



UNIVERSITY OF LEEDS

This is a repository copy of *Model representation and decision-making in an ever-changing world: The role of stochastic process models of transportation systems..*

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/84517/>

Version: Accepted Version

Article:

Watling, DP and Cantarella, GE (2013) Model representation and decision-making in an ever-changing world: The role of stochastic process models of transportation systems. Networks and Spatial Economics. 1 - 40. ISSN 1566-113X

<https://doi.org/10.1007/s11067-013-9198-2>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

SUBMISSION TO THE DTA2012 SPECIAL ISSUE

**MODEL REPRESENTATION & DECISION-MAKING IN AN EVER-CHANGING WORLD:
THE ROLE OF STOCHASTIC PROCESS MODELS OF TRANSPORTATION SYSTEMS**

David P. Watling, Institute for Transport Studies, University of Leeds, U.K.

d.p.watling@its.leeds.ac.uk

Giulio E. Cantarella, Department of Civil Engineering, University of Salerno, Italy

g.cantarella@unisa.it

Abstract – We review and advance the state-of-the-art in the modelling of transportation systems as a stochastic process. The conceptual and theoretical basis of the approach is explained in detail. A variety of examples are given to motivate its use in the field. While the examples cover a wide range of modelling philosophies, in order to provide focus they are restricted to modelling a special class of problems involving driver route choice in networks. Our overall objective is to establish the applicability of this approach as a ‘unifying framework’ for modelling approaches involving dynamic and stochastic elements, developing further the ideas put forward in Cantarella & Cascetta (1995). Directions for further development and research are identified.

1. INTRODUCTION

The global economic downturn, the ‘Arab spring’ and the turmoil in currencies are recent reminders that we live in an ever-changing world. Economic and social factors have profound influences on the level and pattern of travel demand and the choices of travellers within a given transport infrastructure. They also impact on the ability of responsible authorities to fund the maintenance and improvement of infrastructure, and to conduct effective travel demand management and control policies. It is just at such stages of major change and uncertainty that those planning future transport policies most need support in making their decisions, but in general this is exactly when most of the modelling tools we adopt fail to offer support, with their assumptions based on either an unchanging world, or one in which the future follows deterministically from the present. Even in periods of relative economic/social stability, such assumptions are increasingly difficult to support; this is most notable in cities where continued demand growth has outpaced the expansion in capacity of the transport infrastructure, with the transport system highly sensitive to daily and seasonal fluctuations in demand and capacities.

The question then arises as to how we might develop modelling approaches to better deal with such situations. One approach to such problems is that of ‘worst-case’ planning, whereby the models suggest actions for a planner to take so as to minimise the impacts under a worst-case scenario. The worst-case scenario itself may be user-defined, or for some methods may be itself generated as part of the modelling approach; examples of the latter are methods based on robust optimization (e.g. Ben-Tal *et al*, 2011), or on game theory in which a fictitious evil entity is at work (e.g. Bell, 2000). Such approaches have the advantage that there is no need to define event probabilities for the factors that affect the transportation system performance, and one can continue to use deterministic methods with tractable solution approaches. Such approaches have a particular advantage for representing extreme but rare events, where we may not have sufficient real-life evidence to make reasonable estimates of the component probability distributions. An obvious disadvantage of such methods is that they provide no information on anything but the extreme-most case. For dealing with typical daily and seasonal fluctuations in demand and supply, it seems that a stochastic model would be much more appropriate, where we aim to explore the full probability distribution of network impacts, not just the extreme-

most point. This is also arguably the case for longer term trends that have major transportation impacts, such as economic development (e.g. GDP) and oil prices, whereby a stochastic model can be used to capture the uncertainty in such future events. In practice, we believe that a combination of scenario-based deterministic methods and fully stochastic methods are appropriate, depending on the nature of the variation under consideration (as discussed above). In the present paper we shall henceforth focus only on stochastic approaches, on the basis that there is considerable empirical evidence of real-life variation that we already routinely collect but make little use of in our modelling.

With this focus in mind, the purpose of the present paper is to raise the profile of a particular class of stochastic approaches to transportation system modelling which was first proposed more than twenty years ago (Cascetta, 1987, 1989), but which has since attracted relatively little attention. Indeed it is commonly misunderstood by researchers in the field, as well as being mistakenly described and interpreted in transportation journal papers, and so we feel that it is timely for a paper to clearly set out the approach and its possibilities, in order to raise its profile. This approach is able to deal with many aspects of both (a) dynamic change and (b) uncertainty/variability, representing the time-evolution of all relevant state variables as a stochastic process. It is very different from the well-known Stochastic User Equilibrium model (so called after Daganzo & Sheffi, 1979), though the appearance of the word 'stochastic' in both can serve to confuse those unfamiliar with the method. It is also very different from the now growing body of research on *deterministic* dynamical system models—for recent examples see Bie & Lo (2010), Han & Du (2012) and He & Liu (2012)—which are able to capture the dynamics in (a), but not the aspects of uncertainty/variability to which we refer in (b). In this context we refer the reader to two companion papers by the authors to the present paper, in one of which we focus entirely on deterministic process models (Cantarella & Watling, 2013), and in the other we explore the relationships between deterministic process, stochastic process and stochastic user equilibrium approaches (Watling & Cantarella, 2013).

At its simplest, most 'stripped down' level, the Stochastic Process (SP) approach could be said to comprise three main elements for representing the epoch-to-epoch changes in a transport system:

1. A learning model, to describe how travellers learn from their travel experiences in past time epochs.
2. A decision model, to describe how travellers make decisions, given their learnt experiences in 1.
3. A supply model, to describe the experiences of travellers in a particular time epoch.

Some or all of these elements are described by probability statements or probability distributions, and when brought together they provide a single, self-consistent framework for representing the mutual interactions between the uncertain components of the transport system. Just as we demand of equilibrium transportation analysis, we can ask to what extent this combination of elements may produce a well-defined and unique 'output' (if the long-run is indeed what interests us), but whereas in equilibrium systems we refer to a unique flow state, in the SP approach we refer to a unique *probability distribution* of flows. That is to say, the result of the modelling approach is to provide the planner with probability distributions, not with single point estimates.

The description given above is deliberately rather general, in that it does not specifically say what we mean by a traveller choice (e.g. which route, which departure time, whether

to travel, where to live), what we mean by a time epoch (e.g. an hour, a day, a week, a month, a year), and what particular combinations of assumptions might make up the learning, decision and supply models. So, at one extreme we might be considering the year-to-year dynamics of residential location choice and the impact of and on the transport system, and at another we might be considering the day-to-day dynamics of route choice and traffic congestion. However, in keeping with the primary focus of most work on this topic to date, we shall focus heavily in our examples (section 3) on the class of problems concerned with the day-to-day dynamics in route choice (though in section 2 we explain the wider context of this work). In some respects this class defines an especially challenging (and therefore interesting) context, since it is subject to both the wider ‘global’ variations described in the opening of this abstract, and the more ‘local’ variations that are always seen between days-of-the week and seasons.

The paper is structured as follows. In section 2 we address a key issue in understanding and applying the approach, namely the possible choices for representing elements of the transportation system, and the implications of these choices, particularly in terms of the state-space representation and the proper specification of probability statements about this system. In section 3 we illustrate these various forms of representation with a range of example models that sit within the stochastic framework. To provide some focus, section 3 considers only the sub-class of problems concerned with driver route choice in networks, in contrast to the rather general treatment of section 2. The transition functions governing the stochastic dynamics are explicitly derived for simple examples of these models. The purpose of these examples is also to illustrate the possibilities for the approach to link to different fields (and philosophies) of transportation modelling, be that behavioural dynamics, dynamic traffic assignment or micro-simulation. In section 4 we address the issue of how such models might be used in a planning environment, either in ‘dynamic’ or ‘stationary’ mode, the latter being an analogue of existing equilibrium methods of planning. For the latter mode, theoretical conditions are set out to guarantee existence and uniqueness of the relevant stationary distributions, as well as indicating efficient computational shortcuts. In section 5, a rather general family of stochastic process models is presented for analysing the class of day-to-day dynamic route choice problems, and the properties of this class analysed. Finally, we conclude by identifying future applications, practical issues and research directions.

2. REPRESENTATION & BASIC NOTATION

At its broadest level, the stochastic process approach allows the modeller many choices as to how to represent the features of the underlying transportation system, how to represent the interactions between these features over space and time, and even how to represent space and time themselves. The ability to represent all these possibilities in a consistent framework is one of the advantages of the approach, giving the modeller a wide range of opportunities depending on the available data/evidence and problem at hand. The purpose of the present section is to give a flavour of these possibilities in a quite general context, illustrating some of its potential as a complete protocol for modelling transportation systems under uncertainty. As we shall show, we have many options for the way in which the system is represented, and so the step of choosing a particular representation is a key part of the modelling process. As we shall see in subsequent sections, the particular form of representation chosen can have important ramifications

for the theoretical properties that may be established for the model, among other things. Table 1 specifies the elementary components that we shall and shall not permit in the present paper. In the remainder of this section we shall explain in detail the meaning of each of these attributes.

Attribute	Possibilities		
<i>Time</i>	Discrete	Continuous	Dual-scale
<i>State</i>	Discrete	Continuous	Mixed Function
<i>Space</i>	Discrete	Continuous	Mixed
<i>Users</i>	Disaggregate	Aggregate	Segmented

Table 1: Possible representation of basic elements (shaded cells are the possibilities permitted within the theoretical framework of the present paper)

2.1 Representation of time

Perhaps the most fundamental decision to be made in representing transportation system dynamics is how to represent *time*. In order to understand this issue we need first to define the application context of the class of modelling problems we shall consider, namely to problems of transport planning over some given future planning horizon of weeks or years (though often it may not be explicitly defined). This application domain distinguishes them in particular from operational models that be used for optimizing short-term system performance over a periods of a few minutes or hours. Within this context, a central element of the models we shall consider will be the adaptive behaviour of travellers over time, as they repeat the requirement to make certain travel decisions. For example, this may be a traveller, or group of travellers, deciding on a residential or work location (a decision perhaps reviewed over periods of years). Alternatively, the traveller may be a commuter choosing a mode and route/service to work each morning, a decision which might be reviewed over days or weeks; even if this choice is stable, the decision to continue with a previous choice is then also being made, at least in our models if not consciously by travellers. In the latter (commuting) example, ‘day-to-day dynamics’ has been suggested as a term to capture this idea of repeated decision-making and adaptation.

Therefore, we shall argue, ‘time’ in some sense naturally divides into discrete ‘epochs’ of time (be they individual days, weeks or years) over which travellers review their travel decisions. Cascetta (1989) when introducing the notion of an epoch in this context noted:

‘... epochs can have either a “chronological” interpretation as successive reference periods of similar characteristics (e.g. the a.m. peak period of successive working days) or they can be defined as “fictitious” moments in which users acquire awareness of path attributes and make their choices’.

This ‘natural’ discretisation of time is therefore still more evident when one considers that it is quite unusual to model travel over a complete day, but rather only a portion of the day containing the most congested periods. Therefore, the term day-to-day dynamics is often used to describe the dynamical adjustments between days of drivers who travel in this sub-period of the day; in this respect, ‘time’ is then ontologically discrete. On the other hand, there are other kinds of interaction—such as the interaction of congested traffic, and the en route revisions to plans/strategies made by private or public transport users—for which there are no such natural reasons and ‘units’ for a discrete division of time. Of course, for computational ease we may choose to discretise time in some way, but this is a later consideration; our point here is that there is no such natural discretisation on what might be

called the ‘within-day scale’ (or, more generally, the ‘within-epoch scale’), in contrast with the comments above on the ‘between-day’ or ‘between-epoch’ scale.

Thus, unless we do truly model 24 hours of each day, our argument will be that it is natural to restrict attention to models in which the between-epoch scale is discrete, and for this reason we only consider such processes in the present paper. This is not a necessary restriction, and indeed there exist counterpart results for continuous-time stochastic processes that we could have considered (see: Fan & Liu, 2007; Hofbauer & Sandholm, 2007). The within-epoch scale is somewhat different, and in this case we might argue either for a discrete- or continuous-time representation, but to simplify the exposition, we suppose the within-epoch time-scale is also discrete. Again, this is not necessary, and we could in principle specify stochastic processes with a ‘dual-scale’ of mixed discrete/continuous for between/within-epoch time respectively, but we would pay significantly in terms of mathematical complexity in that the random state variables we introduce in section 2.2 below would all then be random functions of time.

2.2 Representation of state & distributions

Having restricted attention to discrete time processes, as explained in section 2.1, we move on to the second element of Table 1, namely the definition of ‘state’, and at this stage we can begin to introduce some notation. We denote the discrete time epochs (“between-epoch time” or “day-to-day time”) by the letter t (for $t = 0, 1, 2, \dots$), the state vector describing epoch t as $\mathbf{x}^{(t)}$, and the state-space to which any state must belong as a set \mathcal{S} , i.e. $\mathbf{x}^{(t)} \in \mathcal{S}$ (for $t = 0, 1, 2, \dots$). This compound state vector may contain several different kinds of ‘entity’—such as choices and memories of travellers, and experiences of travel times—measured at some chosen level of aggregation, and describing within-epoch (“within-day time”) variations in time and space. The details are important, and we shall delve into them in the remainder of the paper, but for the moment the key issue is that $\mathbf{x}^{(t)}$ is a *sufficient* description in two respects:

- A1 if we know $\mathbf{x}^{(t)}$ then we know (or can infer) everything we might want to know from the model for the purposes of design or evaluation; and
- A2 if we know $\mathbf{x}^{(t)}$ then we have sufficient information to write down a probability law that determines the probabilities of all future states of the modelled transportation system for times $t + 1, t + 2, \dots$

Assumption A1 implies that model outputs are sufficient for the intended purpose, both in terms of what they measure and their level of aggregation, but does not preclude the subsequent application of sub-models to infer further outputs; for example, it may be that $\mathbf{x}^{(t)}$ itself does not contain a pollution level variable, but contains information on explanatory variables (flows, speeds, etc.) that might subsequently be fed into a pollution sub-model for evaluation purposes. On the other hand, assumption A2 implies that if the future evolution of the transport system were somehow dependent on the pollution levels (e.g. affecting the choice of residential location), then they would need to be in $\mathbf{x}^{(t)}$. However, the most important point to take from these assumptions is that from A2, it is only time t (and not earlier states) that affects times $t + 1, t + 2, \dots$ —the so-called Markov property. This property is important for establishing theoretical results, as well as being attractive for computation, but brings with it various restrictions, especially as we shall wish to restrict attention to fixed, finite-dimensional state spaces S . In particular, this combination of requirements rules out the possibility of us simply defining one component

of $\mathbf{x}^{(t)}$ to be the ‘history’ $\mathbf{h}^{(t)} \equiv (\mathbf{x}^{(t-1)}, \mathbf{x}^{(t-2)}, \dots, \mathbf{x}^{(0)})$ in order to preserve the Markov property, since then our state space is either finite-dimensional but evolving as the size of $\mathbf{h}^{(t)}$ (and hence $\mathbf{x}^{(t)}$) expands with time t , or is infinite-dimensional in order to incorporate *a priori* all histories of any dimension (including as $t \rightarrow \infty$). We shall see that aside from the standard device of requiring any such histories to have a fixed, finite length, we can also (by appropriate choice of state variables) also incorporate what are apparently infinite histories by a judicious choice of state variable. We return to illustrate this with examples in section 3, but for the moment our purpose is simply to highlight the key nature of Assumptions A1 and A2, and the care needed in ensuring that they are satisfied.

A quite separate issue is what *kinds* of variables might be in the space \mathcal{S} . We shall assume that *either* \mathcal{S} is finite and formed from integer n -tuples (i.e. $\mathcal{S} \subseteq \mathbb{Z}^n$), *or* that \mathcal{S} is part of m -dimensional Euclidean space ($\mathcal{S} \subseteq \mathbb{R}^m$), *or* that \mathcal{S} is a combination of these two kinds of variable (i.e. $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \subseteq \mathbb{Z}^n \times \mathbb{R}^m$). Thus we permit discrete, continuous and mixed discrete/continuous state-spaces; we provide examples of each of these later, in section 3.

The assumptions made thus far allow us to introduce some basic, general notation to describe the evolution of the stochastic system. In order to allow different kinds of (discrete, continuous or mixed) state-space, we adopt a particular form of notation that might apply to either case. In particular:

- at any given time epoch t , let $\{q^{(t)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ denote the (epoch t) joint probability/ probability-density function across the possible states $\mathbf{x} \in \mathcal{S}$ (for $t = 0, 1, 2, \dots$) (we shall refer to this as “the state probability distribution at time t ”); and
- for any given state $\mathbf{y} \in \mathcal{S}$ and given parameter vector $\boldsymbol{\theta} \in \Theta$, let $\{\phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) : \mathbf{x} \in \mathcal{S}\}$ denote the *conditional* joint probability mass/density function across possible states $\mathbf{x} \in \mathcal{S}$ in the current time epoch, *given* that \mathbf{y} was the state in the previous time epoch (we shall refer to this as the “transition function”).

It should be noted that we assume the transition function ϕ to be time-independent, and the parameter vector $\boldsymbol{\theta} \in \Theta$ to be fixed and time-independent. Together, these imply that the resulting process is *time-homogeneous*. The assumptions above imply that the joint probability/ probability-density function of \mathbf{x} varies in time only due to one factor, namely due to the state \mathbf{y} in the previous time epoch. While this assumption may seem somewhat restrictive, we shall illustrate in section 3 how it can accommodate just about any form of model that has been proposed to date in the transportation literature, and provides a framework for many more not yet proposed. As we show in section 3 (e.g. example 3.2), we can easily accommodate a dependence on a finite number of past states (i.e. not just ‘yesterday’), by making a component of the state variable a ‘finite history up to that day’, and by judicious choice of state variables can even in some cases include apparently infinite histories (e.g. section 3.3).

Although in the present paper we shall not consider control/information systems, we note that these may also be represented in such a framework by including the decisions/recommendations of the control/information system as a component of the state variable. These decisions/ recommendations, even if they concern predictions or plans for the future, must be based on historic information, and it seems reasonable to believe that such systems themselves follow a time-homogeneous law; given the same sequence of past information and the same predictions of future events, they will provide

the same decisions/recommendations regardless of the current clock-time t . This is true even for so-called ‘anticipatory systems’ since even they cannot know the future, they too must be driven by forecasts of or plans for the future, based on historic information. For the interested reader, examples of a stochastic process model in interaction with some responsive control systems (signal control and demand-responsive bus operations) are discussed in Watling (1996).

Based on the assumptions made to date, we may then write our stochastic process as one of the following¹, depending on the nature of the state-space:

For any given initial distribution $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, then for $t = 1, 2, \dots$:

i) Markov Process:

$$q^{(t)}(\mathbf{x}) = \int_{\mathbf{y} \in \mathcal{S}} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) q^{(t-1)}(\mathbf{y}) d\mathbf{y} \quad (\mathbf{x} \in \mathcal{S} \subseteq \mathbb{R}^m; \boldsymbol{\theta} \in \Theta)$$

ii) Markov Chain²:

$$q^{(t)}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{S}} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) q^{(t-1)}(\mathbf{y}) \quad (\mathbf{x} \in \mathcal{S} \subseteq \mathbb{Z}^n; \boldsymbol{\theta} \in \Theta)$$

iii) Markov Process/Chain

$$\begin{aligned} q^{(t)}(\mathbf{x}) &\equiv q^{(t)}((\mathbf{x}_{[1]}, \mathbf{x}_{[2]})) = \\ &\sum_{\mathbf{y}_{[1]} \in \mathcal{S}_{[1]}} \int_{\mathbf{y}_{[2]} \in \mathcal{S}_{[2]}} \phi((\mathbf{x}_{[1]}, \mathbf{x}_{[2]}), (\mathbf{y}_{[1]}, \mathbf{y}_{[2]}); \boldsymbol{\theta}) q^{(t-1)}((\mathbf{y}_{[1]}, \mathbf{y}_{[2]})) d\mathbf{y}_{[2]} \\ &(\mathbf{x} \equiv (\mathbf{x}_{[1]}, \mathbf{x}_{[2]}), \mathbf{x}_{[1]} \in \mathcal{S}_{[1]} \subseteq \mathbb{Z}^n; \mathbf{x}_{[2]} \in \mathcal{S}_{[2]} \subseteq \mathbb{R}^m; \boldsymbol{\theta} \in \Theta). \end{aligned}$$

All cases require a distribution $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ to initialise the process, and this may arise from one of several sources. One possibility is that it has been estimated by observation of some “current” conditions (if $t = 0$ is “the present”). In this way, it acts more like an additional set of parameters, and indeed we may even choose to parameterise the initial distribution itself, with the task then to estimate the parameters of the initial distribution from observation. A second possibility with discrete state-space is to specify $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ by setting all probability at a single point in the state-space; the justification might be that ‘the past is certain, the future is uncertain’. A third possibility is that $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ is generated by ‘the end-point’ of an earlier application of the stochastic process approach (where ‘end-point’ may mean the distribution at some given time $t = T$ or an “infinite-time” stationary distribution). This idea of using an earlier application sits particularly well with the idea of a before-and-after study of some hypothetical scheme, for example, whereby the process is first used to replicate the time prior to the implementation of the scheme, and the resulting distribution then used to initialise a model of the situation after the scheme implementation.

Aside from the initial distribution, the three specifications given also have the common feature that each includes the functions ϕ and q , and in each specification these functions perform the same role: in simple terms we might say that ϕ is “the model” and (for time $t = 1, 2, \dots$) then q is “the unknown”, as we now explain. The function ϕ is parameterised by the

¹ Note that we could combine these cases as a single case by use of a Riemann-Stieltjes integral.

² If our interest had only been in models with a discrete, finite state space, then a simpler, standard specification would be as $\mathbf{q}^{(t)} = \mathbf{P}\mathbf{q}^{(t-1)}$, with the states in \mathcal{S} labelled $1, 2, \dots, |\mathcal{S}|$, the transition probabilities in a $|\mathcal{S}| \times |\mathcal{S}|$ matrix \mathbf{P} , and the time-dependent state probabilities in a $|\mathcal{S}| \times 1$ column vector $\mathbf{q}^{(t)}$.

vector θ which we assume to be constant³, and something that the modeller specifies at the start of the process; it is a collection of all parameters across all sub-models. $\phi(\mathbf{x}, \mathbf{y}; \theta)$ also depends on the vector \mathbf{y} , “yesterday’s state”, and whenever we apply ϕ in the formulae above, we do so in a recursive way in which \mathbf{y} is effectively known as we consider all the possible yesterday states in \mathcal{S} . In this way, \mathbf{y} also acts as a kind of parameter, although we do not assign it a particular value but consider all its values across \mathcal{S} . $\phi(\mathbf{x}, \mathbf{y}; \theta)$ may then effectively be thought of as a function of \mathbf{x} only, given knowledge of \mathbf{y} and θ , a function that expresses the spread of likely values of \mathbf{x} for “today” when we are given that \mathbf{y} occurred “yesterday” (and assuming that the model parameters are given in θ). In order to specify this spread of likely values (formally, conditional probabilities), we shall typically specify a series of sub-models, so that the function ϕ is a rather complex, composite function that arises from the combination of these sub-models. Typically we will not want or need to actually write down ϕ but will instead specify it by implication, by setting out a series of conditional statistical assumptions for the model components, as we shall see in the examples in section 3.

In this way, like the initial distribution $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, the function ϕ is also something that is assumed to be known and specified by the modeller, before application of the equations above. What the equations above allow is for $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ to be combined with ϕ to produce $\{q^{(1)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, and then for $\{q^{(1)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ to be combined with (the same function) ϕ to produce $\{q^{(2)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, and so on. Thus the objective of the modelling process can be viewed to be that of recursively generating a sequence of probability distributions $\{\{q^{(t)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\} : t = 1, 2, \dots\}$, given knowledge of $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ and ϕ . This sequence of probability distributions is the ‘model output’, in theory at least, although in practice we may find it sufficient to only store/record some summary measures of the distribution, or indeed (in a spirit similar to equilibrium analysis) may not refer to the temporal evolution at all – an issue to which we shall return later (see section 4.2).

2.3 Representation of space

Given the possibilities presented in sections 2.1. and 2.2 for how to represent time and state, the modeller then has several choices for how to represent the detailed elements of the transportation system, within the framework of possibilities given by section 2.1/2.2. In Table 1, we have suggested that the third such element is the question of how to represent space. To be clear, we are not intending to suggest that the question of spatial representation is something unique to adopting a stochastic process approach, rather the particular issue faced is deciding on the *type* of representation in conjunction an appropriate state representation, since this may have profound effects on the theoretical properties that may be established for the chosen model.

Discrete space models are undoubtedly the most proliferous in the transportation literature. In public and private transport models, the dominant representation of the

³ At this stage this is not a necessary assumption, and in fact we may wish to consider models in which endogenous factors vary over the time of the process, e.g. economic factors, seasonal changes in demand. The theoretical properties described later (section 4.2) are established under the assumption of a constant parameter vector (leading to a time-homogenous process with time-independent transition probabilities), but it is certainly possible to consider and model time-inhomogeneous processes. This is an interesting possibility left for future research to consider.

physical infrastructure is through a discrete graph. On the other hand, diachronic networks may be used to represent feasible temporal connections, especially in the case of public transport services (Wilson & Nuzzolo, 2009). A different spatial question is how vehicles move within the infrastructure. For ease of implementation we may choose to approximate the dynamic movements in continuous space along the arcs of a discrete spatial graph, by adopting a further discretisation scheme as in the cellular automaton (Rickert *et al*, 1996) or cell transmission (Daganzo, 1994) approaches. In spite of the dominance of discrete-graphs, alternatives based on a *continuous* representation of space have existed for many years (e.g. Dafermos, 1980), and may be justified as an approximation to discrete infrastructure needing less parameters or less data. For cases such as pedestrian route choices the available space within which routes are chosen is ontologically continuous (and so there is no obvious discrete graph with which to begin), meaning that a continuum model is an obvious option to consider (Wong, 1998; Hoogendoorn & Bovy, 2004; Huang *et al*, 2009). For many land-use/transport interaction problems a continuum representation is also rather natural, for example given the continuous nature of new residential locations (Ho & Wong, 2007). A third possibility is a *mixed* spatial representation, such as one in which a continuum representation of demand is combined with a discrete graph of the infrastructure, or where different levels of detail are appropriate for different parts of a street network (Guo & Liu, 2012). A particularly familiar, mixed spatial approach is one in which vehicles are moved continuously in space (though discretely in time) along a discrete street network, as in car-following-type micro-simulation models (see the example of section 3.6).

2.4 Representation of users

Completing our consideration of Table 1, the final issue of representation is concerned with the users of the transportation system. In a similar way to the treatment of space, at least *some* of the issues in representing users are familiar ones faced in all transportation modelling exercises, and in such cases the key point is again to be aware that the choice of representation may imbue the process with particular theoretical properties. However, the representation of users differs to that of spatial representation in that there are special modelling considerations that only arise because we are adopting a dynamic process approach (be it deterministic or stochastic). So the purpose of the present section is also to indicate these special considerations, and the options that are available in those cases.

The most familiar issue to consider in the representation of users is the level of detail at which we represent a) users' decisions, and b) the stimuli that motivate those decisions. In the study of transportation networks, for example, aggregate approaches dominate, though even within this class one can distinguish models where all users think alike (Wardropian approaches) from models incorporating diversity (proportionate splits in behaviour) through expectations of statistical distributions, e.g. distributed values-of-time (Leurent 1998), stochastic user equilibria (Sheffi, 1985). Though these approaches are familiar, subtle conceptual issues arise when incorporating these latter ideologies within a stochastic process approach, in terms of how one interprets the underlying distributions from which conventional approaches take only expected values. For example, when we aim to incorporate Random Utility Models (RUMs) into a stochastic process framework, is the statistical variability and the notion of probability in RUMs describing a kind of day-to-day variability in the behaviour of individuals? Or are individuals not so variable in their

behaviour, and ‘probability’ is a device for describing the uncertainty of *modellers* in explaining individuals’ behaviour? Aside from these conceptual points, we have a possibility commonly used in traditional models of some kind of fixed segmentation of users, applying the models described above within a segment/user-class.

As noted at the outset, there are also special issues that arise in representing users within a dynamic process framework. These concern the dynamics of how users remember and learn from their experiences in order to make predictions (upon which their choices are made), and also the extent to which they display inertia/habit in their propensity to reconsider previous choices. In the present section we are not concerned with the detailed assumptions of these model components (such details will emerge in the remainder of the paper, and the references therein), but rather to indicate that exactly the same issues of aggregation/disaggregation/segmentation arise for these elements as were identified above for the modelling of the decision behaviour. While we might typically expect the same level of aggregation to be chosen for all such elements (experience, learning, habit, choice), there may be also be cases in which different levels of aggregation are appropriate. For example, in modelling the longer-term impacts of driver information systems, we might models users making choices and gaining personal travel experiences at the individual level, but part of the effect of the information system might be represented by pooling the experiences of all users after journeys have been completed.

3. MOTIVATING ILLUSTRATIVE EXAMPLES

Having introduced a rather general framework for stochastic process modelling in a transportation systems context, it is now possible to move on to specifying particular families within this framework in section 5, but before moving on to these general families, we provide in the present section some deliberately simple examples. We will derive explicit transition functions (as defined in section 2.2) for particular models. While it would be possible to implement the models without explicit derivation of the transition function (e.g. through Monte Carlo simulation), we wish to highlight here the theoretical equations behind the model rather than its implementation. As we shall see in later sections, understanding the transition function is the key to establishing overall theoretical properties of the modelling approach. A further objective is to demonstrate the generality of the stochastic process approach in its possibility to link to fields such as dynamic traffic assignment and micro-simulation; it should not be considered as a particular model, but rather a way of modelling which encompasses many particular models and approaches.

All the examples are readily generalised to networks of arbitrary size and complexity; such generalisations can be found in the citations given in each section, and we explicitly demonstrate how such a generalisation is made later in the paper, for a particular model family. However, the notation for such cases is sufficiently complex that it can mask the features that we wish to illustrate here. Therefore in all the examples, we consider the simplest possible network of a single origin-destination movement connected by two parallel arcs, and make simplifying assumptions that allow explicit transitions functions to be derived using standard statistical distributions. We shall imagine that we wish to examine the evolution of a system over days, and so we shall refer to an epoch as a day or better, a given peak-period of the day.

3.1 Example 1: Myopic learning with discrete state space

Cascetta (1989) introduced a family of *discrete* state space stochastic process models to the transportation field. The simplest example within this family is that of a network consisting of two parallel arcs/routes serving a given integer OD demand d , such that we may represent the state of the network by an integer scalar x denoting the flow on arc 1 (with the flow on arc 2 then clearly $d - x$). Thus, in this example our state space $\mathcal{S} \subseteq \mathbb{Z}$ is given by $\mathcal{S} = \{0, 1, 2, \dots, d\}$, and we are appealing to specification ii) in section 2.2, that of a Markov Chain. Let $c_i(\cdot; \lambda)$ be the performance function on arc i ($i = 1, 2$), parameterised in total by the vector λ , with the performance function denoting the cost to traverse the arc as a function of the flow on that arc only⁴. Thus, given state x , the arc costs are $c_1(x; \lambda)$ and $c_2(d - x; \lambda)$, and for ease of notation it will be convenient to define a scalar function for the difference in actual arc costs:

$$c(x; \lambda, d) \equiv c_1(x; \lambda) - c_2(d - x; \lambda).$$

Suppose that drivers as a group remember only the costs experienced (as a group) on the previous day. Then, on any given day, conditionally on the remembered costs, each of the d drivers is supposed to choose between the routes independently and at random according to a logit choice probability (with scale parameter $\xi > 0$) evaluated at the remembered costs. The parameters of the overall model may be collected together in the vector $\Theta = (\lambda, d, \xi)$. The assumptions together imply that the *transition function*—i.e. the conditional probability distribution of the flow x on arc 1 on any one day, given that the flow on arc 1 was y yesterday—is given by the Binomial expression:

$$\phi(x, y; \Theta) = \{d!/(x!(d-x)!) \cdot \{1/\{1 + \exp(\xi c(y; \lambda, d))\}\}^x \cdot \{1 - \{1/\{1 + \exp(\xi c(y; \lambda, d))\}\}\}^{d-x}\} \quad (\text{for } x = 0, 1, \dots, d; y = 0, 1, \dots, d; \Theta = (\lambda, d, \xi)).$$

with the expression above we have all that is needed to generate the state probability distributions for times $t = 1, 2, \dots$, given an initial $t = 0$ state distribution. Denoting the initial state probability distribution is $\{q^{(0)}(x) : x = 0, 1, 2, \dots, d\}$, then the state probability distribution at time $t = 1$ is (applying the Markov Chain expression in section 2.1):

$$q^{(1)}(x) = \sum_{y \in \{0, 1, 2, \dots, d\}} \phi(x, y; \Theta) q^{(0)}(y) \quad (x = 0, 1, 2, \dots, d).$$

That is to say, it is a mixture distribution of Binomial variables with mixture probabilities given by the initial state distribution. Given $q^{(1)}(x)$, we may then compute $q^{(2)}(x)$ by the same process, and so on.

3.2 Example 2: Learning processes and retaining the Markov assumption

While Example 1 may be extended in many ways, a basic limitation it has is the assumption that drivers remember only the previous day's travel costs, which was later termed 'myopic behaviour'. From early evidence, the assumptions regarding how drivers 'learn' was known to be both (a) highly influential on the plausibility of predictions of the overall model (e.g. Horowitz, 1984) and (b) something for which there was evidence of non-myopic behaviour (e.g. Chang & Mahmassani, 1988; Iida et al, 1992). The simplest extension of Example 1 within the discrete-state framework proposed by Cascetta (1989) is to assume that drivers now remember some finite number of previous days' costs, and form their predictions of costs for the forthcoming day based on a weighted moving average. As a simple example, suppose that drivers remember only costs on the previous

⁴ In fact this is trivially generalised to non-separable cost functions if desired.

two days, weighting yesterday's costs by ϖ and the costs from two days ago by $1 - \varpi$, for given $0 < \varpi \leq 1$. Our state variable now is extended to be a two-dimensional vector \mathbf{x} with x_1 and x_2 respectively denoting the flow on arc 1 that occurred today and yesterday. Thus, now our state space $\mathcal{S} \subseteq \mathbb{Z}^2$ is given by $\mathcal{S} = \{(x_1, x_2) : x_i \in \{0, 1, 2, \dots, d\} \text{ for } i=1,2\}$, and again we are able to appeal to (the Markov Chain) specification ii) in section 2.2. Retaining all other assumptions made in Example 1, the parameters of the overall model may now be collected together in a slightly extended vector $\boldsymbol{\theta} = (\lambda, d, \xi, \varpi)$.

Now we aim to write down the transition function. For example, we must describe how a pair of states on days 1 and 2 transforms into a pair of states on days 2 and 3. Now, given knowledge of the states on days 1 and 2, then there is no uncertainty regarding the state on day 2, and so the only pairs of days 2/3 states with non-zero conditional probability of being transformed into are those that have the same given state on day 2. Now in general in the notation given, the sequence of consecutive states that interest us (i.e. give non-zero conditional probability) are over a period of 3 consecutive days: $y_2, y_1 = x_2, x_1$ (corresponding respectively to the day before yesterday, yesterday and today). The conditional probability that the states (today, yesterday) are (x_1, x_2) given that the states (yesterday, day-before-yesterday) were (y_1, y_2) is therefore zero if $x_2 \neq y_1$, and is otherwise equal to the conditional probability that today's state is x_1 given that the states (yesterday, day-before-yesterday) were (y_1, y_2) . Overall, this implies that the transition function, while now a function of vector states, turns out to be a quite simple modification of that given in Example 1, namely the combination of a Binomial expression and an indicator function:

$$\phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = \{d!/(x_1!(d-x_1)!)\} \cdot \{1/\{1 + \exp(\xi(\varpi c(y_1; \lambda, d) + (1-\varpi)c(y_2; \lambda, d)))\}\}^{x_1} \cdot \\ \{1 - \{1/\{1 + (\xi(\varpi c(y_1; \lambda, d) + (1-\varpi)c(y_2; \lambda, d)))\}\}\}^{d-x_1} \delta(x_2, y_1) \\ (\text{for } \mathbf{x} \in \mathcal{S}; \mathbf{y} \in \mathcal{S}; \boldsymbol{\theta} = (\lambda, d, \xi, \varpi))$$

where $\delta(a,b)$ is the Kronecker delta function:

$$\delta(a,b) = 1 \text{ if } a = b \\ = 0 \text{ if } a \neq b.$$

As well as having a difference from Example 1 in the manner that state transitions are defined (in that we must consider pairs of states), there is also a difference in the state distribution, which likewise pertains to pairs of states on consecutive days. So our initial state probability distribution at time $t = 0$ is now actually a specification of the joint probability distribution of the arc 1 flow at time $t = -1$ and time $t = 0$.

A point often misunderstood is that this truly is a *joint* distribution, and so we may have correlations between the same variable on consecutive days (so-called auto-correlations). These auto-correlations persist and evolve as the process evolves, just as the marginal probability distributions evolve of the flows on any one day; looking ahead to section 3, they persist even when the process is stationary (an issue we shall discuss later).

The initial state distribution⁵ is thus now denoted $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S} = \{(x_1, x_2) : x_i \in \{0, 1, 2, \dots, d\} \text{ for } i=1,2\}\}$, each element giving the joint probability that the day 0 arc 1 flow is x_1 and the day -1 arc 1 flow is x_2 . However the initial joint state distribution is generated, the

⁵ In practice, we may wish to simplify the specification by using a model of the form given in Example 1 to generate the flow on day 1, given the flow on day 0, and then apply the model given here starting from day 2, given knowledge of the probabilities of the pair of states on days 0 and 1.

state distribution on day 1 (a bivariate distribution of the flows on days 1 and 0) is generated according to:

$$q^{(1)}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{S}} \phi(\mathbf{x}, \mathbf{y}; \theta) q^{(0)}(\mathbf{y}) \quad (\mathbf{x} \in \mathcal{S})$$

and so on for the state distributions on days 2 and beyond.

3.3 Example 3: Exponentially-weighted learning with continuous state space

While the formulation in Example 2 can be extended in the obvious way to weighted moving average learning processes of any finite order, or indeed many other forms of learning process that need not be based on averages, ultimately there has to be a finite limit specified on the number of days of remembered experiences. This is in order that the process has a fixed state-space and retains the Markov assumption, which (as we shall see later) becomes a key element in establishing important theoretical properties. A negative aspect to moving average processes of high order, however, is that they are relatively computationally expensive to implement, since one must retain and update a moving window of experiences. In this respect, an exponentially-weighted process is much more attractive, but if analysed within the framework of Cascetta (1989) it yields a problem of infinite memory. This is a particular case in which being able to adopt different forms of state-space becomes especially convenient. The simplest example of this arises from an instance from the family considered by Cantarella & Cascetta (1995), a model used for analysing a deterministic dynamical system by Watling (1999). We consider the same example as Example 1, except that now we suppose that drivers using an updating process for their *predicted costs* for their forthcoming journey. Indeed these predicted costs become so central that we define a scalar state variable x now to be the difference in predicted cost given by the predicted cost on arc 1 minus the predicted cost on arc 2, with our state space now $\mathcal{S} = \mathbb{R}$. Given that yesterday's predicted cost difference was y , and given that yesterday's actual cost difference was z , then we suppose that today's predicted cost difference x would simply be a weighted average of these two, $x = \beta z + (1 - \beta)y$ for some parameter β typically assumed to satisfy $0 < \beta \leq 1$. (It is noted in passing that the special case $\beta = 1$ corresponds to the simple model considered in Example 1.)

Now, given that yesterday's predicted cost difference was y , then according to the assumption that drivers choose independently and at random according to logit choice probabilities based on the predicted costs, we know the probability distribution of yesterday's flow on arc 1:

$$\Pr(\text{yesterday's arc 1 flow} = f \mid \text{yesterday's predicted costs were } y) = \\ \{d!/(f!(d-f)!)\} \cdot \{1/\{1 + \exp(-\xi y)\}\}^f \cdot \{1 - \{1/\{1 + \exp(-\xi y)\}\}\}^{d-f} \quad (\text{for } f = 0, 1, \dots, d; y \in \mathbb{R}).$$

and so this also gives a distribution of the actual cost differences conditional on y :

$$\Pr(\text{yesterday's actual cost difference} = c(f) \mid \text{yesterday's predicted costs were } y) = \\ \{d!/(f!(d-f)!)\} \cdot \{1/\{1 + \exp(-\xi y)\}\}^f \cdot \{1 - \{1/\{1 + \exp(-\xi y)\}\}\}^{d-f} \quad (\text{for } f = 0, 1, \dots, d; y \in \mathbb{R}).$$

Thus the actual cost differences are discrete variables, and thus so are the predicted cost differences, since they are weighted sums of actual cost differences. However, the discretisation becomes unwieldy to keep track of as t grows, since the number of possible states for the predicted cost multiplies at a rate of d per time step; indeed the discretisation is on a limiting path towards a continuous representation. Instead, then, we shall seek a continuous representation from the outset. There are several ways this might

be achieved, but the most convenient to illustrate in the present context is one in which we begin by making the flow variables continuous.

In particular, we assume that d is sufficiently large and that the probability of choosing any arc is sufficiently far from 0 or 1, that we may use the Normal approximation to the Binomial distribution, so that if F and Y denote the random variables corresponding to f and y , then:

$$F | Y = y \sim \text{Normal}(d/\{1 + \exp(\xi y)\}, \{d/\{1 + \exp(\xi y)\}\}\{1 - \{1/\{1 + \exp(\xi y)\}\}\}).$$

For the sake of this simple example, let us also suppose that the travel cost functions are affine, leading to an actual cost difference function of:

$$\begin{aligned} c(f; \lambda, d) &\equiv c_1(f; \lambda) - c_2(d - f; \lambda) = (\lambda_1 + \lambda_2 f) - (\lambda_3 + \lambda_4(d - f)) \\ &= (\lambda_1 - \lambda_3 - \lambda_4 d) + (\lambda_2 + \lambda_4)f. \end{aligned}$$

Thus, as an affine transformation of a Normal random variable, yesterday's actual cost difference conditional on yesterday's predicted cost difference is also Normal (using Z to denote the random variable corresponding to z as introduced earlier):

$$\begin{aligned} Z | Y = y &\sim \text{Normal}((\lambda_1 - \lambda_3 - \lambda_4 d) + d(\lambda_2 + \lambda_4)/\{1 + \exp(\xi y)\}, \\ &\quad \{d(\lambda_2 + \lambda_4)^2/\{1 + \exp(\xi y)\}\}\{1 - \{1/\{1 + \exp(\xi y)\}\}\}). \end{aligned}$$

It follows finally that today's predicted cost difference $X = \beta Z + (1 - \beta)Y$, when conditioned on $Y = y$, is also Normal:

$$\begin{aligned} X | Y = y &\sim \text{Normal}(\beta((\lambda_1 - \lambda_3 - \lambda_4 d) + d(\lambda_2 + \lambda_4)/\{1 + \exp(\xi y)\}) + (1 - \beta)y, \\ &\quad \{d\beta^2(\lambda_2 + \lambda_4)^2/\{1 + \exp(\xi y)\}\}\{1 - \{1/\{1 + \exp(\xi y)\}\}\}) \end{aligned}$$

and so our transition function can finally be defined in the Markov Process form i) from section 2.2, as the Normal density:

$$\begin{aligned} \phi(x, y; \theta) &= (2\pi\{d\beta^2(\lambda_2 + \lambda_4)^2/\{1 + \exp(\xi y)\}\}\{1 - \{1/\{1 + \exp(\xi y)\}\}\})^{-0.5} \\ &\quad \exp(-0.5\{x - \{\beta((\lambda_1 - \lambda_3 - \lambda_4 d) + d(\lambda_2 + \lambda_4)/\{1 + \exp(\xi y)\}) + (1 - \beta)y\}^2/\{d\beta^2(\lambda_2 + \lambda_4)^2/\{1 + \exp(\xi y)\}\}\{1 - \{1/\{1 + \exp(\xi y)\}\}\}) \\ &\quad (\text{for } -\infty < x < \infty; -\infty < y < \infty; \theta = (\lambda, d, \xi, \beta)). \end{aligned}$$

Thus we have managed to retain the Markov property with a complete representation of the stochastic process over \mathbb{R} , in contrast with Examples 1, 2 which were defined over \mathbb{Z} . The state distributions $\{q^{(t)}(x) : -\infty < x < \infty\}$ generated by this process, as well as the initial state distribution at $t = 0$, are then clearly probability density functions as opposed to probability mass functions as they were in Examples 1 and 2. Note also that in the special case $\beta = 1$ we obtain a continuous approximation to the discrete model presented in Example 1.

3.4 Example 4: Modelling user habit with a mixed state space

The examples above provide illustrations of the usefulness of state space representations in \mathbb{Z}^n and \mathbb{R}^n ; in the following example, we illustrate the usefulness of the final, mixed representation iii) specified in section 2.2, again referring to a simple example from the family proposed by Cantarella & Cascetta (1995). In particular, starting from the two-route example so far considered, we now introduce the notion of 'habit', given the quite strong evidence that such inertia can be observed in practice (e.g. Chen & Mahmassani, 2004). As a highly simplistic model of such behaviour, we suppose that with probability α travellers

reconsider their previous day's choice, and make choices according to a logit model, as in earlier examples, with a β -type learning model (as used in Example 3) used to update the predicted travel costs. With probability $1 - \alpha$ travellers choose between the available routes with probabilities equal to the fraction of travellers that actually chose each of those routes on the previous day. Unlike Example 3, we shall retain the discrete nature of the flow variables (and so not make a Normal approximation to the Binomial). However, we will face a similar difficulty effectively in how to generate continuously-distributed actual costs from discrete flows, but here we shall adopt a different strategy of assuming:

- a) the travel cost functions provide a *mean* for a given level of flows, and that the actual cost for a particular day, given the flows for that day, is distributed according to a *continuous* random variable about that mean; and
- b) the predicted travel cost for the forthcoming day also has a random residual term associated with it, with mean equal to the learning filter based on the actual costs above.

As an example of assumption a), we specifically assume stationary Multivariate Normal random errors for the actual arc costs, and this will imply a univariate Normal random error in the actual cost *difference* between the two arcs/routes—let us suppose that this resulting, univariate Normal random error has mean 0 and variance σ^2 . By a similar logic, in view of assumption b) we shall assume that the predicted travel cost is distributed about the learning filter according to a stationary Normal distribution with mean 0 and variance ν^2 . We note that exactly the same strategy of introducing these error terms could have been adopted in Example 3, avoiding the need for the Normal approximation; the only reason to adopt the Normal approximation and affine cost functions in that case was in order to choose an example that allowed us to illustrate an analytic form for the transition function. (It is noted in passing that in the degenerate case $\nu = \sigma = 0$, the corresponding normal density functions appearing in the transition function below are replaced by Kronecker delta functions.)

In the present example, we shall adopt a three-dimensional, mixed discrete/continuous state variable $\mathbf{x} = (x_1, x_2, x_3)$, where $x_1 \in \{0, 1, 2, \dots, d\}$ is the flow on route 1, $x_2 \in \mathbb{R}$ is the predicted cost difference (at the beginning of the day) and $x_3 \in \mathbb{R}$ is the actual cost difference experienced (on that day). Now our aim will be to write down the joint conditional probability mass/density function of the random variable $\mathbf{X} = (X_1, X_2, X_3)$, given that yesterday's state was $\mathbf{Y} = \mathbf{y} = (y_1, y_2, y_3)$. It is first worth noting that there is *conditional dependence* between the component variables in \mathbf{X} given \mathbf{Y} , and therefore we cannot simply hope that the joint distribution of $\mathbf{X} | \mathbf{Y}$ is the product of the marginal distributions of $X_1 | \mathbf{Y}$, $X_2 | \mathbf{Y}$ and $X_3 | \mathbf{Y}$. On the other hand, we can see there is logical sequence (dependence structure) in considering the components of \mathbf{X} . This can be seen since $X_2 | \mathbf{Y}$ depends only on \mathbf{Y} (and not additionally on $X_1 | \mathbf{Y}$ and $X_3 | \mathbf{Y}$), $X_1 | \mathbf{Y}$ depends only on \mathbf{Y} and $X_2 | \mathbf{Y}$ (not additionally on $X_3 | \mathbf{Y}$), and $X_3 | \mathbf{Y}$ depends only on $X_1 | \mathbf{Y}$ (not additionally on $X_2 | \mathbf{Y}$). Therefore we may instead write (with, by a slight abuse of notation, 'Pr' denoting probability density or probability mass function as appropriate):

$$\text{Pr}((X_1, X_2, X_3) | \mathbf{Y}) = \text{Pr}(X_2 | \mathbf{Y}) \cdot \text{Pr}(X_1 | \mathbf{Y}, X_2) \cdot \text{Pr}(X_3 | \mathbf{Y}, X_1).$$

We may determine each of the component distributions in this product. Considering firstly $X_2 | \mathbf{Y} = (y_1, y_2, y_3)$, then this distribution has two influences. On the one hand, its mean is determined by the learning filter as a weighted combination of yesterday's predicted cost difference y_2 and yesterday's actual cost difference y_3 , such that:

$$E[X_2 | \mathbf{Y} = (y_1, y_2, y_3)] = (1 - \beta) y_2 + \beta y_3.$$

The second influence is the assumed error distribution of the predicted cost about this mean, yielding:

$$X_2 | \mathbf{Y} = (y_1, y_2, y_3) \sim \text{Normal}((1 - \beta) y_2 + \beta y_3, \nu^2).$$

Now we move on to consider the distribution of $X_1 | \mathbf{Y} = (y_1, y_2, y_3), X_2 = x_2$. With the extended behavioural model, travellers now are assumed to have two reasons for choosing route 1 (over route 2): either they choose it out of habit based on a probability of y_1/d (the fraction of drivers that actually chose that route yesterday), or they choose it according to a logit model based on today's predicted cost difference x_2 . The probability of choosing for the first reason is $1 - \alpha$ and for the second reason is α . This combination of assumptions implies:

$$X_1 | \mathbf{Y} = (y_1, y_2, y_3), X_2 = x_2 \sim \text{Binomial}(d, \{(1 - \alpha) y_1/d\} + \alpha/\{1 + \exp(\xi x_2)\}).$$

Finally we consider the distribution of $X_3 | \mathbf{Y} = (y_1, y_2, y_3), X_1 = x_1$. In an analogous way to the predicted cost difference, the distribution of the actual cost difference was two influences: its mean is determined by the link performance functions and its variation about the mean by a stationary Normal random error term. Assuming as in Example 1 that the mean arc cost difference is given by a difference in performance functions denoted by $c(\cdot; \lambda, d)$, then we have:

$$X_3 | \mathbf{Y} = (y_1, y_2, y_3), X_1 = x_1 \sim \text{Normal}(c(x_1; \lambda, d), \sigma^2).$$

Thus we are in a position to construct the complete transition function (i.e. the complete conditional joint probability density/mass function) as a product of the three component distributions derived above:

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = & \{d!/(x_1! (d - x_1)!) \} \cdot \{(1 - \alpha) y_1/d\}^{x_1} \cdot \\ & \{1 - \{(1 - \alpha) y_1/d\} - \alpha/\{1 + \exp(\xi x_2)\}\}^{d - x_1} \cdot \\ & (2\pi\nu)^{-0.5} \exp(-0.5((x_2 - (1 - \beta) y_2 - \beta y_3)/\nu)^2) \cdot \\ & (2\pi\sigma)^{-0.5} \exp(-0.5((x_3 - c(x_1; \lambda, d))/\sigma)^2) \\ & (\text{for } x_1 = 0, 1, \dots, d; -\infty < x_2 < \infty; -\infty < x_3 < \infty; y_1 = 0, 1, \dots, d; \\ & -\infty < y_2 < \infty; -\infty < y_3 < \infty; \boldsymbol{\theta} = (\lambda, d, \xi, \beta, \alpha, \nu, \sigma)). \end{aligned}$$

To derive the state distributions we now must use formulation iii) in section 2.2, in which the state space is split into a discrete and continuous component – in this case, the discrete component corresponding to x_1 and the continuous component to (x_2, x_3) . Thus, the state distributions that arise for $t = 1, 2, \dots, \{q^{(t)}(\mathbf{x}) : x_1 \in \{0, 1, 2, \dots, d\}; -\infty < x_2 < \infty; -\infty < x_3 < \infty\}$, as well as the initial $t = 0$ distribution, have both a discrete and continuous component. In order to facilitate the specification of the initial joint distribution, a decomposition strategy analogous to that carried out above should be performed.

3.5 Example 5: Dynamic traffic assignment in a stochastic process context

In the analysis so far we have presumed that the within-day scale is unimportant, and have used only steady-state link performance functions to represent congestion. Such an assumption is not necessary, and indeed any of the models so far described can be extended to permit within-day dynamic interactions (see, for example, Cascetta & Cantarella, 1991; Balijepalli & Watling, 2005). For example, let us consider the simplest example we considered so far, namely Example 1, and now suppose that traffic flows are propagated along each link according to a Cell Transmission Model (CTM, Daganzo, 1994),

based on a discretisation of the origin time-period modelled within each day into n departure periods, for some given n . Assuming first-in-first-out to hold, then a travel time may be imputed for each departure period, by comparison of the cumulative in-flows to each link with the cumulative out-flows from that link. Let us suppose that the origin-destination demand flow in each departure period i is given by the constant integer d_i for $i = 1, 2, \dots, n$, and let x_i denote the flow that choose to use route 1 when departing in time period i , for $i = 1, 2, \dots, n$. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ thus be the state vector, with state space:

$$\mathcal{S} = \{(x_1, x_2, \dots, x_n) : x_i \in \{0, 1, 2, \dots, d_i\} \text{ for } i = 1, 2, \dots, n\}.$$

For any given \mathbf{x} the flow profile entering route 2 is $\mathbf{d} - \mathbf{x}$, and so given \mathbf{x} we may apply the CTM as described above to obtain a unique travel time for each route for each entry period (forming a weighted average across the entry period when traffic exits in different exit time periods). For each entry period, suppose that we record the travel time on route 1 minus the travel time on route 2. The relationship between \mathbf{x} and these travel time differences may therefore be represented as a vector function $\mathbf{c}(\mathbf{x}; \lambda, \mathbf{d})$, where $\mathbf{c}: \mathcal{S} \rightarrow \mathbb{R}^n$ and where λ denotes the vector of parameters of the CTM across both links.

Assuming that, as in Example 1, drivers consider only the previous day's travel time relevant to the departure period in which they are travelling, that for simplicity we simply assume all parameters hold for all time periods, and that conditionally on the past drivers choose independently between time periods, then the transition function for this model is a simple extension of that presented in Example 1:

$$\phi(\mathbf{x}, \mathbf{y}; \theta) = \prod_{i \in \{1, 2, \dots, n\}} \{d_i! / (x_i! (d_i - x_i)!)\} \cdot \{1 / \{1 + \exp(\zeta c_i(\mathbf{y}; \lambda, \mathbf{d}))\}\}^{x_i} \cdot \{1 - \{1 / \{1 + \exp(\zeta c_i(\mathbf{y}; \lambda, \mathbf{d}))\}\}\}^{d_i - x_i} \quad (\text{for } \mathbf{x}, \mathbf{y} \in \mathcal{S}; \theta = (\lambda, \mathbf{d}, \zeta)).$$

The fact that such an apparently simple model arises is due to the conditional independence assumption, yet this is not such a strong assumption; effectively it means that while drivers travelling at the same time and at different times of day will have had commonalities and correlations in their past experiences, that once this experience is gained they do not then collude in making decisions on any particular day (analogous to the 'selfish routing' principle behind the Wardrop equilibrium concept). Certainly travel times do not separate by time-of-departure (note that c_i above depends on the vector \mathbf{y}). Indeed when this transition function is used within the Markov chain model ii) in section 2.2, the resulting state distributions $\{q^{(t)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ will in general be joint distributions that are correlated across departure time periods, and for more general learning models auto-correlated over between-day time (see Balijepalli *et al*, 2007). We remark that the same approach could be used for extending examples 2, 3 and 4 given earlier.

3.6 Example 6: A theoretical basis to micro-simulation

The examples considered so far all assumed drivers to act as an aggregate, but we may also disaggregate drivers to any desired level. At an extreme, we may model individual drivers with their own individual-specific attributes, thus providing a theoretical basis for the computational method known as (discrete) micro-simulation. For illustration, we shall adapt Example 4 (simplifying to no habitual effect, i.e. $\alpha = 1$ in the notation of Example 4) to follow the spirit of a model presented in Liu *et al* (2006). We shall suppose three inter-related aspects to the micro-simulation: in the propagation of individual vehicles through the network along given routes, in the way in which individuals learn from their own (but not others') experience, and in the individual-specific attributes that motivate the decision to select a route. The OD demand of d is thus split into individual drivers.

Let $\mathbf{f} = (f_1, f_2, \dots, f_d)$ be a vector of 0/1 indicator variables, with $f_i = 1$ only if individual i selects route 1 (rather than route 2) for $i = 1, 2, \dots, d$. Suppose that each individual possesses a vector of individual-specific, traffic-related behavioural parameters which determine factors such as that individual's preferred departure time, their 'aggressiveness' in accepting gaps when car-following or lane-changing, and the performance characteristics of the vehicle they drive. Suppose that all such parameters are collected together across all individuals in a single parameter vector λ . All of these factors account for drivers having a systematically individual-specific experience of travel time on the route that they choose to follow. Let $c_i(\mathbf{f}; \lambda)$ denote the mean travel time individual i would experience on their chosen route (i.e. on route 1 if $f_i = 1$ in the argument \mathbf{f} or on route 2 otherwise), when the population of drivers choose routes according to \mathbf{f} and when the population has individual traffic-related behavioural parameters given by λ . Suppose that the actual travel time experienced by each individual i is a random variable given by $c_i(\mathbf{f}; \lambda) + \varepsilon_i$ (for $i = 1, 2, \dots, d$), where $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d)$ are independent and identically distributed Normal random variables with mean 0 and variance σ^2 . (In realistic micro-simulation models the randomness would have a more complex source, but we adopt these simplistic assumptions here purely so we can write down an analytic expression for illustration.)

Now an important distinction from the aggregate models considered earlier is that when drivers learn, they learn only of the travel time on the route they actually chose, whereas for the unchosen alternative they leave their prediction of travel time unchanged (presumably in the real world an intermediate behaviour occurs). The fact that individuals learn individually means that we cannot utilise the strategy used in the aggregate examples, whereby only the difference in travel time between the routes is updated, but instead must separately record/update the predicted travel time for *each* route, for each individual. We shall use a weighted average learning model for the chosen route, as was used in Example 4, but here applied to each individual. This learning model provides the mean predicted cost for each route for each individual i , and to this we add a zero mean, Normally distributed and individual-specific error term, that is independent and identically distributed across individuals with variance v_k^2 for route $k = 1, 2$. Based on these predicted costs, each driver chooses a route based on a logit model with common scale parameter ξ .

In order to represent this model we shall use a state variable $\mathbf{x} \in \mathcal{S} = (\{0,1\} \times \mathbb{R}^3)^d$ which contains four items of information for each individual, stacked in order of individual number. Thus individual i 's variables are the elements $x_{4i-3}, x_{4i-2}, x_{4i-1}$ and x_{4i} (for $i = 1, 2, \dots, d$) denoting:

- $x_{4i-3} = 1$ if individual i chose route 1, and = 0 if route 2 chosen
- $x_{4i-2} =$ individual i 's actual travel time on their chosen alternative
- $x_{4i-1} =$ individual i 's predicted travel time on route 1
- $x_{4i} =$ individual i 's predicted travel time on route 2.

Conditional on yesterday's state \mathbf{y} , then we can first compute the new mean predicted travel times on each route. We do this by first noting that we may use y_{4i-3} and $1 - y_{4i-3}$ for routes 1 and 2 respectively as a form of indicator variable, each equal to 1 only if that route was the chosen route, and equal to 0 otherwise. Thus our learning model only updates the route that was chosen. For each individual $i = 1, 2, \dots, d$ we then have:

$$x_{4i-1} = y_{4i-1} + \beta y_{4i-3} (y_{4i-2} - y_{4i-1})$$

$$x_{4i} = y_{4i} + \beta(1 - y_{4i-3})(y_{4i-2} - y_{4i}).$$

Thus, when we consider the predicted travel times as random variables (by adding the iid error terms to the mean predicted travel times above) we obtain the two sets of conditionally independent distributions:

$$\begin{aligned} X_{4i-1} | \mathbf{Y} = \mathbf{y} &\sim \text{Normal}(y_{4i-1} + \beta y_{4i-3} (y_{4i-2} - y_{4i-1}), \nu_1^2) \text{ independently for } i=1,2,\dots,d \\ X_{4i} | \mathbf{Y} = \mathbf{y} &\sim \text{Normal}(y_{4i} + \beta(1 - y_{4i-3})(y_{4i-2} - y_{4i}), \nu_2^2) \text{ independently for } i=1,2,\dots,d. \end{aligned}$$

Given these predicted costs, each driver's choice of route for the present day is a Bernoulli trial, for each individual $i=1,2,\dots,d$ independently:

$$X_{4i-3} | X_{4i-1} = x_{4i-1}, X_{4i} = x_{4i}, \mathbf{Y} = \mathbf{y} \sim \text{Bernoulli}\left(1/\{1 + \exp(\xi(x_{4i-1} - x_{4i}))\}\right).$$

Finally, based on the choices made by all individuals, the actual cost for each individual on their chosen route is the sum of an individual-specific systematic component, and an individual-specific iid random error, such that for each $i=1,2,\dots,d$ independently:

$$X_{4i-2} | \{\mathbf{X}_{4i-1} = \mathbf{x}_{4i-1}, \mathbf{X}_{4i} = \mathbf{x}_{4i}, \mathbf{X}_{4i-3} = \mathbf{x}_{4i-3}, \mathbf{Y} = \mathbf{y}\} \sim \text{Normal}(c_i(\mathbf{x}_{4i-3}; \boldsymbol{\lambda}), \sigma^2).$$

Collected together our joint conditional probability mass/density function leads us to an overall transition function of:

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = \prod_{i \in \{1,2,\dots,d\}} & \{1/\{1 + \exp(\xi(x_{4i-1} - x_{4i}))\}\}^{x_{4i-3}} \cdot \\ & (2\pi\nu_1)^{-0.5} \exp(-0.5((x_{4i-1} - y_{4i-1} - \beta y_{4i-3} (y_{4i-2} - y_{4i-1}))/\nu_1)^2) \cdot \\ & (2\pi\nu_2)^{-0.5} \exp(-0.5((x_{4i} - y_{4i} - \beta(1 - y_{4i-3})(y_{4i-2} - y_{4i}))/\nu_2)^2) \cdot \\ & (2\pi\sigma)^{-0.5} \exp(-0.5((x_{4i-2} - c_i(\mathbf{x}_{4i-3}; \boldsymbol{\lambda}))/\sigma)^2) \\ & (\text{for } \mathbf{x}, \mathbf{y} \in \mathcal{S}; \boldsymbol{\theta} = (\boldsymbol{\lambda}, d, \xi, \beta, \nu_1, \nu_2, \sigma)). \end{aligned}$$

As for Example 4 earlier, we have a mixed discrete/continuous state-space, and so we use formulation iii) of section 2.2 to model the state transitions.

3.7 Example 7: Incorporation of 'rule-based' approaches

Some micro-simulation models are based on a series of 'if ... then' logical rules, rather than smooth, continuous response functions, and it is interesting to see that these also may be incorporated within the framework proposed. As an example, we examine a behavioural rule that has received some considerable attention in the research literature, namely that of bounded rationality (e.g. Hu & Mahmassani, 1997; Guo & Liu, 2010). As a simple illustration, we shall adapt the model presented in section 3.6.

Specifically we will adapt the learning process presented in section 3.6, by assuming that drivers only update their predicted travel times if their latest experience of travel time is significantly different (in absolute terms) from the travel time they anticipated. This provides the following rule:

```
If  $|y_{4i-1} - y_{4i-2}| < \varepsilon$  then
   $x_{4i-1} = y_{4i-1}$ 
   $x_{4i} = y_{4i}$ 
else
   $x_{4i-1} = y_{4i-1} + \beta y_{4i-3} (y_{4i-2} - y_{4i-1})$ 
   $x_{4i} = y_{4i} + \beta(1 - y_{4i-3})(y_{4i-2} - y_{4i}).$ 
```

Here, we assume a simple case of a deterministic threshold ε , though more generally it may be stochastic; also, we put no constraint on the relative improvement, only the absolute improvement is considered in the rule.

This rule can be written more concisely by introducing the heavy-side function $h(\cdot)$ given by:

$$\begin{aligned} h(z) &= 0 \text{ if } z < 0 \\ &= 1 \text{ if } z \geq 0 \end{aligned}$$

and then we can subsume these cases with the expressions:

$$\begin{aligned} x_{4i-1} &= y_{4i-1} + h(|y_{4i-1} - y_{4i-2}| - \varepsilon) \beta y_{4i-3} (y_{4i-2} - y_{4i-1}) \\ x_{4i} &= y_{4i} + h(|y_{4i-1} - y_{4i-2}| - \varepsilon) \beta (1 - y_{4i-3})(y_{4i-2} - y_{4i}) . \end{aligned}$$

The remainder of the model presented in section 3.6 is unaffected by this modification, and so using the results of that section we are able to write down the new transition function of the model as a slightly modified form of that given in section 3.6:

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) &= \prod_{i \in \{1, 2, \dots, d\}} \{1 / \{1 + \exp(\xi(x_{4i-1} - x_{4i}))\}\}^{x_{4i-3}} \cdot \\ &\quad (2\pi\nu_1)^{-0.5} \exp(-0.5((x_{4i-1} - y_{4i-1} - \beta h(|y_{4i-1} - y_{4i-2}| - \varepsilon) y_{4i-3} (y_{4i-2} - y_{4i-1}))/\nu_1)^2) \cdot \\ &\quad (2\pi\nu_2)^{-0.5} \exp(-0.5((x_{4i} - y_{4i} - \beta h(|y_{4i-1} - y_{4i-2}| - \varepsilon) (1 - y_{4i-3})(y_{4i-2} - y_{4i}))/\nu_2)^2) \cdot \\ &\quad (2\pi\sigma)^{-0.5} \exp(-0.5((x_{4i-2} - c_i(\mathbf{x}_{4i-3}; \boldsymbol{\lambda}))/\sigma)^2) \\ &\quad (\text{for } \mathbf{x}, \mathbf{y} \in \mathcal{S}; \boldsymbol{\theta} = (\boldsymbol{\lambda}, d, \xi, \beta, \nu_1, \nu_2, \sigma, \varepsilon)) . \end{aligned}$$

3.8 Example 8: Capturing stochasticity through Brownian motion

As well as constructing stochastic processes through what are essentially adaptations of traditional transport modelling methods, we may also look to the wider literature of stochastic processes and consider ways in which the general techniques developed therein might be applied to transportation problems. It is rather difficult to find examples of transportation papers that have followed such an approach, but one exception is that of Friesz *et al* (2008). This paper exploits the theory of stochastic differential equations to propose a model based on Brownian motion, which distinguishes separate notions of deterministic change and stochastic variation. Unfortunately for our purposes of constructing a simple illustrative example, the transportation application it considers is extremely complex, quite aside from the main feature that we wish to illustrate here. Moreover, the model finally produced is effectively a deterministic (rather than stochastic) process, since conditionally on the past at any one time, complete statistical expectations of the relevant stochastic variables are used. Therefore in the present section we present a much simpler example which is motivated by many elements of the Friesz *et al* paper, but which considers a highly-simplistic pricing mechanism, defined as a stochastic process. In keeping with the presentation so far, we also convert their continuous-time formulation into a discrete-time process with ‘simple’ state variables (i.e. the state variables are not *functions* of within-day time, but are discretized by within-day time).

As in all examples in this section, we suppose that drivers are travelling on a single origin-destination movement with two parallel routes available. As noted above, we maintain the focus of the present paper on discrete-time (day-to-day) systems in which the state variables are simple variables rather than functions of (within-day) time, and so modify and the continuous-time approach of Friesz *et al* (2008). Thus we suppose that within-day time is divided into n discrete time periods, with travellers assumed to make a joint choice of route and departure time among the $2n$ alternatives therefore possible in our two-route

network. In fact, we add a further no-travel alternative, which Friesz *et al* use to represent the decision to telecommute. We suppose that one of the routes (route 1) is subject to paying a price which varies between and within days, and is a price that users do not know with certainty until the end of their journey.

Our state variable $\mathbf{x} \in \mathbb{R}^n \times [0, \infty)^{2n+1}$ will be defined as:

x_j = price paid at destination to use route 1 when departing in interval j ($j = 1, 2, \dots, n$)

x_{rn+j} = travel time experienced on path r when departing in interval j
 $(r = 1, 2 \text{ and } j = 1, 2, \dots, n)$

x_{3n+1} = number of users who telecommute .

We suppose that the total origin-destination demand flow on any one day is given by the constant $d \in (0, \infty)$, and that $\gamma \in [0, 1]$ is a constant parameter defining an upper bound on the fraction of telecommuters permitted on any one day (say, defined by the employer at the destination).

Then, as throughout this paper, using \mathbf{y} to denote the 'yesterday' version of the state variable \mathbf{x} , the number of telecommuting travellers are assumed to vary according to a simple deterministic dynamical system with transitions given by:

$$x_{3n+1} = \max\{\min\{y_{3n+1} + \beta((2n)^{-1}(\sum_{j \in \{1, 2, \dots, n\}} \sum_{r \in \{1, 2\}} y_{rn+j}) - \chi), \gamma d\}, 0\}$$

where $\chi \in (0, \infty)$ is a parameter representing a congestion threshold and where $\beta \in [0, \infty)$ is a parameter controlling the rate of adjustment. The term inside the inner parentheses, against which χ is compared, is simply yesterday's average experienced travel time across all paths and departure times. The min and max operations simply ensure that the result is constrained to be on the interval $[0, \gamma d]$, where the parameter γ was defined above. Thus we have defined the first component of the overall model, namely the telecommuting sub-model. The second component of the model concerns the congestion pricing element of the model, and is the component in which the stochastic elements arise. This is the place in which we depart most from the source paper, representing only a highly simplistic pricing system. We suppose that only one of the routes (route 1) is priced; the other route is free. For the priced route, the probability distribution of the price X_j paid (at the destination) when departing in departure interval j (given the travel time y_{n+j} on route 1 and departure time j on the previous day) is constructed recursively across departure intervals according to a discretised form of geometric Brownian motion:

$$X_j = X_{j-1} + \Delta\mu_j + B_j \sigma_j(y_{n+j}; \boldsymbol{\alpha}) \quad B_j \sim \text{Nor}(0, \xi_j^2) \quad (j = 1, 2, \dots, n; X_0 = 0)$$

for constant parameters $\Delta, \mu_1, \mu_2, \dots, \mu_n$ and $\xi_1, \xi_2, \dots, \xi_n$, and for functions $\sigma_j(\cdot; \boldsymbol{\alpha})$ ($j = 1, 2, \dots, n$) which are parameterised in total by the vector $\boldsymbol{\alpha}$. The three characteristic components of this Brownian model are: the parameters $\mu_1, \mu_2, \dots, \mu_n$ which capture the *deterministic drift* that depends on time of day through j (Friesz *et al* proposed a simple linear relationship with time); $\sigma_j(y_{n+j}; \boldsymbol{\alpha})$ which is the *volatility* which depends on both time of day j and the particular day (through the previous day's travel time y_{n+j}); and B_j which represents *white noise* with variance ξ_j^2 . This way of constructing the prices provides a simple mechanism for implicitly defining correlation across time periods, through a series of conditionally independent statements, i.e:

$$X_j | X_{j-1} \sim \text{Nor}(X_{j-1} + \Delta\mu_j, (\xi_j \sigma_j(y_{n+j}; \boldsymbol{\alpha}))^2) \quad (\text{independently for } j = 1, 2, \dots, n; X_0 = 0).$$

Thus the joint probability density function $f(\cdot)$ of the prices (X_1, X_2, \dots, X_n) is given by the product of these conditionally independent distributions:

$$f(x_1, x_2, \dots, x_n; \alpha, \xi, \mu, \Delta, x_0 = 0, \{y_{n+1}, y_{n+2}, \dots, y_{2n}\}) = \\ \prod_{j \in \{1, 2, \dots, n\}} (2\pi)^{-0.5} (\xi_j \sigma_j(y_{n+j}; \alpha))^{-1} \exp(-0.5((x_j - x_{j-1} - \Delta \mu_j) / (\xi_j \sigma_j(y_{n+j}; \alpha)))^2).$$

The final element of the model is to describe how travellers respond to the prices. This could be captured through learning processes of the kinds described in sections 3.2 and 3.3, whereby over time travellers build up their perceptions of the mean prices; but for simplicity of illustration we will adopt the simple myopic learning rule of section 3.1, and leave it to the reader to imagine the quite simple extension to more realistic learning behaviour. Thus we assume drivers base their route and departure time choice decisions on the prices incurred at the destination on the previous day. In keeping with the spirit of Friesz *et al's* approach (and unlike the other models described above), we shall assume that given the price signal, drivers are able to organise themselves on any one day into a Dynamic User Equilibrium (DUE) route and departure time pattern. The DUE model is based on assuming utility for any (route, departure time) combination is a linear combination of the schedule disutility, the travel time and the price. We suppose our DUE model is parameterised by some parameter vector λ , and (by the logic above) is a function of the departure-time-dependent prices on the priced route in the previous day, $\{y_1, y_2, \dots, y_n\}$ (i.e. the 'yesterday' counterpart of the state variables $\{x_1, x_2, \dots, x_n\}$). We assume there to be a unique DUE solution for any given λ and $\{y_1, y_2, \dots, y_n\}$. At this unique solution we may compute the travel times on any route and departure time combination. We thus introduce the functions $g_{rj}: \mathbb{R}^n \rightarrow \mathbb{R}$ to be the DUE mapping from (yesterday's) route 1 prices $\{y_1, y_2, \dots, y_n\}$ to (today's) DUE travel time x_{rn+j} for route r and departure time j (for $r = 1, 2$; and $j = 1, 2, \dots, n$), i.e.:

$$x_{rn+j} = g_{rj}(y_1, y_2, \dots, y_n; \lambda) \quad r = 1, 2; j = 1, 2, \dots, n.$$

Combining the expressions above, with state space $\mathcal{S} = \mathbb{R}^n \times [0, \infty)^{2n+1}$, the resulting transition function for the process is:

$$\phi(\mathbf{x}, \mathbf{y}; \theta) = \prod_{j \in \{1, 2, \dots, n\}} (2\pi)^{-0.5} (\xi_j \sigma_j(y_{n+j}; \alpha))^{-1} \exp(-0.5((x_j - x_{j-1} - \Delta \mu_j) / (\xi_j \sigma_j(y_{n+j}; \alpha)))^2) \\ \times \prod_{j \in \{1, 2, \dots, n\}} \prod_{r \in \{1, 2\}} \delta(x_{rn+j}, g_{rj}(y_1, y_2, \dots, y_n; \lambda)) \\ \times \delta(x_{3n+1}, \max\{\min\{y_{3n+1} + \beta((2n)^{-1}(\sum_{j \in \{1, 2, \dots, n\}} \sum_{r \in \{1, 2\}} y_{rn+j}) - \chi), \gamma d\}, 0\}) \\ (\text{for } \mathbf{x}, \mathbf{y} \in \mathcal{S}; \theta = (\alpha, \xi, \mu, \Delta, x_0 = 0, \lambda, \beta, \chi, \gamma, d))$$

where as earlier $\delta(\cdot, \cdot)$ denotes the Kronecker delta function.

4. PLANNING CONTEXT & THEORETICAL PROPERTIES

4.1 Dynamic Planning

So far in the paper we have defined a theoretical basis for modelling stochastic dynamics in transportation systems (section 2), and have illustrated this theory by describing particular models consistent with this theoretical basis (section 3). The approach traces the change over time in a joint probability distribution (state distribution) that fully describes the transportation system, when given an initial state distribution, and we have discussed (in section 2.2) how such a state distribution might be defined in practice. By the way we construct the method, we know that from any given initial state distribution at

time $t = 0$, there will exist a unique future trajectory of state distributions at times $t = 1, 2, 3, \dots$, provided that we can associate a transition function with the defined model. We have shown in section 3 how such transition functions may be explicitly derived. Therefore, we have a basis for suggesting this approach as a planning tool, e.g. to test hypothetical scenarios or anticipated future events. In conventional dynamic or static equilibrium analysis we might hope to guarantee the existence of a unique solution for any given scenario, and in the stochastic process context we might claim that the analogy is the existence of a unique time trajectory of probability distributions given the combination of any scenario and an initial state probability distribution.

In terms of numerical estimation, carrying out the computations to recursively calculate the state distributions (from the equations in section 2.2) might appear cumbersome for realistic scale networks, especially given the potential dimension of the state vector and size of the state space. However, one approach that could be said to circumvent such problems is Monte Carlo simulation. In the present context this would involve pseudo-random sampling of a particular initial state, and then subsequent pseudo-random sampling from the conditional distribution given by the transition function. This would provide a single trajectory of particular states. By replicating this process a number of times with different random number seeds, we would build up a collection of trajectories, and as the number of replications grows so the relative frequencies with which particular trajectories occur provide increasingly precise, unbiased estimates of the theoretical probability of such trajectories occurring in the underlying model (by the strong law of large numbers). Therefore, from these replicated Monte Carlo draws, we can compute summary statistics which can serve as estimates of the means and variances in flows and travel times at any particular time t , as well as examining temporal correlations and changes in such measures across time.

Therefore, in principle at least, we have both a theoretical basis and a solution method (suitable for large scale applications) for analysing the impacts of alternative scenarios, e.g. the impact of new infrastructure, or a driver information system, or the consequences of a hypothetical failure/incident in the network. In this context we might draw parallels with the approaches commonly used in short-term traffic forecasting (Okutani & Stephanedes, 1984; Vlahogianni *et al*, 2004; Ghosh *et al*, 2007), where historical information is used to calibrate stochastic models in order that near-term forecasts can be made, and control measures derived to influence future traffic states in a desirable way. The analogy in the present context might be with a planner making short-term decisions on a day-to-day basis; this may be useful, for example, in the days immediately following some major planned or unplanned event that leads to capacity reductions over a period of days.

We may also draw parallels at an entirely different temporal scale, with the ‘time-marching’ models used in dynamic land-use/transport interaction models (Wegener, 2004; Shepherd *et al*, 2006; Pfaffenbichler *et al*, 2008). While generally deterministic rather than stochastic processes, these models nevertheless operate over scales of perhaps 20-30 years, and in making appraisals of hypothesised measures must take into account the transient dynamics of the modelled system from some given initial conditions. An analogous problem of how to appraise and plan with transient systems has also been considered on the day-to-day scale, Friesz *et al* (2004) proposing techniques by which dynamic congestion pricing may be designed and implemented in such an environment. Indeed there exists a whole range of techniques developed for application fields such as

financial planning, which have only just begun to be considered in transportation (Zhao *et al.*, 2004; De Palma *et al.*, 2009), and transferring more such insights into the field must surely be a fruitful area for future research.

Therefore, there are promising possibilities for utilising the models so-defined in a dynamic planning environment. Indeed, we might argue that such a direction is not only promising but necessary, given empirical evidence suggesting that many of the choice processes we typically consider stabilise rather too slowly for the assumptions of steady-state analysis to hold (Goodwin, 1998). However, while we would wish to promote this as a research agenda, we believe a twin-track approach should be pursued, with the alternative approach one that we shall term ‘stationary planning’ (as defined in detail in section 4.2). This suggestion is based on several reasons:

1. Adopting a dynamic planning approach places great emphasis on the assumptions made about the assumed behavioural adaptation of travellers and the rates of adjustment implied by the parameter values. Any evaluation (such as cost-benefit analysis) performed with such a model will be potentially highly sensitive to such assumptions, especially during transient periods. However, our empirical understanding of such processes is still in its infancy, and the technology that might yield the fidelity and quantity of data available to inform such understanding (e.g. from mobile communications traces) is only just becoming available.
2. Linked to point 1, a dynamic planning approach places great emphasis on what we actually *mean* by ‘time’. For example, although there is a growing literature on day-to-day dynamic models, it is typically less clear what the ‘days’ actually represent; invariably, a simulation of 60 days is not intended to represent (say) a particular 12-week period of weekdays, but rather the days are meant to represent the *kinds* of travel experiences faced by commuters on daily level. Except in the case of very short-term responses over a few days, we would have difficulty conceptually matching a model run of n days with any particular n days; again this is a serious problem for any evaluation performed with the model, which would be sensitive to the value of n .
3. It is impossible to ignore the fact that there has been enormous experience in using equilibrium analyses for planning real-world transportation systems, and many measures and policies were justified in the past on the basis of such analyses. It would be naïve to believe that planners have no political investment in approaches; therefore, one cannot expect them to immediately alter their appraisal methods to use such dynamic, stochastic methods. Rather, some kind of transition in methodology is needed, whereby equilibrium analyses can be seen to play a part.
4. Point 3 suggests that understanding the connection of stochastic process methods to equilibrium analyses is a practical necessity, but also we can hypothesise that there can be useful theoretical and methodological issues to learn from equilibrium analyses. An example, though admittedly connected with deterministic dynamical systems, is the crossover between certain more ‘behaviourally-motivated’ algorithms for estimating equilibrium, and the stability analysis of dynamical adjustment processes (e.g. Smith, 1984). A second example is the study of Jacobian asymmetry in non-separable cost functions for equilibrium models, which has been found to be indicative of certain kinds of pseudo-stable convergence behaviour in Monte Carlo simulation of stochastic process systems (Watling, 1996).

5. From a purely computational point of view, it is desirable to seek more efficient methods for implementing such processes. As noted earlier, a dynamic planning analysis would potentially need a large number of Monte Carlo replications to estimate the state distributions over time. This is not a unique problem faced in transportation, and in this respect there appears to be potential in transferring more ideas from other fields, such as ‘scenario generation’ (Di Domenica *et al*, 2007). Nevertheless, avoiding the need to replicate Monte Carlo simulations is highly desirable, and we shall explore non-dynamic planning situations where this is feasible.

Therefore for these reasons, in the remainder of the paper we shall focus on an alternative interpretation of the stochastic process approach, which is indeed the interpretation that has been adopted in virtually all transportation research papers on this topic to date. This is not to suggest, however, that we advocate it as the only approach; we believe that both dynamic planning and stationary planning should be considered for their possibilities.

4.2 Stationary Distributions as a means for ‘Stationary Planning’

In this section we define concepts that utilise the techniques described in sections 2 and 3, but within a context that is more akin to traditional equilibrium analysis. In order to do so, we first must define some standard notions from stochastic process theory, which we will then relate to the theory described in earlier sections (and notably the theory and notation for the three types of stochastic process defined in section 2.2). For more details of the theory the reader is referred to Stokey & Lucas (1989), and for its application to transportation systems to Cantarella & Cascetta (1995).

Definition 1

Let $\mathfrak{E} \subseteq \mathcal{S}$ be a minimal subset of the state-space with the property that there is zero probability of leaving it once entered, and where *minimal* means that \mathfrak{E} does not contain any proper subset with this property. Then \mathfrak{E} is called an *ergodic* subset.

This definition implies that, if we neglect the steps by which a process arrives at some ergodic subset, then we can consider ergodic subsets in isolation, since always the process will transform back into states into the same subset. It gives the opportunity to think in terms of something analogous to a fixed point of a deterministic dynamical system within any such ergodic subset, but instead here in probability terms:

Definition 2

A state distribution $\{q(\mathbf{x}) : \mathbf{x} \in \mathfrak{E} \subseteq \mathcal{S}\}$ is termed a *stationary distribution* over an ergodic subset \mathfrak{E} if any only if (as appropriate)

$$q(\mathbf{x}) = \int_{\mathbf{y} \in \mathfrak{E}} \phi(\mathbf{x}, \mathbf{y}; \theta) q(\mathbf{y}) d\mathbf{y} \quad (\forall \mathbf{x} \in \mathfrak{E}; \theta \in \Theta) \quad [\text{Markov process, } \mathcal{S} \subseteq \mathbb{R}^m]$$

$$q(\mathbf{x}) = \sum_{\mathbf{y} \in \mathfrak{E}} \phi(\mathbf{x}, \mathbf{y}; \theta) q(\mathbf{y}) \quad (\forall \mathbf{x} \in \mathfrak{E}; \theta \in \Theta) \quad [\text{Markov chain, } \mathcal{S} \subseteq \mathbb{Z}^n]$$

$$q(\mathbf{x}) = q((\mathbf{x}_{[1]}, \mathbf{x}_{[2]})) = \sum_{\mathbf{y}_{[1]} \in \mathfrak{E}_{[1]}} \int_{\mathbf{y}_{[2]} \in \mathfrak{E}_{[2]}} \phi((\mathbf{x}_{[1]}, \mathbf{x}_{[2]}), (\mathbf{y}_{[1]}, \mathbf{y}_{[2]}); \theta) q((\mathbf{y}_{[1]}, \mathbf{y}_{[2]})) d\mathbf{y}_{[2]}$$

$$(\forall \mathbf{x} = (\mathbf{x}_{[1]}, \mathbf{x}_{[2]}), \mathbf{x}_{[1]} \in \mathfrak{E}_{[1]}; \mathbf{x}_{[2]} \in \mathfrak{E}_{[2]}; \mathfrak{E} = \mathfrak{E}_{[1]} \times \mathfrak{E}_{[2]}; \theta \in \Theta)$$

$$[\text{Markov process/chain, } \mathcal{S} \subseteq \mathbb{Z}^n \times \mathbb{R}^m].$$

A stationary distribution is thus a kind of invariant point in the (infinite-dimensional) space of probability distributions, since whenever the process takes on the probabilities/probability-densities from a stationary distribution at some time t , then the transition function will map it to exactly the same probabilities/probability-densities at times $t + 1$, and so on for all future times.

We are then in a position to make a definitions which apply to the whole state space of a stochastic process.

Definition 3

A stochastic process is called *stationary* if at least one stationary probability distribution exists (existence), *ergodic* if stationary and exactly one stationary probability distribution exists (uniqueness), and *regular* if ergodic and its probability distribution converges⁶ to the one stationary probability distribution whatever the initial state distribution (global convergence).

That is to say, for a regular stochastic process there exists a unique stationary distribution $\{q(\mathbf{x}) : \mathbf{x} \in \mathcal{E} \subseteq \mathcal{S}\}$ over a unique ergodic subset \mathcal{E} of state-space \mathcal{S} , for which $\{q^{(t)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\} \rightarrow \{q(\mathbf{x}) : \mathbf{x} \in \mathcal{E}; \text{ with } q(\mathbf{x}) = 0 \text{ for } \mathbf{x} \notin \mathcal{E}\}$ as $t \rightarrow \infty$, regardless of the initial state distribution $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$. Therefore it is very interesting to know what conditions on the process might be sufficient to guarantee regularity, which the theorem below summarises.

Theorem 1

Let $\{\phi^{(n)}(\mathbf{x}, \mathbf{y}) : \mathbf{x}, \mathbf{y} \in \mathcal{S}\}$ denote the *n-step transition function* of a Markov process or chain, namely the *conditional* joint probability mass/density function for the possible states $\mathbf{x} \in \mathcal{S}$ in the current time epoch, given that \mathbf{y} was the state that occurred n time epochs previously.

(a) Markov Chain:

Let $\varepsilon^{(n)} = \sum_{\mathbf{y} \in \mathcal{S}} \min\{\phi^{(n)}(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in \mathcal{S}\}$. Then the stochastic process is regular if for some integer $N \geq 1$, $\varepsilon^{(N)} > 0$.

(b) Markov Process:

The stochastic process is regular if there exists $\varepsilon \geq 0$ and an integer $N \geq 1$ such that for any subset $\mathcal{A} \subseteq \mathcal{S}$,

$$\begin{aligned} \text{either: } & \int_{\mathbf{y} \in \mathcal{A}} \phi^{(n)}(\mathbf{x}, \mathbf{y}) d\mathbf{y} \geq \varepsilon \quad \text{for all } \mathbf{x} \in \mathcal{S} \\ \text{or: } & \int_{\mathbf{y} \notin \mathcal{A}} \phi^{(n)}(\mathbf{x}, \mathbf{y}) d\mathbf{y} \geq \varepsilon \quad \text{for all } \mathbf{x} \in \mathcal{S}. \end{aligned}$$

Proof: See Theorem 11.4 of Stokey & Lucas (1989, pp 332-333) for Markov Chains, and Theorem 11.12 (pp 350-351) for Markov Processes.

⁶ For general Markov processes or Markov chains/processes, the notion of what we mean by convergence is a study in itself, since we may define a variety of norms over which convergence may be studied. The results established here concern what is termed *strong convergence* (see Stokey & Lucas, 1989, pp 338-344).

Since the Theorem only requires there to exist some $N \geq 1$ where the property holds, then it is a natural first step to consider whether it holds for $N = 1$, i.e. by checking whether it holds for the defining transition function of the process. In all of the examples considered in section 3 the properties can be seen to hold, since the transition function associates positive probability mass/density for transitions between all pairs of states across the entire state space. Thus in all these cases, by construction, we can associate a unique stationary distribution with the model, to which the process will converge regardless of the initial conditions.

This existence of a unique stationary distribution holds not only theoretical interest, since for a regular process, all the relevant statistics for the single stationary distribution can be computed from just one pseudo-realization of the process, whatever the starting state, since clearly by definition any stationary distribution we obtain is the only one that can arise. We no longer face the problem noted with dynamic planning of needing to simulate replicated pseudo-random realizations of the process, and so stationary distributions also hold great interest from the viewpoint of computational tractability.

The possibility, then, to guarantee the existence of a unique and readily-computable stationary distribution provides an obvious opportunity to use such a distribution for appraising hypothetical policies. To be clear, this distribution contains rather rich information, not only means, variances and covariances of flows, travel times and other factors, but also between-day covariances (autocorrelations) which occur even in the stationary regime. It provides the basis for what we might call ‘stationary planning’, a stochastic process analogue of familiar methods used with equilibrium models.

4.3 Relationships between SP, DP and equilibrium models

A proper treatment of the relationship between deterministic process, stochastic process and equilibrium models would require a paper in its own right, and such a treatment is not the focus of the present paper. However, as was remarked in section 4.1, there are compelling practical and theoretical reasons for seeking a full understanding of such relationships, and so in this section we briefly note some of the main issues that are understood in this relationship.

Intuition would suggest that a likely relationship between the models might run thus. Imagine first that a stochastic process model is specified with transition function $\phi(\mathbf{x}, \mathbf{y})$, and to pick an example let us imagine this is on a discrete state space, so that our process is:

$$q^{(t)}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{S}} \phi(\mathbf{x}, \mathbf{y}) q^{(t-1)}(\mathbf{y}) \quad (\mathbf{x} \in \mathcal{S} \subseteq \mathbb{Z}^n).$$

Let us now ‘determinise’ this process by taking expectations with respect to the conditional distribution given by the transition function at a particular given value of \mathbf{y} . This amounts to constructing the vector function $\psi(\cdot)$ given by:

$$\psi(\mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{S}} \mathbf{x} \phi(\mathbf{x}, \mathbf{y}).$$

Now consider the deterministic dynamical system:

$$\mathbf{x}^{(t)} = \psi(\mathbf{x}^{(t-1)}) \quad (t = 1, 2, \dots).$$

Actually we are being somewhat informal here since the state space will no longer be discrete, and so it is not entirely clear that we are permitted to apply ψ in this way, but for the purposes of this informal discussion this is not a critical issue.

Now consider a stationary point of this dynamical system, i.e. a state \mathbf{x}^* such that:

$$\mathbf{x}^* = \psi(\mathbf{x}^*) .$$

Thus we have a connection between an equilibrium state \mathbf{x}^* , a deterministic dynamical system defined by ψ , and the original stochastic process with transition function ϕ . It is tempting then to imagine this construction process in reverse, i.e. to ask what does the deterministic dynamical system add to a study of equilibrium, and then what does the study of a stochastic process add to a stochastic process? In this case, a natural line of reasoning would be that the deterministic system adds a description of transient, non-equilibrium or pre-equilibrium behaviour, and the stochastic process essentially adds 'noise' to this process to reflect some uncertainty.

Tempting as such logic might be, the difficulty with such reasoning is that the dynamics generated by the stochastic process are of rather complex form, especially given the nonlinear nature of the processes involved. For example, even when the process is stationary it is not the case that each 'day' is then like an independent random sample; even when stationary, complex dynamics persist in terms of the auto-correlations between days. Furthermore, it is not the case that such a determinised system would correctly predict the mean evolution; neglecting variability not only affects variance but affects the mean evolution of the process. On the other hand, if the deterministic system evolves towards a periodic or a chaotic attractor, it also has a variance and a correlation structure whose relationship with their stochastic counterparts is not pre-determined. In this case, the flow distribution can be expected to exhibit several modes for both deterministic and stochastic processes, and the auto correlogram of the stochastic process will show the same structure as the deterministic one but with less extreme values.

For such reasons relationships between deterministic and stochastic process models are an open research field in the theory of non-linear dynamic systems, and we cannot expect general results to connect any such processes. On the other hand, some special results exist for the kinds of model we consider here (and will consider further in section 5). Davis and Nihan (1993) show that a particular form of stochastic process model converges, as both the OD demands and capacities grow in tandem, to the sum of a deterministic non-linear process (with mean equal to a stochastic user equilibrium) and Gaussian white noise (that is a non stationary linear stochastic process defined by a 0-mean normal vector). From the practical point of view, then, it can be concluded that if O-D demand and capacity values are large enough, deterministic and stochastic descriptions of system evolution should be, according to the law of large numbers, increasingly similar. Hazelton & Watling (2004) built on Davis & Nihan's result to develop a practical, analytic approximation method for estimating the stationary distribution of such a process without needing to simulate its evolution, but only given an SUE solution. This approximation was developed for within-day static cost-flow functions, but was later extended to the within-day dynamic case by Balijepalli & Watling (2005).

Cascetta (1989) also analyzed the relationships between equilibrium and stochastic stationary distribution for a simplified model, confirming that the equilibrium link or path flows are generally different from corresponding expected values of the stochastic process. Moreover, if the deterministic process converges to a fixed-point attractor coincident with an equilibrium state and the system dimensionality is large enough, the same can be said for the stochastic process; in this case, flow variances should be increasingly small and average stationary values should become close to those of equilibrium.

On the other hand, in cases where the conditions for existence of a unique stationary distribution almost break down, examples can be found of ill-conditioned, seemingly paradoxical behaviour. Watling (1996) studied a series of simple examples of problems which were known to produce multiple equilibria in conventional models, but which nevertheless theoretically satisfy the conditions for existence of a unique stationary distribution in a stochastic process sense. It was found that, depending on the behavioural parameter assumed in the model, the process could nevertheless exhibit kinds of ‘pseudo-stable’ behaviour, where it was trapped within a region of stable point equilibrium. Thus, in practical simulations, the link between the stochastic process, deterministic process and equilibrium approaches could be stronger than might be anticipated from the theory.

In summary, a growing body of evidence is emerging that establishes connections between the stochastic process, deterministic process and equilibrium approaches, but these connections are rather complex in nature and would benefit from more theoretical and empirical investigation. For further discussion and insights into this issues, see Watling & Cantarella (2013).

5. A GENERAL FAMILY OF STOCHASTIC PROCESS TRAFFIC ASSIGNMENT MODELS

In section 3 we presented a variety of simple examples on a network consisting of only two arcs, though as we noted at the time all of the examples provided in that section may be extended to general networks. In the present section we explicitly set out such a general network approach, with a family of models that incorporate many of the features set out in those examples. For this family we will derive the transition function for a particular state representation, and (based on the results set out in section 4.2) identify the assumptions on the underlying sub-models that guarantee the existence of a unique stationary distribution.

The context is a single (private car) mode, with constant demand based on a segmented aggregate model of choice, behaviour and learning, with the supply-side restricted to within-day static cost-flow relationships. Unlike with the examples in section 3, in general networks it would be unwieldy to attempt to use the notation \mathbf{x} to describe all state variables. Instead the notation follows, for the most part, that in Cascetta (2009). Since demand flows are assumed constant, path choice is the only user choice behaviour affected by network performance, or more properly by congestion. Users travelling between a common origin-destination pair with common behavioural features are grouped into N user classes, with each user class i assumed to have a set of (elementary) available paths \mathcal{K}_i (assumed non-empty and finite) and a constant, integer *demand flow* $d_i \geq 0$, $d_i \in \mathbb{Z}$. That

is to say the user class index is used both to distinguish different types of user and also to distinguish different origin-destination movements.

5.1 Modelling transportation supply

Transportation supply is modelled through a network with a transportation cost c_a and a flow f_a associated to each arc a . Let:

$\mathbf{c} \geq \mathbf{0}$ be the vector of arc costs, with entries c_a for $a = 1, 2, \dots, n$;

$\mathbf{f} \geq \mathbf{0}$ be the vector of arc flows, with entries f_a for $a = 1, 2, \dots, n$;

$\boldsymbol{\Gamma}_i$ be the arc-path incidence matrix for user class i (for $i = 1, 2, \dots, N$), with entries $\gamma_{iak} = 1$ if arc a belongs to path k and $k \in \mathcal{K}_i$, otherwise $\gamma_{iak} = 0$;

$\mathbf{w}_i \geq \mathbf{0}$ be the vector of path costs for user class i , with entries w_k , $k \in \mathcal{K}_i$;

$\mathbf{h}_i \geq \mathbf{0}$, with $\mathbf{1}^\top \mathbf{h}_i = d_i$, be the vector of path flows for user class i , with entries h_k , $k \in \mathcal{K}_i$.

Omitting the possibility of path-specific costs (for simplicity's sake), the *arc-path cost consistency* condition is expressed by:

$$\mathbf{w}_i = \boldsymbol{\Gamma}_i^\top \mathbf{c} \quad \forall i.$$

Moreover, the *arc-path flows consistency* condition is expressed by:

$$\mathbf{f} = \sum_i \boldsymbol{\Gamma}_i \mathbf{h}_i \in \mathcal{F} \subseteq \mathbb{Z}^n$$

where the *feasible arc flow set* is given by

$$\mathcal{F} = \{ \mathbf{f} \in \mathbb{Z}^n : \mathbf{f} = \sum_i d_i \mathbf{h}_i \text{ where } \mathbf{h}_i \in \mathbb{Z}^{m_i}, \mathbf{h}_i \geq \mathbf{0} \text{ and } \mathbf{1}^\top \mathbf{h}_i = d_i \forall i = 1, 2, \dots, N \}.$$

Let us now consider the role of day-to-day random variation. In order to avoid misunderstanding, it should be made clear that the random variation we consider here is entirely different from the "inter-personal" random variation posited in random utility models, such as logit and probit. This inter-personal variation is considered later, in section 5.3t. In the present section we consider a different source of variation, namely the kind of real variation from day-to-day that travellers may experience in their travel times. This kind of variation is something that happens in real-life (it is not only a model construct), in the sense that we may observe it on street, and indeed it may form part of some evaluation of the actual performance of the network. We assume that part of this daily variation in travel times is demand-related and is explained by variation in the flows, and that the remaining daily variation in travel times (e.g. due to weather affecting free-run speeds) acts as an additional, additive component. Specifically, conditionally on the arc flows $\mathbf{F}^t = \mathbf{f}^t$ on any day t , there are assumed to be two components to the arc costs \mathbf{C}^t for that day. On the one hand, congestion is simulated assuming that *mean* arc costs depend on the given arc flows, through the *arc cost function* $\mathbf{c}(\mathbf{f}; \boldsymbol{\mu})$ ($\mathbf{f} \in \mathcal{F}$), a vector function assumed to take non-negative values, where $\boldsymbol{\mu}$ is the vector of all relevant supply-side parameters, such as capacity and sensitivity to congestion. Thus, $E[\mathbf{C}^t | \mathbf{F}^t = \mathbf{f}^t] = \mathbf{c}(\mathbf{f}^t; \boldsymbol{\mu})$. The second component to the arc costs \mathbf{C}^t is the random variation about the mean. While there is no great restriction on the assumptions, we shall adopt a quite simple model, namely that the random arc cost disturbances are independent between links, with the disturbance for arc a following a Normal distribution with mean 0 and constant variance σ_a^2 ($a = 1, 2, \dots, n$).

5.2 Modelling memory, learning and forecasting

Generally users do not know in advance the costs they will actually incur during their trip. Thus it is assumed that they make their choices according to forecasted path costs, resulting from their memory and learning processes, and generally different from actual path costs not yet known. Personal experience is usually complemented by information exchanged with other users and possibly provided by an information system. As was noted in section 2.4, such learning processes may be modelled at different levels of aggregation. In the example in section 3.6, the learning and forecasting processes of each individual were explicitly modelled by weighting differently the information relative to experienced and non-experienced path costs (disaggregate memory). While the specification given so far could permit such an approach, here we shall restrict attention to the aggregate memory case. That is to say, the distribution of forecasted arc costs across users is obtained following an approach in which individual differences among users of the same class are taken into account in the choice behaviour model through random residuals with respect to common, average forecasted path costs.

Generally, forecasted costs depend on costs incurred on previous days. Hence, learning and forecasting processes can be modelled through filters applied to costs incurred on previous days. In the following, forecasting filters will be assumed time-invariant, that is their functional form and parameters are independent of day t . In an analogous way to the treatment of actual costs in the previous subsection, the (random) forecasted arc cost vector \mathbf{Z}^t is assumed to have two components. On the one hand, the mean of \mathbf{Z}^t conditional on the previous day's state is assumed to be a function of the previous day's actual arc costs $\mathbf{c}(\mathbf{f}^{t-1}; \mu)$ and the previous day's forecasted arc costs \mathbf{z}^{t-1} , through the *cost updating recursive equation*:

$$E[\mathbf{Z}^t | \mathbf{C}^{t-1} = \mathbf{c}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}] = \varphi(\mathbf{c}^{t-1}, \mathbf{z}^{t-1}) \quad (t = 1, 2, \dots)$$

where $\varphi: \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$ is the learning filter. A simple example of such a learning filter is the exponential smoothing filter, in which $\beta \in]0, 1]$ is the weight given to yesterday's actual costs when forecasting today's costs, whereby:

$$\varphi(\mathbf{c}, \mathbf{z}) = \beta \mathbf{c} + (1-\beta) \mathbf{z} \quad \text{with } \mathbf{z}^0 = \mathbf{c}(\mathbf{f}^0; \mu), \mathbf{f}^0 \in \mathcal{F}.$$

However, many other types of filter may be cast within this framework, such as those described in Chang & Mahamasani (1988), Iida *et al* (1992), Cascetta & Cantarella (1993) and Davis & Nihan (1993).

The second component to the forecasted arc cost vector \mathbf{Z}^t is the random variation about the mean. As for the actual costs, there is no great restriction on the assumptions, and for ease of explanation we shall again adopt a quite simple model, namely that the random forecasted arc cost disturbances are independent between links, with the disturbance for arc a following a Normal distribution with mean 0 and constant variance ν_a^2 ($a = 1, 2, \dots, n$).

5.3 Modelling user choice behaviour and habit

A final and very important element to the model is to specify the stochastic mechanism by which users are assumed to choose between the available paths for their user class. Since habit is a little tricky to represent, let us first consider a simpler case where there are no

habitual effects, and then secondly we will generalise this so as to be able to include habitual factors.

Now, users' choices are basically related to paths, and in general these paths may be physical or they may be fictitious (in a hyper-network sense, useful in some cases only) representing other choice dimensions such as mode, or whether to travel. The fraction of users of class i following path $k \in \mathcal{K}_i$ on day t is modelled through random variables with expected values given by choice probabilities. Let:

$$p_{i,k}^t = \text{probability a user of class } i \text{ chooses path } k \in \mathcal{K}_i \text{ on day } t$$

and let \mathbf{p}_i^t denote the vector of such probabilities for user class i on day t , with $\mathbf{1}^\top \mathbf{p}_i^t = 1$.

In the first, simplest model, we assume that \mathbf{p}_i^t depends only on the forecasted path costs $\Gamma_i \mathbf{z}^t$ for class i on day t , but not additionally on factors such as the choice a user actually made on the previous day. In such a case path choice behaviour can be modelled through a random utility model. Such a model assumes that each user in class i associates to each path $k \in \mathcal{K}_i$ a *perceived utility*, the vector of such perceived utilities modelled by a vector random variable that depends on the forecasted path costs $\Gamma_i \mathbf{z}^t$ and with a distribution parameterised by η_i (say). This distribution represents several sources of uncertainty from the perspective of both the users and the modeller. Under a random utility model, $p_{i,k}^t$ is then equal to the probability that path $k \in \mathcal{K}_i$ has the maximum perceived utility among all the paths in \mathcal{K}_i . To reflect these dependencies in the notation we may write:

$$\mathbf{p}_i^t = \zeta(\Gamma_i \mathbf{z}^t; \eta_i)$$

where the function ζ may be used to represent (for example) a logit or probit choice model (NB: in the case of logit, ζ is not itself a logit choice function with argument the systematic utility, yet we may represent a logit model by defining $\zeta_k(\mathbf{c}; \eta) = \exp(-\eta c_k)/(\sum_j \exp(-\eta c_j))$). Assuming that, conditionally on the forecasted costs, individuals choose independently and at random according to the probabilities above, then the random path flow vector \mathbf{H}_i^t for each class i on day t is distributed as:

$$\mathbf{H}_i^t \mid \mathbf{Z}^t = \mathbf{z}^t \sim \text{Multinomial}(d_i, \zeta(\Gamma_i \mathbf{z}^t; \eta_i)) \quad \text{independently for } i = 1, 2, \dots, N.$$

In the second, more complex model we consider how to generalise this approach to represent user habit. In the simple model above, the whole demand of d_i was treated the same (i.e. followed the same distribution), regardless of the choices made in time $t - 1$. However, in practice the fact that they made different choices yesterday may affect their prevalence to choose alternative routes today. Although we describe this generically as 'habit', in fact we may capture several real-life aspects of behaviour by considering such a dependency on past choices, not only habit in the sense of conservative behaviour in disliking change. For example, the converse to habit might be a desire for variety, which still suggests a form of dependence on past choices. Alternatively, the dependence on past choices might be through the information available to the user. In all such cases the probability to choose a path may depend additionally on previous choices and/or previous experiences, over and above the dependencies reflected in the predicted path cost vector.

Looking to the literature there have been several approaches proposed for capturing such effects (more than we mention here). In deterministic or stochastic threshold models,

users following boundedly rational behaviour may base their switching choice probabilities on the difference between (a) today's forecasted cost on their maximum utility alternative and (b) the actual or forecasted cost of the alternative chosen the previous day (Chang & Mahamassani, 1988). In so-called 'extra utility models', the path chosen on the previous day is given an extra utility, expressing the so-called transition cost to a different alternative (Cascetta and Cantarella, 1991). In addition to such features motivated by previous choices, there may also exist a kind of general inertia to change. For example, Cantarella & Cascetta (1995) proposed a simple model in which users each day have a constant probability $\alpha \in [0,1]$ of reconsidering yesterday's choices; those that do not reconsider simply follow the choice made on the previous day, regardless of any actual or predicted costs.

We thus aim to capture all such possible approaches through a generic approach, in which the choice probabilities of a user in class i at the start of day t depend on:

- the forecasted path costs $\Gamma_i \mathbf{z}^t$ for class i on day t ;
- the forecasted path costs $\Gamma_i \mathbf{z}^{t-1}$ for class i on day $t - 1$;
- the actual path costs $\Gamma_i \mathbf{c}^{t-1}$ experienced by class i on day $t - 1$;
- the path actually chosen by that user on day $t - 1$;
- a vector of parameters η_i relevant to class i .

To consider the dependence on the path previously chosen we first amend our consideration of path choice probability, to focus on the conditional path choice probabilities:

$$p_{i,k|j}^t = \text{probability a user of class } i \text{ chooses path } k \in \mathcal{K}_i \text{ on day } t \text{ given that they chose path } j \in \mathcal{K}_i \text{ on day } t - 1$$

and we collect these probabilities together across all paths for user class i in vectors $\mathbf{p}_{i|j}^t$ ($j \in \mathcal{K}_i$).

To reflect the dependencies of these conditional path choice probabilities on the factors identified above, we introduce functions ρ_j :

$$\mathbf{p}_{i|j}^t = \rho_j(\Gamma_i \mathbf{z}^t, \Gamma_i \mathbf{z}^{t-1}, \Gamma_i \mathbf{c}^{t-1}; \eta_i) \quad (j \in \mathcal{K}_i; i = 1, 2, \dots, N).$$

Conditional on the choices made by users on the previous day, we are now able to apply these distinct probabilities to each subset of class i , depending on the path they followed on day $t - 1$, yielding:

$$\mathbf{G}_{i|j}^t \mid \{\mathbf{Z}^t = \mathbf{z}^t, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}, \mathbf{C}^{t-1} = \mathbf{c}^{t-1}, \mathbf{H}_i^{t-1} = \mathbf{h}_i^{t-1}\} \sim \text{Multinomial}(h_{ij}^{t-1}, \rho_j(\Gamma_i \mathbf{z}^t, \Gamma_i \mathbf{z}^{t-1}, \Gamma_i \mathbf{c}^{t-1}; \eta_i)) \quad \text{independently for each } j \in \mathcal{K}_i; i = 1, 2, \dots, N$$

$$\mathbf{H}_i^t = \sum_{j \in \mathcal{K}_i} \mathbf{G}_{i|j}^t.$$

5.4 The transition function

Recall that \mathbf{c} denotes the vector of actual arc costs (actually experienced on a particular day), that \mathbf{z} denotes the vector of forecasted arc costs (predicted at the start of a particular day), and that \mathbf{h}_i denotes the vector of path flows for user class i ($i = 1, 2, \dots, N$), then we consider the following state variable:

$$\mathbf{x} = (\mathbf{c}, \mathbf{z}, \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N)$$

with mixed continuous/discrete state space $\mathcal{S} = \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{H}_1 \times \mathcal{H}_2 \times \dots \times \mathcal{H}_N$ where

$$\mathcal{H}_i = \{\mathbf{h}_i \in \mathbb{Z}^{m_i} : h_{ik} \geq 0 \ \forall k \in \mathcal{K}_i, \sum_{k \in \mathcal{K}_i} h_{ik} = d_i\} \quad (i = 1, 2, \dots, N).$$

For the most general model defined in sections 5.1–5.3, the transition function is a somewhat messy to write down, so let us first consider the simplified model presented at the start of section 5.3 in which they are not habitual factors.

Denoting the state we transform from (yesterday's state) by:

$$\mathbf{y} = (\mathbf{c}', \mathbf{z}', \mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_N)$$

we find:

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) &= \prod_{i=1,2,\dots,N} \{ \{d_i! / (h_{i1}! h_{i2}! \dots !)\} \prod_{k \in \mathcal{K}_i} \{\zeta_k(\Gamma_i \mathbf{z}; \boldsymbol{\eta}_i)\}^{h_{ik}} \} . \\ &\quad \prod_{a=1,2,\dots,n} \{2\pi\sigma_a)^{-0.5} \exp(-0.5((c_a - c_a(\Sigma_i \Gamma_i \mathbf{h}'_i; \boldsymbol{\mu}))/\sigma_a)^2) . \\ &\quad \prod_{a=1,2,\dots,n} \{2\pi\nu_a)^{-0.5} \exp(-0.5((z_a - \varphi_a(\mathbf{c}', \mathbf{z}'))/\nu_a)^2) \\ &\quad (\text{for } \mathbf{x}, \mathbf{y} \in \mathcal{S}; \boldsymbol{\theta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_N, \mathbf{d}, \mathbf{v}, \boldsymbol{\sigma})). \end{aligned}$$

Provided the random utility model $\zeta(\cdot)$ assigns positive probability to all alternatives (as is the case for regular random utility models – see Cantarella & Cascetta, 1995), and assuming well-behaved learning filter and cost functions, then the transition function assigns positive (one-step) probability mass/density across the whole state space, and so by the results of section 5.2 a unique stationary distribution is guaranteed to exist.

For the most general model presented in section 5.3, the explicit transition function is somewhat tedious to set out, primarily as we will have to consider a combinatorial problem for assessing the probability of today's path flows from combinations of transfers from routes chosen yesterday. But even without writing down the explicit function, it can be inferred that again, following the same regularity conditions, there will be some positive probability assigned to one-step transitions across the entire state-space, and so again we are guaranteed a unique stationary distribution.

6. CONCLUSIONS & FUTURE DIRECTIONS

Transportation system modelling has traditionally been dominated by modelling paradigms that postulate a stable, deterministically-predictable, unchanging world, or at least one in which change only occurs due to the intervention of a policy measure. However, the theoretical and practical tools now exist to model transportation systems as dynamic stochastic processes, providing the opportunity for our models to better reflect the real-world change and uncertainty that frequently occurs.

In our conclusions we would like to highlight several areas of practical and research interest where we believe this approach can make a particular contribution, and which warrant further research attention:

1. **Modelling variability** and uncertainty of various kinds **in an internally consistent manner**, while providing a complete statistical description. For instance, SP models provide estimates of variance in any relevant measures—such as travel times, flows or

benefits. They also allow us to understand the skewness in factors such as the distribution of travel times, which is important for computing measures of travel time reliability and in understanding user responses to the risk of arriving late at their destination. The variances produced are a composition of many internal elements of uncertainty and variability. SP models are also able to identify distributions with multiple modes, for which standard measures of central tendency and dispersion are not so useful, and which are known to arise theoretically in cases where conventional models would give multiple equilibria. In this case SP models also provide a “basin analysis” useful to understand how a given equilibrium may be reached, avoiding the others. Some initial work towards this objective may be found in Watling & Cantarella (2013), though there is considerable potential for further research to be done beyond this paper.

2. **Modelling transient situations.** SP models allows us to model the adaptive behaviour of travellers, the transient dynamics of the demand-supply interactions, and dynamics in the underlying internal and external driving forces. In this way, SP models allow us to look at transient situations, not just stationary systems, which is especially relevant for decision processes that do not adjust fast enough for an ‘equilibrium’ to be relevant. Analysis of transients becomes even more relevant when multiple equilibria may occur, to support “dynamic” policies (see previous comments).
3. **Modelling adaptation over multiple time scales.** The SP approach allows various decisions that travellers make to be reviewed and adapted over different embedded time-scales, e.g. mode choice reviewed less often than route choice. One reason that mobility choices (e.g. car licence/ownership, generation, distribution) and transport choices (mode, route, etc.) occur over different time-scales is that the socio-economic and demographic driving forces for mobility choices change over a different time-scale to the level-of-service driving factors for transport choices. This multi-scale adaptation may also be extended to the transport authorities and operators, who may (say) adapt traffic signal settings every week, while bus frequencies might only be adjusted every month.
4. **Incorporating alternative theories and levels of aggregation within a unified framework.** The approach is equally valid whether the various decision models and traffic model components are specified at the aggregate level, the individual level or some intermediate. The ability to model disaggregation down to individuals if needed may be particularly relevant, for example, to modelling ATIS and providing *personal* information to users. In addition, the approach provides a single consistent framework for incorporating and testing alternative theories of traffic flow, user behaviour or other factors.
5. **Supporting project appraisal/design in a variable environment,** e.g. selecting a project with small variance, defining a “dynamic” traffic control policy aimed at influencing the system toward a preferred stable regime. This connects to robust planning, where both the mean and variance of relevant factors are pertinent to the planning process. An important point is that SP variance is not neutral to policy decisions; we cannot just specify it exogenously, and then assume it to be constant, invariant to any policy measure.

A final point we would make is that the whole topic of inter-periodic dynamics should be carefully re-considered from the empirical point-of-view. While relatively limited, the empirical evidence that exists seems to support the existence of significant day-to-day dynamics in users' behaviour. However, stronger evidence is needed to support the assumed models of learning and habit. Also, further work is needed to consider the 'whole system' effect of change and 'shocks', and to understand the nature of seasonal and other forms of systematic and non-systematic sources of variability. Technological advances in tracing individual movements, and the widespread instrumentation of our highways, both promise a rich future source of data for a better understanding of such phenomena.

Acknowledgments: We would like to thank the anonymous reviewers, as well as colleagues at DTA2012, for constructive comments that helped us to improve an earlier version of this paper. This work was partially supported by UNISA local grant ORSA091208 (financial year 2009) and ORSA118135 (financial year 2011), and by UK EPSRC grant refs. EP/I00212X/1 (2011-12) and EP/I00212X/2 (2012-16). The financial assistance of Prof Terry Friesz is also gratefully acknowledged, in supporting the visit of the first-named author to deliver the keynote paper at the DTA 2012 International Symposium on Dynamic Traffic Assignment (Martha's Vineyard, USA), upon which the present paper is based.

References

- Bie J. & Lo H.K. (2010). Stability and attraction domains of traffic equilibria in a day-to-day dynamical system formulation. *Transportation Research Part B* 44, 90–107.
- Balijepalli, N.C. and Watling, D.P. (2005). Doubly Dynamic Equilibrium Approximation Model for Dynamic Traffic Assignment. In: *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction* (ed.: H.Mahmassani), Elsevier, Oxford, UK, 741-760.
- Balijepalli N.C., Watling D.P. & Liu R. (2007). Doubly Dynamic Traffic Assignment: Simulation Modelling Framework and Experimental Results. *Transportation Research Record* 2029, 39-48.
- Bell M.G.H. (2000). A game theory approach to measuring the performance reliability of transport networks. *Transportation Research B* 34, 533-545.
- Ben-Tal A., Chung B.D., Mandala S.R. & Yao T. (2011). Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation Research B* 45, 1177-1189.
- Cantarella G.E & Cascetta E. (1995). Dynamic Process and Equilibrium in Transportation Networks: Towards a Unifying Theory. *Transportation Science* 29, 305-329.
- Cantarella G.E. & Watling D.P. (2013). Modelling road traffic assignment as a day-to-day dynamic process: State-of-the-art and research perspectives. Submitted for publication.
- Cascetta E. (1987) Static and dynamic models of stochastic assignment to transportation networks. In: *Flow control of congested networks* (eds: Szaego G., Bianco L. & Odoni A.), Springer Verlag, Berlin.
- Cascetta E. (1989). A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research Part B* 23, 1-17.
- Cascetta E. (2009). *Transportation Systems Analysis: Models and Applications*. Springer.
- Cascetta E. & Cantarella G.E. (1991). A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research Part A* 25, 277-291.

- Cascetta E. & Cantarella G.E. (1993). Modelling Dynamics in Transportation Networks. *Journal of Simulation Practice and Theory* 1, 65-91.
- Chang G.L. & Mahmassani H.S. (1988). Travel time prediction and departure time adjustment behaviour dynamics in a congested traffic system, *Transportation Research Part B* 22, 217-232.
- Chen R. & Mahmassani H.S. (2004). Travel Time Perception and Learning Mechanisms in Traffic Networks. *Transportation Research Record* 1894, 209-221.
- Dafermos S.C. (1980). Continuum modelling of transportation networks. *Transportation Research Part B* 14, 295-301.
- Daganzo C.F. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B* 28, 269-287.
- Davis G. & Nihan N. (1993). Large population approximations of a general stochastic traffic assignment model. *Operations Research* 41, 169-178.
- De Palma A., Picard N. & Andrieu L. (2009). Risk in Transport Investments. *Networks & Spatial Economics*, published on-line at DOI: 10.1007/s11067-009-9109-8.
- Di Domenica N., Mitra G., Valente P. & Birbilis G. (2007). Stochastic programming and scenario generation within a simulation framework: An information systems perspective. *Decision Support Systems* 42, 2197-2218.
- Friesz T.L., Bernstein D. & Kydes N. (2004). Dynamic Congestion Pricing in Disequilibrium. *Networks & Spatial Economics* 4, 181-202.
- Friesz T.L., Mookherjee R. & Yao T. (2008). Securitizing congestion: The congestion call option. *Transportation Research B* 42, 407-437.
- Ghosh B., Basu B. & O'Mahony M. (2007). Bayesian Time-Series Model for Short-Term Traffic Flow Forecasting. *Journal of Transportation Engineering* 133, 180-189.
- Goodwin, P.B. (1998). The end of equilibrium. In: Theoretical foundations of travel choice modelling (eds: Garling T., Laitila T. & Westin K.), Elsevier, Amsterdam.
- Guo X. & Liu H. (2010). Bounded rationality and irreversible network change. *Transportation Research B* 45, 1606-1618.
- Guo X. & Liu H. (2012) A day-to-day dynamic model in discrete/continuum transportation networks. *Transportation Research Record*, in press.
- Han L. & Du L. (2012). On a link-based day-to-day traffic assignment model. *Transportation Research Part B* 46, 72-84.
- Hazelton M.L. & Watling D.P. (2004). Computation of equilibrium distributions of Markov traffic assignment models. *Transportation Science* 38, 331-342.
- He X. & Liu H.X. (2012). Modeling the day-to-day traffic evolution process after an unexpected network disruption. *Transportation Research Part B* 46, 50-71.
- Ho H.W. & Wong S.C. (2007). Housing allocation problem in a continuum transportation system. *Transportmetrica* 3, 21-39.
- Hofbauer J. & Sandholm W.H. (2007). Evolution in games with randomly disturbed payoffs. *Journal of Economic Theory* 132, 47-69.
- Hoogendoorn S.P. & Bovy P.H.L. (2004). Dynamic user-optimal assignment in continuous time and space. *Transportation Research Part B* 38, 571-592.
- Horowitz J.L. (1984). The stability of stochastic equilibrium in a two-link transportation network. *Transportation Research Part B* 18, 13-28.
- Hu T.-Y. & Mahmassani H.S. (1997). Day-to-day evolution of network flows under real-time information and reactive signal control. *Transportation Research C* 5, 51-69.

- Huang L., Wong S.C., Zhang M., Shu C.W. & Lam W.H.K. (2009). A reactive dynamic user equilibrium model for pedestrian flows: A continuum modeling approach. *Transportation Research Part B* 43, 127-141.
- Iida Y., Akiyama T. & Uchida T. (1992). Experimental analysis of dynamic route choice behaviour. *Transportation Research Part B* 26, 17-32.
- Leurent F.M. (1998). Sensitivity and error analysis of the dual criteria traffic assignment model. *Transportation Research Part B* 32, 189-204.
- Lim Y., Lee S. J. & Kim J. (2007). A Stochastic Process Model for Daily Travel Patterns and Traffic Information. *Agent and Multi-agent Systems: Technologies and Applications* 4496, 102-110.
- Liu R., Van Vliet D. & Watling D. (2006). Microsimulation models incorporating both demand and supply side dynamics. *Transportation Research* 40A(2), 125-150.
- Okutani I. & Stephanedes Y.J. (1984). Dynamic prediction of traffic volume through Kalman Filtering Theory. *Transportation Research Part B* 18, 1-11.
- Pfaffenbichler P, Emberger G & Shepherd S (2008). The integrated dynamic land use and transport model MARS. *Networks & Spatial Economics* 8, 183-200.
- Rickert M., Nagel K. & Schreckenberg M. (1996). Two lane traffic simulation using cellular automata. *Physica A* 231, 534-550.
- Sheffi Y. (1985). *Urban transportation networks*. Prentice Hall, New Jersey.
- Shepherd SP, Zhang X, Emberger G, May AD, Hudson M & Paultey N (2006). Designing optimal urban transport strategies: the role of individual policy instruments and the impact of financial constraints. *Transport Policy* 13, 49-65.
- Smith M. J. (1984). The stability of a dynamic model of traffic assignment: An application of a method of Lyapunov. *Transportation Science* 18, 245-252.
- Stokey N.L. & Lucas R.E. (1989). *Recursive methods in economic dynamics*. Harvard University Press, Cambridge, Massachusetts.
- Vlahogiannia E.I., Golias J.C. & Karlaftisa M.G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews* 24, 533-557.
- Watling D.P. (1996). Asymmetric problems and stochastic process models of traffic assignment. *Transportation Research Part B* 30, 339-357.
- Watling D.P. (1999). Stability of the stochastic equilibrium assignment problem: a dynamical systems approach. *Transportation Research Part B* 33, 281-312.
- Watling D.P. & Cantarella G.E. (2013). Modelling sources of variation in transportation systems: theoretical foundations of day-to-day dynamic models. *Transportmetrica B: Transport Dynamics* 1(1), 3-32.
- Wegener (2004). Overview of land-use transport models. In: *Transport Geography and Spatial Systems* (eds Hensher D.A. & Button K.), Pergamon, Kidlington, UK, 127-146.
- Wilson N.H.M. & Nuzzolo A. (2009). *Schedule-Based Modeling of Transportation Networks*. Springer, New York.
- Wong S.C. (1998). Multi-commodity traffic assignment by continuum approximation of network flow with variable demand. *Transportation Research Part B* 32, 567-581.
- Yang F. & Zhang D. (2009). Day-to-day stationary link flow pattern. *Transportation Research Part B* 43, 119-126.
- Yang F. & Liu H.X. (2007). A New Modeling Framework for Travelers' Day-to-Day Route Adjustment Processes. In: *Transportation and Traffic Theory* (eds: Allsop R.E., Bell M.G.H., & Heydecker, B.G.), Elsevier, 813-837.
- Zhao T., Sundararajan S.K. & Tseng C.-L. (2004). Highway development decision-making under uncertainty: A real options approach. *Journal of Infrastructure Systems* 10, 23-32.