absolute peak values and then lower down successively to a steady state. The time dependence is not necessarily uniform along the route. This dynamic behaviour has been assumed in several other works (e.g. Jackson and Wren, 1972; Gerrard and Brook, 1972 and Bly and Jackson, 1974) and has, at least partly, been observed in real data (Chapman, Gault and Jenkins, 1976, where the time of day variations are studied for link travel times).

The *arrival intensity of passengers* at a stop varies over such a time period. Motivated by data aspects and by simulation demands for tractable models, we avoid estimating separate time functions for each queue by concentrating all the time variation associated with the number of passengers to the total arrival intensity $\lambda$ at the stop (summing up over all queues). In consequence, we assume that each queue intensity amounts to a time constant (stationary) proportion of $\lambda$.

Another hypothesis is that each day has a specific *day-level*, due to its position in the week or to the weather, in addition to the daily peak time evolution. The type of such an influence on the arrival intensities is hard to predict, so we study both additive models, where the day effect enters as an additive constant in the polynomial describing the time variation of the intensity, and models where the day parameter affects the time evolution through a multiplicative effect on the intensity in separate subperiods of the studied time interval.

As for the *link travel times*, we believe in a multiplicative day effect: if traffic is increased at off-peak a certain day, the effect grows to congestion in the rush period. We also look for a statistical distribution which is suitable both for parameter estimation and for testing. We have not found any theoretical approach, similar to the well-known theory underlying the use of exponential service-times in queue theory (see Cox and Smith, 1961), which would imply a definite form for the link travel time distribution. Our choice is the lognormal distribution, similar in form to the previously used gamma distribution and to the normal distribution. For the use of different distributions see Jenkins (1976). The logarithms of link travel times then become normally distributed and the day effect becomes additive, leading to a straightforward analysis. Since the daily evolution of link travel time parameters (mean and dispersion) is easily observable in real traffic and thus important to imitate in a realistic simulation model, the mean of the logarithms is described by a polynomial in the time of day.

In reality, the *alighting procedure* may also be time dependent. But since the stop times are often dominated by the boarding procedure, we are content with estimating a stationary average proportion $\mu$ of the bus load, getting off at a stop. Thus, the time variations in the mean number of alighting passengers will be a consequence of the variations in arrival intensity at the previous stops. The alighting proportion $\mu$ refers to those passengers who are determined to alight in the current route section.

Since data on individual passengers' waiting times and travel pattern are not available for this study, these features are not explicitly accounted for in the model.

Because of short headways, we assume that the passengers arrive independently and at random.

The *stop times* are assumed to depend on the number of alighting and boarding passengers and possibly on current bus load. Although marginal boarding and alighting times may vary over the day, due to variations in passenger type, it is assumed that they remain constant during the rush-hours.

Finally, we remark that, although the simulation is intended to run over one peak period, the estimation procedures applied here do not exclude the computation of statistics for longer time intervals.

## 2. ROUTE STRUCTURE AND PASSENGER DESTINATION

A bus route may consist of several service variants in each run direction, each variant specified by its own sequence of bus stops. There are several reasons for a strict definition of the route structure: it gives a theoretical framework for the formulation of mathematical models and the analysis of data and also a basis for the appropriate mechanisms in the intended simulation. (For an account of similar and other approaches to this problem see e.g. Jenkins, 1976.) We regard the two run directions of a route as separate entities (one is shown in Fig. 1), i.e. our structure-model splits up the physical termini when these are common to both directions.

First we consider a stop belonging to one single variant, like stops 1, 2, 6, 7 or 8 in Fig. 1.

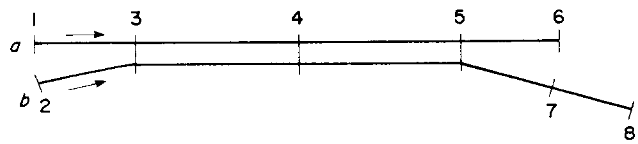


Fig. 1. A small bus route. *a*,*b*: service variants 1–8: bus stops.

Apart from termini, where either alighting or boarding is absent, such a one-variant bus stop (stop 7 in Fig. 1) is characterized by two quantities: the passenger arrival intensity $\lambda$—a function of the time of day—and the alighting proportion $\mu$. Let $H$ equal the headway (in minutes) between two consecutive buses at the stop. If $\lambda$ is approximately constant over the period $H$, then $\lambda \cdot H$ is the expected number of passengers arriving during $H$. Now consider a bus stop common to several service variants, like stops 3, 4 or 5 in Fig. 1. A passenger boarding here, intending to get off in the common route section, may choose an arbitrary variant ($a$ or $b$ in Fig. 1); he takes the first bus after his arrival.

Consider a stop where boarding occurs. For such a stop, we introduce destination sections and queue categories. Let the "set of common variants" at a subsequent stop be those variants to which both stops belong (i.e. the set of variants available for travelling between the boarding and the subsequent stop). A *destination section* for the boarding stop is composed of consecutive stops with identical "sets of common variants". At this boarding stop we define one *queue category* for each destination section.

Thus, at stop 3 and 4 in Fig. 1 we distinguish three queue categories—$a$, $b$ and $ab$—each a subset of $ab$, the variant set at the boarding stop. For stop 3, the corresponding destination sections are $\{6\}$, $\{7, 8\}$ and $\{4, 5\}$, respectively, since for stops 3 & 6 the "set of common variants" is $\{a\}$, for stops 3 & 7 and 3 & 8 it is $\{b\}$, for stops 3 & 4 and 3 & 5 it is $\{ab\}$.

We further assume that the arrival intensity $\lambda^c$ at queue category $c$ is a stationary proportion $p_c$ of the total arrival intensity, i.e.

$$\lambda^c = \lambda \cdot p_c \quad \text{and} \quad \sum_c p_c = 1$$

at each stop. For a bus running on variant $v$, calling the boarding stop $H^c$ minutes after the previous $c$-loading bus, the expected number of new arrivals within this category is $\lambda^c \cdot H^c$ and the total expected number of new passengers is

$$\hat{B} = \lambda \cdot \sum_{c \in C(v)} p_c H^c = \lambda \cdot \bar{H},$$

where $\bar{H}$ is called the *weighted headway* and $C(v)$ denotes the queue category set of variant $v$, i.e. those queue categories at the actual stop that can board a bus running on variant $v$:

$$C(a) = \{ab, a\},\ C(b) = \{ab, b\} \text{ at stops 3 or 4, } C(v) = \{v\} \text{ otherwise.}$$

In an intended simulation run, the passengers boarding from the different queues at a stop are put into corresponding *category boxes* on the bus for appropriate handling, say when alighting. The number of boxes will change along the typical bus route due to the connection or disconnection of parallel service variants. Let us study variant $a$ (see Fig. 1), reproduced in Fig. 2.

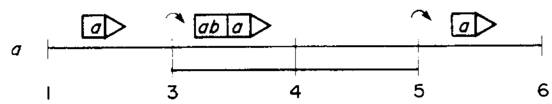


Fig. 2. Bus load categories on the small bus route.

At stop 1 all the boarding passengers come from queue category $a$, and accordingly form the contents of a box $a$. Here the box notation $a$ simply means that the passengers will get off at the following stops on variant $a$.

The boarding passengers at stop 3 are distributed among the boxes $ab$ and $a$, according to their destination sections. Box $a$ now carries the load only to stop 6. This declaration of destination should include also the "old" box $a$-passengers remaining after stop 3. By the introduction of stationary *redistribution proportions* $r_{c_f,c}$ at each connection stop, the load of an "old" box category $c_f$ is spread over the different destination sections $c$, i.e.

$$\sum_{c} r_{c_f,c} = 1 \quad \text{for all} \quad c_f.$$

With this redistribution policy at each connection stop (and the elimination of inadequate boxes at each disconnection stop, e.g. the $ab$-box at stop 5), the alighting at a stop takes place from one load-box.

Thus each stop/variant pair corresponds to a unique *alighting category* $c$ and $\mu = \mu_c$ is the expected proportion of the load in box $c$ which alights there. (On variant $a$, the alighting at stop 4 and 5 comes from the $ab$-box, otherwise from the $a$-box.)

Although the presentation here is confined to a small bus route of two service variants, the model is applicable to any bus route network. (An example of more complexity is given in Section I.4 of Andersson *et al.*, 1979.) The DEC10-implemented estimation program (*ibid.*) automatically determines all necessary categories for an arbitrary number of service variants, given only the sequences of (unique) stop numbers which define the service variants involved: first (by regarding the stops in logical—i.e. forward—order, starting at a stop with no predecessor) the alighting category for each stop and each defining service variant is settled; then (by treating the stops in opposite—i.e. backward—order, starting at a stop with no successor) the queue categories for each stop and variant is determined; at last the redistribution categories at each connection stop (i.e. beginning route section) are fixed for each concerned variant.

## 3. ESTIMATION OF BOARDING, ALIGHTING AND REDISTRIBUTION PROPORTIONS

The number of alighting passengers $A$ at a stop is assumed to vary according to the *binomial distribution* $Bi(L^c, \mu_c)$, where $L^c$ denotes the load in the alighting category $c$-box just before the stop.

This assumption is made in accordance with our scope of primarily representing bus movements rather than travel patterns. Since we assume independence between alighting passengers, i.e. no travelling in groups, the binomial distribution becomes the most natural and efficient model tool. We also assume that the alighting proportion $\mu_c$ is stationary and independent of the passengers' various origin stops. The stop attracts the whole $c$-load, remaining from the preceding stops, with common force. As a consequence of this prehistory ignorance we also expect the $\mu$'s at a stop in a common section to be equal for all frequent variants, except at a connection stop (e.g. stop 3 in Fig. 1).

Since the distribution variance is proportional to $L^c$ we seek the *least squares* solution:

$$\underset{\{\mu_c\}}{\text{minimize}} \sum_{j=1}^{J} (A_j - \mu_c L_j^c)^2 / L_j^c,$$

where $j$ is a trip index. The same $\mu_c$ applies to all trips $j$ satisfying $c \in C(v_j)$, i.e. those taken on a variant $v = v_j$ carrying box category $c$. The *alighting proportion* estimate therefore becomes (the summations are over $j$):

$$\hat{\mu}_c = \sum_{j \mid c \in C(v_j)} A_j \Big/ \sum_{j \mid c \in C(v_j)} L_j^c. \tag{1}$$

Notice that (1) is also the *maximum likelihood* (ML-)estimator of $\mu_c$.

The category load $L_j^c$ may be unknown since load data usually include only the total bus load

$$\sum_c L_j^c$$

on each trip $j$. However, the $L_j^c$ values are consequences of the assumptions made and can be reconstructed as follows:

Place a "*total bus*" on each service variant to accumulate trip data, i.e. total boarding $B$, load $L$ and alighting $A$ at each stop on the variant. We will drive these "total buses" backwards from the last variant stop.

The route structure indicates the category boxes to be used in the various route sections, and the unique alighting category at a stop/variant pair decides from which box the passengers should alight. Finally, the origin independence assumption implies that the boarding passengers at a stop are distributed among the destination sections proportionally to the box load distribution before the stop—at least on average, i.e. on the "total bus". Alternatively, moving backwards and proceeding from the load division state after the boarding event at a stop, we should subtract the same proportion from each load box (up to the boarding total) to get the consistent "old" box-load division.

Figure 3 shows an example of the backing of a "total bus" on service variant $a$. The box contents are derived from data and the previous assumptions. To determine the

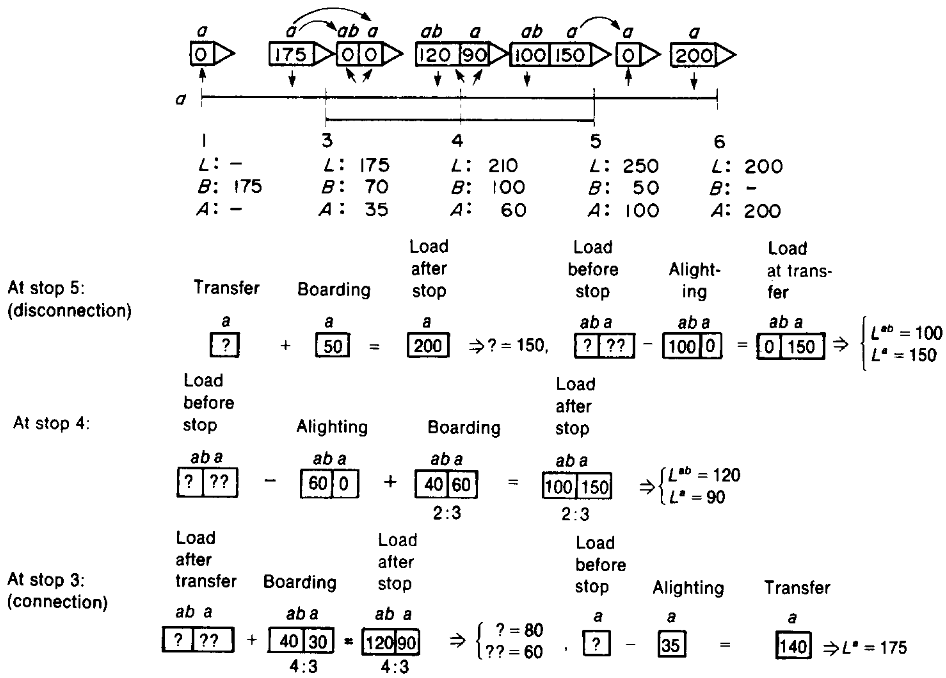


Fig. 3. A "total bus" on the small bus route. Estimation of box contents.
$L$: accumulated passenger load before stop, $B$: boarding total, $A$: alighting total.

alighting proportion $\mu$ according to estimation formula (1), we need information about all the variants at the stop. In any case, the "total buses" are driven independently of the estimation method being used; to test formula (1) the alighting proportions can, e.g. be estimated for each variant separately, even in a common route section.

The *boarding proportions* at a common stop are based upon data from all the variants running on the section. Suppose that the "total bus" computation on variant $b$ (see Fig. 1) gives the boarding sums

| *ab* | *b* |
|---|---|
| 80 | 20 |

at stop 4.

Adding the variant $a$ figures

| *ab* | *a* |
|---|---|
| 40 | 60 |

determined in Fig. 3 the overall totals become

| *ab* | *a* | *b* |
|---|---|---|
| 120 | 60 | 20 |

resulting in $p_{ab} = 0.6$, $p_a = 0.3$ and $p_b = 0.1$.

The redestination of passengers, performed at the first stop in a common route section (stop 3 in our example), is also made in accordance with the "total bus" computations—here involving one variant at a time. Since the transfers from box $a$ on variant $a$ amount to

| *ab* | *a* |
|---|---|
| 80 | 60 |

the estimated *redistribution proportions* are $r_{a,ab} = 0.57$ and $r_{a,a} = 0.43$. In other situations, say the transfer from an $ab$-section to an occasional $abc$-section, the redistribution proportions—now being $r_{ab,abc}$ and $r_{ab,ab}$—are based upon the redestinations on several variants ("total buses") from a common "old" box category $c_f = ab$.

## 4. ARRIVAL INTENSITIES

Due to the expected variations in passenger behaviour along the route each stop is regarded separately at the estimation of arrival intensities. It is assumed that the passenger arrivals are not observed directly, only the headways and the numbers of boarding passengers are known—with no information about the number of passengers left waiting at a stop when a bus gets full. The estimation methods will therefore be based on such data. Properly speaking, only "boarding intensities" can be estimated with this kind of data. However, boarding differs from arrival only when buses become full (loaded up to maximum capacity) before all potential passengers have boarded. This restriction is not important in the present study since we are not primarily interested in passengers' waiting times. However, an attempt is made to recognize such overload situations and to correct the observed boarding intensities. In this correction, the trip observations at the actual stop are arranged in time-order for each day $i$. The number $B'_{ij}$ of arriving passengers during the time interval $\bar{H}_{ij}$ since the previous bus departure (trip No. $j$–1) should be the observed number $B_{ij}$ of passengers boarding trip No. $j$ at the stop—corrected if the load value, registered on the link just after the stop, indicates that the bus is full. (These observations of critically heavy load are rare on the analysed Stockholm route and no corrections are therefore made in that application (Andersson *et al.*, 1979), but the general model is presented here for the sake of completeness.)

In general, there is a whole sequence $S$ of consecutive full buses at the stop. In this case the observed numbers $B_{ij}, j \in S$, of boarding passengers give very little information about the momentary arrival intensities, and a registered difference within $S$, i.e. $B_{ij}/\bar{H}_{ij} \neq B_{ij'}/\bar{H}_{ij'}$ for $j,j' \in S$, may be occasional. To supress any unverified variation, the estimation model uses corrected data $B'_{ij} = \lambda' \cdot \bar{H}_{ij}$ for $j \in S$, based on the intensity average

$$\lambda' = \sum_{j \in S} B_{ij} \Big/ \sum_{j \in S} \bar{H}_{ij}.$$

(In general, the $\bar{H}_{ij}$'s here are weighted headways—defined in Section 2.) If the next non-full bus (trip No. $j''$) satisfies $B_{ij''}/\bar{H}_{ij''} > \lambda'$, where the increased intensity is an indication of people left waiting, then the average $\lambda'$ includes $j''$ as well, and also $B'_{ij''} = \lambda' \cdot \bar{H}_{ij''}$ is corrected.

Full information about all passing buses during the studied time period is assumed. Corrections for missing values must therefore be made, when actually using the model.

The passengers are assumed to arrive according to a *Poisson process* with time dependent intensity (see Cox and Lewis, 1966, for definitions). This means that passengers arrive at random and independently at the stop, which is a reasonable assumption since headways are short during the rush-hours (see Chapman, Gault and Jenkins, 1976, for a detailed study of the validity of such assumptions). We propose several models for the total arrival intensity $\lambda$, to be compared later, at the actual estimation.

At first, we try some *additive models* with a day-specific parameter $\alpha_i$ for each day $i$ and a set of polynomial coefficients $\beta_k$, $k = 1(1)K$, taking care of the daily time evolution. Since the change in arrival intensity over one headway is believed to be small, we *linearize the intensity function*, as follows:

$$\hat{B}_{ij} = \lambda_{ij} \cdot \bar{H}_{ij} = \left[\alpha_i + \sum_{k=1}^{K} \beta_k \cdot \left(t_{ij}^k - \frac{\tau_{ik}}{\tau_{i0}}\right)\right] \cdot \bar{H}_{ij} \tag{2}$$

where $j$ is a trip index, $\hat{B}_{ij}$ is the expected number of arrivals between buses $j-1$ and $j$ on day $i$, and $\bar{H}_{ij}$ is the previously defined weighted headway. The constants $\tau_{ik}$ (defined below) are introduced for statistical and computational reasons. This model is suited for simulation since it only requires the calculation of $\bar{H}_{ij}$ and powers of the bus departure time $t_{ij}$. The importance of day-specific levels $\alpha_i$ will be determined by considering the corresponding model with day-common intensity, i.e.

$$\hat{B}_{ij} = \left[\alpha + \sum_{k=1}^{K} \beta_k \cdot \left(t_{ij}^k - \frac{\tau_{\cdot k}}{\tau_{\cdot 0}}\right)\right] \cdot \bar{H}_{ij}. \tag{3}$$

We also compute an *exactly integrated intensity* for all relevant category $c$ headways $H_{ij}^c$:

$$\begin{aligned}
\hat{B}_{ij} &= \sum_{c \in C(v_{ij})} p_c \cdot \int_{t_{ij} - H_{ij}^c}^{t_{ij}} \lambda(t)\,\mathrm{d}t \\
&= \sum_{c \in C(v_{ij})} p_c \cdot \int_{t_{ij} - H_{ij}^c}^{t_{ij}} \left\{\alpha_i + \sum_{k=1}^{K} \beta_k \cdot \left[(k+1)t^k - \frac{\Delta_{ik}}{\Delta_{i0}}\right]\right\} \mathrm{d}t \\
&= \left[\alpha_i - \sum_{k=1}^{K} \beta_k \cdot \frac{\Delta_{ik}}{\Delta_{i0}}\right] \cdot \bar{H}_{ij}^{(0)} + \sum_{k=1}^{K} \beta_k \cdot \bar{H}_{ij}^{(k)}
\end{aligned} \tag{4}$$

where

$$\bar{H}_{ij}^{(k)} = \sum_{c \in C(v_{ij})} p_c \cdot [t_{ij}^{k+1} - (t_{ij} - H_{ij}^c)^{k+1}],$$

i.e. $\bar{H}_{ij}^{(0)} = \bar{H}_{ij}$ (the weighted headway), and $\Delta_{ik}$ are constants defined below. The corresponding day-common model is also formulated.

Since $\bar{H}_{ij}^{(0)} = \bar{H}_{ij}$ is a linear factor in all the expressions (2)–(4) for $\hat{B}_{ij}$, and since $\hat{B}_{ij}$ equals the expected number of arriving passengers as well as the variance of this number for the Poisson distribution, we will estimate the coefficients by solving the weighted least-squares problem

$$\underset{\{\alpha_i\},\{\beta_k\}}{\text{minimize}} \sum_{i,j} (B_{ij} - \hat{B}_{ij})^2 / \bar{H}_{ij},$$

where $B_{ij}$ denotes the observed number of passengers boarding bus $j$ on day $i$, corrected for missing registrations and for observations of full buses.

Nothing prevents the essentially estimated polynomials from being negative in parts of the time period considered or outside it. We will therefore try some models where the nonnegativity problem does not arise. These models will also be compared with the polynomial models with regard to the goodness of fit to data. First, however, some comments on the coefficients and constants introduced in the above defined models:

The purpose of introducing the constants $\tau_{ik}$ and $\Delta_{ik}$ in formulas (2)–(4) is to provide uncorrelated sets of estimates $\{\alpha_i\}$ and $\{\beta_k\}$. They also reduce the numerical range in the calculations. Hence the day-parameters are determined as

$$\alpha_i = \frac{1}{\Delta_{i0}} \cdot \sum_j B_{ij}$$

in the day-specific models, and as

$$\alpha_. = \frac{1}{\Delta_{.0}} \cdot \sum_{i,j} B_{ij}$$

in the day-common models. The $\Delta$'s are defined as

$$\Delta_{ik} \equiv \sum_j \bar{H}_{ij}^{(k)} = \sum_j \sum_{c \in C(v_{ij})} p_c \cdot [t_{ij}^{k+1} - (t_{ij} - H_{ij}^c)^{k+1}] = \sum_c p_c \cdot [t_{i,J_c}^{k+1} - t_{i,\phi_c}^{k+1}]$$

and

$$\Delta_{.k} \equiv \sum_{i,j} \bar{H}_{ij}^{(k)} = \sum_i \Delta_{ik},$$

where $\phi_c$, $J_c$ denote the first and last trip carrying category $c$ on day $i$. In a specific (non-common) section there is only one queue category, i.e.

$$\Delta_{ik} = t_{i,J}^{k+1} - t_{i,\phi}^{k+1}.$$

The polynomial coefficients are the unique solutions to the normal equations

$$\sum_{l=1}^{K} g_{kl}\beta_l = d_k, \quad k = 1(1)K,$$

where the Gram matrix $(g_{kl})$ and the resource vector $(d_k)$ in model (2) fulfil

$$g_{kl} = \sum_i \left( \tau_{i,k+l} - \frac{\tau_{ik}}{\tau_{i0}} \cdot \tau_{il} \right)$$

$$d_k = \sum_{i,j} B_{ij} t_{ij}^k - \sum_i \frac{\tau_{ik}}{\tau_{i0}} \cdot \sum_j B_{ij}$$

with

$$\tau_{ik} \equiv \sum_j \bar{H}_{ij} t_{ij}^k \quad (\text{i.e. } \tau_{i0} = \Delta_{i0})$$

and, in model (4),

$$g_{kl} = \sum_{i,j} \frac{\bar{H}_{ij}^{(k)}}{\bar{H}_{ij}^{(0)}} \bar{H}_{ij}^{(l)} - \sum_i \frac{\Delta_{ik}}{\Delta_{i0}} \Delta_{il}$$

$$d_k = \sum_{i,j} \frac{\bar{H}_{ij}^{(k)}}{\bar{H}_{ij}^{(0)}} B_{ij} - \sum_i \frac{\Delta_{ik}}{\Delta_{i0}} \cdot \sum_j B_{ij}.$$

Since we are to represent one peak period, from and to a steady level, a polynomial of degree $\leq 4$ ought to be sufficient.

Let us now consider a *multiplicative model* with *piecewise constant intensities*, where the day-parameter $\alpha_i$ is a factor in the total intensity expression. Because of the nonlinearity involved in such a model, least-squares objectives are unsuitable for computation. We will therefore seek estimates which maximize the log-likelihood function $F$, satisfying

$$F = \log \prod_{i,j} (\hat{B}_{ij})^{B_{ij}} \cdot \exp(-\hat{B}_{ij})/(B_{ij}!) = \sum_{i,j} B_{ij} \log \hat{B}_{ij} - \sum_{i,j} \hat{B}_{ij} + \text{constant}$$

on our Poisson assumption.

Now letting $\{\beta_k\}$ denote the intensity-level factors in *separate subperiods* $k = 1(1)K$, we propose

$$\hat{B}_{ij} = \alpha_i \cdot \beta_{s(i,j)} \cdot \overline{H}_{ij}, \tag{5}$$

where $k = s(i, j)$—the period to which trip $j$ on day $i$ belongs.

The maximum likelihood estimates are the solutions of

$$\begin{aligned} \alpha_i &= \sum_j B_{ij} \Bigg/ \left( \sum_k \beta_k \cdot \sum_{j|s(i,j)=k} \overline{H}_{ij} \right) \\ \beta_k &= \sum_{i,j|s(i,j)=k} B_{ij} \Bigg/ \left( \sum_i \alpha_i \cdot \sum_{j|s(i,j)=k} \overline{H}_{ij} \right) \end{aligned} \tag{6}$$

In practice, $\{\alpha_i\}$ and $\{\beta_k\}$ are determined iteratively, in a balancing procedure, with subperiod limits $T_k, k = 0(1)K$, kept fixed.

We shall also vary the interior interval limits $T_k$, one at a time, looking for subdivisions of the studied time period that fit data better. A shift in $T_k$ affects the sums in subperiods $k$ and $k + 1$ if at least one trip changes intervals. For each new subdivision, the log-likelihood function is maximized over all variables according to (6). When no further improvement of the maximum value is attained, we have found a locally optimal period subdivision.

In real traffic we expect a smoother behaviour of the total arrival intensity than model (5) provides. For the sake of completeness we therefore define a *stepwise linear model*

$$\hat{B}_{ij} = \alpha_i \cdot \left[ \beta_k \cdot \frac{T_{k+1} - t_{ij}}{T_{k+1} - T_k} + \beta_{k+1} \cdot \frac{t_{ij} - T_k}{T_{k+1} - T_k} \right] \cdot \overline{H}_{ij}, \quad k = s(i,j). \tag{7}$$

The included linear factor makes the likelihood function unattractive as an objective and we turn to least-squares, once again. The day-parameters $\{\alpha_i\}$ are treated as given constants, taken from model (6). The optimum solution for this model is also used as an initial solution for model (7), linearized as in Fig. 4 (where $\beta'_k = (\beta_k + \beta_{k+1})/2$, $k = 1(1)K - 1$, denote the breakpoint levels).

By the same arguments that were used in connection with models (2)–(4), but now accepting $\alpha_i \cdot \overline{H}_{ij}$ as a given element of the variance, we formulate our objective as to

$$\underset{\{\beta_k\}}{\text{minimize}} \sum_{i,j} (B_{ij} - \hat{B}_{ij})^2/(\alpha_i \cdot \overline{H}_{ij}).$$

Since only one interior breakpoint $T_k$ is shifted at a time (thus getting a fast iterative procedure), we only have to minimize over the corresponding level $\beta'_k$. Trip sums in the surrounding subperiods $k$ and $k + 1$ are affected. By evaluating the objective function

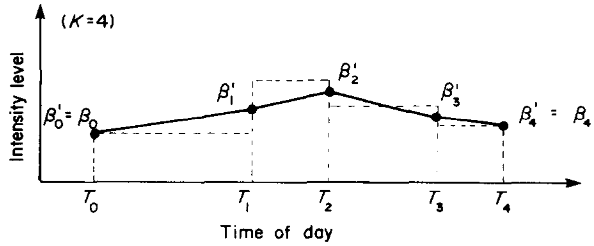


Fig. 4. A solution of model (6) and the corresponding linearized initial solution to (7).

after each $T_k$-step and the corresponding $\beta'_k$-optimisation, we arrive at a stepwise linear function, defined by $\{(T_k, \beta'_k)\}$, which is again a local optimum.

Day-common subperiod models ($\alpha_i = \alpha_.$ for each $i$ in (5) and (7)) are also tried.

## 5. LINK TRAVEL TIMES

To accomplish a realistic simulation, it is necessary to have an accurate description of all the different phases of a bus tour: link travel times including penalties for bus deceleration and acceleration at stops, and stop times caused by passengers boarding and alighting and by time table halts at timing points, termini and driver exchange stops along the route.

The estimation of a *halting penalty* at a bus stop should normally be based on the travel time reduction resulting when a bus passes by a stop without stopping. Due to the lack of precision in the available BPM-data—observations being rounded off to tenths of a minute in Stockholm traffic—such comparisons would be inaccurate here. An overall halting penalty is determined separately instead. In case of more accurate registrations and stops with different characteristics it may be worthwhile to estimate a specific penalty for each kind of stop. The estimation of *link travel times* will be based on observations of buses stopping at both ends of the link. The link travel times are measured from acceleration with closed doors to standstill and door-opening at the next stop. Since the buses in Stockholm are of two equivalent types (one-man operated with front entrance and backdoor exit, and equal in performance), all the observations are assumed to come from one population. If different types of buses are used on the route, separate link travel time functions should be estimated for each type. Missing data are ignored.

The travel time variations are regarded on a "macro" level, i.e. the individual delays on a link—caused by occasional fluctuations in the traffic and signal settings—can be summarized into a single link travel time distribution. Such a view is taken in this study since the overall bus movements are to be studied in the simulation and the various sources of delays are both undistinguishable and out of the user's control, in this case.

When the *bunching* of buses is reliably registered, the observations of bus-delayed buses should be excluded before estimation if the buses are not allowed to overtake each other, since an unwanted dependence between observations otherwise would be introduced. Such an exclusion is also consistent with the simulation logic, since bunching and overtaking rules are separately taken care of in the program. Overtaking is prohibited in Stockholm for buses running on the same variant, but since time registrations on different buses are not perfectly synchronized and may differ by more than a minute from correct time, bunching events cannot be reliably detected. However, these events amount to a negligible quantity in Stockholm traffic.

Another feature not included in the model is the *driver's behaviour*, when influenced by deviations from the schedule, trying to drive faster if late. We presume that the driver either has small possibilities to speed up in peak hour traffic or simply that he(she) ignores the clock except at timetable halts (see Chapman, Gault and Jenkins, 1976).

Three factors are of importance for a model of link travel times: *specific day-levels*, describing the difference between mean travel times on different days, the *daily time evolution*, describing how link travel times depend on variations in traffic during the day and a *random mechanism*, describing the variations around the mean behaviour.

The day-level is not believed to have an additive effect over the whole peak period (the time evolution of the mean link travel time being translated by a constant amount from one day to another), but a *multiplicative influence*, i.e. a somewhat raised mean running time at off-peak is followed by a reinforced, more increased peak value.

The statistical distribution of the link travel times should be skewed to the right since there is a (small) probability of getting very large values but a positive lower running time limit depending on bus efficiency and speed regulations. This intrinsic asymmetry makes the normal distribution unsuitable, although used by, e.g. Lesley (1975). Since the gamma(Erlang)-distribution, adopted by, e.g. Dekindt and Griffe (1970), Oliver and Uren (1972) and others, does not allow an easy analysis of mean value functions, we propose the *lognormal distribution*, which is similar both to the normal and the gamma distributions for appropriate parameter values. The lognormal distribution is related to the normal distribution in the following way:

$$\xi(i, t) = \ln[x(i, t) - x^0],$$

where $x(i, t)$ denotes the link travel time at time $t$ on day $i$, and $x^0$ is the ideal (lower limit) value. If the quantity $x(i, t)$ is assumed to be log-normally distributed, $\xi(i, t)$ becomes normally distributed: $\xi(i, t) \sim N[\mu(i, t), \sigma^2]$, with $\mu(i, t)$ and $\sigma^2$ denoting mean and variance, respectively.

Letting $t_{ij}$ be the departure time for bus No. $j$ on day $i$, we abbreviate

$$x_{ij} \equiv x(i, t_{ij}) \quad \text{and} \quad \mu_{ij} \equiv \mu(i, t_{ij}).$$

The log-likelihood function for all observations is

$$L[\{\mu_{ij}\}, \sigma^2, x^0] = -\frac{1}{2\sigma^2} \cdot \sum_{i,j} [\ln(x_{ij} - x^0) - \mu_{ij}]^2 - J \cdot \ln \sigma - \sum_{i,j} \ln(x_{ij} - x^0),$$

where $J$ is the total number of observations.

If $x^0$ is known, the ML-estimation of $\mu_{ij}$ coincides with the least squares estimation

$$\underset{\{a_i\},\{b_k\}}{\text{minimize}} \sum_{i,j} (\xi_{ij} - \mu_{ij})^2,$$

where $\{a_i\}$ and $\{b_k\}$ are parameters, defined below, describing the mean value function $\mu(i, t)$.

Once $\{a_i\}$, $\{b_k\}$ and $x^0$ have been fixed, the ML-estimate of the variance $\sigma^2$ becomes

$$\hat{\sigma}^2 = \frac{1}{J} \cdot \sum_{i,j} (\xi_{ij} - \hat{\mu}_{ij})^2.$$

The presence of the unknown *ideal time* $x^0$ makes the log-likelihood function highly non-linear. For practical computation, the ML-estimates $\hat{\mu}(x^0)$ and $\hat{\sigma}^2(x^0)$ are determined for different $x^0$-values, and the univariate function

$$L^0(x^0) \equiv L[\{\hat{\mu}_{ij}(x^0)\}, \hat{\sigma}^2(x^0), x^0]$$

is maximized in an iterative procedure, leading to a stationary point $\hat{x}^0$ such that $[\{\hat{\mu}(\hat{x}^0)\}, \hat{\sigma}^2(\hat{x}^0), \hat{x}^0]$ also maximizes $L$.

In the simulation, lognormally distributed random numbers greater than $x^0$ are generated. Since $x^0$ is the lower limit of link travel times, it should be non-negative. In the estimation,

$$x^0 < \min_{i,j} x_{ij} \equiv x_{\min}$$

must hold, since the smallest value observed is a pole (discontinuity point) of $L^0$ and $L$. Hill (1963) has shown that the path adopted in the iterative procedure theoretically fulfils

$$\hat{\mu}(x^0) \to -\infty, \hat{\sigma}^2(x^0) \to +\infty, L^0(x^0) \to +\infty \quad \text{when} \quad x^0 \to x_{\min}.$$

Instead of these unsuitable ML-estimates, Hill looks for a region of high posterior probability. In general, $L$ has another maximum $[\{\mu\}, \sigma^2, x^0]$ in such a region (to the left of the singularity $x_{\min}$) whenever the prior joint density function varies slowly there, and by maximizing $L^0$ numerically we find the corresponding $x^0$-value. For this purpose, a rational interpolation procedure (well suited near the singularity) is applied to $L^0$ and its derivative, in combination with a robust (but slow) linear search optimisation routine—see, e.g. Zangwill (1969), also for general iterative optimisation methods. The $x^0$ so fixed becomes an operationally defined ideal time, chosen to fit the $\mu$-expression being used.

We will test models for $\mu$ of varying complexity although of a common structure. The simplest model is to have a *constant level* for all days and points of time. The corresponding estimate of the mean value function $\mu(i, t)$ is

$$\mu(i, t) = a. = \xi_{..} \equiv \frac{1}{J} \cdot \sum_{i,j} \xi_{ij} \quad \text{(grand mean).} \tag{8}$$

Introducing a *time dependent polynomial* of degree $K$ to take care of the daily time evolution, assumed to be identical for all days, the model becomes

$$\mu(i, t) = a. + \sum_{k=1}^{K} b_k \cdot (t^k - t^k_{..}). \tag{9}$$

The next step is to differentiate between days by introducing *specific day-levels* $a_i$, leading to

$$\mu(i, t) = a_i + \sum_{k=1}^{K} b_k \cdot (t^k - t^k_{i.}). \tag{10}$$

The purpose of the quantities

$$t^k_{..} \equiv \frac{1}{J} \cdot \sum_{i,j} t^k_{ij}$$

and

$$t^k_{i.} \equiv \frac{1}{J_i} \sum_{j=1}^{J_i} t^k_{ij},$$

$J_i$ being the number of observations on day $i$, is to provide uncorrelated sets of estimates $\{a_i\}$ and $\{b_k\}$, where the day-levels $\{a_i\}$ fulfil $a_i = \xi_{i.}$ (day-means).

Since, in the previous models (8)–(10), the time evolution around the mean was equal for all days, the final step is to introduce day-dependence in the polynomials, thereby getting

$$\mu(i, t) = a_i + \sum_{k=1}^{K} b_{ik} \cdot (t^k - t^k_{i.}).$$

As a consequence of the lognormal assumption, each transformed observation

$$\xi_{ij} = \mu_{ij} + e_{ij}$$

shall have the residual $e_{ij}$ normally distributed with mean 0 and variance $\sigma^2$.

The mean and variance for the transformed variable $\xi_{ij}$ are related to the corresponding quantities $m_{ij}$ and $s^2_{ij}$ for the untransformed link travel time $x_{ij}$ through

$$\begin{aligned} m_{ij} &= x^0 + \exp(\mu_{ij} + \sigma^2/2) \\ s^2_{ij} &= \exp(2\,\mu_{ij} + \sigma^2)[\exp(\sigma^2) - 1] = (m_{ij} - x^0)^2 \cdot [\exp(\sigma^2) - 1] \end{aligned}$$

Since $m - x^0$ is an exponential, it is automatically positive, and $\sigma^2 > 0$ implies $s^2_{ij} > 0$. The time evolution of $\mu$ is accompanied by an exponentially increased evolution of $m$ and $s^2$. The hypothesis of a time-dependent variance is supported by the findings of Chapman, Gault and Jenkins (1976). The polynomial degree $K = 4$ will probably be sufficient for the description of the daily time-evolution over one peak period.

The normal assumption justifies the use of several *test quantities*. The residuals $e_{ij}$ remaining after the estimation of a model $M$ and the corresponding sum of squares

$$R_M = \sum_{i,j} e^2_{ij}$$

are used in an *analysis of covariance* (see Scheffé, 1959) to evaluate the model itself and to compare it with another, simpler model $M'$. More specifically: if $M$ is an acceptable model, the significance of weakening the model by eliminating some of the parameters, thereby reducing it to the model $M'$, can be established in an $F$-test by considering the quantity

$$\frac{(R_{M'} - R_M)/(df_{M'} - df_M)}{R_M/df_M} \sim F(df_{M'} - df_M, df_M),$$

where $df_M$ and $df_{M'}$ stand for the number of degrees of freedom (number of observations minus number of parameters) in the respective models.

The *coefficient of determination*

$$\rho^2 = 1 - \frac{R_M}{R_O}$$

indicates the extent to which the total variation in data is explained by model $M$, compared to the "explanation" afforded by a basic, simple model "$O$". Here

$$R_O = \sum_{i,j} (\xi_{ij} - \xi_{..})^2$$

is the sum of squares corresponding to the simplest model (8) with constant level for all days and at all points of time. Since the last part of the analysis is an ordinary normal polynomial regression, confidence intervals for the coefficients $\{b_k\}$ can be calculated and tests of significance be performed.

The *normalized residuals* $\{e_{ij}/\hat{\sigma}\}$ are expected to have an almost unit normal distribution. Their frequency histogram can therefore be compared to the unit normal curve, as a further test of the normality. The variance $\hat{\sigma}^2$ in an accepted model is here taken as the unbiased estimate $R_M/df_M$ (rather than the pure ML-estimate).

To test for any remaining trends in the observations, the normalized residuals can be plotted as a function of time, both for each day separately and for all days together.

## 6. TIME AT STOPS

The stop time is the time needed at a stop for letting passengers on and off; more specifically, from door-opening (at standstill) to acceleration (with closed doors). Stop events involving other activities, e.g. lay-over at a timing point or terminus or change of drivers along the route, are treated separately. In case of low data precision—tenths of a minute in Stockholm—all the remaining observations are assumed to belong to one homogeneous population, with no differentiation between individual stops and buses. In general, i.e. with more accurate registrations, a separate estimation may be performed for each main stop (where many passengers get on and off) and for each bus type, because of the few specifying parameters in the below proposed stop time models.

The passengers move in *two one-way streams*: one boarding through the front door and paying the driver or showing a prepaid month-ticket and the other alighting through the rear-door (which is also the entrance for prams). In terms of queueing theory: the entrance/driver on one hand and the exit on the other are two service stations for the passengers queueing to get on and off the bus, respectively. Although the model can be formulated in more general terms, we will assume that the individual components of the service times can be described by random variables having the *gamma distribution* (this assumption is a generalization from the exponential distribution in classical queueing theory). The $\Gamma$-distribution will be denoted by $\Gamma(m, v)$, where $m$ is the mean and $v$ the variance.

The boarding stream service time $x$ consists of a bus/driver dependent *dead-time* $x_0 \sim \Gamma(m_{0b}, v_{0b})$ for door opening and checking, and a passenger dependent pure boarding time. Since all the $B$ passengers line up at the front-door for boarding, it seems natural to add identically distributed *marginal boarding times* $x_i \sim \Gamma(m_{1b}, v_{1b})$, $i = 1(1)B$, for the total service time of the boarding queue. Analogously, the alighting stream service time $y$ is composed of a dead-time $y_0 \sim \Gamma(m_{0a}, v_{0a})$ and *marginal alighting times* $y_j \sim \Gamma(m_{1a}, v_{1a})$, $j = 1(1)A$, if $A$ passengers alight at the stop.

The resulting stop time $z$ is determined by the slowest stream as

$$z = \max\left[x_0 + \sum_{i=1}^{B} x_i,\, y_0 + \sum_{j=1}^{A} y_j\right].$$

We will now introduce some hypotheses and simplifying assumptions. Their validity will be examined by comparing estimates and goodness of fit of the resulting models.

In the case of independent marginal times, the pure boarding and alighting times satisfy

$$\sum_{i=1}^{B} x_i \sim \Gamma(m_{1b}\cdot B, v_{1b}\cdot B) \quad \text{and} \quad \sum_{j=1}^{A} y_j \sim \Gamma(m_{1a}\cdot A, v_{1a}\cdot A).$$

Thus, the stop time is determined by four $\Gamma$-distributed numbers, which is impractical for simulation. One way to reduce this problem is to assume that $x_0$ and the $x_i$'s have a *common intensity* $\lambda_b$ (for definitions see Cox and Miller, 1968), and similarly for the alighting stream. The parameters of the distributions must then fulfil

$$\lambda_b = m_{0b}/v_{0b} = m_{1b}/v_{1b} \quad \text{and} \quad \lambda_a = m_{0a}/v_{0a} = m_{1a}/v_{1a}.$$

The two stream service times $x$ and $y$ satisfy

$$x = x_0 + \sum_{i=1}^{B} x_i \sim \Gamma[m_{0b} + m_{1b}\cdot B, (m_{0b} + m_{1b}\cdot B)/\lambda_b]$$

$$y = y_0 + \sum_{j=1}^{A} y_j \sim \Gamma[m_{0a} + m_{1a}\cdot A, (m_{0a} + m_{1a}\cdot A)/\lambda_a] \tag{11}$$

and the stop time $z$ becomes the greater of two $\Gamma$-distributed quantities. The number of parameters to be estimated is also reduced.

Many authors have assumed that the mean stop time equals the greater of the two mean stream service times. This is generally an underestimate since, for any two variates $x$ and $y$, the mean of $\max[x, y]$ will be greater than any of the means of $x$ and $y$, unless one of them dominates the other with probability one. For instance, let $x$ and $y$ be exponentially distributed with means $1/\lambda_x$ and $1/\lambda_y$, respectively. If $x$ and $y$ are independent it can be shown that the mean of $\max[x, y]$ equals $1/\lambda_x + 1/\lambda_y - 1/(\lambda_x + \lambda_y)$, which is greater than both $1/\lambda_x$ and $1/\lambda_y$.

Let us denote the mean service times for the boarding and alighting streams in the above defined *stream models by* $\mu_x$ and $\mu_y$, respectively, and the mean stop time by $\mu_z$. Analogously, we denote the corresponding variances by $\sigma_x^2$, $\sigma_y^2$ and $\sigma_z^2$. The analytical expressions for $\mu_z$ and $\sigma_z^2$ are very complex, even in the simplified case (11) with constant stream intensities $\lambda_b$ and $\lambda_a$. The expected stop time can always be represented as

$$\mu_z = \max[\mu_x + \mu_{xy}, \mu_y + \mu_{yx}], \tag{12}$$

where the *corrections* $\mu_{xy}$ and $\mu_{yx}$ depend on the parameters of the $\Gamma$-distributions. We will approximate these corrections with

$$\mu_{xy} = m_{2b}\cdot \delta_{BA} + m_{3b}\cdot BA, \tag{13}$$

where $\delta_{BA} = 1$ indicates simultaneous boarding and alighting, $\delta_{BA} = 0$ otherwise; a corresponding expression ascribed to $\mu_{yx}$.

Given these corrections, $\mu_z$ will be determined by one of the "corrected streams". In order to let this stream determine $\sigma_z^2$ as well, we also need correction terms for $\sigma_x^2$ and $\sigma_y^2$. These will be approximated by

$$\sigma_{xy}^2 = v_{2b}\cdot \delta_{BA} + v_{3b}\cdot BA, \tag{14}$$

and by a corresponding expression for $\sigma_{yx}^2$. If $x$ and $y$ once more are assumed to be exponentially distributed (i.e. $\sigma_x^2 = 1/\lambda_b^2$, $\sigma_y^2 = 1/\lambda_a^2$) and the "corrected $x$-stream" determines $\mu_z$, then

$$\sigma_z^2 = \frac{1}{\lambda_b^2} + \left[\frac{1}{\lambda_a^2} - \frac{3}{(\lambda_b + \lambda_a)^2}\right] \equiv \sigma_x^2 + \sigma_{xy}^2. \tag{15}$$

Thus, in this case the variance correction $\sigma_{xy}^2$ is positive if $\lambda_b > \lambda_a(\sqrt{3} - 1)$ and otherwise negative.

These corrections are also chosen to represent a possible real *interaction between* the two *passenger streams.* The dead-time correction $m_{2b} \cdot \delta_{BA}$ is the mean extra time needed for the driver to operate and check also the rear-door, $m_{3b} \cdot BA$ is the mean contribution from $A$ alighting passengers to the pure boarding time, through their blocking the way and drawing the driver's attention.

Only in the case of definite dominance, e.g. the dead-time dominating the pure boarding time and boarding dominating alighting, the stop time definitely has a $\Gamma$-distribution, viz. the distribution of the dominating event. In the general case, when one event does not clearly dominate all the others, the distribution of the stop time is complicated and certainly non-standard. Anyhow, we will approximate it with a $\Gamma(\mu_z, \sigma_z^2)$-distribution, taking the parameters $\mu_z$ and $\sigma_z^2$ as in (12)–(15), combined with stream parameters [e.g. from (11)]. Other arguments in favour of this assumption are the non-negativity of the $\Gamma$-distribution and its good descriptive power in other queueing applications.

The evaluation of the proposed $\Gamma$-models is partly based on comparisons with similar mean and variance models, where our earlier assumptions are dropped:

*Interaction between the boarding passengers*, e.g. a "slow" passenger often followed by quicker or more prepared ones, means deviations from the assumed additive marginal contributions to the stream service time mean and variance and may be revealed by the introduction of quadratic terms into the $\mu_x$- and $\sigma_x^2$-expressions.

*Bus load influence on the marginal boarding time*, because of congestion in the bus, is examined by the introduction of a term proportional to $B \cdot L$ ($L$ = bus load).

*Bus load causing a delay of the start of the boarding process*, also due to congestion in the bus, is regarded as a dead-time increase and is represented by a term proportional to $L$ or $L^2$.

*Time of day variations*, because of varying passenger types (passengers moving slowly, like shopping parents with prams and old people vs passengers going to or from work, usually with prepaid month-tickets), are not explicitly represented in the model since the passenger population during rush-hours is thought to be relatively stable. This assumption will, however, be checked through residual plots.

To summarize, the most advanced boarding time mean and variance model is

$$\begin{aligned} \mu_x &= (m_{0b} + m_{4b} \cdot L + m_{5b} \cdot L^2) \cdot \delta_B + m_{1b} \cdot B + m_{6b} \cdot B^2 + m_{7b} \cdot BL \\ \sigma_x^2 &= v_{0b} \cdot \delta_B + v_{1b} \cdot B + v_{4b} \cdot B^2 \end{aligned} \tag{16}$$

together with analogous expressions for the alighting parameters.

For the estimation of the *variance* $\sigma_z^2$, the observations are divided into separate classes $c$, each defined by a given $(B,A)$-combination where $B$ and $A$, as before, denote the number of boarding and alighting passengers. The variance parameters are adjusted to fit the sample variances $s_c^2$, in a least squares sense ($\sigma_{zc}^2$ is the estimated population variance):

$$\underset{\{v_k\}}{\text{minimize}} \sum_c w_c \cdot (e_c - \sigma_{zc}^2/s_c^2)^2. \tag{17}$$

Let $\nu_c$ denote the degrees of freedom of $s_c^2$ (number of observations in class $c$ minus 1—due to the mean-estimate). The constants $w_c$ and $e_c$ will be chosen on the assumption that the sample variance in each class approximately fulfills that $\nu_c \cdot s_c^2/\sigma_{zc}^2$ has the chi-square distribution with $\nu_c$ degrees of freedom.

We choose ($E[\cdot]$ stands for mathematical expectation):

$$e_c = E[\sigma_{zc}^2/s_c^2] = 1/(1 - 2/\nu_c)$$

and

$$w_c = 1/E[(e_c - \sigma_{zc}^2/s_c^2)^2] = (v_c - 4)\cdot(1 - 2/v_c)^2 \quad \text{for} \quad v_c > 4,$$

since a small expected class residual should carry great weight, to get the class contributions equally sized. The $\chi^2$-approximation is reasonable in classes where either $A$ or $B$ is large or a large number of observations is available. An alternative method (MINQUE) for variance component estimation is given in Rao (1973), but has not been used here for computational reasons. When the variance $\sigma_z^2$ has been estimated, the *mean value* parameters are determined by weighted least squares (the summation runs over all stop time observations):

$$\underset{\{m_k\}}{\text{minimize}} \sum (z - \mu_z)^2/\sigma_z^2. \tag{18}$$

In practice, the estimation is performed stepwise:

The marginal distributions, i.e. boarding only ($A = 0$) and alighting only ($B = 0$), are first studied. They represent more than 40% of the Stockholm data and are used to estimate the variance and mean parameters for the two one-way streams (see (16)), by minimizing (17) and (18). The remaining observations, representing simultaneous boarding and alighting, form the basis in an iterative search for the best corrections $\mu_{xy}$, $\sigma_{xy}^2$ etc. The iteration starts with finding new trial values $\sigma_{xy}^2$—see (14)—and $\sigma_{yx}^2$, i.e. modifications $\Delta\sigma_{xy}^2$ and $\Delta\sigma_{yx}^2$ of the old ones: For current values of $\mu_x + \mu_{xy}$ and $\mu_y + \mu_{yx}$ (specific to each class), expression (12) is used to determine the dominating stream in the class under consideration. Given the appropriate variance, e.g. (15) if boarding dominates, the least squares system corresponding to (17) is updated. Then the normal equations are solved for new modifications $\Delta\sigma_{xy}^2$ and $\Delta\sigma_{yx}^2$, to be used in the next iteration. Before another variance iteration is performed, the latest variance values are used to determine new values of $\mu_{xy}$ (see (13)) and $\mu_{yx}$, by updating and solving (18) for minimizing modifications $\Delta\mu_{xy}$ and $\Delta\mu_{yx}$ in a similar way.

The iterative procedure, starting with 0-correction terms, can be terminated after one iteration or may go on until convergence, i.e. 0-modifications. Because of the non-smoothness of the problem, only local optimality can be guaranteed. (If an updating leaves dominance unaffected for each class, the method leads straight to a local optimum. In fact the least squares solution (LSS) is formulated on this stability assumption, and when that fails the LSS may be worse for the objective (18) than the basic iteration point. In that case the method looks for the next iteration point along the direction to the LSS, using (18) for interpolation. The whole iterative procedure is interrupted when no point along a chosen direction reduces the objective more than a predetermined proportion.)

The constant stream intensity model (11) is treated somewhat differently: Since mean and variance are coupled, their quotient being the stream intensity, we first estimate the means $\mu_x$ and $\mu_y$ from the marginal distributions (where the stream-dominance is obvious, e.g. $\mu_z = \mu_x$ for boarding-only events) by unweighted least squares, ignoring the variances.

The corrections $\mu_{xy}$ and $\mu_{yx}$ are then calculated as described for the general models and the variances are updated according to their relation to the means, i.e. new stream intensities $\lambda_b$, $\lambda_a$ are determined by minimizing (17) (with inserted class variances equal to

$$\frac{1}{\lambda_b}\cdot(\mu_x + \mu_{xy}) \quad \text{and} \quad \frac{1}{\lambda_a}\cdot(\mu_y + \mu_{yx}),$$

respectively), yielding

$$\frac{1}{\lambda_b} = \frac{\Sigma e_c w_c \cdot \mu_{zc}/s_c^2}{\Sigma w_c \cdot (\mu_{zc}/s_c^2)^2}$$

where the sums include those classes $c$ where $\mu_{zc} = \mu_x + \mu_{xy}$ (boarding dominance); a corresponding calculation is performed for $1/\lambda_a$.

Since the $\mu$-corrections objective is an unweighted sum of squares, it is unaffected by the variances, i.e. $\lambda_b$ and $\lambda_a$ are computed once—after the final $\mu$-iteration.

## 7. FURTHER RESEARCH AND DEVELOPMENT

As with most models, there are two main directions of development: while retaining the present scope, the complexity of the model may be increased or else, with the risk of having to drop some details, the model may be extended to represent more general situations. Which of these courses should be followed, is for the potential user to decide. One must keep in mind that the use of the model should dictate its form and not the other way round. We will only point at some possible applications and refinements of the present mathematical model, as part of an interactive simulation program. In this connection we will also discuss its limitations.

As it is, the model is fit for descriptive and didactic purposes, since it represents the route from the operator's or traffic controller's point of view. In order to increase the realism, a dependence of, e.g. link travel times on signal settings, a smoother arrival intensity variation during the peak and a time evolution of alighting proportions at certain stops might be introduced into the model. The stop time models can also be further developed, especially the theoretical validity of the approximations made and the generalization to routes with heavy bus loads. However, the necessity of these and other refinements can only be assessed by extensive use of the model. We hope that the future experiences of the model will give answers to these questions.

The simulation model is also intended for the planning and evaluation of control strategies. With the present data, neither the Poisson distribution for arrival intensities nor the simplified travelling assumption can be tested; hence only collective passenger waiting and travelling time has been described in the simulation model. However, individual passenger movement is of interest in assessing the service level of the route. We would therefore like to introduce an origin-destination matrix, describing in detail the passengers' mean travel pattern (for earlier attempts in this direction see e.g. Dekindt and Griffe, 1970; Jackson and Wren, 1972 and Bly and Jackson, 1974), and to keep track of individuals, following them from their arrival to the queue to their final alighting. This would be particularly interesting if the model included several bus routes, with the possibility of changing buses. These extensions would allow more passenger-oriented measures of performance, and an evaluation of more general strategies, affecting several bus routes simultaneously.

It is uncertain, however, whether such models would be manageable for the human operator, without comprehensive and relevant measures of performance. The objectives (quality numbers) chosen in connection with the present model are defined in Andersson *et al.* (1979). We feel that the correspondence between such measures and the operator's aims should be studied much further. Another important, but still open, question is how imperfections in a simulation model affect the possibility of evaluating control strategies for operative application.

Apart from the above mentioned possible extensions of the model, there is a need for suitable models of the correlation between travel times for different buses on one link as well as for one bus on different links (see e.g. Chapman, Gault and Jenkins, 1976; Gerrard and Brook, 1972 and Jackson and Wren, 1972). We believe that a closer study of the underlying conditions, such as the variations of traffic density and the duration and extension of local traffic conditions and their effects in combination with varying bus driver behaviour, is needed in order to formulate appropriate mathematical models and to decide how data should be collected in order to substantiate such models.

Also, in trying to apply larger and more complex models, we are faced with two major problems. The need for larger amounts of accurate data implies high costs and technical problems. The computational cost also increases. These costs should be weighed against the increase of realism and utility.

Better tests of validity are also needed, since the models will be too large for checking by inspection. Although the models can be decomposed into smaller submodels, which can be tested separately by our or other methods, it becomes more difficult to analyse and evaluate these larger total models which also contain all the possible interactions between the submodels.

Despite all the problems still unsolved, it is our hope that the techniques and ideas presented in this paper shall be of some help in the development of the field.

## REFERENCES


Andersson P.-Å., Hermansson Å., Tengvald E. and Scalia-Tomba G.-P. (1979) Analysis and simulation of an urban bus route. *Transpn Res.* **13A**, 439–466.

Bly P. H. and Jackson R. L. (1974) Evaluation of bus control strategies by simulation. *Rep. 637*, Department of the Environment, Transport and Road Research Laboratory, Crowthorne.

Chapman R. A., Gault H. A. and Jenkins I. A. (1976) Factors affecting the operation of urban bus routes. *Working paper No. 23*, Transport Operations Research Group, University of Newcastle upon Tyne.

Cox D. R. and Lewis P. A. W. (1966) *The Statistical Analysis of Series of Events.* Methuen, New York.

Cox D. R. and Miller H. D. (1968) *The Theory of Stochastic Processes.* Chapman & Hall, London.

Cox D. R. and Smith W. L. (1961) *Queues.* Chapman & Hall, London.

Dekindt H. and Griffe P. (1970) Simulation d'une ligne d'autobus. *Proc. Symp. Traffic Control, Versailles.*

Gerrard J. P. D. and Brook D. (1972) A computer simulation model for bus operations planning. *Rep. UKSC 0040*, UK Scientific Centre, IBM United Kingdom Ltd., Peterlee, Co. Durham.

Hill B. (1963) The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *J. Am. Statist. Assoc.* **58**, 72–84.

Jackson H. S. and Wren A. (1972). Simulation of urban bus operations. *Rep. ULORU-10*, Operational Research Unit, University of Leeds.

Jenkins I. A. (1976) A comparison of several techniques for simulating bus routes. *Working paper No. 14*, Transport Operations Research Group, University of Newcastle upon Tyne.

Lesley L. J. S. (1975) The role of the timetable in maintaining bus service reliability. *Proc. Symp. Operating Public Transport*, University of Newcastle upon Tyne.

Oliver A. M. and Uren M. B. (1972) A computer simulation model of the E3 bus route: a detailed description. *Tech. Note TN42*, Department of Operations Research, London Transport Executive.

Rao C. R. (1973) *Linear Statistical Inference and its Applications.* 2nd Edn. Wiley, New York.

Scheffé H. (1959) *The Analysis of Variance.* Wiley, New York.

Zangwill W. I. (1969) *Nonlinear Programming: A Unified Approach.* Prentice-Hall, Englewood Cliffs, New Jersey.