Validation of TASHA: A 24-h activity scheduling microsimulation model

Matthew J. Roorda [*], Eric J. Miller [1], Khandker M.N. Habib [2]

Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ont., Canada M5S 1A4

Abstract

The objective of this paper is to verify/validate the results of an application of the Travel Activity Scheduler for Household Agents (TASHA) in the Greater Toronto Area (GTA), Canada. Activity generation and scheduling components of TASHA are validated using 1996 and 2001 travel survey data for the GTA. This validation proceeds in two parts: (a) verification that TASHA replicates the 1996 base case upon which the model was originally built; and (b) comparison of TASHA's forecasts of 2001 daily travel behaviour with observed travel survey data for 2001. TASHA activity generation and scheduling model components replicate observed activities with good accuracy and precision for the base year. Although TASHA is not able to predict an observed increase in activity participation rate in a five year forecast, the distribution of activities in the day is forecast with greater success. Predictions of average travel distance are oversimulated in the base year by 0.9%. Increased average distances are underpredicted for school, shopping, and other activities, and are overpredicted for work and return home activities. These validation results are promising, although there exist opportunities to improve model performance, and to further validate other elements of the TASHA model.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Activity based modeling; Validation; Activity scheduling; Travel demand forecasting; TASHA; Microsimulation

1. Introduction

Activity-based approaches to modeling travel demand are beginning to enter into practice in many jurisdictions across North America (Davidson et al., 2007), including Ohio (MORPC) (Vovsha et al., 2004), South Florida (FAMOS) (Pendyala et al., 2005), San Francisco (Jonnalagadda et al., 2001) and Portland (Bowman et al., 1998). Although there is growing consensus among travel behaviour researchers that such activity-based approaches are behaviourally superior to conventional approaches (early examples include Kitamura, 1988;

---

[*] Corresponding author. Tel.: +1 416 978 5976; fax: +1 416 978 5054.
   E-mail addresses: roordam@ecf.utoronto.ca (M.J. Roorda), miller@ecf.utoronto.ca (E.J. Miller), khandker.nurulhabib@utoronto.ca (K.M.N. Habib).
   [1] Tel.: +1 416 978 4076; fax: +1 416 978 5054.
   [2] Tel.: +1 416 978 5049; fax: +1 416 978 5054.

Jones et al., 1990; Pas, 1990), there is a need to verify these models' ability to replicate base year travel behaviour with good accuracy and precision and to validate their forecasting ability.

As with conventional modeling efforts, activity-based models used in practice usually involve a serious effort to test a model's ability to replicate base year behaviour. At the very least this includes statistical goodness of fit tests of various model components, and it may involve testing overall performance of the modeling system as a whole, for example, through comparison of predicted model volumes with road counts (e.g. Vrtic et al., 2007). Unfortunately, forecasting tests are more rarely undertaken. When data for a subsequent year becomes available, emphasis is usually placed on model updating rather than validation of the previous model. "Back-casting" techniques are also a potential tool for testing longitudinal stability of a model without having to wait for subsequent data to become available, but few examples of such tests can be found.

Arguably, less conventional modeling approaches require more rigorous validation efforts because such models have a smaller body of experience. In particular those that attempt to model new elements of human behaviour, such as the activity scheduling process cannot credibly claim to be better transportation policy analysis tools unless they can be shown to out-perform conventional models.

The objective of this paper is to verify/validate the results of an application of the Travel Activity Scheduler for Household Agents (TASHA) (Miller and Roorda, 2003) in the Greater Toronto Area (GTA), Canada. This modeling system attempts to replicate the sequence of decisions that are made in the formation of a travel/activity schedule. It builds a person's activity schedule from the bottom-up, beginning with the generation of individual and multi-person activity episodes, and entering those activities into a feasible activity schedule, shifting and shortening activities as necessary to resolve situations where activities overlap in time. Trip chains (tours) arise naturally out of this process, and TASHA includes a tour-based mode choice model that explicitly evaluates ridesharing, vehicle allocation and joint travel to joint activities (Roorda et al., 2006).

While the first implementation of TASHA was applied in a study of urban form changes on greenhouse gas emissions in the GTA (Miller and Roorda, 2002), it is fair to say that this first version of TASHA was not fully validated. Validation of the activity generation and scheduling components of TASHA, discussed in this paper, are done using 1996 and 2001 travel survey data for the GTA. This validation proceeded in two parts: (a) verification of the ability of TASHA to replicate the 1996 base case upon which the model had been originally built; and (b) comparison of TASHA's forecasts of 2001 daily travel behaviour with observed travel survey data for 2001.

The 1996 verification test is clearly an essential first step: if a model is unable to replicate its base year behaviour well it has little hope of forecasting the future adequately. As a result of early tests a number of changes were made in the original TASHA prototype to improve its ability to generate activities by time of day correctly. The 2001 test is a much stronger validation test. It can indicate the extent to which the model can successfully predict travel behaviour in a forecast context that is different from the calibration context.

2. Overview of the TASHA method

The Travel/Activity Scheduler for Household Agents (TASHA) is a model of activity scheduling and mode choice. It is designed to improve upon current four-stage modelling systems used in the Toronto Area in a number of ways, most importantly, the behavioural representation of human decision-making, the spatial and temporal precision of outputs, and the sensitivity to demand-oriented policies. It is designed either to operate as a stand-alone model or to be embedded within the Integrated Land Use Transportation Environment (ILUTE) modelling framework for forecasting long-term and short-term decisions of households and firms that lead to the evolutionary development of urban areas (Salvini and Miller, 2005). Key features of TASHA are as follows:

- It is *activity-based*. Travel arises out of the need to participate in out-of-home activities. Thus, in order to model trip-making, we need to understand how people organize their lives, since it is the interplay between the desire/need to participate in activities and the mobility/accessibility options provided by the transportation system that ultimately determines travel behaviour.

- It is *household-based*. Although individuals travel, their travel decisions are made to a considerable extent within the context of household-level interactions, constraints and needs. Availability of household vehicles, need to chauffeur children, joint household activities, etc. all have a significant influence on an individual's daily activity (and, hence, travel) pattern.
- It is an *agent-based microsimulation* model, in that persons and households are represented explicitly in the model as "intelligent objects" or "agents" who are capable of perceiving their environment, making decisions and acting into their environment. Microsimulation permits the full power of the disaggregate, activity-based approach to be exploited. As operationalized to date, the microsimulation is run for a 5% sample of households in the Greater Toronto Area, corresponding to the sampling rate of the GTA's travel survey.

The full conceptual design, methodology and preliminary results of an operational prototype of the TASHA activity scheduling model are given by Miller and Roorda (2003). A conceptual representation of the methods used in the TASHA model is shown in Fig. 1. As shown in Fig. 1, the activity generation model
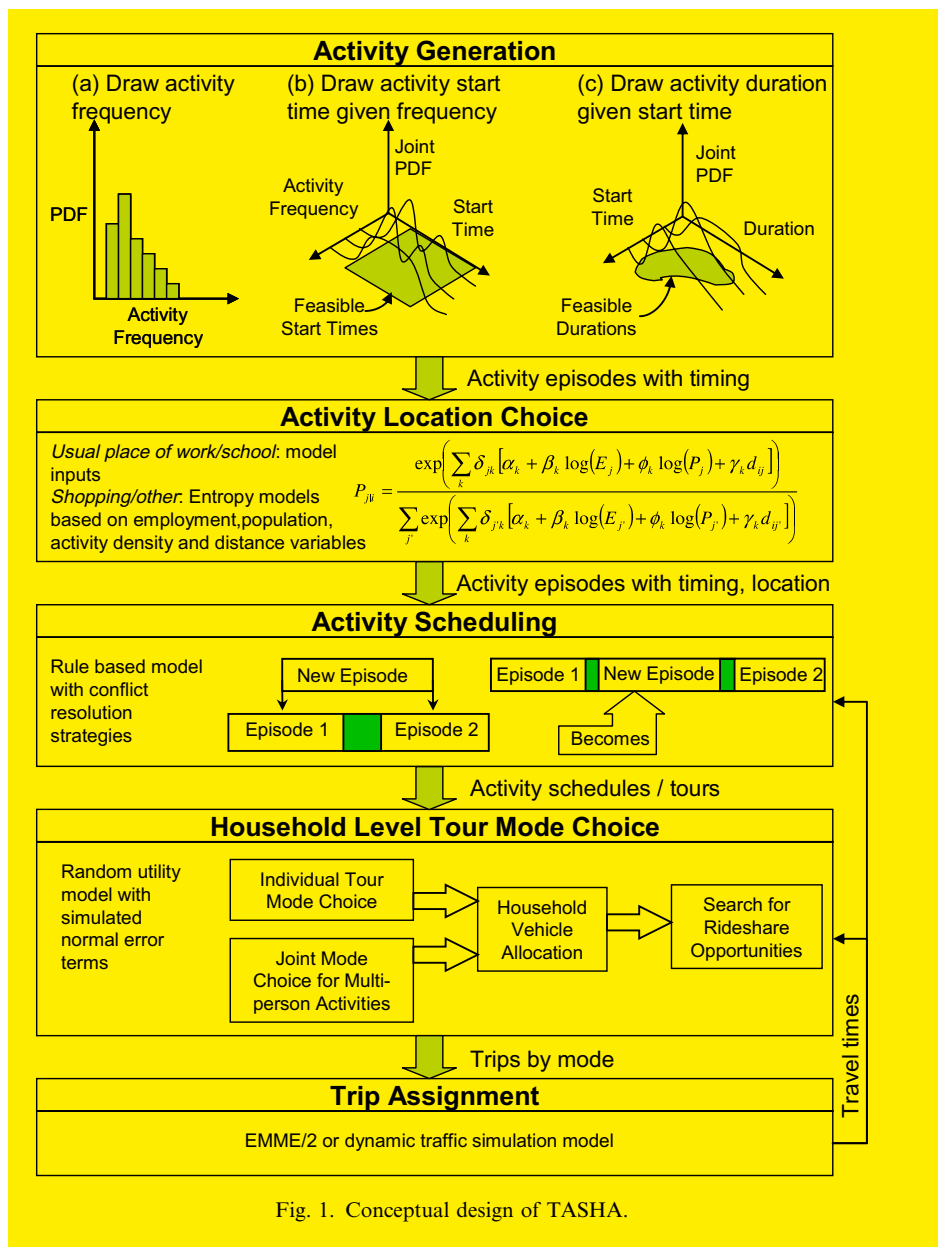

Fig. 1. Conceptual design of TASHA.

is based on random draws of activity attributes from observed joint probability distribution functions of frequency, start time and duration. A total of 262 distributions are developed from the 1996 TTS data, cross-classified by activity type, person, household and schedule attributes.[3] Activity start time is then randomly drawn from a joint frequency–start time probability distribution function, conditional upon activity frequency. Finally, activity duration is randomly drawn from a joint start time–duration probability distribution function, recognizing that these two activity attributes are closely interconnected (Ettema et al., 2007). Activity types in the model are limited to broad categories. The activity type categorization used is as follows:

(1) Work business (work episodes at a location other than the usual workplace).
(2) Work-at-home business (out-of-home work episodes for people whose usual workplace is home).
(3) Primary work (work episodes occurring at the usual workplace that begin before 3:00pm).
(4) Secondary work (returning to the usual workplace after 3:00pm).
(5) Return home from work (at home episodes embedded between primary work episodes, for example, returning home for lunch).
(6) School (including full time, part time, night courses, etc.).
(7) Joint other (non-work/school/shopping activities done with other household members).
(8) Joint shopping (shopping episodes done with other household members).
(9) Individual other (non-work/school/shopping activities done alone).
(10) Individual shopping (shopping activities done alone).

The locations of home and usual place of work/school are given as model inputs. The activity location choice for other activities is based on a series of entropy models, documented by Eberhard (2002). The probability that a person who lives in zone i (or for work-business activities, works in zone i) chooses location zone j is defined as:

$$P_{j|i} = \frac{\exp\left(\sum_k \delta_{jk}[\alpha_k + \beta_k \log(E_j) + \phi_k \log(P_j) + \gamma_k d_{ij}]\right)}{\sum_{j'} \exp\left(\sum_k \delta_{j'k}[\alpha_k + \beta_k \log(E_{j'}) + \phi_k \log(P_{j'}) + \gamma_k d_{ij'}]\right)} \tag{1}$$

where

δ_{jk}	=1, if zone j belongs to zone activity category k; 0 otherwise
E_j	=employment in zone j
P_j	=population in zone j
d_{ij}	=distance from zone i to zone j
$\alpha_k, \beta_k, \phi_k, \gamma_k$	=parameters to be estimated
k	=1, if the zone is the city core
	=2, if employment density > 3000workers/km^2 (work), shopping mall floor space > 100,000 sq. ft. (shopping), >50 retail stores/km^2(other)
	=3, otherwise.

The activity scheduling model uses a rule based method, in which activities are first added to "project agendas", in which they are placed into a preliminary time sequence with other activity episodes that have a common purpose (e.g. the work project). From these projects, activities are then moved into the person's activity schedule, one-by-one, in the order of precedence shown above. This order is consistent with the order of precedence observed from an interactive computer survey of activity scheduling (Doherty et al., 2004), as documented in (Roorda and Miller, 2005). The activity schedule begins with a single, default, at-home activity. As activity episodes are added to the schedule, conflicts (overlaps in time) can occur between out-of-home activities that are already in the schedule and those that are being inserted. These conflicts are handled by shifting, shortening, and sometimes rejecting activities, according to a detailed system of rules outlined in (Miller and

---

[3] The definition of these distributions was adjusted over previous versions of TASHA (Miller and Roorda, 2003) as the only action taken to improve base year verification results. The revised distributions reflect an improved specification of the classification structure relative to the original version.

Roorda, 2003). In the example shown in Fig. 1, a New Episode conflicts with Episode 1 and Episode 2. First, the travel between Episodes 1 and 2 is removed and new trips are defined from Episode 1 to the New Episode to Episode 2. Episodes 1 and 2 are moved earlier and later in the schedule, respectively, as necessary to make room for the New Episode and its associated travel. If other adjacent out-of-home activities are already scheduled prior to Episode 1 or after Episode 2, then Episodes 1, 2 and the New Episode are shortened to a minimum of 50% of their original duration. Episodes are rejected and regenerated up to a maximum of 10 times if the insertion procedure fails because of a lack of room in the schedule. Finally, a clean-up algorithm is applied to represent rescheduling/ fine tuning before execution of the activity schedule. This fine tuning algorithm rearranges activities in the schedule to remove unrealistically short work episodes with duration of less than or equal to 30 min.

The following step is a random utility tour-based mode choice model, that incorporates joint mode choice for joint activities, vehicle allocation and an explicit search for rideshare opportunities within the household. This model is not subject to verification and validation in this paper, therefore the reader is referred to the detailed description of the model in Roorda et al. (2006).

In a forecasting environment, this modelling system interacts with a traffic assignment model (to date EMME/2 has been used, although the system is well suited to feed a dynamic traffic assignment or simulation model). The system is set up to run iteratively, including travel time feedback from EMME/2 into the activity scheduling process. For trip assignment purposes, the trips generated by the 5% microsimulation sample are expanded to reflect travel by the full population of the Greater Toronto Area, based on geographically stratified sampling rates established in the travel survey.

This validation exercise focuses on the activity generation, location choice and activity scheduling components. For both the base year (1996) and the forecast year (2001) verification/validation exercises presented in this paper, the mode choice and trip assignment models are not run using TASHA activity schedules/tours. Rather, EMME/2 model runs are conducted using observed schedules/auto tours from travel survey data in the GTA, and the travel times from these EMME/2 model runs are inputs into the activity scheduling component of TASHA.

It is noted that no parameter adjustments have been made in this verification/ validation exercise that would be considered "calibration" actions, as would typically be required prior to operational use of a model for practical planning applications. This practical step is intentionally left out in order to honestly demonstrate the strengths and short-comings of the underlying behavioural approach embodied in TASHA, and to identify the areas in which behavioural methodological improvements are required. Clearly, model fit can be dramatically improved through the use of calibration factors in an operational context.

3. Data

The TASHA model is based on the 1996 implementation of the Transportation Tomorrow Survey (TTS) (Data Management Group, 1996). The TTS is a large-scale cross-sectional trip diary survey that has been conducted on approximately 5% of GTA households every five years since 1986. Prior to the use of the 1996 TTS database for model development, it has been cleaned and transformed, such that it is suitable for activity-based model development and tour-based mode choice model estimation. The transformation/cleaning process is documented by Eberhard (2002).

The activity scheduling component of TASHA makes use of the 1996 TTS data for building activity generation distributions and estimating the location choice models (see Eberhard, 2002). The 4.9% sample of GTA households and persons surveyed in the 1996 TTS (after data cleaning, the sample is reduced to 4.5%) also form the base year sample that is microsimulated by TASHA (81,554 households, and 219,773 persons). The 1996 model verification involves the comparison of model generated activity schedule attributes to those observed in the 1996 TTS database. It is called model verification, rather than model validation, simply because no new data are being introduced for testing the model results beyond those data that are used for model development (*validation* traditionally makes use of a "hold-out sample").

The 2001 model validation involves running the TASHA model on the 5.8% sample of Toronto Area households and persons surveyed in the 2001 TTS (113,608 households, and 315, 202 persons).

4. Base year (1996) verification

The base year verification is a test of the activity generation (frequency, start time, duration), activity location choice, and activity scheduling model components of TASHA. Since these individual model components are each estimated on the base data we have strong confidence that, as individual models, they reasonably reflect reality. However, what is not clear is whether the *system* of model components, *in combination*, is able to replicate observed outcomes of the sequence of decisions in reality. Furthermore, the activity scheduling model component is a non-parameterized rule-based system. The scheduling rules, although simple and logical, are nonetheless a hypothesized set of rules, *informed by* but not *calibrated to* observed activity scheduling data from the CHASE© survey (Doherty et al., 2004; Roorda and Miller, 2005). It is only possible to validate this type of scheduling model by observing aggregate results and searching for systematic biases that are the outcome of the simplified activity scheduling process. The base year verification is also a generalized test of the microsimulation approach, to ascertain whether disaggregate microsimulated outcomes give sensible aggregate results. In this verification exercise, 10 replications of the TASHA simulations have been completed. The mean and standard deviation of the model outcomes for these 10 replications are presented.

Overall, in the process of modelling activity/travel behaviour it is necessary to make many simplifying assumptions and abstractions of reality. One of the questions that we seek to answer in the base year model verification is whether the appropriate balance of modelling decisions has been made. Good modelling practice avoids modelling one component of behaviour with a high degree of detail and precision, only to have that precision consumed by gross inaccuracies in another poorly modelled component of behaviour. By systematically testing a range of indicators of model performance, such imbalances can be identified. In this verification exercise, combinations of activity frequency, activity start time, activity duration and travel distance by activity type are assessed.

4.1. Verification results – activity frequency

Table 1 shows the frequency of activities generated and successfully scheduled by TASHA, in comparison with the TTS database. This is a test of both the activity generation and the activity scheduling model components. Overall the total number of activities that are generated by TASHA match the observed number to within 0.2%. The process of activity scheduling involves, in some cases, the rejection of activities that conflict with other activities that have already been inserted into a person's activity schedule. As shown in Table 1, this results in an underestimation of observed school, shopping and "other" activities of 4.8%, 7.3% and 2.9%, respectively. These underestimations are many times greater than the standard deviations over 10 Monte Carlo simulations, indicating that the difference is not due only to simulated random variation. This result is expected because school, shopping and other activities are considered to be lower priority activities than work, they are inserted into the activity schedules later in sequence, and therefore they are more likely to be rejected. By contrast, work activities are very closely simulated. Return home activities are not generated explicitly by TASHA, rather, they are the "default" activity that is done if there are no other out of home activities generated. The fact that we replicate the number of home activities with only a 2.7% oversimulation is a strong indicator of model quality. The oversimulation here is due to chaining behaviour as will be discussed later in this paper.

4.2. Verification results – activity start time

Clearly, for transportation infrastructure and policy analysis it is crucial to properly replicate activities/travel by time of day. Fig. 2 shows that, averaged over the sample of persons in the GTA, activities of different types are generated by TASHA with the correct distribution over time throughout the day (within 2%, nominally, for all but 1 h for return home). Kolmogorov Smirnov tests were applied to verify this, and for all activity types this test resulted in $P > 0.97$ indicating a very high probability that the modelled and observed distributions are the same. This is not surprising, since activities are generated based on the observed joint distributions of activity frequency and start time. However, these graphs are evidence that the rules of activity scheduling assumed in TASHA do not introduce any major biases in activity start time. First, it is noted that

Table 1
Activity frequency comparison, TASHA vs TTS, 1996[a]

Activity type	Model average total activities (TASHA)[b]	Model std. dev. total activities (TASHA)[b]	Observed total activities (TTS)	Model ±activities (#)	Model ±activities (%)	Model average distance (km)[b]	Model std. dev. distance (km)[b]	Observed average distance (km)	Model ±distance (km)	Model ±distance (%)
Work	95,173	224	95,152	21	0.0	12.41	0.02	12.08	0.33	2.7
School	28,881	60	30,352	−1471	−4.8	5.58	0.02	5.05	0.53	10.5
Shopping	32,891	134	35,469	−2578	−7.3	5.62	0.05	4.85	0.77	15.9
Other	58,998	260	60,741	−1743	−2.9	7.64	0.03	7.48	0.16	2.1
Home	180,796	271	175,966	4830	2.7	8.43	0.02	8.8	−0.37	−4.2
Total	396,739	405	397,680	−941	−0.2	8.83	0.02	8.74	0.08	0.9

[a] Comparison excludes trips external to the Greater Toronto Area, facilitate passenger trips, and trips made by children younger that 11 years old.
[b] Model results are the average (standard deviation) of 10 replications of the Monte Carlo simulation.
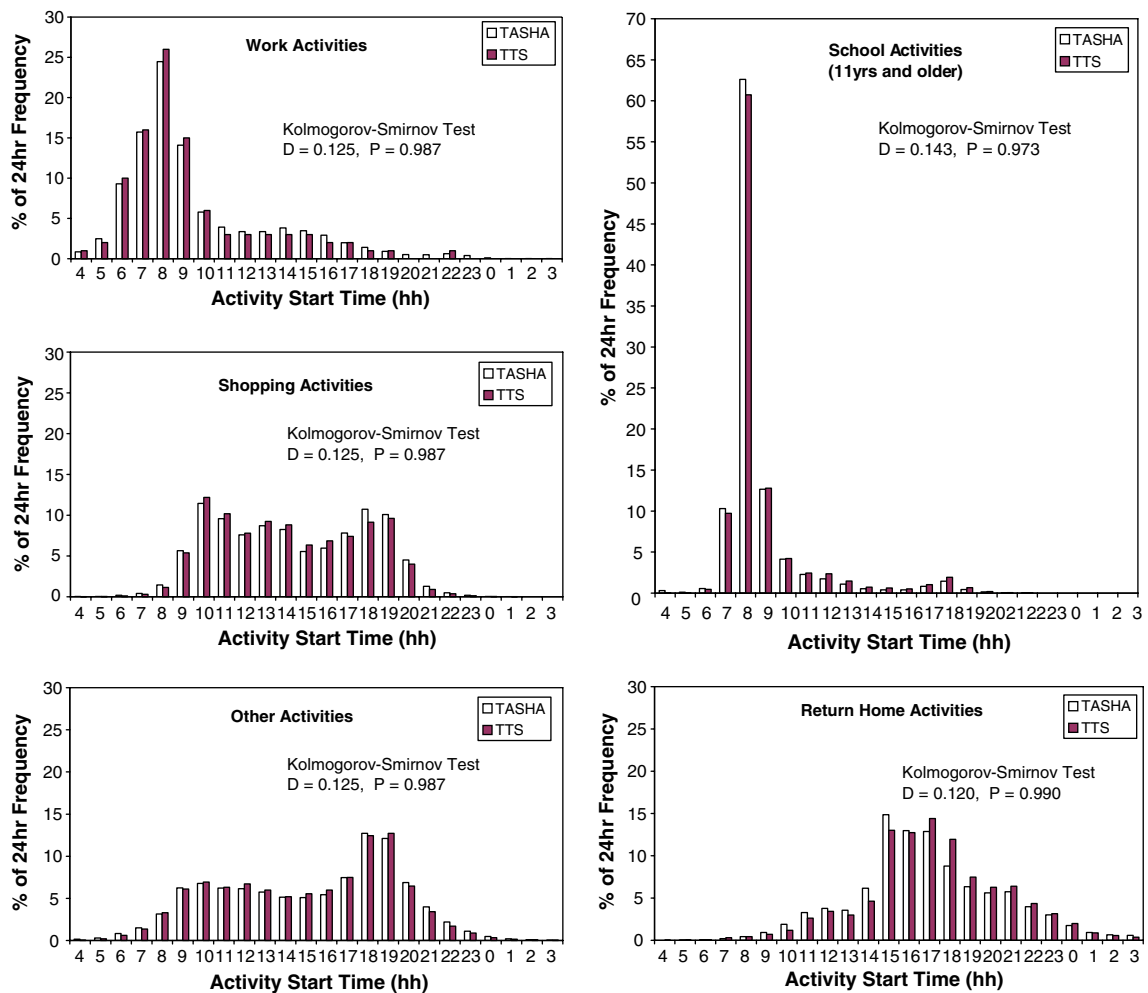

Fig. 2. Base year verification – weekday activity start time distribution.

even though school, shopping and other activities are undersimulated in total, the rejections do not add a major bias to the distribution over the day. If anything, there is a slight undersimulation of such activities during work hours of 8:00 am to 6:00 pm and a slight oversimulation in the evening hours, when conflicts with work are less likely to occur. Second, the temporal distribution of return home activity start times shows the most discrepancy, with bias towards returning home between the hours of 9:00 am to 4:00 pm and an undersimulation from 5:00 pm to 7:00 pm. Since return home activities are not generated directly, but are rather a by-product of the generation of the frequency, start time, duration and scheduling of out-of-home activities, this is considered to be a stronger indicator of overall model quality.

4.3. Verification results – activity duration

Average activity durations for each activity type, by activity start time, are shown in Fig. 3. These graphs show that the average activity duration for 10 TASHA simulations follows a similar pattern through the day as observed data (TTS) for each activity type. For all activity types, the average activity duration is lower for activities that start later in the day, an expected result. The Kolmogorov Smirnov tests for these distributions are inconclusive for all activity types except for "other" activities, for which the null hypothesis that the distributions are the same is rejected with 90% confidence. However, the parts of these graphs that show differences between observed and simulated average durations are all at times of day when there are low activity frequencies, as noted on the graphs by the band in which 90% of trips take place. Within the bands of day in which 90% of activities take place, the modelled activity duration is within 10 min or 10% of the observed average for all cases except work (2:00–3:59 pm) and "other" activities (2:00–5:59 pm and 8:00–9:59 pm).

4.4. Verification results – activity location

The overall average distance travelled to each activity type is shown in Table 1. Distances to all activity types except for the return home are being oversimulated and the oversimulation exceeds the random variation of the model as measured by the standard deviation by many times. The source of this error can be attributed
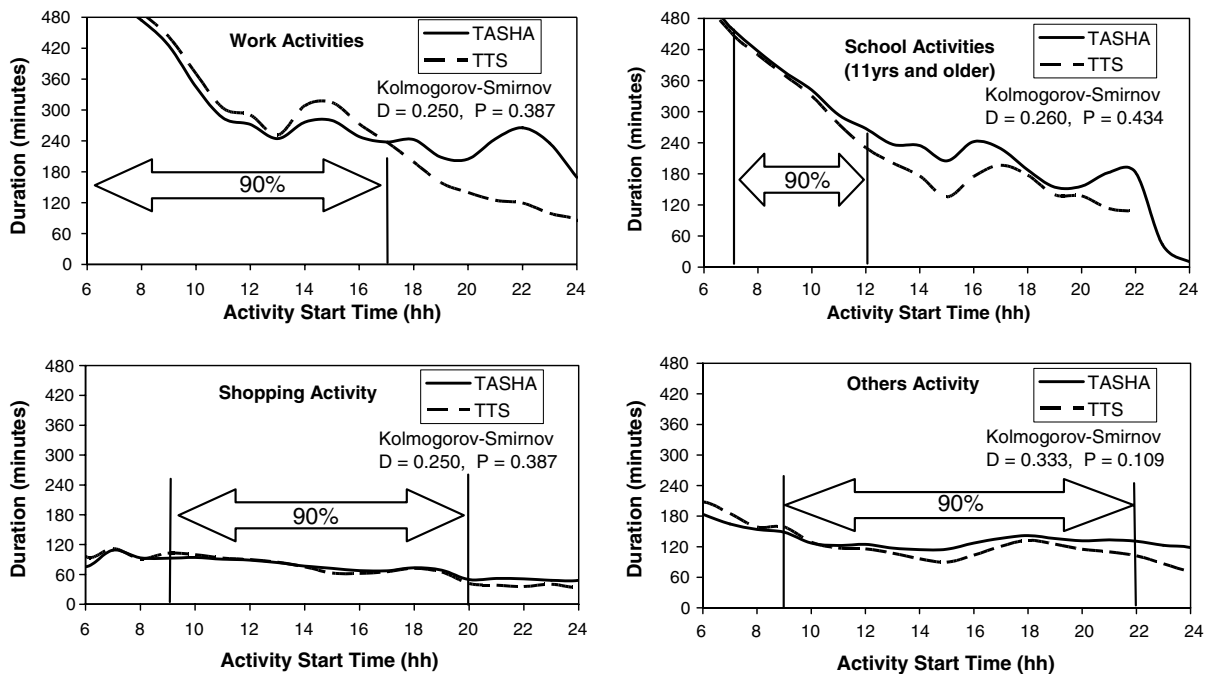

Fig. 3. Base year verification – weekday activity duration distribution.

to two possible problems. First, for activities that are not at the usual place of work/school, the activity location choice model is a spatial interaction model that is *either* home-based (shopping and other) *or* work-based (for work business meetings at locations other than the usual place of work). Activity locations are not chosen relative to the locations of other activities generated for that day. For example, in the model, shopping and other activities are generated using distance from home as the only spatial measure, yet it is probable that many shopping and other activity locations are chosen that are close to a person's work, and it is also probable that shopping and other activities done in sequence have "clustered" locations. The simplified approach used in TASHA does not account for these interactions. Second, TASHA synthesizes activities individually before arranging them into schedules and developing trip chains as a result. While this can be considered a behavioural strength of TASHA, it can result in activity chains that are less "optimized" than is observed in reality. As shown in Fig. 4, TASHA tends to oversimulate trips chains with only two trips, and tends to undersimulate three-trip chains. This contributes to the oversimulation of trip distances.

It is noted that activity location choice in TASHA is made irrespective of the time of day that the activity occurs, and vice versa. This decoupling of activity location choice and time of day choice by activity type is a source of concern in our model. However, as shown in Fig. 5, the trip distances aggregated over all activity types vary by time of day insofar as the different activity types, with different average trip distances are each modelled at appropriate times of day.

4.5. Base year verification conclusions

(1) Overall, TASHA replicates the aggregate observed frequency of work, return home and other activities with reasonable accuracy ($\pm 3\%$), whereas school and shopping activities are undersimulated by 4.8% and 7.3% respectively. The percentage distribution of activities by hour throughout the day is simulated within 2% over all hours, except the evening peak for return home activities. This is a very significant
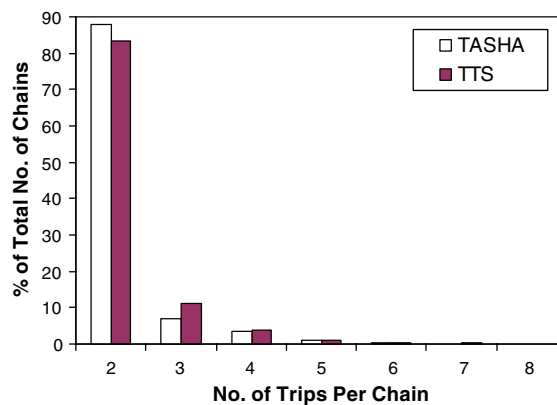

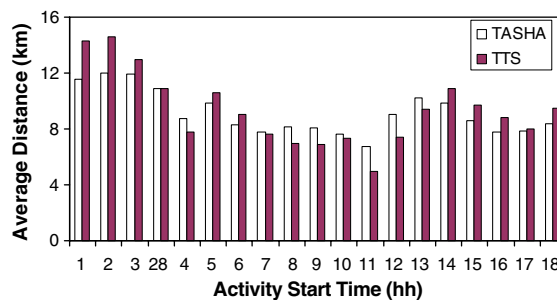Fig. 4. Base year verification – Distribution of number of trips per chain.

Fig. 5. Base year verification – average distance to activities of all types.

step above conventional modelling systems, which simply do not have a detailed representation of activity start time and duration, and often do not model for a full 24 h period.

(2) The activity location choice model component is the weakest link of TASHA, resulting in a systematic oversimulation of travel distance for all out-of-home activity purposes, especially school and shopping (10.5% and 15.9% oversimulations, respectively). An insensitivity of activity generation/scheduling to the distances travelled to particular activity types by time of day has been found, however, the aggregate travel distances over all trip purposes vary by time of day because location choices are made in a disaggregate fashion.

5. Model validation – testing a 5-year forecast to 2001

A form of validation that is rarely done in practice is a test of the forecasting ability of a modelling system. The concern is that, while a modelling system may succeed in replicating the data upon which the model was originally built, it may perform poorly when applied in future years. Just as "model overfitting" can take place in a very simple regression model by adding too many explanatory variables, it is possible to "overcalibrate" a model to current conditions resulting in an overly precise model that forecasts poorly. If that is the case, the modelling exercise is of limited value because the ultimate purpose of such a model is to forecast behaviour in response to changes in transportation infrastructure or policy.

TASHA is structured using a "bottom up" procedure, in which activities are generated and then scheduled, rather than a "top down" procedure where patterns of activities are chosen. The "bottom up" structure was chosen because it was considered to model at a deeper level of human behaviour than that of the "top down" structure. TASHA attempts to model the desire to engage in activities of different types at different times, and then arranges those activities in consideration of the constraints imposed by household members, travel congestion and the limited amount of time available in a day. The expectation is that while travel patterns may change over time because of changing constraints, this deeper level of human behaviour remains more constant. It is not possible at this stage to test this hypothesis fully. However, by applying the TASHA model system that has been estimated for base year (1996) conditions to the forecast year of 2001, it is possible to at least see whether or not the high degree of modelling precision achieved for the base year is actually "false precision" when the model is applied in a five year forecast.

The validation test undertaken here does not include a forecast of the population, but rather uses the observed 2001 population as input into the model. This allows us to isolate inaccuracy due to model prediction (assuming correct inputs) from inaccuracy due to incorrect population forecasts. Of course, conventional models and TASHA alike require good forecast input values to accurately predict reality. The input values of TASHA are not much different from other disaggregate modelling approaches, including a modest range of person and household characteristics. Therefore, we leave the question of the sensitivity of the model to the accuracy of population forecasts for future research.

5.1. Year 2001 validation results – activity frequency

Table 2 shows the frequency of activities forecast by TASHA in comparison with 2001 TTS observed activity frequencies by activity type. The overall number of activities is undersimulated by 3.2%, which is significantly worse than the 1996 simulation, which was very close to observed total activity generation rates. This change can be explained by an increase in the per household trip rate observed in the GTA from an average of 5.60 trips/day in 1996 to 5.83 trips/day in 2001 (Data Management Group, 2003). Had the trip rate remained constant over time, then the model would have oversimulated by approximately 0.7%. TASHA is not able to predict the trend towards an increase in the per household trip rate. This trend is typical of industrialized nations (e.g. Susilo and Kitamura, 2007; McGuckin et al., 2005) and appears to be occurring because of a baby boom generation that is more mobile in old age than the pre-boom generation, a decreasing difference between in the travel behaviour among males and females (who have traditionally been less mobile than males), and a decrease in household size (leading to fewer economies of scale in such activities as shopping and household obligations) (Roorda et al., 2008).

Table 2
Activity frequency comparison, TASHA vs TTS, 2001[a]

Activity type	Model average total activities (TASHA)[b]	Model std. dev. total activities (TASHA)[b]	Observed total activities (TTS)	Model ±activities	Model ±%	Model average distance (km)	Model std. dev. distance (km)	Observed average distance (km)	Model ±distance (km)	Model ±distance (%)	Increase in model average distance 1996–2001 (%)	Increase in observed average distance 1996–2001 (%)
Work	143,990	329	145,123	−1133	−0.8	13.20	0.02	12.78	0.42	3.2	6.3	5.8
School	41,987	62	43,930	−1943	−4.4	5.58	0.02	5.30	0.28	5.3	0.0	5.0
Shopping	46,844	357	53,989	−7145	−13.2	5.80	0.03	5.40	0.40	7.4	3.2	11.3
Other	84,577	360	93,771	−9194	−9.8	7.88	0.03	8.02	−0.14	−1.8	3.1	7.2
Home	26,5031	364	264,588	443	0.2	8.92	0.01	9.22	−0.30	−3.2	5.9	4.8
Total	582,429	1131	601,401	−18,972	−3.2	9.34	0.01	9.26	0.07	0.8	5.8	5.9

[a] Comparison excludes trips external to the Greater Toronto Area, facilitate passenger trips, and trips made by children younger that 11 years old.
[b] Model results are the average (standard deviation) of 10 replications of the Monte Carlo simulation.
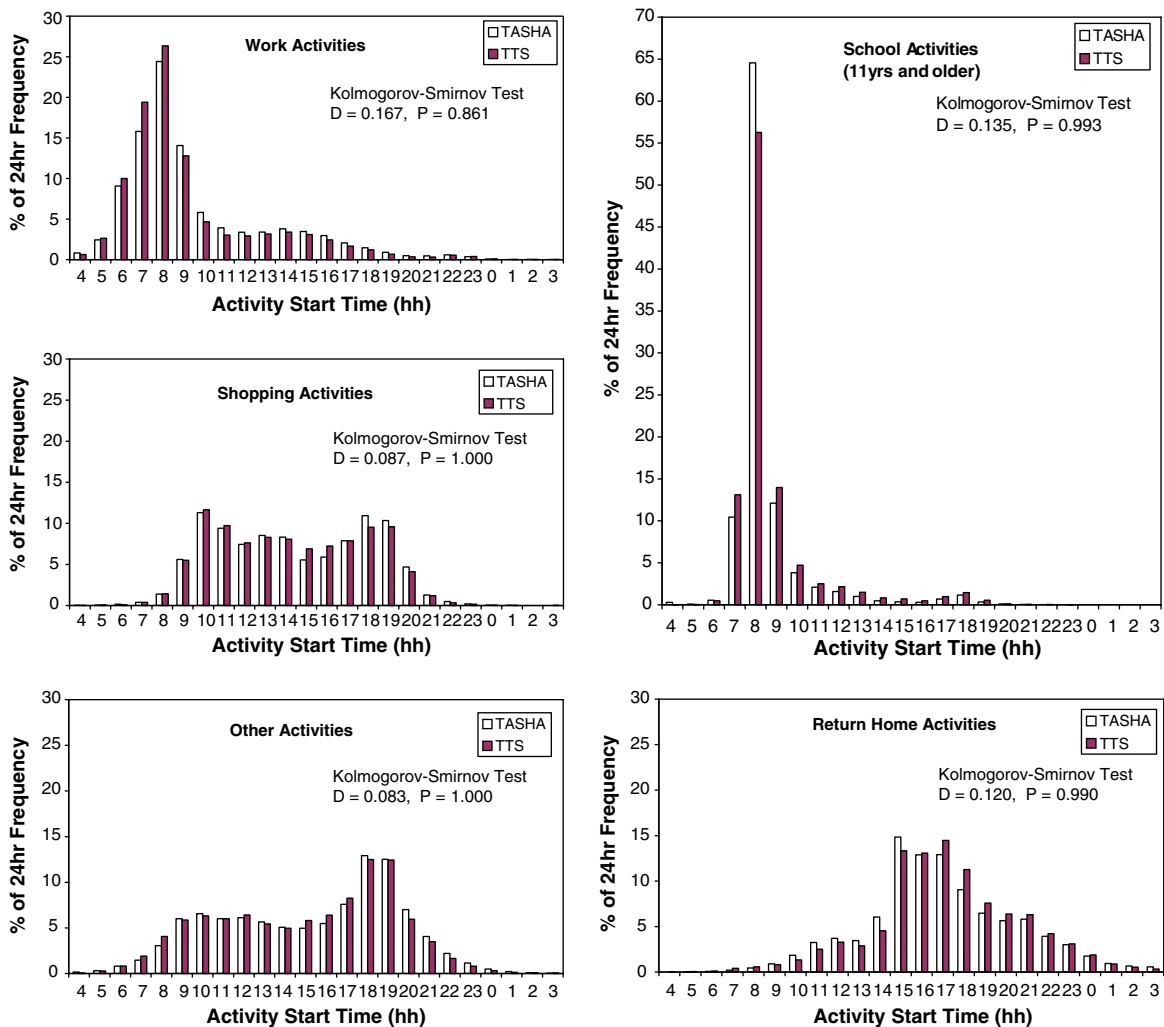
Fig. 6. Year 2001 validation – Weekday activity start time distribution.

5.2. Year 2001 validation results – activity start time

The distribution of start times by activity type for the year 2001 model forecast are shown in Fig. 6. Clearly, the activity start time distribution in 2001 did not change dramatically from that of 1996. In general, TASHA predictions of the average year 2001 activity start time distributions follow the same general shape of the 2001 observed distributions. Slight trends in observed behaviour have not been predicted well, such as a slight shift in work activity start times from the mid-day (9:00am–4:00 pm) to the early am peak period (5:00 am–8:00 am) and a slight widening of the peak of school start times from the 8:00–9:00 am period to the previous and the following hour. TASHA models 2001 shopping, other and return home activity start time distributions as well as it models 1996. Kolmogorov Smirnov tests, applied to confirm the goodness of fit, resulted in $P > 0.86$ for all activity types indicating a high probability that the modelled and observed distributions are the same.

5.3. Year 2001 validation results – activity duration

Activity duration forecasts by activity type are shown in Fig. 7. For all activity types, activity duration forecasts for 2001 are slightly worse than the base year predictions for "other" activities, but slightly better than
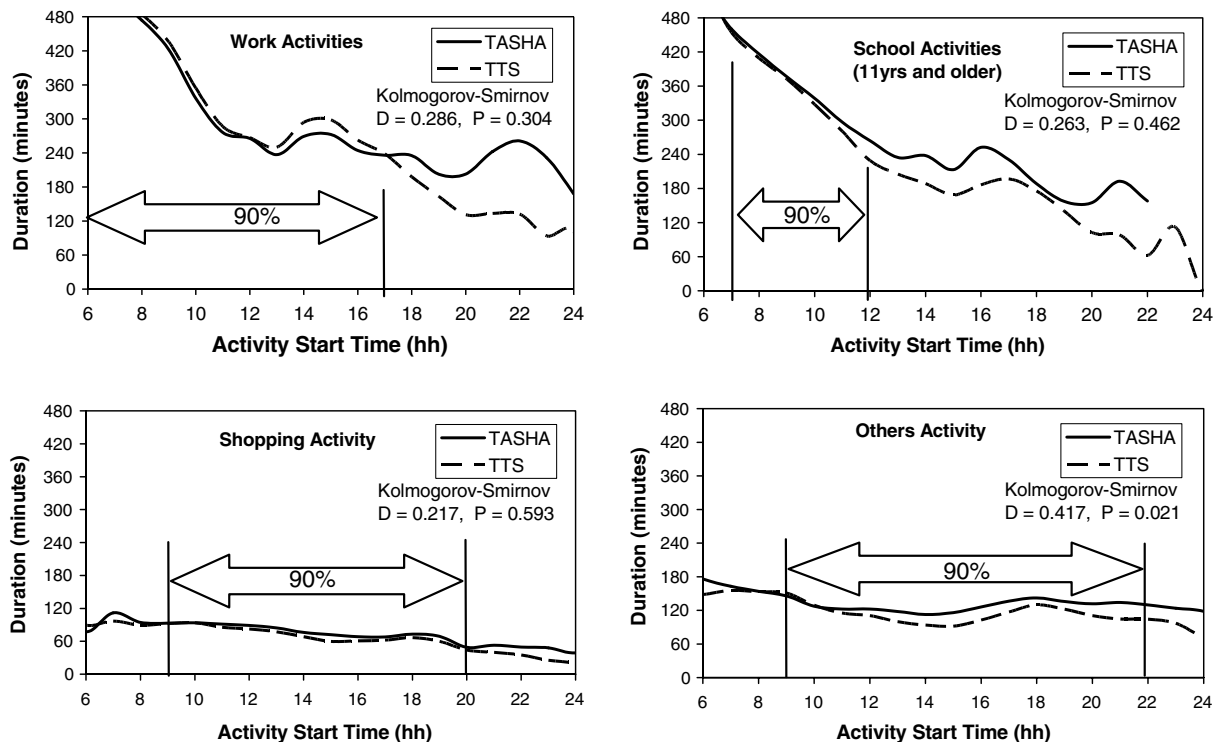
Fig. 7. Year 2001 validation – Weekday activity duration distribution.

base year predictions for work activities. Within the bands representing times of day in which 90% of activities take place, the modelled activity duration is within 10 min or 10% of the observed average for all cases except shopping (3:00–3:59 pm) and other activities (12:00–5:59 pm and 7:00–9:59pm). Kolmogorov Smirnov test do reject the hypothesis that the modelled and observed "other" activity distributions are the same ($P = 0.021$), but are inconclusive for work, school, and shopping activities (P ranges from 0.304 to 0.593).

5.4. Year 2001 validation results – activity location

As shown in Table 2, the average observed distance to each activity type has increased from 1996 to 2001 by 5.8%, 5.0%, 11.3%, 7.2% and 4.8% for work, school, shopping, other and return home activities, respectively. TASHA predicts increases in average travel distance of 6.3% for work, 3.2% for shopping, 3.1% for other activities, and 5.9% for return home activities, but predicts no increase in average trip distance for school trips. In summary, although the TASHA prediction of the average growth rate in travel distance of 5.8% is very close to the observed growth rate of 5.9%, TASHA oversimulates growth in work distance and underestimates growth in distance to other activities. Given that there are no built-in mechanisms in TASHA to predict increases in trip distance to shopping, other and return home activities, the fact that TASHA predicts any increase at all is an emergent outcome that on the whole is somewhat encouraging.

It is also of note that trip distances in 1996 were oversimulated by TASHA. Therefore, by undersimulating the *increase* in trip distance for school, shopping and other activities, the TASHA travel distances are actually closer to observed values in 2001 than in 1996. This is reflected in Fig. 8, which show an improved correspondence over the 1996 comparison, between simulated and observed trip distances for most times of day.

5.5. Year 2001 validation conclusions

(1) The overall impression from the year 2001 forecast validation is that TASHA aggregate forecasts of activity frequency are very good for work and return home activities (±1%), but undersimulate the
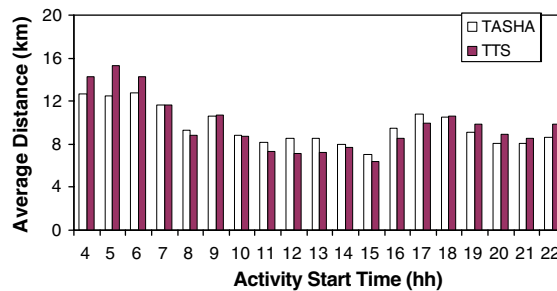
Fig. 8. Year 2001 validation – average distance to activities of all types.

frequency of shopping, school and other activities by 4.4%, 13.2% and 9.8%, respectively. The minor trends that have occurred, that could not be fully forecast by TASHA, include an overall increase in the number of activities generated, and slight shifts in the work and school start time distributions. The reason for the trend towards increased activity participation requires further investigation, but is expected to be due to a combination of a baby boom generation that is more mobile than the previous generation, reduced differences in the travel behaviour of males and females, and a trend toward decreasing household size that leads to increases in household maintenance travel.

(2) TASHA is able to predict the substantial increase in overall average trip distance observed from 1996 to 2001 with good accuracy (5.8% modeled growth compared to 5.9% observed growth). TASHA overpredicts the increase in average trip distance to work activities and to return home activities, captures a smaller proportion of the average increase in trip distance to shopping and other activities, and predicts none of the increase in school trip distance. The fact that TASHA predicts shifts in the right direction for all activity types except school is encouraging, but there is clearly a need to further explore the activity location choice model to better capture such trends.

(3) There is no evidence that there is a "false precision" in the model forecasts delivered by TASHA, or that the model is "over-fitted" to 1996 data. For the most part, activity frequency, start time and duration remain quite stable from 1996 to 2001, including minor hour-to-hour variations. Hence, at least for short-term forecasts, continued development of TASHA is warranted.

6. Summary of validation results and directions for further model improvement

Overall, the validation results presented in this paper are strong enough to warrant the consideration of TASHA as an alternative to conventional modeling systems currently in use in the Greater Toronto Area, and to warrant continued research and model development. It is clear that TASHA is able to reproduce activity/travel patterns in the GTA with a level of precision, by time of day, that is unavailable using conventional modeling approaches. This precision has been shown to be appropriate, at least for short-term forecasts, since patterns of activity distribution over time have remained fairly stable. Improvements, of course, can still be made in the model and will be addressed in subsequent phases of the on-going, evolutionary development of the model. Based on our validation findings, priority items for model improvement include:

- Developing improved models of non-work/school activity location that better account for multiple "anchor points" (e.g., home and work/school), as well as activity schedule constraints.
- Developing activity episode generation models that are sensitive to activity schedule constraints.
- Integrating the scheduling and mode choice procedures within a simultaneous decision process in which mode preference/feasibility is considered during episode scheduling.
- Elaborating the non-work projects in terms of the complexity of activity types they are capable of generating; and
- Incorporating distance effects in work-based mid-day "return home" episode generation.

It is also recognized that the verification and validation exercises discussed in this paper are aggregate in nature, both in terms of time and space, and only cover the activity scheduling/location choice and tour formation components of TASHA. Additional validation tests need to be conducted to assess the ability of TASHA to:

- Reproduce differences in travel behaviour in different parts of the GTA, for example, the differences between the central city, inner suburbs and the outer suburban areas.
- Reproduce differences in travel behaviour for different demographic groups.
- Reproduce base year and forecast mode choice decisions accurately at all times of day.
- Reproduce base year and forecast screenline traffic counts throughout the GTA.
- Produce realistic responses to the transportation infrastructure scenarios, land use scenarios, and demand-oriented policy scenarios that TASHA was designed to address; and
- Outperform conventional modelling systems on all of the above.

Acknowledgements

References

Bowman, J.L., Bradley, M., Shiftan, Y., Lawton, T.K., Ben-Akiva, M.E., 1998. Demonstration of an activity-based model system for Portland. Presented at the 8th World Conference on Transport Research, Antwerp.

Data Management Group, 1996. Transportation Tomorrow Survey: Design and Conduct of the Survey. Joint Program in Transportation. University of Toronto, Toronto.

Data Management Group, 2003. 2001, 1996 and 1986 Travel Survey Summaries for the Greater Toronto Area. Joint Program in Transportation. University of Toronto, Toronto.

Davidson, W., Donnelly, R., Vovsha, P., Freedman, J., Ruegg, S., Hicks, J., Castiglione, J., Picado, R., 2007. Synthesis of first practices and operational research approaches in activity-based travel demand modeling. Transportation Research Part A: Policy and Practice 41, 464–488.

Doherty, S.T., Nemeth, E., Roorda, M.J., Miller, E.J., 2004. Design and assessment of the Toronto area computerized household activity scheduling survey. Transportation Research Record: Journal of the Transportation Research Board 1894, 140–149.

Eberhard, L.K., 2002. A 24-h household-level activity based travel demand model for the GTA. Masters Thesis, University of Toronto, Canada.

Ettema, D., Bastin, F., Polak, J., Ashiru, O., 2007. Modelling the joint choice of activity timing and duration. Transportation Research Part A: Policy and Practice 41, 827–841.

Jones, P., Koppelman, F., Orfeuil, J.P., 1990. Activity analysis: state of the art and future directions. In: Jones, P. (Ed.), Developments in Dynamic and Activity Based Approaches to Travel Analysis. Oxford Studies in Transport, Avebury, pp. 34–55.

Jonnalagadda, N., Freedman, J., Davidson, W.A., Hunt, J.D., 2001. Development of microsimulation activity-based model for San Francisco: destination and mode choice models. Transportation Research Record: Journal of the Transportation Research Board 1777, 25–35.

Kitamura, R., 1988. An evaluation of activity based travel analysis. Transportation 15 (1–2), 9–34.

McGuckin, N., Zmud, J., Nakamoto, Y., 2005. Trip chaining trends in the US – understanding travel behavior for policy making. Transportation Research Record: Journal of the Transportation Research Board 1917, 199–204.

Miller, E.J., Roorda, M.J., 2002. Estimating CO_2 emissions and other transportation impacts of alternative urban forms, report II: summary of base network model test results. Growing Together: Prospects for Renewal in the Toronto Region, Background Reports. Report to the City of Toronto, Toronto: GHK Canada, pp. 17–40.

Miller, E.J., Roorda, M.J., 2003. A prototype model of 24-h household activity scheduling for the Toronto Area. Transportation Research Record: Journal of the Transportation Research Board 1831, 114–121.

Pas, E., 1990. Is travel demand analysis and modelling in the doldrums? In: Jones, P. (Ed.), Developments in Dynamic and Activity Based Approaches to Travel Analysis. Oxford Studies in Transport, Avebury, pp. 3–27.

Pendyala, R.M., Kitamura, R., Kikuchi, A., Yamamoto, T., Fujii, S., 2005. FAMOS: The Florida activity mobility simulator. In: CD Proceedings of the 84th Annual Meeting of the Transportation Research Board, Washington DC, January.

Roorda, M.J., Miller, E.J., 2005. Strategies for resolving activity scheduling conflicts: an empirical analysis. In: Timmermans, H.J.P. (Ed.), Progress in Activity-Based Analysis. Elsevier, Oxford, pp. 203–222.

Roorda, M.J., Miller, E.J., Kruchten, N., 2006. Incorporating within-household interactions into a mode choice model using a genetic algorithm for parameter estimation. Transportation Research Record: Journal of the Transportation Research Board 1895, 171–179.

Roorda, M.J., Morency, C., Woo, K., 2008. Two cities, two realities? A closer look at the evolution of trip rates in Toronto and Montreal. In: CD Proceedings of the 87th Annual Meeting of the Transportation Research Board, Washington DC, January.

Salvini, P.A., Miller, E.J., 2005. ILUTE: an operational prototype of a comprehensive microsimulation model of urban systems. Networks and Spatial Economics 5, 217–234.

Susilo, Y.O., Kitamura, R., 2007. Structural changes in commuters' daily travel: the case of auto and transit commuters in the Osaka metropolitan area of Japan, 1980–2000. Transportation Research Part A: Policy and Practice 42 (1), 95–115.

Vovsha, P., Petersen, E., Donnely, R., 2004. Impact of intra-household interactions on individual daily activity travel patterns. In: CD Proceedings of the 83rd Annual Meeting of the Transportation Research Board, Washington DC, January.

Vrtic, M., Fröhlich, P., Schüssler, N., Axhausen, K.W., Lohse, D., Schiller, C., Teichert, H., 2007. Two-dimensionally constrained disaggregate trip generation, distribution and mode choice model: theory and application for a Swiss national model. Transportation Research Part A: Policy and Practice 41, 857–873.