Networks and Spatial Economics, 5: (2005) 21–40

A Hybrid Model for Driver Route Choice Incorporating En-Route Attributes and Real-Time Information Effects

SRINIVAS PEETA*
School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA
email: peeta@purdue.edu

JEONG WHON YU
Division of Environmental and Urban Engineering, Ajou University, Suwon 443-749, Korea
email: jeongwhon@ajou.ac.kr

Abstract

The en-route driver behavior problem under information provision is characterized by subjective and linguistic variables, in addition to situational factors. Fuzzy modeling provides a robust mechanism to capture subjectivity and/or the linguistic nature of the causal variables. This motivates the development of a hybrid en-route route choice model that combines quantitative and fuzzy variables to more robustly predict driver routing decisions under information provision. Simulation experiments are conducted to analyze the ability of the hybrid model to capture en-route driver behavior effects in the within-day and day-to-day contexts.

Keywords: Hybrid route choice model, subjective and linguistic variables, en-route driver behavior

1. Introduction

The deployment and operational efficiency of Advanced Traveler Information Systems (ATIS) entail the accurate modeling of driver route choice behavior under real-time information provision and the calibration of the associated model parameters. Driver en-route routing decisions are influenced by personal attributes, response attitude to the supplied information, and situational factors such as time-of-day, weather conditions, trip purpose, and ambient traffic conditions. The latent preferences of drivers towards possible routes are typically difficult to capture accurately because they are significantly affected by past experience, subjective interpretation of the traffic information provided, and personal attitudes vis-à-vis dynamically changing traffic conditions. Most existing models are limited in their ability to capture the interacting effects of various situational factors, and typically cannot adjust model parameters in a within-day context. The latter capability is critical for consistency-checking procedures for the real-time operational deployment of advanced information systems.

*Corresponding author.

Driver route choice models under information provision have traditionally adopted the econometric theory of random utility maximization (Ben-Akiva and Lerman, 1985). Mahmassani and Liu (1999) used a multinomial probit framework to model and calibrate the commuter joint pre-trip departure time and route-switching behavior in response to ATIS, based on data from a laboratory interactive dynamic simulator. The study suggests that commuters switch routes if the expected travel time savings exceed an indifference band which varies with the remaining trip time to destination. Abdel-Aty et al. (1997) developed logit models to capture the effect of traffic information on commuter route choice using stated preference data. They analyzed the influence of travel time variability, and the effect of information on it. The choice between a longer route with reliable travel time and a shorter route with an uncertain travel time is investigated based on notions of risk aversion and risk-taking in route choice. Peeta et al. (2000) developed logit models to predict drivers' route diversion decisions under traffic information provided via variable message signs. They showed that a strong correlation exists between message content and driver route diversion decisions, which can be a control variable in operational strategies to enhance network performance (Peeta and Gedela, 2001). Khattak and de Palma (1997) use ordered probit models to investigate the effect of weather on traveler behavior, and suggest that commuters change their travel patterns systematically under adverse weather.

In the context of en-route driver behavior under information provision, qualitative phenomena such as inertia, compliance, delusion, freezing, and perception of traffic information, have recently been identified. Srinivasan and Mahmassani (2000) developed a multinomial probit model with a nested choice structure to examine inertia and compliance. Inertia represents the propensity to remain on the current path, while compliance represents the tendency to choose the path recommended by the traffic information system. They show that a driver's past experience with traffic information and the ambient network conditions dynamically influence the inertial tendency and the compliance propensity.

Probabilistic discrete choice models assume well-defined probability distributions to treat the randomness in driver route choice behavior. However, these assumptions may be restrictive in modeling qualitative phenomena in general. There are several studies that use other modeling approaches to address various aspects of driver route choice behavior under information provision. Nakayama and Kitamura (2000) investigated the notions of delusion and freezing under information provision by examining driver behavior using inductive reasoning. A driver supplied with limited or incorrect information on a route may form a delusion, that is, a biased perception of that route. If this delusion continues over time, the driver develops the habit of excluding that route from consideration. This leads the delusion to become a habit called freezing. Fujii and Kitamura (2000) examined the effects of information acquisition and driving experience using two hypotheses: information dominance and experience dominance. Information dominance implies that information effects become larger as the driver acquires more information, while experience dominance states that the influence of generic information weakens as the driver gains more driving experience. Lotan (1997) adopted a fuzzy modeling approach to analyze driver network familiarity and its influence on route choice behavior, using a driver simulator. The study indicates that familiarity with a route is dynamic, and can vary with the time-of-day, trip purpose, and under specific situations. Chen et al. (2001) use a dynamic weight adjustment method to investigate the effects of information on route choice behavior under three hypothetical traveler information scenarios: congestion, incident, and guidance. They show that the information content influences the relative weights of the route choice criteria.

Factors such as inertia, compliance, delusion, freezing, perception of traffic information, and familiarity vary over time, and are more meaningful in a day-to-day context. Nakayama and Kitamura (2000) address delusion and freezing through day-to-day evolution by updating *if-then* rules. Jha et al. (1998) develop a Bayesian updating model to update drivers' travel time perceptions on a day-to-day basis under information provision that reflects driver confidence in information. Individual drivers are assumed to have their own travel time distributions that indicate their travel time perceptions. These perceptions are updated based on travel time information provided on a specific day and past experience.

The en-route driver route choice problem under information provision is characterized by subjective/linguistic attributes, situational factors, and limited data availability. The standard approach in probabilistic discrete choice models to incorporate subjective or linguistic variables, also called qualitative variables, is to use ordinal and/or dummy variables. Both imply a discrete representation of the attribute values. This is adequate for variables (such as gender) which are discrete by nature. However, they are restrictive for qualitative variables that are continuous. For example, as illustrated by the membership functions in figure 1 (where μ denotes possibility), familiarity is a continuous variable implying that discretizations such as "very familiar" (ordinal value 5) and "familiar" (ordinal value 4) may not have clearly demarcated boundaries for an individual as implied in a discrete choice model, and indeed can have an overlapping region. This is further exacerbated by the fact that the boundary between "very familiar" and "familiar" can vary across individuals, implying

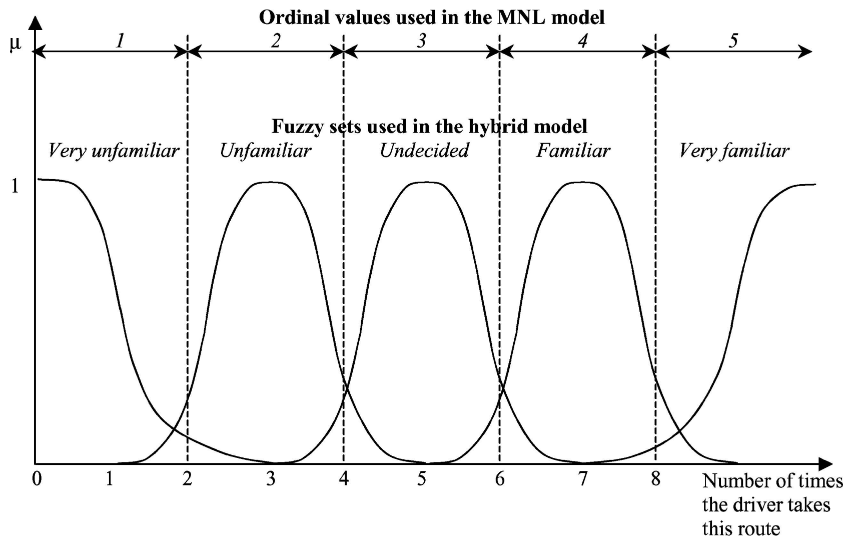

Figure 1. Modeling of driver familiarity with a route.

subjectivity. For some linguistic attributes, discretization may not even be relevant. For example, messages displayed by a traffic information system to describe a traffic situation can be "congestion ahead" or "expect delays", precluding meaningful discretization. In such instances, an ordinal ranking for the associated attribute (traffic information) is not meaningful, and subjectivity exists as well.

The issues discussed heretofore highlight the limitations of probabilistic discrete choice models, especially in the context of driver route choice under information provision. Fuzzy logic provides capabilities to address some of these issues. First, it enables the continuous modeling of relevant attributes using membership functions (Peeta and Yu, 2002a). This precludes rigid boundaries to demarcate attribute values and allows for overlaps between adjacent attribute values. Thereby, it provides a robust modeling capability to capture the uncertainties in the driver perception of the various attribute values. Second, the membership functions can be different for different drivers, accounting for subjectivity. Third, the parameters of the membership functions can be updated over time to reflect the day-to-day evolution of driver characteristics and perceptions, and illustrate some qualitative phenomena for driver route choice behavior under information provision. Finally, since fuzzy modeling is based on *if-then* rules, discretization and ordinal ranking can be circumvented where irrelevant, and rules consistent with the characteristics of that attribute can be developed directly. Several fuzzy logic based models (Peeta and Yu, 2002a; Lotan and Koutsopoulos, 1999; Lotan, 1998; Pang et al., 1999) have been proposed for driver route choice decisions under information provision. However, the justifications for the use of fuzzy modeling have previously been generic and primarily focused on the properties of fuzzy logic rather than on the barriers to modeling the underlying behavioral characteristics in the problem context. A comprehensive review of the literature for fuzzy logic based behavior models is provided in Peeta and Yu (2002a).

Probabilistic discrete choice models suffice for many route choice problems. However, as discussed earlier, they may not be able to capture qualitative variables and/or phenomena effectively under information provision. On the other hand, fuzzy models dilute the value of precise quantitative data as they entail the transformation of quantitative data into qualitative *if-then* rules, implying the use of real data in a proxy manner. That is, the actual data is converted into fuzzy data with a potential loss in precision. The limitations of probabilistic discrete choice and fuzzy models vis-à-vis modeling qualitative and quantitative variables, respectively, are problematic for the route choice problem under information provision where both variable types exist. In this paper, we propose a hybrid model that has a probabilistic discrete choice model form and contains quantitative and fuzzy variables to predict driver en-route routing decisions under real-time information provision. Thereby, variables that are naturally amenable to quantitative measurements are modeled as random variables, and those that are subjective or linguistically oriented are characterized using fuzzy rules. For example, travel time, travel distance, and quantitative traffic information can be characterized as quantitative (random) variables, and qualitative traffic information, familiarity, and inertia can be treated as qualitative (fuzzy) variables. The probabilistic discrete choice model form is retained for the hybrid model because it is widely used, has robust solution procedures, and can incorporate variables defined through other modeling mechanisms seamlessly.

The next section discusses the hybrid model structure and details the fuzzy modeling component. Then, simulation experiments are conducted to analyze the effects of en-route attributes and explore qualitative phenomena. In addition, prediction tests are used to compare the hybrid and probabilistic discrete choice models, and investigate the effect of heterogeneity levels in driver behavior characteristics. Finally, some concluding comments are presented.

2. Methodology

2.1. The hybrid model

The hybrid model is constructed by classifying each en-route driver behavior attribute as a quantitative or qualitative variable. It has the form of a probabilistic discrete choice model where the quantitative variables are directly incorporated and the qualitative variables are transformed fuzzy variables. The multinomial logit model (MNL) structure is used to construct the hybrid model in this study. If quantitative measurements are available for a variable, it is treated as a quantitative variable. If a variable contains linguistic information and/or requires subjective interpretation, it is treated as a qualitative variable. An associated data point is transformed using *if-then* fuzzy logic rules to a corresponding fuzzy variable value. Sometimes, interactions among quantitative variables can imply a qualitative phenomenon. In such instances, fuzzy or other modeling approaches can be used as an adjustment procedure to reflect those interactions. For example, travel time and quantitative traffic information can be used to capture the perceived travel time based on driver confidence in the traffic information system. However, even if fuzzy modeling is used for this adjustment, the associated mechanism is different from that for a qualitative variable in that *if-then* rules and non-quantitative data are not used here. Hence, such variables can be viewed as adjusted quantitative variables. The hybrid model combines utility contributions that are directly obtained for the quantitative variables and captured using fuzzy modeling for the qualitative variables, to represent its systematic utility component.

The random or disturbance component of the hybrid model utility function can be interpreted in light of its systematic component. For the quantitative variables, it incorporates the traditional sources of randomness (Ben-Akiva and Lerman, 1985). For the qualitative variables, the random component contribution is potentially due to errors introduced by fuzzy modeling, in addition to the traditional sources of randomness. However, the fuzzy modeling component itself captures some of the randomness in the problem context more robustly, mitigating its contributions due to the traditional sources of randomness to the disturbance term. This is enabled by its ability to better address qualitative variables that arise in the context of information provision, in terms of subjectivity, discretization issues, and consistent representation of variables for whom discretization and ordinal ranking are irrelevant. Hence, there are trade-offs in terms of errors introduced and mitigated due to the fuzzy modeling component.

In the context of en-route driver route choice problem under information provision, the utility of a route is determined using the quantitative variables and the transformed continuous representations of the qualitative variables. The transformation procedure is discussed in the next section. The hybrid model is as follows:

$$U_{\text{in}} = V_{\text{in}} + \varepsilon_{\text{in}} = \sum_{l} \beta^{l} X_{\text{in}}^{l} + \sum_{m} \gamma^{m} \Omega_{m}\left(Y_{\text{in}}^{m}\right) + \varepsilon_{\text{in}} \tag{1}$$

where,

U_{in} = utility of route i for driver n
V_{in} = systematic utility of route i for driver n
β^{l}, γ^{m} = coefficients of variables
X_{in}^{l} = value of quantitative (or adjusted quantitative) variable l on route i for driver n
Y_{in}^{m} = value of qualitative variable m on route i for driver n
$\Omega_{m}(\cdot)$ = transformation function to determine the fuzzy value of qualitative variable m
ε_{in} = disturbance term for route i for driver n

2.2. *Fuzzy model component*

The fuzzy model transforms qualitative data into continuous fuzzy variable values by assuming that driver route choice decisions are based on simple *if-then* rules, as shown in Table 1. Peeta and Yu (2002a) provide a detailed description of the three steps to transform qualitative data to fuzzy variable values. They are: (i) construction of *if-then* rules and corresponding membership functions, (ii) use of an implication operator for approximate reasoning, and (iii) a defuzzification step to generate a fuzzy variable value. S-shaped membership functions are constructed to represent the fuzzy rules and the associated qualitative variables. An *if-then* rule m is defined as: if A_m, then B_m. As illustrated in Table 1, the left-hand side of the *if-then* rules, A_m, indicates the linguistic label of an attribute for a driver and/or route. The right-hand side of the *if-then* rules, B_m, represents the propensity to take a route based on the linguistic label of the left-hand side.

The implication operator step is necessary because the attribute data value may not correspond directly to any *if-then* rules specified. For example, in a traffic information system, the message "slow traffic" may be provided to drivers. However, based on the "Qualitative traffic information" attribute in Table 1, this message does not correspond to any linguistic label on the left-hand side of the associated *if-then* rules. The implication operator determines the amount of overlap between the attribute value and each rule on the left-hand side of the *if-then* rules. The choice of an appropriate implication operator for approximate reasoning is problem-dependent. Here, the Larsen Product implication is used because it preserves the original shape of the membership function. In this approach, the use of each *if-then* rule m is based on χ_m, the amount of overlap between A_m and the attribute value A^* (such as "slow traffic") for an individual driver. Let B_m^* denote the modified B_m value based on χ_m. The χ_m andB_m^* are determined as follows (Tsoukalas and Uhrig, 1997):

$$\chi_m = \max_{x \in X} \min\left(\mu_{A_m^*}(x), \mu_{A_m}(x)\right) \tag{2}$$

$$\mu_{B_m^*}(y) = \chi_m \cdot \mu_{B_m}(y) \tag{3}$$

All *if-then* rules associated with a qualitative attribute are applied to its value, A^*. The associated B_m^* values are aggregated into one fuzzy set B^*.

Table 1. Fuzzy *If-then* rules.

Attribute	LHS	RHS
Familiarity	If a driver is very familiar with a route	He/she will take the route
	If a driver is familiar with a route	He/she will probably take the route
	If a driver's familiarity is undecided	He/she will be neutral
	If a driver is unfamiliar with a route	He/she will probably not take the route
	If a driver is very unfamiliar with a route	He/she will not take the route
Complexity	If a route is simple	He/she will probably take the route
	If a route is normal	He/she will be neutral
	If a route is complex	He/she will probably not take the route
Qualitative traffic information	If traffic condition is good	He/she will probably take the route
	If traffic condition is normal	He/she will be neutral
	If traffic condition is poor	He/she will probably not take the route
Compliance		
Weather	If weather is good	He/she will probably follow the recommended route
	If weather is not good	He/she will probably not follow the recommended route
Time-of-day	If time-of-day is daytime	He/she will probably follow the recommended route
	If time-of-day is not daytime	He/she will probably not follow the recommended route
Trip purpose	If driver is on a business trip	He/she will probably follow the recommended route
	If driver is on a leisure trip	He/she will probably not follow the recommended route
Inertia		
Weather	If weather is good	He/she will probably switch from the current route
	If weather is not good	He/she will probably not switch from the current route
Time-of-day	If time-of-day is daytime	He/she will probably switch from the current route
	If time-of-day is not daytime	He/she will probably not switch from the current route
Trip purpose	If driver is on a business trip	He/she will probably switch from the current route
	If driver is on a leisure trip	He/she will probably not switch from the current route

B^* is converted into a fuzzy variable value y^* through a process called defuzzification. The Center of Sums method (Tsoukalas and Uhrig, 1997) counts the overlapping areas to determine y^*:

$$y^* = \frac{\sum_l y_l \sum_m \mu_{B_m^*}(y_l)}{\sum_l \sum_m \mu_{B_m^*}(y_l)} \tag{4}$$

Hence, the attribute value A^* is transformed to y^* which serves as the input for the corresponding qualitative variable in the hybrid model. It should be noted that two levels of membership functions are used in the hybrid model: aggregate and disaggregate. The same aggregate membership functions for A_m, the left-hand sides of *if-then* rules, are used for a homogenous driver group. The membership functions for the attribute value A^* are determined uniquely for individual drivers, even within a homogenous group, based on the qualitiative variable attribute relevant for that driver. For example, one driver may encounter the message "slow traffic" while another driver in that homogenous driver group may see the message "congestion ahead". The use of unique membership functions for individual drivers enables capturing taste variations across drivers in terms of the qualitative variables, an important feature of the hybrid model given that subjectivity complicates the interpretation of these variables.

3. Analysis of the hybrid model

Experiments are conducted for the en-route driver route choice problem under information provision to address three primary objectives. They are: (i) to provide insights on the hybrid model, including its ability to capture qualitative phenomena, (ii) to compare the hybrid model and a MNL model, used as a representative discrete choice model, in terms of prediction capability, ability to capture qualitative phenomena, and day-to-day evolution of behavioral characteristics, and (iii) to investigate the effects of heterogeneity levels in driver behavior characteristics on the route choice prediction capability.

3.1. Study network

Figure 2 shows the network used to generate the en-route route choice data. It consists of 11 nodes and 21 links. Some links are one-way as indicated by the directional arrow. Only one origin-destination pair, between nodes 1 and 11, is considered. The pre-trip route choice decisions are made at node 1, and drivers are assumed to make en-route route choices at any nodes en-route to node 11 in response to real-time information provision. The link free-flow speeds range from 88.5 km/h to 104.6 km/h, and the link distances range from 3.2 km to 3.7 km.

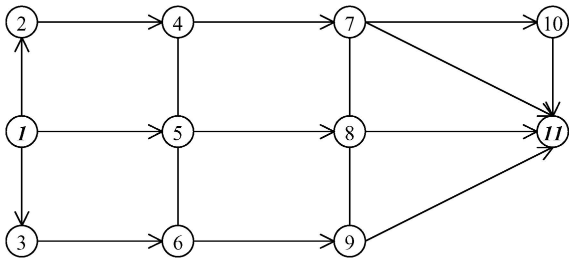

Figure 2. Test network.

3.2. En-route attributes

The three main categories of criteria for driver route choice decisions under information provision are: (i) driver attributes: socio-economic characteristics, network familiarity, confidence in information, sensitivity to delay, and personal preferences; (ii) route characteristics: travel time, travel distance, toll, facility type, route complexity, and location type; and (iii) situational factors: weather, time-of-day, and trip purpose. In the study experiments, the quantitative variables considered are: travel distance, travel time, and quantitative real-time traffic information. The travel time on route i is generated using the Greenshields model on the associated links. The travel distance on each route is the sum of the associated link lengths. The qualitative variables considered are: familiarity, route complexity, lingustic real-time traffic information, inertia, compliance, weather condition, time-of-day, and trip purpose. It is assumed here that a driver's familiarity with a link is based on the number of times he/she has traveled on it. The familiarity with a route is then computed as the average of the familiarities across all links constituting that route. For simplicity, route complexity is assumed to be based on the number of nodes in a route although two different routes with the same number of nodes may have different levels of complexity depending on other route characteristics such as the number of turning movements and stops. The real-time traffic information provided to drivers consists of three categories: (i) descriptive and qualitative, (ii) descriptive and quantitative, (expected travel time), and (iii) prescriptive (route recommendation). Five linguistic labels are used for the descriptive and qualitative information: (i) long delay expected, (ii) incident ahead, (iii) slow traffic, (iv) normal, and (v) free flow. Compliance is addressed only when the traveler information system recommends a route. Inertia is addressed in the context of a driver's current route. The weather conditions are simplified here into two linguistic labels: (i) good, and (ii) bad. The time-of-day is represented using two categories: (i) daytime, and (ii) nighttime. The trip purpose consists of two categories: (i) business, and (ii) leisure.

3.3. Route choice data generation

To the extent that real data is not available, the en-route decisions of drivers are generated here based on two reasonable route choice decision rules: lexicographic and utility maximization. Driver classes are constructed to represent different route choice behaviors. Table 2 illustrates four driver classes based on the two route choice decision rules and a behavioral characteristic (risk-averse and risk-willing). The lexicographic rules assume that the attributes can be rank-ordered by importance for a driver class. The driver chooses the most attractive route by eliminating inferior alternative(s) at each stage based on threshold values for the quantitative attributes and specific elimination rules for qualitative attributes. At the first stage, the most important attribute is compared among alternative routes. If the most attractive route is not chosen yet, the second most important attribute is evaluated next, and this process continues until a single route remains. The utility maximization rule assumes that the utility of a route can be expressed as a single value based on the notion of trade-offs among route attributes. A driver chooses the route with the maximum utility after evaluating the utilities of all alternative routes. The en-route path-switching decision

Table 2. En-route route choice decision rules.

Driver class	Rank	Attribute		Threshold/Elimination rule
Lexicographic rules				
Risk-averse	1	Perceived travel time		10% of the best route
	2	Travel distance		10% of the best route
	3	Familiarity		Take familiar routes
	4	Complexity		Avoid complex routes
	5	Qualitative (descriptive) information		Avoid delay
	6	Compliance	Weather condition Time-of-day Trip purpose	If weather is good “and” time-of-day is daytime “and” driver is on a business trip, follow the recommended route
	7	Inertia	Weather condition Time-of-day Trip purpose	If weather is bad “or” time-of-day is nighttime “or” driver is on a leisure trip, do not switch from the current route
Risk-willing	1	Perceived travel time		15% of the best route
	2	Qualitative (descriptive) information		Avoid delay
	3	Compliance	Weather condition Time-of-day Trip purpose	If weather is good “or” time-of-day is daytime “or” driver is on a business trip, follow the recommended route
	4	Inertia	Weather condition Time-of-day Trip purpose	If weather is bad “and” time-of-day is nighttime “and” driver is on a leisure trip, do not switch from the current route
	5	Travel distance		15% of the best route
	6	Familiarity		Take familiar routes
	7	Complexity		Avoid complex routes
Utility maximization rule*				
Risk-averse	$U_{\text{in}} = -0.15\frac{D_i}{10} - 0.15\frac{T_{\text{in}}}{30} - 0.05\frac{K_{\text{in}}}{30} + 0.20\frac{F_{\text{in}}}{5} - 0.15\frac{P_i}{3} + 0.05\frac{Q_{\text{in}}}{5} + 0.05C_{\text{in}} + 0.20I_{\text{in}}$			
Risk-willing	$U_{\text{in}} = -0.15\frac{D_i}{10} - 0.15\frac{T_{\text{in}}}{30} - 0.15\frac{K_{\text{in}}}{30} + 0.05\frac{F_{\text{in}}}{5} - 0.05\frac{P_i}{3} + 0.20\frac{Q_{\text{in}}}{5} + 0.20C_{\text{in}} + 0.05I_{\text{in}}$			

*Notation defined in Section 3.4.

is influenced, among others, by the willingness of the driver to take risks. This behavioral characteristic is modeled here as consisting of two groups to represent heterogeneity in driver behavior: risk-averse and risk-willing. For each driver class, the corresponding driver route choice behavior model is used to generate the en-route routing decisions of drivers through Monte Carlo simulation using the corresponding decision rules. 1000 drivers are considered for each decision rule. We assume equal numbers of risk-averse and risk-willing drivers under both rules.

The articulation of the experiment set-up is important for consistently interpreting the experiment objectives and associated results. The data generation discussed heretofore implies the real-world situation. This knowledge is used later to analyze the prediction rates of the hybrid and MNL models discussed in Section 3.4. However, only the chosen alternative and the associated attribute values are assumed to be known to the analyst, consistent with real-world data availability expectations. Therefore, the hybrid and MNL models use only the data available to the analyst to infer on the driver route choice behavior, and are unaware of the hypothetical driver route choice decision process used for route choice data generation. The hybrid model constructs *if-then* rules (Table 1) using insights from past route choice behavioral studies and survey data (Peeta et al., 2000), and then processes the data available to the analyst. Note that these rules do not share structural linkages with the lexicographic elimination rules used for generating the hypothetical route choices of some drivers. The MNL model directly uses the analyst data. To avoid potential bias, five randomly generated datasets are used to compare these two models. For each generated dataset, part of the data is used to estimate the model parameters and the rest is used to test the models. Hence, the results (other than Table 3 which shows the estimated model for a specific dataset) are based on averaging across these five datasets.

Table 3. Model estimation results.

Model			Lexicographic rules				Utility maximization			
			Hybrid		MNL		Hybrid		MNL	
Variable	Symbol		Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Route 1 constant	α_1		−3.557	−14.995	−2.183	−3.744	0.053	0.296	0.025	0.047
Route 2 constant	α_2		−2.981	−17.463	−2.125	−3.728	0.130	0.970	0.220	0.420
Route 3 constant	α_3		−2.912	−9.462	−2.137	−3.563	0.136	0.554	0.367	0.663
Travel distance	D		−0.028	−0.226	−0.149	−1.534	−0.204	−1.941	−0.107	−1.142
Experienced travel time	$\Psi(\cdot)$	T	−0.039	−2.984	0.039	3.295	−0.153	−11.980	−0.055	−4.476
Quantitative information		K			−0.041	−10.764			−0.065	−15.170
Familiarity	$\Omega_P(\cdot)$	F	0.092	0.431	0.001	0.019	0.964	5.078	0.122	3.575
Route complexity	$\Omega_F(\cdot)$	P	−0.577	−1.721	−0.036	−0.459	−1.178	−4.011	−0.223	−2.757
Qualitative information	$\Omega_Q(\cdot)$	Q	11.746	18.786	1.439	17.138	6.292	14.192	0.770	10.039
Compliance	$\Omega_C(\cdot)$	C	−0.138	−0.561	0.063	0.633	4.540	19.284	1.525	16.357
Inertia	$\Omega_I(\cdot)$	I	0.142	0.587	0.088	0.893	1.236	6.095	0.374	4.050
Weather	W				−0.073	−0.381			−0.049	−0.252
Time-of-day	G				0.113	1.016			−0.081	−0.417
Trip purpose	S				−0.021	−0.115			−0.041	−0.210
Number of observations			1000		1000		1000		1000	
$L(0)$			−1386.29		−1386.29		−1386.29		−1386.29	
$L(\hat{\beta})$			−476.65		−758.59		−621.42		−791.82	
ρ^2			0.656		0.453		0.552		0.429	
$\bar{\rho}^2$			0.655		0.450		0.550		0.426	
Correct prediction rate (%)			74.6		62.1		66.3		55.7	

*The units for travel time and distance are minutes and kilometers, respectively.

3.4. Model structure

The hybrid model approach is used to analyze en-route route choice decision-making at node 8. The en-route choice set of each driver at node 8 consists of four alternatives: (i) route 1 ($8 \rightarrow 7 \rightarrow 10 \rightarrow 11$), (ii) route 2 ($8 \rightarrow 7 \rightarrow 11$), (iii) route 3 ($8 \rightarrow 11$), and (iv) route 4 ($8 \rightarrow 9 \rightarrow 11$). In general, since alternative routes can share common links, a multinomial probit (MNP) structure may be reasonable in the en-route route choice problem context. However, it is computationally intensive precluding real-time applicability for traffic control architectures that require consistency-checking procedures to update model parameters. For instance, the MNP structure with 10 parameters and 500 observations needs approximately a 40-minute run of the LIMDEP software on a 500 MHz computer, while the MNL structure needs only a fraction of a second. Also, based on a preliminary analysis, the corresponding estimation results are not perceptibly different here as the overlaps across alternative en-route routes are minimal. Hence, the MNL model structure is used for the hybrid model in this study:

$$\begin{aligned} V_{\text{in}} = {} & \alpha_i + \beta_1 D_i + \beta_2 \Psi(T_{\text{in}}, K_{\text{in}}) + \beta_3 \Omega_F(F_{\text{in}}) + \beta_4 \Omega_P(P_i) + \beta_5 \Omega_Q(Q_{\text{in}}) \\ & + \beta_6 \delta_{\text{in}} \Omega_C(W_n, G_n, S_n) + \beta_7 \kappa_{\text{in}} \Omega_I(W_n, G_n, S_n) \end{aligned} \tag{5}$$

where,

α_i = alternative specific constant for route i,
β_j = coefficient of variable/function j,
D_i = travel distance on route i,
$\Psi(\cdot)$ = adjustment function to capture the perceived travel time,
T_{in} = travel time experienced by driver n on route i,
K_{in} = quantitative traffic information on route i for driver n,
$\Omega_F(\cdot)$ = transformation function to determine the fuzzy value of familiarity,
F_{in} = the number of times driver n took route i in the past,
$\Omega_P(\cdot)$ = transformation function to determine the fuzzy value of route complexity,
P_i = the number of nodes in route i,
$\Omega_Q(\cdot)$ = transformation function to determine the fuzzy value of descriptive qualitative traffic information,
Q_{in} = descriptive qualitative traffic information on route i for driver n,
δ_{in} = 1 if route i is the recommended route to driver n; 0 otherwise,
$\Omega_C(\cdot)$ = transformation function to determine the fuzzy value of compliance vis-à-vis recommended route i,
W_n = weather condition for driver n,
G_n = time-of-day for driver n,
S_n = trip purpose of driver n,
κ_{in} = 1 if route i is the current route of driver n; 0 otherwise,
$\Omega_I(\cdot)$ = transformation function to determine the fuzzy value of inertia vis-à-vis current route i.

Among the three quantitative variables, only travel distance is used directly in the above utility function. Travel time and quantitative traffic information serve as inputs to a fuzzy modeling based adjustment function $\Psi(\cdot)$ to capture the travel time perceived by a driver vis-à-vis his/her route choice decision. It is assumed here that the driver's perceived travel time is based on all prior experience and the quantitative traffic information (in terms of route travel time) currently provided. The parameters of the membership functions for T_{in} and K_{in} are determined based on the data available for these variables. For example, the mean and variance of the experienced travel time determine the parameters of the membership function for T_{in}. Driver confidence in real-time information is characterized through the membership function of K_{in}. For example, a driver with high confidence has a more narrow membership function than a driver with low confidence. Since the mean and variance of T_{in}, and the driver's confidence in traffic information potentially change after each trip, the membership function parameters are also updated. The perceived travel time $\Psi(T_{\text{in}}, K_{\text{in}})$ is obtained according to the weighted combination scheme suggested in Dubois and Prade (1988), as shown in figure 3. Each transformation function $\Omega(\cdot)$ is generated using the corresponding *if-then* rules illustrated in Table 1. The number of times a driver uses a route is used to infer on familiarity, and the number of nodes in a route is used to represent route complexity. We assume that inertia and compliance are influenced en-route by situational factors W_n, G_n, and S_n which vary everyday.

The pure MNL model, which is used to benchmark the hybrid model performance, uses ordinal values for the various linguistic expressions of a qualitative variable. Also, the driver experienced travel time and quantitative traffic information are treated as separate variables. Thus, the model does not explicitly incorporate the associated interaction effects. The situational factors, compliance, and inertia are treated as dummy variables. The model is as follows:

$$V_{\text{in}} = \alpha_i + \beta_1 D_i + \beta_2 T_{\text{in}} + \beta_3 K_{\text{in}} + \beta_4 F_{\text{in}} + \beta_5 P_i + \beta_6 Q_{\text{in}} + \beta_7 C_{\text{in}} + \beta_8 I_{\text{in}} + \beta_9 W_n + \beta_{10} G_n + \beta_{11} S_n \tag{6}$$

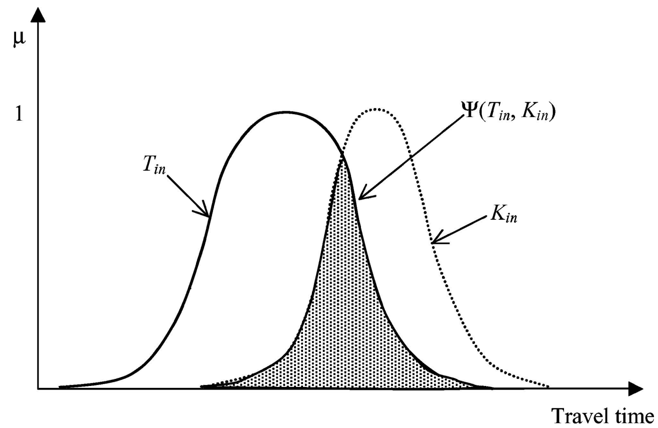

Figure 3. Weighted combination scheme to determine perceived travel time.

where,

β_j = coefficient of variable j,
C_{in} = compliance for prescriptive information on route i for driver n; 1 if route i is the recommended route; 0 otherwise,
I_{in} = inertia effect on route i for driver n; 1 if route i is the current route; 0 otherwise,
W_n = weather condition for driver n; 1 if good; 0 otherwise,
G_n = time-of-day for driver n; 1 if daytime; 0 otherwise,
S_n = trip purpose of driver n; 1 if business trip; 0 otherwise.

3.5. Results

3.5.1. Effects of qualitative en-route attributes. Table 3 summarizes the coefficient estimates and the asymptotic *t* values for the model estimations of the hybrid and MNL models for the lexicographic and utility maximization rules. The coefficient of the experienced travel time attribute for the MNL model is positive under lexicographic rules. This may seem counterintuitive, but can occur because lexicographic rules are non-compensatory unlike the compensatory basis for the MNL estimation model. It implies that the experienced travel time and the quantitative traffic information provided may be divergent. Since the hybrid model updates the travel time perceptions using both these variables, it is able to capture their contributions more consistently, leading to the expected negative coefficient. The model coefficients suggest that the explanatory power of familiarity vis-à-vis en-route route choice is minimal under the MNL model. However, the hybrid model indicates that familiarity has a relatively higher contribution to the utility as is expected from the experiment set-up. This is because familiarity is used as an ordinal variable in the MNL model whereas the hybrid model avoids rigid demarcation of boundaries between various familiarity levels, and better addresses subjectivity. A similar conclusion can be drawn for the route complexity attribute as well.

3.5.2. Effects of qualitative phenomena. Here, the relative abilities of the hybrid and MNL models to capture qualitative phenomena such as delusion, freezing, information effects, inertia and compliance are investigated. The traffic information provided affects drivers' perceptions of the route travel time in a within-day context. The driver experience with a route affects the confidence in the traffic information system in a day-to-day context. Hence, incorrect information on a route reduces the driver's propensity to consider that route in the short-term (delusion), and potentially excludes that route as an alternative in the long-term (freezing). These phenomena and associated information effects are explored by observing the route choices of 100 homogeneous drivers from day 1 to day 30. Each driver generated here is given a unique identification number so that his/her route choices can be tracked from one day to the next. Before simulating driver route choices for a day, the driver behavior characteristics are updated based on the previous day's experience. From day 1, drivers are provided with correct and detailed travel time information on route 3, and incorrect and limited information on route 4. No information is provided on routes 1 and 2. Here, correct information on a route for a driver implies that the travel time provided by the traveler information system on that route is within ±5% of the estimated travel time for that driver. The corresponding range for the incorrect information is either +20% to +40% or −20% to −40% from the estimated travel time. Since drivers update their perceptions of route travel time and the traffic information system on a daily basis, the average probability (across the 100 drivers) of choosing each route is affected by their updated perceptions. Figures 4(a) and (b) compare the abilities of the hybrid and MNL models, respectively, to capture the delusion and freezing phenomena. The hybrid model captures the delusion effect robustly as indicated by the rapid reduction in the average probability of choosing route 4 within the

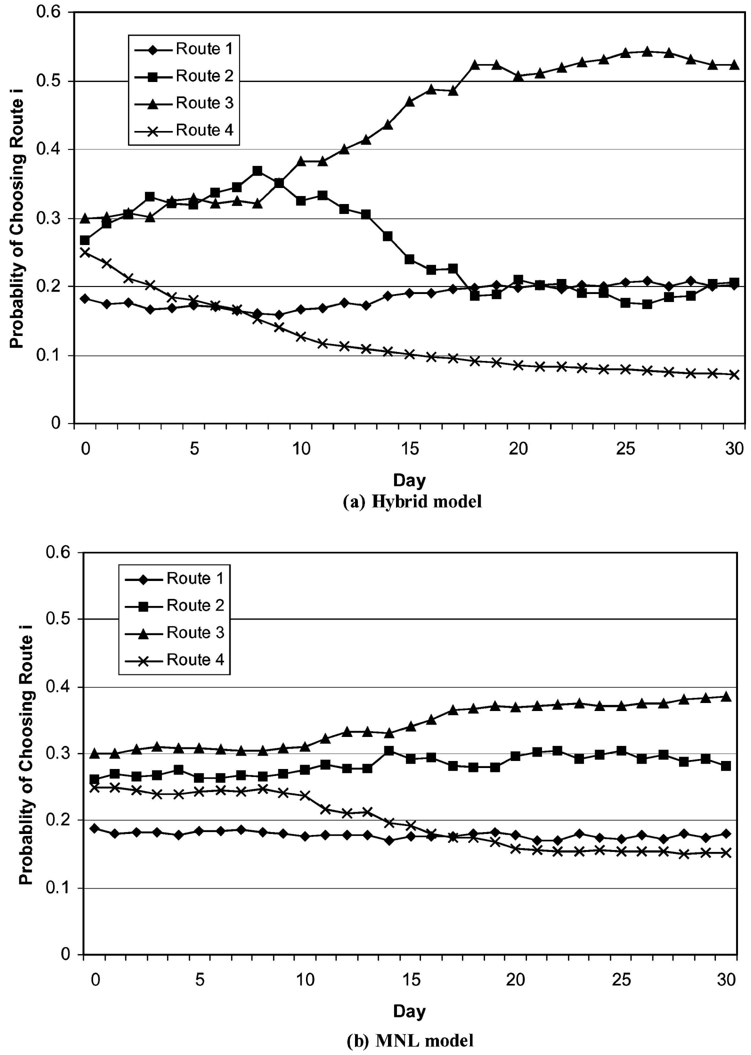

Figure 4. Delusion and freezing effects.

first few days. It further emphasizes the freezing effect as the choice probability for route 4 falls below 0.1 by day 15. Also, the value of reliable information is highlighted by the rapid increase in the choice probability for route 3 to beyond 0.5. The MNL model results suggest a diminished capability to capture these phenomena. The choice probabilities are relatively unchanged for the first 10 days implying the lack of sensitivity to information reliability. This implies that the MNL model is unable to capture the delusion phenomenon. Beyond day 10, the choice probability of route 3 increases and that of route 4 decreases, but at reduced rates compared to those of the hybrid model. That is, route 3 choice probability is below 0.4 on day 30. This indicates that the MNL model is less sensitive to information effects compared to the hybrid model, and less robust in capturing the associated qualitative phenomena.

The en-route compliance and inertia have higher coefficients for the hybrid model under the utility maximization rule implying greater explanatory power than for the MNL model. They can also be partly analyzed indirectly while comparing the specifications of the hybrid and MNL models to investigate their relative suitability. This is because a major difference in their specifications is in the representation of compliance and inertia. The hybrid model specifies driver inertia and compliance as functions of the situational variables, while the MNL model views them as dummy variables. The adjusted likelihood ratio index $\bar{\rho}^2$ is used to compare the models. Under the null hypothesis that the MNL model is a better specification, the following holds asymptotically (Ben-Akiva and Lerman, 1985):

$$Pr\left(\bar{\rho}_H^2 - \bar{\rho}_M^2 > z\right) \leq \Phi\{-[2NzlnJ + (K_H - K_M)]^{1/2}\}, \quad z > 0 \tag{7}$$

where,

$\bar{\rho}_H^2$ = the adjusted likelihood ratio index for the hybrid model,
$\bar{\rho}_M^2$ = the adjusted likelihood ratio index for the MNL model,
N = the number of observations,
J = the number of alternative routes,
K_H = the number of parameters in the hybrid model,
K_M = the number of parameters in the MNL model,
Φ = the standard normal cumulative distribution function.

The right-hand side of the equation is an upper bound on the probability that the MNL model is better than the hybrid model. From Table 3, the $\bar{\rho}_H^2$ and $\bar{\rho}_M^2$ for the lexicographic rule drivers are 0.655 and 0.450, respectively. The number of parameters in the hybrid and MNL moldels are 10 and 14, respectively. Thus:

$$\begin{aligned} Pr\left(\bar{\rho}_H^2 - \bar{\rho}_M^2 > 0.205\right) &\leq \Phi\{-[2 \times 1000 \times 0.205 \times \ln 4 + (10 - 14)]^{1/2}\} \\ &= \Phi(-23.8) \approx 0.00 \end{aligned} \tag{8}$$

Based on the test result, we can reject the null hypothesis. The claim is rejected for utility maximization rule drivers as well. This implies that the hybrid model merits further consideration than the MNL model since the MNL model has lower $\bar{\rho}^2$ though it has more parameters.

3.5.3. Prediction tests. Table 3 shows that the $\bar{\rho}^2$ of the hybrid model is greater than that of the MNL model. While this is favorably indicative of the performance of the hybrid model, a practical performance indicator is the correct prediction rate. The hybrid and MNL model probabilities are transformed into discrete route choices through Monte Carlo

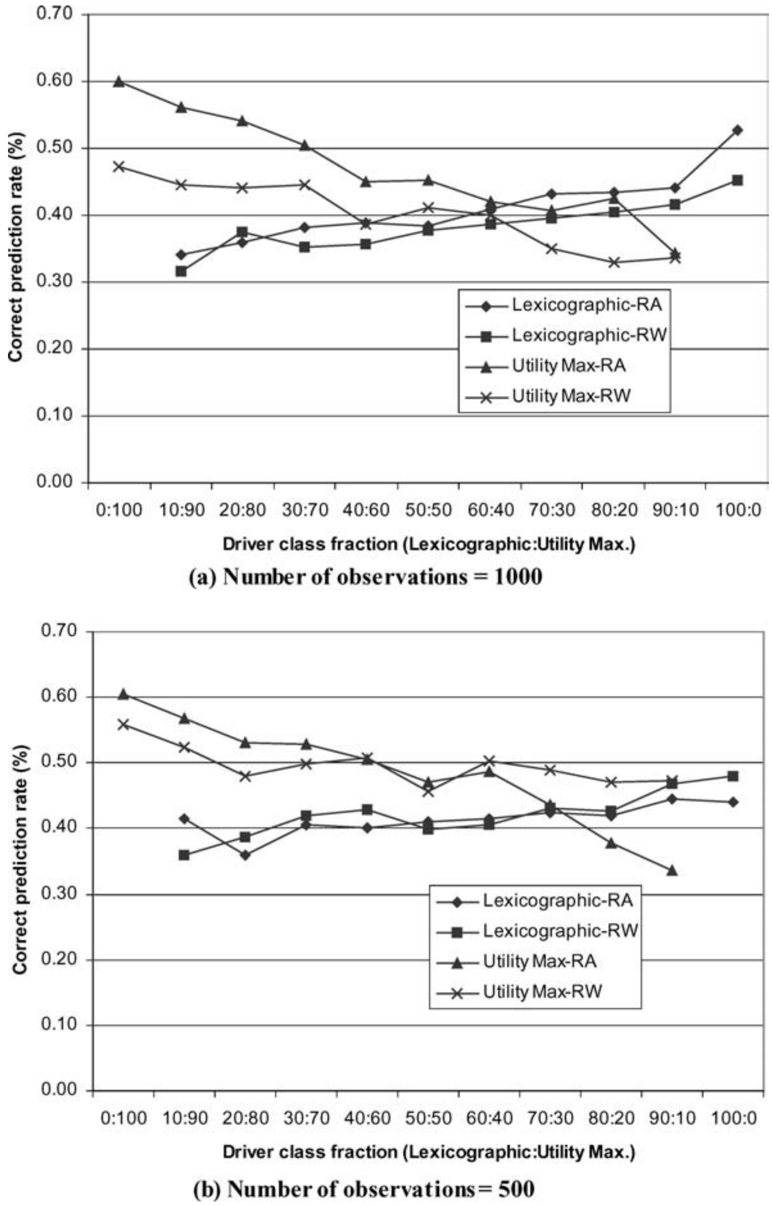

Figure 5. Effect of the level of heterogeneity in driver behavior.

simulation (Peeta and Yu, 2002a). First, a uniform random number generator is used to generate values between 0 and 1. In addition, the probability range between 0 and 1 is demarcated into smaller ranges according to the probabilities of choosing various routes by a driver. For example, if there are three routes with choice probabilities 0.3, 0.2, and 0.5, the ranges associated with each route are 0.0–0.3, 0.3–0.5, and 0.5–1.0, respectively. If the random number generated falls in the range of choosing a specific route, the driver is assumed to choose that route. This process is repeated for all drivers, and the prediction rate is determined by comparing the actual (from the generated dataset) and predicted choices across all drivers. The entire Monte Carlo simulation procedure is repeated 10 times to generate a set of potential route choice scenarios consistent with the probabilities predicted by a model, and an average prediction rate is computed. This average prediction rate is used to compare the two models. The prediction rates for the lexicographic and utility maximization rule drivers using the hybrid models are 74.6% and 66.3%, respectively. By comparison, the corresponding values for the MNL models are 62.1% and 55.7%. This suggests that the hybrid models have significantly better prediction power than the MNL models.

There are very few studies that analyze aggregate prediction rates for homogenous groups. Lotan (1997) analyzed aggregate prediction rates for familiar and unfamiliar drivers. However, heterogeneous driver classes in terms of decision rules and driver behavior characteristics are not considered. Here, four driver classes are considered to analyze heterogeneity in terms of driver behavior. This is done by testing various scenarios by varying the fractions of lexicographic and utility maximization rule drivers in the traffic stream. Traffic streams dominated by drivers of one decision rule are labeled as being more homogeneous. Here, we assume equal numbers of risk-averse and risk-willing drivers under each decision rule.

Figures 5(a) and (b) show the aggregate prediction rates as a function of the fractions of lexicographic and utility maximization rule drivers. Fractions towards either end (for example, 10:90, 20:80, 90:10) on the x-axis imply more homogeneous groupings. The results indicate that the level of heterogeneity is a significant factor that influences the prediction capability for individual driver classes. For example, higher prediction rates are obtained for lexicographic rule drivers when more of them are present in the traffic stream. However, in that situation, the prediction rates for the utility maximization rule drivers are lower. This implies that the market segments of driver classes need to be accurately predicted to generate better prediction rates. The number of observations also affects this trend. More observations yield steeper increases/decreases in the correct prediction rates as seen in the figure.

4. Concluding comments

This paper presents a hybrid model to predict en-route driver route choice decisions under real-time information provision. This problem is governed by several subjective/linguistic variables and qualitative phenomena such as delusion, freezing, information effects, inertia, and compliance. Probabilistic discrete choice models may not suffice in this context as they entail the discretization of continuous qualitative variables, and rigid boundaries that are restrictive vis-à-vis modeling subjectivity. Fuzzy modeling provides flexibility in modeling qualitative variables by generating membership functions that avoid rigid boundaries. The hybrid model has a probabilistic discrete choice model form, and consists of quantitative and fuzzy variables. Quantitative data is used directly while qualitative data is transformed into fuzzy variable values using *if-then* rules. A positive characteristic of the hybrid model compared to probabilistic discrete choice models is in terms of model estimation for problems with qualitative variables. The discrete choice models entail first identifying the appropriate model structure and then a trial-and-error procedure for the best specification. By contrast, the fuzzy modeling component addresses the qualitative variables using a standard procedure.

The performance of the hybrid model is compared with that of a pure MNL model. The hybrid model has better prediction capability, can more robustly capture qualitative phenomena, and has better explanatory power for qualitative attributes. The experiments highlight the importance of the level of heterogeneity in driver behavior on the prediction capabilities for individual driver classes. This has practical implications for the deployment of route choice prediction models. The study insights are based on hypothetical data for a single network, and may not be generalizable to all situations. However, they highlight the capabilities of the hybrid model for problems with qualitative variables. Detailed experiments using a real-world network that focus more on parameter sensitivity analysis, demand variability, supply variability (through incidents), and route characteristics analysis, are discussed in Peeta and Yu (2004a).

The hybrid model is motivated by the need to predict en-route driver route choice behavior for implementing control strategies in the context of real-time operations under information provision. Unlike current tools such as dynamic traffic assignment models, it does not assume rigid behavioral tendencies to specify driver classes such as user optimal, system optimal, stochastic user optimal, and boundedly-rational. Thereby, it can be used to represent driver classes based on survey data. It also provides capabilities to seamlessly represent day-to-day learning and within-day dynamics in a single consistent framework (Peeta and Yu, 2004b). These characteristics of the hybrid model provide a significant practical application capability. It is the key constituent of a behavior-based consistency-seeking model (Peeta and Yu, 2002b) that provides a practical alternative to dynamic traffic assignment models for real-time traffic management under information provision. The current study highlighted that a key source of incorrect prediction of driver route choices is the incorrect projection of driver class fractions. However dynamic traffic assignment models assume that the time-dependent driver class fractions are known a priori, in addition to assuming behaviorally restrictive driver classes. The behavior-based consistency-seeking model determines the time-dependent traffic flow pattern consistent with the real-time traffic flow measurements and driver behavior, while explicitly determining the driver class factions. In future work, we plan to extend this model to enable the determination of the specific driver classes in the ambient traffic stream through a two-stage model.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 9702612. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abdel-Aty, M.A., R. Kitamura, and P.P. Jovanis. (1997). "Using Stated Preference Data for Studying the Effect of Advanced Traffic Information on Drivers' Route Choice." *Transportation Research* 5C, 39–50.

Ben-Akiva, M. and S.R. Lerman. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Chen, T., H. Chang, and G. Tzeng. (2001). "Using a Weight-Assessing Model to Identify Route Choice Criteria and Information Effects." *Transportation Research* 35A, 197–224.

Dubois, D. and H. Prade. (1988). "Default Reasoning and Possibility Theory." *Artificial Intelligence* 35, 243–257.

Fujii, S. and R. Kitamura. (2000). "Anticipated Travel Time, Information Acquisition, and Actual Experience." *Transportation Research Record* 1725, TRB. Washington, D.C: National Research Council, pp. 79–85.

Jha, M., S. Madanat, and S. Peeta. (1998). "Perception Updating and Day-to-Day Travel Choice Dynamics in Traffic Networks with Information Provision." *Transportation Research* 6C, 189–212.

Khattak, A.J. and A. de Palma. (1997). "The Impact of Adverse Weather Conditions on the Propensity to Change Travel Decisions: A Survey of Brussels Commuters." *Transportation Research* 31A, 181–203.

Lotan, T. (1997). "Effects of Familiarity on Route Choice Behavior in the Presence of Information." *Transportation Research* 5C, 225–243.

Lotan, T. (1998). "Modeling Discrete Choice Behavior Based on Explicit Information Integration and its Application to the Route Choice Problem." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 28(1), 100–114.

Lotan, T. and H.N. Koutsopoulos. (1999). "Modeling Default Behavior in the Presence of Information and its Application to the Route Choice Problem." *International Journal of Intelligent Systems* 14, 501–533.

Mahmassani, H.S. and Y. Liu. (1999). "Dynamics of Commuting Decision Behavior Under Advanced Traveller Information Systems." *Transportation Research* 7C, 91–107.

Nakayama, S. and R. Kitamura. (2000). "Route Choice Model with Inductive Learning." *Transportation Research Record* 1725. Washington, D.C.: TRB, National Research Council, pp. 63–70.

Pang, G.K.H., K. Takahashi, T. Yokota, and H. Takenaga. (1999). "Adaptive Route Selection for Dynamic Route Guidance System Based on Fuzzy-Neural Approaches." *IEEE Transactions on Vehicular Technology* 48(6), 2028–2040.

Peeta, S. and S. Gedela. (2001). "Real-Time Variable Message Sign-Based Route Guidance Consistent with Driver Behavior." *Transportation Research Record* 1752. Washington, D.C.: TRB, National Research Council, pp. 117–125.

Peeta, S., J.L. Ramos, and R. Pasupathy. (2000). "Content of Variable Message Signs and On-line Driver Behavior." *Transportation Research Record* 1725. Washington, D.C.: TRB, National Research Council, pp. 102–108.

Peeta, S. and J.W. Yu. (2002a). "A Data-Consistent Fuzzy Approach for On-line Driver Behavior Under Information Provision." *Transportation Research Record* 1803. Washington D.C.: TRB, National Research Council, pp. 76–86.

Peeta, S. and J.W. Yu. (2002b). "On Behavior-Based Consistency-Seeking Models as Alternatives to Dynamic Traffic Assignment Models for Real-Time Deployment." Submitted to *Transportation Research Part C*.

Peeta, S. and J.W. Yu. (2004a). "Experimental Analysis of a Hybrid Route Choice Model to Capture Dynamic Behavioral Phenomena Under Advanced Information Systems." Submitted to *Transportation Research Part B*.

Peeta, S. and J.W. Yu. (2004b). "Adaptability of a Hybrid Route Choice Model to Incorporating Driver Behavior Dynamics Under Information Provision." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 34(2), 243–256.

Srinivasan, K.K. and H.S. Mahmassani. (2000). "Modeling Inertia and Compliance Mechanisms in Route Choice Behavior Under Real-Time Information." *Transportation Research Record* 1725. Washington, D.C.: TRB, National Research Council, pp. 45–53.

Tsoukalas, L.H. and R.E. Uhrig. (1997). *Fuzzy and Neural Approaches in Engineering*. New York: John Wiley and Sons, Inc.