Asymptotic Approximations of Transient Behaviour for Day-to-Day Traffic Models

David P. Watling[a], Martin L. Hazelton[b,*]

[a]*Institute for Transport Studies, University of Leeds, U.K.*

[b]*Institute of Fundamental Sciences, Massey University, New Zealand*

Abstract

We consider a wide class of stochastic process traffic assignment models that capture the day-to-day evolving interaction between traffic congestion and drivers' information acquisition and choice processes. Such models provide a description of not only transient change and 'steady' behaviour, but also represent additional variability that occurs through probabilistic descriptions. They are therefore highly suited to modelling both the disturbance and subsequent 'drift' of networks that are subject to some systematic change, be that a road closure or capacity reduction, new policy measure or general change in demand patterns. In this paper we derive analytic results to probabilistically capture the nature of the transient effects following such a systematic change. This can be thought of as understanding what happens as a system moves from varying about one equilibrium state to varying about a new equilibrium state. The results capture analytically the changes over time in descriptors of the system, in terms of link flow means, variances and covariances. Formally, the analytic results hold asymptotically as approximations, as we imagine demand increasing in tandem with capacities; however, our interest is in general cases where such tandem increases do not occur, and so we provide conditions under which our approximations are likely to work well. Numerical results of applying the methods are reported on several examples. The quality of the approximations is assessed through comparisons with Monte Carlo simulations from the true underlying process.

Keywords: Markov process, network change, route choice, stochastic process, Stochastic User Equilibrium, transportation network

1. Introduction

There is substantial evidence that real-life transport networks are subject to considerable variation, disturbance and change. Major efforts have been made in recent years to reflect such aspects in the models used to predict, control and analyse such networks. These efforts have advanced the standard practice of 'comparative statics', whereby static network equilibria are compared before and after some systematic change in the model inputs. In particular, we have seen substantial advances in modelling:

*Corresponding author

Email addresses: d.p.watling@its.leeds.ac.uk (David P. Watling), m.hazelton@massey.ac.nz (Martin L. Hazelton)

- *Dynamic Traffic Assignment*, in which a consistent treatment is made of time-varying flows and travel delays, aiming to reflect the typical within-day variations of demands and congestion patterns (e.g. Peeta and Ziliaskopoulos 2001, Szeto and Wong 2012).

- *Network Performance Variability*, where elements of the transportation system are represented through (steady state) random variables, aiming to capture phenomena such as between-day variations in demands, capacities and travel times (e.g. Castillo et al. 2014, Nakayama and Watling 2014).

- *Network Reliability*, a broad field in which the risk-averse strategies of either travellers (e.g. in their route or departure time decisions) or planners (e.g. in their decisions regarding capacity allocation) are represented, in the face of unreliable performance on a route-, trip-, or network-level (e.g. Bell and Cassir 2000, Lam et al. 2014, Chen et al. 2011).

- *Dynamic Processes*, in which the trip-to-trip learning process of drivers is explicitly modelled, allowing the study of the stability of point equilibrium solutions as well as other kinds of emergent behaviour, a rapidly growing field as evidenced by the literature in the recent reviews of deterministic Cantarella and Watling (2016) and stochastic process models Watling and Cantarella (2015).

In the present paper we will be addressing the issue of network performance variability through the use of a dynamic process of drivers' trip-to-trip learning, so in a sense we simultaneously address two of the research areas above. This is achieved by the use of a (day-to-day dynamic) stochastic process model, for which there is now a growing literature (e.g. Cascetta 1989, Cantarella and Cascetta 1995, Hazelton 2002, Hazelton and Watling 2004, Watling and Cantarella 2013, Parry et al. 2016). This focus on trip-to-trip learning and adaptation of route choice over days is distinctive from other models that consider the junction-by-junction adaptation of travellers as they traverse a network on a particular day (e.g. Boyer et al. 2015), or models that represent stochastic queuing networks without route choice (e.g. Flötteröd and Osorio 2017), or models that represent route adaptation by a continuous-time adaptation (e.g. Zhang et al. 2015, Smith and Watling 2016).

The particularly distinctive feature of our work is that we will focus on the *transient* stage of such a day-to-day dynamic process, as it adjusts between (stochastically) stable regimes, following some systematic network change to long-run capacities, tolls, demands or some policy measure. A particular practical motivation is the increased incidence of major innovations in mobility services (such as ride ride-sharing), typically implemented over a short time frame and leading to marked changes in patterns of traffic flow. While there has been an analysis of transient phenomena in other forms of traffic or queuing system (e.g. Huang et al. 2010, Jabari and Liu 2013, Osorio and Yamani 2017), we are not aware of any theoretical work existing on transient stochastic processes of day-to-day dynamic route choice in the transportation literature.

Although there is paucity of theoretical work on this issue of transience in a day-to-day dynamic context, there do exist a handful of empirical studies to stimulate our analysis. In particular Zhu et al. (2010) analysed several sources of data for evidence of the traffic and behavioural impacts of the I-35W bridge collapse in Minneapolis. Most pertinent to the present paper is their location-specific analysis of link flows at 24 locations. By computing the root mean square difference in flows between successive weeks, and comparing the trend for 2006 with that for 2007 (the latter with the bridge collapse), they observed an apparent transient impact of the bridge collapse. They also showed there was no statistically-significant evidence of a difference in the pattern of flows in the period September–November 2007 (a period starting 6 weeks after the bridge collapse), when compared with the corresponding period in 2006. They suggested that this was indicative of the length of a re-equilibration process in a conceptual sense.

A second such empirical study is that reported in Watling et al. (2012). They analysed the impacts of two capacity interventions in the city of York, one a bridge closure and the other a capacity reduction for maintenance works. Through registration plate surveys conducted for a series of days before, during and after such interventions, the aim was to separate ambient daily variations from systematic changes, and they were indeed able to identify statistically significant impacts in this way. Calibrating an equilibrium model to the 'before' data only, for the case of the capacity reduction, the model was found to be broadly successful in predicting the impacts as seen in the 'after' data, with route choice in the neighbourhood of the intervention seemingly re-stabilising extremely quickly (i.e. the transience seemed very short).

From these two empirical studies, it is relevant to ask to what extent the length and nature of a transient period may be specific to the typical network conditions, level of ambient variation and the nature/level of the systematic change to the network. In principle day-to-day dynamic stochastic process models are very well suited to this task. However, the analysis of the properties of such models can present formidable challenges. A major breakthrough was provided by Davis and Nihan (1993), who focused on approximations derived (in essence) from an asymptotic regimen in which travel demand and network capacity increase in tandem. (This is of course just a mathematical device: in practice we will apply the results to general and non-asymptotic cases where such tandem increases do not occur). They showed that a vast range of stochastic day-to-day models can be approximated by a form of (discrete time) Gaussian autoregressive process.

In theory these Gaussian processes can provide an excellent approximation to the properties of a more general stochastic model, both when the process is following its stationary distribution and also during transient periods. What is more, because the (multivariate) Gaussian distribution is specified by its mean vector and covariance matrix, the dynamics of the process are completely captured by the temporal variation of these quantities. Davis and Nihan (1993) described the evolution of the mean as a nonlinear process, and described the dynamics of the covariance matrix as an iterative updating scheme written in terms of Jacobians of cost and probability functions. However, these Jacobians need to be recalculated at every iteration. This will be computationally expensive for large networks, since the Jacobians are square matrices of dimension equal to the number of routes. Perhaps more importantly, the nonlinearity of the mean process and the temporal inhomogeneity of the Jacobians significantly reduces the mathematical tractability of the approximation model, and hence its utility for theoretical analyses.

In order for the approximation to work over the entire space of feasible route flows, the need to update the Jacobians is unavoidable. However, we show that by using a single pair of time homogeneous Jacobian matrices, it is nevertheless possible to provide an accurate Gaussian process approximation that works within a relatively large neighbourhood of stochastic user equilibrium. It follows that our approximation has the capacity to describe transient properties of the underlying stochastic model over a range of states in a highly convenient manner. In essence our methodology works because the asymptotic order of the approximation error for the Jacobian matrices differs from the order of the purely stochastic variation. As a consequence, the fixed Jacobians continue to be applicable at flow patterns that differ from the mean of the stationary distribution by more than random variation; in other words, during transient periods.

To facilitate exposition we develop our results incrementally, beginning with simple types of network and route choice model. Section 2 introduces the notation and basic model elements, and then gives our main theoretical results (with proofs) for networks with a single origin-destination pair, and in which travellers' route choices arise from a simple exponential learning model. Section 3 extends these results to general networks and learning mechanisms. Illustrative numerical results are presented in Section 4. Finally, section 5 contains conclusions and directions for future research.

2. Modelling Framework and Initial Results

Consider a transport network with m origin-destination (OD) pairs. The jth of these is serviced by n_j routes. We model the flow on these routes over a sequence of disjoint time periods indexed by t, which we will often refer to as a 'day' although it need not correspond to a full 24 hour period. The total traffic volume on route j at day t is denote X_j^t, and the volumes on all routes is concatenated into the vector $\boldsymbol{X}^t$. The traffic at time t generates a vector of route-specific travel costs $\boldsymbol{c}$.

At day t, travellers base their route choices on a disutility that is defined in terms of costs of route costs and disutilities over a history of τ earlier days. Let u_j^t denote the (measured) disutility of route j at time t, and let $\boldsymbol{u}^t$ be the corresponding vector of disutilities. Based on the disutilities $\boldsymbol{u}^t$, each traveller makes a route choice at time t. It is assumed that these choices are independent (conditional on the disutility), with the probability of selecting route j at time t being denoted p_j^t. These route choice probabilities are generated by a vector-valued route choice probability function $\boldsymbol{p}(\boldsymbol{u}^t)$.

To facilitate exposition, we will focus initially on networks with a single origin-destination (OD) pair. We write ζ for the travel demand for that pair. This is assumed to be constant through time. However, our model can incorporate variable realized demand by introducing a dummy route that corresponds to the decision not to travel. If the number of potential travellers is large but the probability of travelling is relatively small, then our model can also approximate closely alternative models with Poisson demand.

Define the vector of standardised route flows by $\boldsymbol{x}^t = \zeta^{-1}\boldsymbol{X}^t$. In order to obtain tractable mathematical results using limiting theorems from probability and statistics, we follow Davis and Nihan (1993) and consider an asymptotic regimen in which $\zeta \to \infty$. Under such a process we can expect $\boldsymbol{x}^t$ to converge to a finite deterministic limit courtesy of the Law of Large Numbers. In order to ensure that the route costs and disutilities remain finite as $\zeta \to \infty$, we assume that the former are function of the standardised flows. That is, $\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{x})$. This is equivalent to assuming that the capacity of the network increases in proportion to the travel demand.

Consider for now a relatively simple exponential learning process, in which disutilities are updated each day based upon experience from the previous day according to

$$\boldsymbol{u}^t = \beta\boldsymbol{c}(\boldsymbol{x}^{t-1}) + (1-\beta)\boldsymbol{u}^{t-1} . \tag{1}$$

These disutilities then give rise to a vector of route choice probabilities $\boldsymbol{p}^t = \boldsymbol{p}(\boldsymbol{u}^t)$. This can be motivated in terms of each traveller minimizing his/her perceived disutility, where the perceptual variation is modelled by adding a vector of subject-specific random variables to $\boldsymbol{u}^t$. Assuming that each traveller at time t makes a route choice independent of the choices of the other travellers at that time, the random vector of unstandardized route flows follows a multinomial distribution conditional on the current vector of disutilities. That is,

$$\boldsymbol{X}^t | \boldsymbol{u}^t \sim \mathsf{Mn}(\zeta, \boldsymbol{p}(\boldsymbol{u}^t)) \ .$$

The state $\boldsymbol{s}^t$ of the system at time t is defined by

$$\boldsymbol{s}^t = \begin{pmatrix} \boldsymbol{u}^t \\ \boldsymbol{x}^t \end{pmatrix} \ .$$

Under our modelling assumptions, $\{\boldsymbol{s}^t \colon t = 0, 1, 2, \ldots\}$ is a Markov process with initial state $\boldsymbol{s}^0$ (e.g. Davis and Nihan 1993). For a given demand parameter ζ we denote the state space by $\mathcal{S}_\zeta$. To accommodate limiting behaviour as $\zeta \to \infty$, we define $\mathcal{S} = \cup_{\zeta=1}^{\infty} \mathcal{S}_\zeta$. This space can be decomposed as $\mathcal{S} = \mathcal{S}^{\boldsymbol{u}} \times \mathcal{S}^{\boldsymbol{x}}$, where $\mathcal{S}^{\boldsymbol{u}}$ and $\mathcal{S}^{\boldsymbol{x}}$ denote subspaces corresponding to the coordinates of $\boldsymbol{s}$ indicated by the superscripts. The space $\mathcal{S}_{\boldsymbol{x}}$ is compact, and so $\mathcal{S}_{\boldsymbol{u}}$ and hence $\mathcal{S}$ are likewise compact assuming (as we do henceforth) that the cost function $\boldsymbol{c}$ is continuous.

The Markov chain process $\{\boldsymbol{s}^t\}$ will have a unique stationary distribution if it is regular. A sufficient condition for regularity is provided in Lemma 1 below, which is due to Davis and Nihan (1993) (Proposition 1). We note that this result also holds for the systems with multiple OD movements and more general utility specifications considered later in this paper.

Lemma 1. [Davis and Nihan, 1993]
Assume

(A1) $0 < \boldsymbol{p}(\boldsymbol{u})$ *for all* $\boldsymbol{u} \in \mathcal{S}_{\boldsymbol{u}}$.

Then the process $\{\boldsymbol{s}^t\}$ *is regular and hence has a unique stationary distribution.*

Remark 1: Assumption (A1) ensures that all routes retain a non-zero probability of being chosen, regardless of the variations in disutility. This property is common. For example, it applies to logit and Probit route choice models.

We let $\boldsymbol{s}^*$ denote the mean of the stationary distribution. This can be partitioned into means for the underlying route flow and disutility components according to $\boldsymbol{s}^* = \left[\begin{smallmatrix} \boldsymbol{u}^* \\ \boldsymbol{x}^* \end{smallmatrix}\right]$. In developing our asymptotic approximations as $\zeta \to \infty$, we note that the pattern of route flows will become increasingly concentrated about the mean courtesy of the Law of Large Numbers. We therefore seek to linearize the dynamics of the process about $\boldsymbol{s}^*$. Assuming that all second derivatives of the route cost function $\boldsymbol{c}$ and the probability function $\boldsymbol{p}$ are continuous, we have

$$\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x}^*) + B(\boldsymbol{x} - \boldsymbol{x}^*) + O(||\boldsymbol{x} - \boldsymbol{x}^*||^2)$$

where B is Jacobian matrix for $\boldsymbol{c}$ evaluated at $\boldsymbol{x}^*$, and

$$\boldsymbol{p}(\boldsymbol{u}) = \boldsymbol{p}(\boldsymbol{u}^*) + D(\boldsymbol{u} - \boldsymbol{u}^*) + O(||\boldsymbol{u} - \boldsymbol{u}^*||^2)$$

where D is Jacobian matrix for $\boldsymbol{p}$ evaluated at $\boldsymbol{u}^*$. In these equations and elsewhere in this paper, the $O(\cdot)$ order terms apply elementwise when added to vectors or matrices. For example, a vector $\boldsymbol{v} = O(a)$ indicates that $\lim_{a \to 0} v_i / a < \infty$ for all coordinates v_i. We define

$$M = \begin{pmatrix} (1-\beta)I & \beta B \\ (1-\beta)D & \beta DB \end{pmatrix}. \tag{2}$$

We are now in a position to give our major results as they apply to a system with a single OD pair (assumed to be connected by two or more paths so as to avoid the trivial instance of a system with no route choice).

Theorem 1. *Define $\rho = ||\boldsymbol{s}^0 - \boldsymbol{s}^*||$. Assume (A1) and*

(A2) All eigenvalues of M have modulus less than one.

(A3) All second derivatives of $\boldsymbol{c}$ and $\boldsymbol{p}$ are bounded on $\mathcal{S}^{\boldsymbol{x}}$ and $\mathcal{S}^{\boldsymbol{u}}$ respectively.

(A4) $\rho\zeta^{1/4} \to 0$ as $\zeta \to \infty$.

Define $\boldsymbol{\mu}^t = \mathsf{E}[\boldsymbol{s}^t]$. Then as $\zeta \to \infty$,

$$\zeta^{1/2}(\boldsymbol{s}^t - \boldsymbol{\mu}^t) \xrightarrow{L} \mathsf{N}(\boldsymbol{0}, \Sigma^t)$$

where $\xrightarrow{L}$ indicates convergence in law (i.e. distribution). The mean vector satisfies

$$\lim_{\zeta\to\infty} \zeta^{1/2} \left\{(\boldsymbol{\mu}^t - \boldsymbol{s}^*) - M(\boldsymbol{\mu}^{t-1} - \boldsymbol{s}^*)\right\} = \boldsymbol{0} \tag{3}$$

and the covariance matrix evolves according to

$$\Sigma^t = M\Sigma^{t-1}M^\mathsf{T} + V \tag{4}$$

where

$$V = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \mathsf{diag}(\boldsymbol{x}^*) - \boldsymbol{x}^*\boldsymbol{x}^{*\mathsf{T}} \end{pmatrix}. \tag{5}$$

Proof

Partition the mean as $\boldsymbol{\mu}^t = \begin{pmatrix} \boldsymbol{\mu}^t_{\boldsymbol{u}} \\ \boldsymbol{\mu}^t_{\boldsymbol{x}} \end{pmatrix}$ using the obvious subscript notation. Then

$$\begin{aligned} \boldsymbol{\mu}^t_{\boldsymbol{x}} = \mathsf{E}[\boldsymbol{x}^t] &= \mathsf{E}[\mathsf{E}[\boldsymbol{x}^t|\boldsymbol{s}^{t-1}]] \\ &= \mathsf{E}\left[\boldsymbol{p}\left(\beta\boldsymbol{c}(\boldsymbol{x}^{t-1}) + (1-\beta)\boldsymbol{u}^{t-1}\right)\right] \\ &= \boldsymbol{p}\left(\beta\boldsymbol{c}(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}}) + (1-\beta)\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}}\right) + O(\zeta^{-1}) \end{aligned}$$

by a standard application of the delta method. Linearizing the functions $\boldsymbol{p}$ and $\boldsymbol{c}$ we obtain

$$\begin{aligned} \boldsymbol{\mu}^t_{\boldsymbol{x}} &= \boldsymbol{p}(\boldsymbol{u}^*) + D\left(\beta\boldsymbol{c}(\boldsymbol{x}^*) + \beta B(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^*) + (1-\beta)\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} - \boldsymbol{u}^*\right) + O(\rho^2 + \zeta^{-1}) \\ &= \boldsymbol{p}(\boldsymbol{u}^*) + \beta DB(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^*) + (1-\beta)D(\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} - \boldsymbol{u}^*) + \beta(\boldsymbol{c}(\boldsymbol{x}^*) - \boldsymbol{u}^*) + O(\rho^2 + \zeta^{-1}). \end{aligned} \tag{6}$$

The appearance of the additional $O(\rho^2)$ term in the remainder follows immediately when $t = 1$ from the definition of ρ and the smoothness conditions. It also applies for $t > 1$ by iteration, noting that the eigenvalue condition on M avoids the remainder term blowing up as t becomes large.

Turning to the disutility,

$$\begin{aligned} \boldsymbol{\mu}^t_{\boldsymbol{u}} = \mathsf{E}[\boldsymbol{u}^t] &= \mathsf{E}\left[\mathsf{E}[\boldsymbol{u}^t|\boldsymbol{s}^{t-1}]\right] \\ &= \mathsf{E}\left[\beta\boldsymbol{c}(\boldsymbol{x}^{t-1}) + (1-\beta)\boldsymbol{u}^{t-1}\right] \\ &= \beta\boldsymbol{c}(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}}) + (1-\beta)\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} + O(\zeta^{-1})\ . \end{aligned} \tag{7}$$

Applying the same arguments as above, it follows that

$$\begin{aligned} \boldsymbol{\mu}^t_{\boldsymbol{u}} &= \beta\boldsymbol{c}(\boldsymbol{x}^*) + \beta B(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^*) + (1-\beta)\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} + O(\rho^2 + \zeta^{-1}) \\ &= \boldsymbol{c}(\boldsymbol{x}^*) + \beta B(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^*) + (1-\beta)(\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} - \boldsymbol{c}(\boldsymbol{x}^*)) + O(\rho^2 + \zeta^{-1}). \end{aligned} \tag{8}$$

When the process is stationary $\boldsymbol{\mu}^{t-1} = \boldsymbol{\mu}^t = \boldsymbol{s}^*$. It then follows from (7) that

$$\boldsymbol{u}^* = \boldsymbol{c}(\boldsymbol{x}^*) + (1-\beta)(\boldsymbol{u}^* - \boldsymbol{c}(\boldsymbol{x}^*)) + O(\zeta^{-1})$$

and hence

$$\boldsymbol{u}^* = \boldsymbol{c}(\boldsymbol{x}^*) + O(\zeta^{-1}). \tag{9}$$

Also,

$$\boldsymbol{x}^* = \boldsymbol{p}(\boldsymbol{u}^*) + O(\zeta^{-1}) \tag{10}$$

by another application of the delta method.

Substituting equations (9) and (10) into (6) and subtracting $\boldsymbol{x}^*$ from both sides gives

$$\boldsymbol{\mu}^t_{\boldsymbol{x}} - \boldsymbol{x}^* = \beta DB(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^*) + (1-\beta)D(\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} - \boldsymbol{u}^*) + O(\rho^2 + \zeta^{-1}).$$

Similarly, from equations (8), (9) and (10) we get

$$\boldsymbol{\mu}^t_{\boldsymbol{u}} - \boldsymbol{u}^* = \beta B(\boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^*) + (1-\beta)(\boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} - \boldsymbol{u}^*) + O(\rho^2 + \zeta^{-1}).$$

Combining these two results gives

$$\begin{pmatrix} \boldsymbol{\mu}^t_{\boldsymbol{u}} - \boldsymbol{u}^* \\ \boldsymbol{\mu}^t_{\boldsymbol{x}} - \boldsymbol{x}^* \end{pmatrix} = \begin{pmatrix} (1-\beta)B & \beta B \\ (1-\beta)DB & \beta DB \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}^{t-1}_{\boldsymbol{u}} - \boldsymbol{u}^* \\ \boldsymbol{\mu}^{t-1}_{\boldsymbol{x}} - \boldsymbol{x}^* \end{pmatrix} + O(\rho^2 + \zeta^{-1}).$$

Collecting terms on the left-hand side and multiplying through by $\zeta^{1/2}$ gives

$$\zeta^{1/2}\left\{(\boldsymbol{\mu}^t - \boldsymbol{s}^*) - M(\boldsymbol{\mu}^{t-1} - \boldsymbol{s}^*)\right\} = O(\zeta^{1/2}\rho^2 + \zeta^{-1/2})$$

when equation (3) follows courtesy of assumption (A4).

Turning to the covariance matrix,

$$\mathsf{Var}(\boldsymbol{s}^t) = \mathsf{Var}\left(\mathsf{E}[\boldsymbol{s}^t|\boldsymbol{s}^{t-1}]\right) + \mathsf{E}\left[\mathsf{Var}(\boldsymbol{s}^t|\boldsymbol{s}^{t-1})\right]. \tag{11}$$

The conditional expectation $\mathsf{E}[\boldsymbol{s}^t|\boldsymbol{s}^{t-1}]$ is a smooth function of $\boldsymbol{s}^{t-1}$ and hence the first term on the right-hand side is amenable to the delta method. In more detail,

$$\begin{aligned} \mathsf{E}[\boldsymbol{s}^t|\boldsymbol{s}^{t-1}] &= \begin{pmatrix} \beta\boldsymbol{c}(\boldsymbol{x}^{t-1}) + (1-\beta)\boldsymbol{u}^{t-1} \\ \beta\boldsymbol{p}(\boldsymbol{c}(\boldsymbol{x}^{t-1}) + (1-\beta)\boldsymbol{u}^{t-1}) \end{pmatrix} \\ &= \begin{pmatrix} \beta B\boldsymbol{x}^{t-1} + (1-\beta)\boldsymbol{u}^{t-1} \\ \beta DB\boldsymbol{x}^{t-1} + (1-\beta)D\boldsymbol{u}^{t-1} \end{pmatrix} + O(\zeta^{-1}) \\ &= M\boldsymbol{s}^{t-1} + O(\zeta^{-1}). \end{aligned}$$

Hence

$$\mathsf{Var}(\mathsf{E}[\boldsymbol{s}^t|\boldsymbol{s}^{t-1}]) = M\mathsf{Var}(\boldsymbol{s}^{t-1})M^\mathsf{T} + O(\zeta^{-2}) \tag{12}$$

where the order of the remainder comes from noting that $\mathsf{Var}(\boldsymbol{s}^{t-1}) = O(\zeta^{-1})$.

For the second term on the right-hand side of equation (11), observe that

$$\begin{aligned} \mathsf{Var}(\boldsymbol{s}^t|\boldsymbol{s}^{t-1}) &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathsf{Var}(\boldsymbol{x}^t|\boldsymbol{u}^t) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \zeta^{-1}\mathsf{diag}(\boldsymbol{p}(\boldsymbol{u}^t)) - \zeta^{-1}\boldsymbol{p}(\boldsymbol{u}^t)\boldsymbol{p}(\boldsymbol{u}^t)^\mathsf{T} \end{pmatrix} \end{aligned}$$

using the fact that $\boldsymbol{u}^t$ is a deterministic function of $\boldsymbol{s}^{t-1}$ and hence $\mathsf{Var}(\boldsymbol{u}^t|\boldsymbol{s}^{t-1})$ is the zero matrix. The form of the block corresponding to $\mathsf{Var}(\boldsymbol{x}^t|\boldsymbol{s}^{t-1})$ follows from the properties of the multinomial distribution.

Now

$$\begin{aligned}\mathsf{E}\left[\mathsf{Var}(\boldsymbol{s}^t|\boldsymbol{s}^{t-1})\right] &= \zeta^{-1}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathsf{diag}(\mathsf{E}[\boldsymbol{p}(\boldsymbol{u}^t)]) - \mathsf{E}[\boldsymbol{p}(\boldsymbol{u}^t)\boldsymbol{p}(\boldsymbol{u}^t)^{\mathsf{T}}]\end{pmatrix}\\ &= \zeta^{-1}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathsf{diag}(\boldsymbol{p}(\boldsymbol{\mu}_{\boldsymbol{u}}^t)) - \boldsymbol{p}(\boldsymbol{\mu}_{\boldsymbol{u}}^t)\boldsymbol{p}(\boldsymbol{\mu}_{\boldsymbol{u}}^t)^{\mathsf{T}}\end{pmatrix} + O(\zeta^{-2})\end{aligned}$$

by further applications of the delta method. Using earlier results on the evolution of $\boldsymbol{\mu}_{\boldsymbol{u}}^t$, and taking account of the continuity of $\boldsymbol{p}$, this leads to

$$\begin{aligned}\mathsf{E}\left[\mathsf{Var}(\boldsymbol{s}^t|\boldsymbol{s}^{t-1})\right] &= \zeta^{-1}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathsf{diag}(\boldsymbol{p}(\boldsymbol{u}^*)) - \boldsymbol{p}(\boldsymbol{u}^*)\boldsymbol{p}(\boldsymbol{u}^*)^{\mathsf{T}}\end{pmatrix} + O(\zeta^{-1}\rho^2 + \zeta^{-2})\\ &= \zeta^{-1}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathsf{diag}(\boldsymbol{x}^*) - \boldsymbol{x}^*\boldsymbol{x}^{*\mathsf{T}}\end{pmatrix} + O(\zeta^{-1}\rho^2 + \zeta^{-2})\ .\end{aligned}$$

Combining this result with equations (11) and (12) we get

$$\mathsf{Var}(\boldsymbol{s}^t) = M\mathsf{Var}(\boldsymbol{s}^{t-1})M^{\mathsf{T}} + \zeta^{-1}V + O(\zeta^{-1}\rho^2 + \zeta^{-2}),$$

and therefore

$$\begin{aligned}\Sigma^t &= \zeta\mathsf{Var}(\boldsymbol{s}^t) + O(\rho^2 + \zeta^{-1})\\ &= M\Sigma^{t-1}M^{\mathsf{T}} + O(\rho^2 + \zeta^{-1}).\end{aligned}$$

Result (4) follows as $\zeta \to \infty$.

Finally, a random variable following a $\mathsf{Mn}(\zeta, \boldsymbol{p})$ for fixed $\boldsymbol{p}$ will converge in distribution to a (multivariate) Gaussian random variable as $\zeta \to \infty$. It hence follows that the conditional distribution of $\boldsymbol{s}^t|\boldsymbol{s}^{t-1}$ is normal in the limit. Moreover, for deterministic $\boldsymbol{s}^0$ it follows that $\boldsymbol{s}^1$ has a marginal normal distribution as $\zeta \to \infty$, and hence the limiting distribution of $\boldsymbol{s}^t$ is also Gaussian by standard properties of the normal distribution, completing the proof.

Remark 2: Assumption (A2) is also noted in Davis and Nihan (1993) as a requirement for the process $\{\boldsymbol{s}^t\}$ to be stationary when $\boldsymbol{s}^0 = \boldsymbol{s}^*$. See Hazelton and Watling (2004) for further comments on the interpretation of this condition.

In principle Theorem 1 can be used to describe the following approximation to the day-to-day model:

$$\boldsymbol{s}^t \sim \mathsf{N}(\boldsymbol{s}^* + M(\boldsymbol{\mu}^{t-1} - \boldsymbol{s}^*), \zeta^{-1}M\Sigma^{t-1}M^{\mathsf{T}} + \zeta^{-1}V)\ . \tag{13}$$

However, this is of limited practical utility because it requires knowledge of the stationary mean $\boldsymbol{s}^*$ of the process. To counter this, in Corollary 1 below we show that the results in Theorem 1 continue to hold when we replace $\boldsymbol{x}^*$ and $\boldsymbol{u}^*$ respectively by $\boldsymbol{x}^\dagger$ and $\boldsymbol{u}^\dagger = \boldsymbol{c}(\boldsymbol{x}^\dagger)$, where $\boldsymbol{x}^\dagger$ is Daganzo and Sheffi's (1977) Stochastic User Equilibrium (SUE) flow pattern. This is important because the SUE flow pattern can be calculated using routine techniques, and so Corollary 1 provides a computationally cheap way of approximating the properties of day-to-day traffic models.

Corollary 1 follows directly from the following Lemma, which describes the proximity of $\boldsymbol{x}^\dagger$ to $\boldsymbol{x}^*$.

Lemma 2. *Assume (A1), (A3) and*

(A5) The equation $\boldsymbol{x} = \boldsymbol{p}(\boldsymbol{c}(\boldsymbol{x}))$ has a unique solution $\boldsymbol{x}^\dagger$.

Define $\boldsymbol{s}^\dagger = \left[\begin{smallmatrix}\boldsymbol{u}^\dagger \\ \boldsymbol{x}^\dagger\end{smallmatrix}\right]$. Then $\boldsymbol{s}^\dagger = \boldsymbol{s}^* + O(\zeta^{-1})$.

Proof

Hazelton and Watling (2004) (Corollary 1) proved that $\boldsymbol{x}^\dagger = \boldsymbol{x}^* + O(\zeta^{-1})$. Now $\boldsymbol{u}^\dagger = \boldsymbol{c}(\boldsymbol{x}^\dagger)$ by the definition of SUE. Then $\boldsymbol{u}^\dagger = \boldsymbol{c}(\boldsymbol{x}^*) + O(\zeta^{-1})$ by an application of Taylor's theorem, and hence $\boldsymbol{u}^\dagger = \boldsymbol{u}^* + O(\zeta^{-1})$ courtesy of equation (9). This completes the proof.

Remark 3: A discussion of sufficient conditions for the existence of a unique SUE flow pattern can be found in Cantarella and Cascetta (1995). In brief, monotonicity of the functions $\boldsymbol{c}$ and $\boldsymbol{p}$ is adequate.

Corollary 1. *Assume (A1), (A2), (A3), (A4) and (A5). Let B and D denote the Jacobian matrices for $\boldsymbol{c}$ and $\boldsymbol{p}$ evaluated at $\boldsymbol{x}^\dagger$ and $\boldsymbol{u}^\dagger$ respectively, and let M be define according to equation (2) in terms of these Jacobians. Then as $\zeta \to \infty$ the results of Theorem 1 continue to hold when $\boldsymbol{x}^*$ and $\boldsymbol{u}^*$ are replaced by $\boldsymbol{x}^\dagger$ and $\boldsymbol{u}^\dagger$ respectively, and M is defined as above.*

In comparison to Theorem 1 and Corollary 1, Davis and Nihan (1993) provide separate asymptotic results for the cases (i) where the system is following its stationary distribution, and (ii) when it is displaying transient behaviour. Their work implies that the marginal distribution of $\boldsymbol{s}^t$ in its stationary state is $N(\boldsymbol{s}^\dagger, \zeta^{-1}\Sigma)$ where Σ is the fixed point for the recursion in equation (4). Under transient conditions the distribution of $\boldsymbol{s}^t$ is $N(\boldsymbol{\mu}^t, \zeta^{-1}\Sigma^t)$ where the mean vector evolves according to the non-linear process

$$\boldsymbol{\mu}^t = \boldsymbol{p}(\boldsymbol{c}(\boldsymbol{\mu}^{t-1})) \tag{14}$$

and Σ^t evolves according to equation (4), but where M is computed in terms of Jacobians B and D evaluated at $\boldsymbol{\mu}_{\boldsymbol{u}}^{t-1}$ and $\boldsymbol{\mu}_{\boldsymbol{x}}^{t-1}$ respectively. See the implications of Proposition 3 as discussed by Davis and Nihan (1993) for details.

Davis and Nihan's (1993) results for transient behaviour are more general than ours in that assumption (A4) is unnecessary. This means in principle their approximation can describe the evolution of the traffic flow pattern for any initial flow pattern $\boldsymbol{s}^0 \in \mathcal{S}$. In order to achieve this, the evolution of the mean process cannot be specified in the linear manner of equation (3), and computation of the dispersion matrix requires new Jacobian matrices to be calculated at each time point.

In contrast, Corollary 1 is based upon an asymptotic regimen in which $\boldsymbol{s}^0$ converges asymptotically to $\boldsymbol{s}^*$, so that (intuitively speaking) the process remains (with probability one) sufficiently close to $\boldsymbol{s}^\dagger$ for a fixed pair of Jacobian matrices to provide an adequate description of the dynamics. Nonetheless, it is critical to note our linear approximation does cover cases in which the initial point is arbitrarily far from the stationary mean in a relative sense. To see this, note that $\mathsf{E}[||\boldsymbol{s}^t - \boldsymbol{s}^*||]/||\boldsymbol{s}^0 - \boldsymbol{s}^*|| = O(\rho^{-1}\zeta^{-1/2})$ when $\boldsymbol{s}^t$ follows its stationary distribution, with the same result holding when $\boldsymbol{s}^*$ is replaced by $\boldsymbol{s}^\dagger$. If we select an asymptotic regimen for which $\rho\zeta^{1/4} \to 0$ and $\rho\zeta^{1/2} \to \infty$ as $\zeta \to \infty$ then Theorem 1 and Corollary 1 hold but $\boldsymbol{s}^0$ is essentially inconsistent with the stationary distribution and is instead a state that one would only observe under non-stationary (transient) conditions. Setting $\rho = \zeta^{-1/3}$ is an example of an asymptotic scheme that captures this behaviour.

It follows that Corollary 1 provides a convenient tool for examining transient behaviour without the computational expense required to implement Davis and Nihan's (1993) approximation.

Furthermore, because we are able to describe the dynamics of the system (within a suitable neighbourhood of $\boldsymbol{s}^{\dagger}$) using a single time-homogenous matrix M, our asymptotic approximation is very convenient for subsequent mathematical analysis. For instance, we are able to confirm that the system will settle down to its stationary distribution in a predictable manner so long as the eigenvalues of the matrix M (computed using Jacobians calculated at SUE) all have modulus less than one. We return to this point in the numerical experiments in Section 4.

Finally, we observe that Theorem 1 and Corollary 1 assume a fixed (deterministic) initial value $\boldsymbol{s}^0$. However, it is straightforward to show that these results will also hold if $\boldsymbol{s}^0$ is normally distributed with covariance matrix of order $O(\zeta^{-1})$.

3. Extension to General Networks and Learning Models

In this section we first consider how to extend the previous results to networks with multiple OD movements. In order to do so, we need to first extend some of the notation presented earlier.

Suppose now that there are $m \geq 1$ origin-destination (OD) movements, and that in total there are n possible routes across all OD movements. Let ζ denote the total demand across all OD movements, and define the m-vector of weights $\boldsymbol{w}$ such that $\zeta\boldsymbol{w}$ is the vector of (integer) OD demands. Let the $n \times m$ matrix Γ denote the route-OD incidence matrix, equal to 1 if a given route serves a given OD pair and 0 otherwise. Then $\mathsf{diag}(\zeta\Gamma\boldsymbol{w})$ is an $n \times n$ diagonal matrix, with diagonal entries equal to the relevant OD demand corresponding to each route.

As before, $\boldsymbol{X}^t$ denotes the random route flow vector at time t, and $\boldsymbol{x}^t = \zeta^{-1}\boldsymbol{X}^t$ is the standardized version thereof. We now partition these vectors by OD pair as $\boldsymbol{X}^t = (\boldsymbol{X}_1^t, \boldsymbol{X}_2^t, ..., \boldsymbol{X}_m^t)^{\mathsf{T}}$ and $\boldsymbol{x}^t = (\boldsymbol{x}_1^t, \boldsymbol{x}_2^t, ..., \boldsymbol{x}_m^t)^{\mathsf{T}}$ respectively, so that $\boldsymbol{x}_k$ is the vector of standardized route flows for OD pair k. The vectors of route costs $\boldsymbol{c}$, disutilities $\boldsymbol{u}$ and route choice probabilities $\boldsymbol{p}$ are partitioned after the same fashion.

Travellers select a route only from those connecting relevant OD pair. We continue to assume that route choices at time t are made independently (conditional on the past), so that

$$\boldsymbol{X}_k^t | \, \boldsymbol{u}^t \sim \mathsf{Mn}(\zeta w_k, \boldsymbol{p}_k(\boldsymbol{u}^t))$$

for $k = 1, \ldots, m$. The conditional expectation of $\boldsymbol{x}^t$ is therefore given by

$$\mathsf{E}[\boldsymbol{x}^t | \, \boldsymbol{u}^t] = \mathsf{diag}(\Gamma\boldsymbol{w})\boldsymbol{p}(\boldsymbol{u}^t) \ .$$

Expanding asymptotically as $\zeta \to \infty$, we obtain

$$\mathsf{E}[\boldsymbol{x}^t | \, \boldsymbol{u}^t] = \mathsf{diag}(\Gamma\boldsymbol{w})D + O(\zeta^{-1})$$

where D continues to denote the Jacobian matrix for $\boldsymbol{p}$ evaluated at the stationary mean disutility $\boldsymbol{u}^*$.

If we continue to work with the simple exponential learning model from (1), then the appropriate form of the matrix M for multiple OD pairs is

$$M = \begin{pmatrix} I \\ \mathsf{diag}(\Gamma\boldsymbol{w})D \end{pmatrix} \begin{pmatrix} (1-\beta)I & \beta B \end{pmatrix} = \begin{pmatrix} (1-\beta)I & \beta B \\ (1-\beta)\mathsf{diag}(\Gamma\boldsymbol{w})D & \beta\mathsf{diag}(\Gamma\boldsymbol{w})DB \end{pmatrix}.$$

The covariance matrix V is adapted for multiple OD pairs so that

$$V = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{V} \end{pmatrix}. \tag{15}$$

where $\tilde{V}$ is a block diagonal matrix formed based on the OD-partitioned stationary mean vector $\boldsymbol{x}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, ..., \boldsymbol{x}_m^*)$ with its m diagonal blocks given by:

$$\tilde{V}_k = {w_k}^{-1} \left(\mathsf{diag}(\boldsymbol{x}_k^*) - \boldsymbol{x}_k^* {\boldsymbol{x}_k^*}^{\mathsf{T}} \right) \qquad (k = 1, 2, ..., m).$$

With these modifications to M and V, Theorem 1 applies to networks with multiple OD pairs.

While the simple exponential learning model from (1) is popular for modelling day-to-day dynamics, far more flexibility is permitted through the disutility formulation

$$\boldsymbol{u}^t = \boldsymbol{f}(\boldsymbol{u}^{t-1}, \boldsymbol{c}(\boldsymbol{x}^{t-1}), \boldsymbol{u}^{t-2}, \boldsymbol{c}(\boldsymbol{x}^{t-2}), ..., \boldsymbol{u}^{t-\tau}, \boldsymbol{c}(\boldsymbol{x}^{t-\tau})) \tag{16}$$

where $\boldsymbol{f}$ is a temporally homogeneous smooth function. See Davis and Nihan (1993). For such a learning model the appropriate state vector becomes

$$\boldsymbol{s}^t = \begin{pmatrix} \boldsymbol{u}^t \\ \boldsymbol{x}^t \\ \boldsymbol{u}^{t-1} \\ \boldsymbol{x}^{t-1} \\ \vdots \\ \boldsymbol{u}^{t-\tau+1} \\ \boldsymbol{x}^{t-\tau+1} \end{pmatrix}.$$

The process $\{\boldsymbol{s}^t \colon t = 0, 1, 2, \ldots\}$ is once again a regular Markov chain under our standard assumptions on the functions $\boldsymbol{c}$ and $\boldsymbol{p}$. It therefore has a unique stationary distribution, the mean of which is denoted $\boldsymbol{s}^*$ as before.

In order to extend our results the new general learning model we must linearize $\boldsymbol{f}$ about the stationary mean. To that end we define $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}^{t-j}}$ to be the Jacobian matrix of $\boldsymbol{f}$ with respect to $\boldsymbol{u}^{t-j}$ using the argument ordering from (16), evaluated at $\boldsymbol{s}^*$. Similarly, $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^{t-j}}$ is the Jacobian matrix with respect to $\boldsymbol{x}^{t-j}$. We then form the matrices

$$F_j = \begin{pmatrix} I \\ \mathsf{diag}(\Gamma \boldsymbol{w}) D \end{pmatrix} \begin{pmatrix} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}^{t-j}} & \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^{t-j}} B \end{pmatrix} \quad (j = 1, 2, ..., \tau)\ . \tag{17}$$

Asymptotically, the dynamics of the moments of the process are now governed by the matrix

$$M = \begin{pmatrix} F_1 & F_2 & F_3 & ... & F_{\tau-1} & F_\tau \\ I & \mathbf{0} & \mathbf{0} & ... & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} & ... & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & I & \mathbf{0} \end{pmatrix}. \tag{18}$$

We are now in a position to generalize Theorem 1 and its corollary.

Theorem 2. *Consider a dynamic traffic model with a general pattern of travel demand, and a learning model specified by (16). Define* $\rho = ||\boldsymbol{s}^0 - \boldsymbol{s}^*||$. *Assume (A1), (A2), (A3), (A4), (A5) above, and also*

(A6) All second derivatives of $\boldsymbol{f}$ *are bounded on* $\mathcal{S}$.

Then the results (3) and (4) specified in Theorem 1 hold with M and V are defined according to (18) and (15) respectively.

The proof is a straightforward extension of the proof of Theorem 1 and so is omitted. We note that Lemma 1 continues to hold for the more general networks and learning models, leading to the following corollary.

Corollary 2. *Assume (A1) to (A6). Let the matrices $F_1, \ldots, F_\tau$ be defined according to equation (17), where all Jacobian matrices are evaluated at the appropriate coordinates of $\boldsymbol{s}^\dagger$, and let M be as defined in equation (18). Then as $\zeta \to \infty$ the results of Theorem 2 continue to hold when $\boldsymbol{x}^*$ and $\boldsymbol{u}^*$ are replaced by $\boldsymbol{x}^\dagger$ and $\boldsymbol{u}^\dagger$ respectively.*

4. Numerical Studies

We first illustrate our results using a numerical study on a simple network with a single OD movement, three parallel links/routes and OD demand $\zeta = 40$. The cost functions for the three routes are respectively

$$c_1(\boldsymbol{x}) = 2 + 8x_1$$
$$c_2(\boldsymbol{x}) = 3 + 10x_2^2$$
$$c_3(\boldsymbol{x}) = 6 + 25x_3^2$$

where $\boldsymbol{x} = \boldsymbol{X}/\zeta$ is the standardized traffic flow. The disutility updates according to the simple convex combination from (1), with $\beta = 0.05$. We employ a logit route choice model, so that the probability that a traveller takes route j at time t is

$$p_j(\boldsymbol{u}^t) = \frac{\exp(-\theta u_j^t)}{\sum_{i \sim j} \exp(-\theta u_i^t)} \tag{19}$$

where $i \sim j$ indicates that routes i and j serve the same OD pair.

The ensuing numerical results were obtained using the software R version 3.4.3 (R Core Team 2017) running under Windows 10 on a computer with 16 GB of memory. In all cases we implemented our linear approximations based on Corollary 2, so that $\boldsymbol{s}^t \sim N(\boldsymbol{\mu}^t, \Sigma^t)$ with

$$\boldsymbol{\mu}^t = \boldsymbol{s}^\dagger + M(\boldsymbol{\mu}^{t-1} - \boldsymbol{s}^\dagger)$$

and

$$\Sigma^t = M\Sigma^{t-1}M^\mathsf{T} + V,$$

where M and V are computed by evaluating the requisite matrices at the SUE vector $\boldsymbol{s}^\dagger$.

To begin with the logit parameter is set to $\theta = 0.3$. The (unstandardized) SUE flow pattern is then $\boldsymbol{X}^\dagger = (15.15, 16.61, 8.24)^\mathsf{T}$ (to two decimal places). The matrix M has maximum eigenvalue 0.95, so condition (A1) holds and we may expect our asymptotic approximation to perform well. To assess this, we generated 1000 simulations of the model, each over a period of $T = 30$ days. These simulations were used to estimate the true mean of the process, and also provide limits for 95% prediction intervals for flows at each day. These results for the true process (plotted in red) are compared with our approximations (black line) in Figure 1. The 95% prediction intervals are derived from the standard 'mean ± 2 standard deviations' courtesy of the limiting normal distribution, so their accuracy is a direct reflection of the quality of the approximations of the corresponding variance terms. A more refined approximation for the mean using Davis and Nihan's (1993) methodology (see equation 14) is plotted as a dotted line, and the associated prediction intervals are likewise plotted with dots. The time plots for the first 10 simulations are also plotted (in light purple) for comparison.

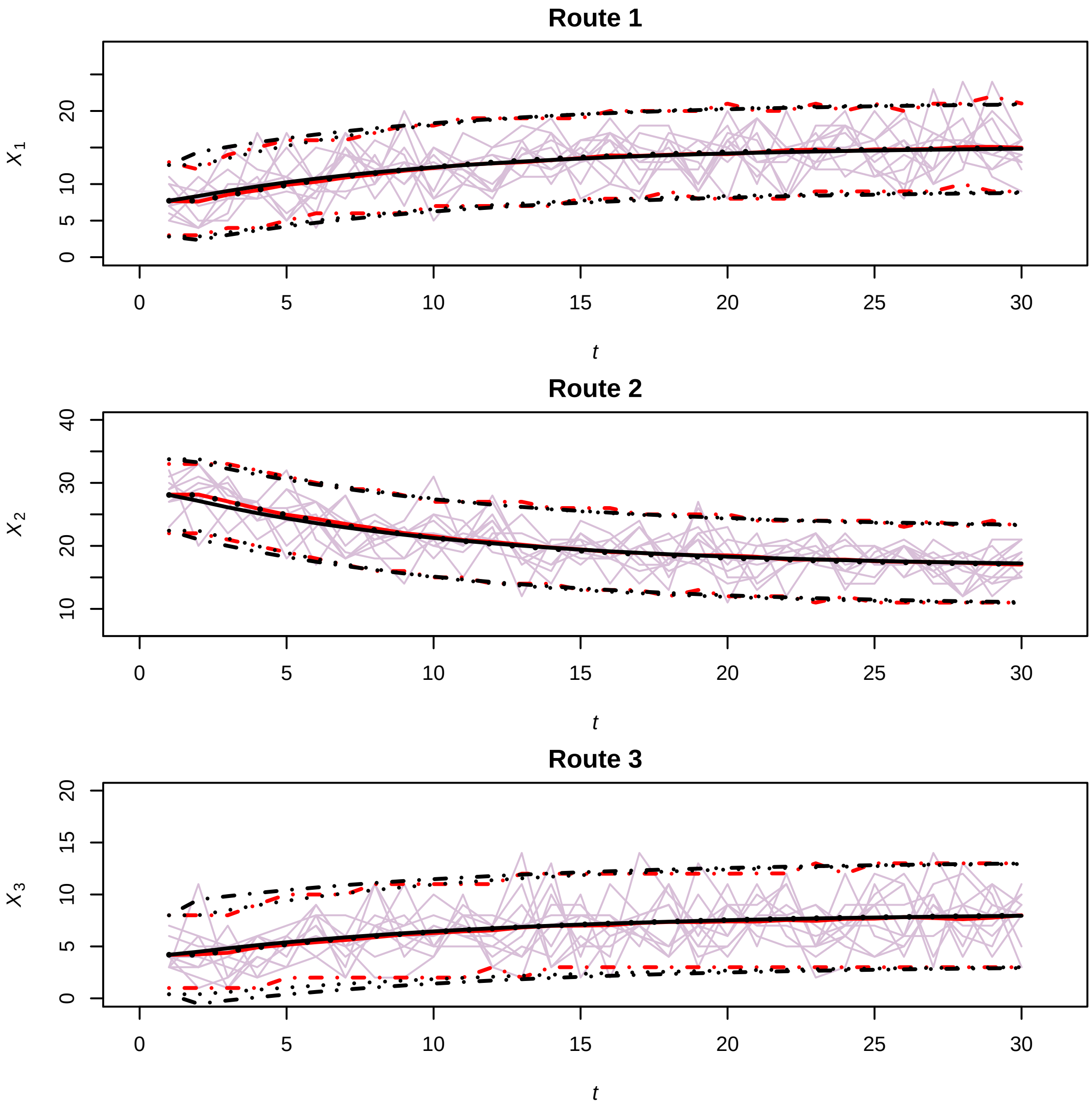

Figure 1: Three-route example with logit parameter $\theta = 0.3$. The unbroken lines depict the true mean flow (red line) and our linear approximation thereof (black line). The dotted central line is the mean approximation using Davis and Nihan's (1993) nonlinear methodology. The outer lines indicate limits of 95% prediction intervals for flows, matched to the mean by plotting symbol and/or colour. The jagged light purple lines show the realized time plots of traffic flows from 10 simulations of the model.

Clearly our approximations have worked well in Figure 1. This is true not just for the period after about time 20 when the process has settled down to its stationary distribution, but also during the transient period. The initial state of the system was obtained by perturbing the SUE disutility vector, so that $\boldsymbol{u}^1 = \boldsymbol{u}^\dagger + (4, 0, 4)^\mathsf{T}$. Importantly, this means that the (understandardized) initial flow $\boldsymbol{X}^1 = (7.72, 28.09, 4.20)^\mathsf{T}$ is outside the standard range of flows that we see when the process is stationary. For example, focusing on route 2, the initial flow $X_2^1 = 28.09$ is well outside the stationary 95% prediction interval, $(10.95, 23.49)$. Despite using fixed Jacobian matrices, our approximations can provide useful guidance regarding transient behaviour, as foreseen in our discussion at the end of Section 2. Davis and Nihan's (1993) nonlinear approximation is a only marginally more precise.

For our second numerical illustration we work with exactly the same network and model, but use a far more extreme initial state. Specifically, we set $\boldsymbol{u}^1 = \boldsymbol{u}^\dagger + (12, 12, 0)^\mathsf{T}$. This means that the initial utility is 3 times more extreme (in comparison to the stationary mean utility) than in the previous case. The results are displayed in Figure 2 in exactly the same manner as for Figure 1.

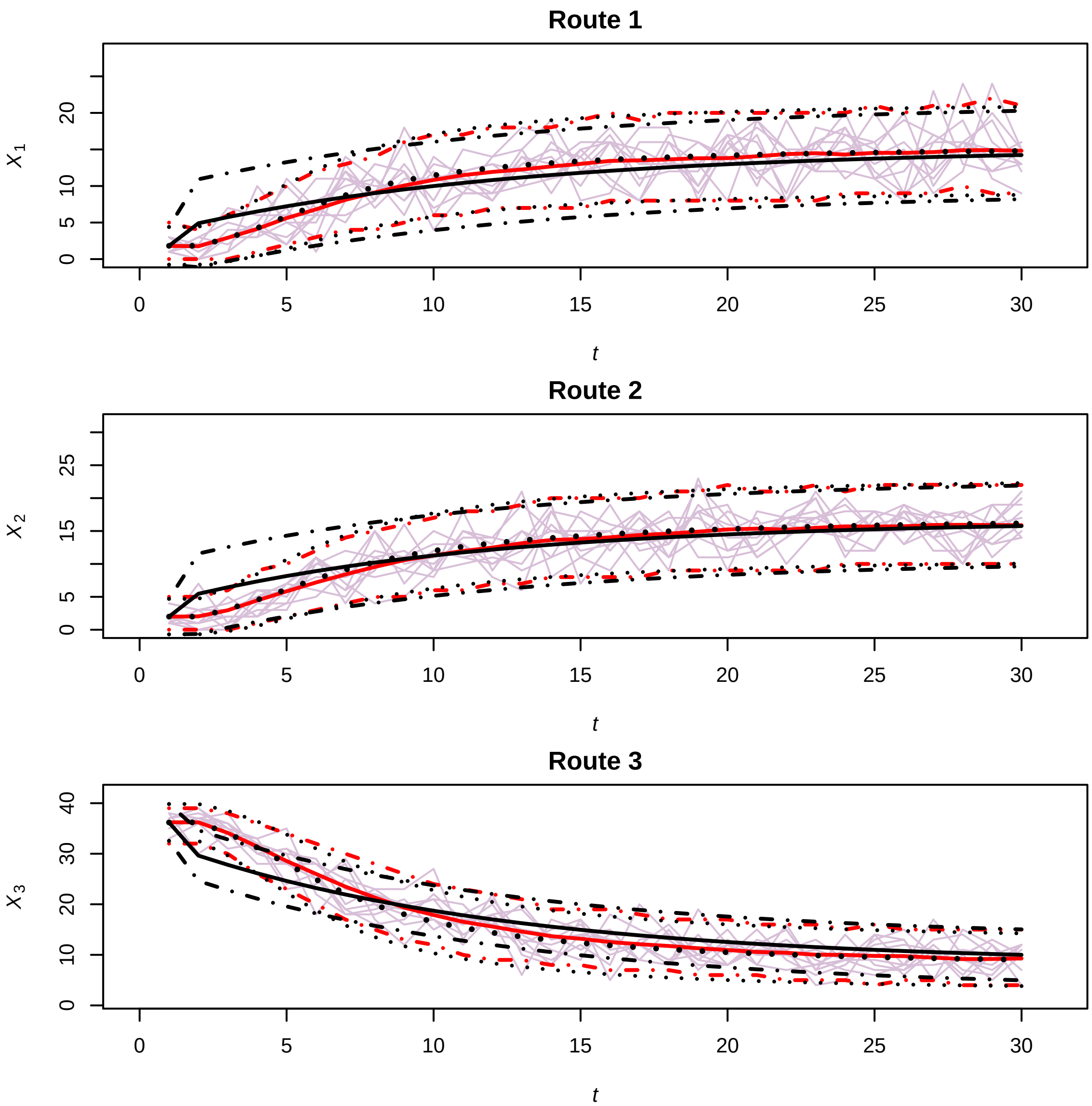

Figure 2: Three-route example with logit parameter $\theta = 0.3$ and very extreme initial state. The unbroken lines depict the true mean flow (red line) and our linear approximation thereof (black line). The dotted central line is the mean approximation using Davis and Nihan's (1993) nonlinear methodology. The outer lines indicate limits of 95% prediction intervals for flows, matched to the mean by plotting symbol and/or colour. The jagged light purple lines show the realized time plots of traffic flows from 10 simulations of the model.

Clearly our linear approximation works less well over the transient period in Figure 2 than in the previous case. In essence, condition (A4) is failing with the more extreme initial state. In this second example, the Jacobian matrices for the cost and probability functions evaluated at SUE provide quite a poor description of rates of change at state $\boldsymbol{s}^1$. What is more, because of the form of the learning model and the relatively small value of β, travellers forget the initial disutilities rather slowly. The result is that the inaccuracy in the mean approximation (and prediction intervals) persists for several days. By contrast, Davis and Nihan's (1993) approximation remains rather accurate.

We continue with the same 3 route network for our third example, but we increase the logit parameter to $\theta = 1.1$ and also increase the coefficient of x^r in each cost function through multiplication by a factor of 5^r. The latter change is equivalent to imposing a five-fold decrease in the nominal capacity of each link. The consequence of these modifications is a far more reactive system, with travellers sensitive to modest changes in disutilities, and route costs sensitive to relatively small changes in traffic volumes. The results for this example are displayed in Figure 3 in the usual manner.

The initial state in this latest case is not far from SUE; it was generated by setting $\boldsymbol{u}^1 = \boldsymbol{u}^\dagger + (2, 2, 0)^\mathsf{T}$. However, it is clear that our linear approximation method breaks down completely, with the approximate mean process (depicted by the black line) oscillating wildly at the later time points. The approximated variances degrade even more swiftly, as evidenced by the rapid divergence of the bounds of the 95% prediction intervals (depicted by the dashed black lines). The explanation for this behaviour is that the largest eigenvalue of the matrix M is $\lambda_{\max} = 1.22$, so that assumption (A1) fails and so Theorem 1 and Corollary 1 do not apply. We note that using Davis and Nihan's (1993) nonlinear form for the evolution of the mean (equation 14), and iteratively updating the Jacobian matrices when computing the covariance matrices, does not provide a remedy. The corresponding dotted line approximations to the mean and prediction intervals are also hopelessly inaccurate.

For our final pair of numerical examples we consider a section of the road network in the English city of Leicester, as abstracted in Figure 4. This network has 85 OD pairs, and a total of 123 plausible routes. We use OD travel demands based on the analysis from Hazelton (2015). Link cost functions are quadratic, so that the cost function for route j can be written as

$$c_j = \sum_{i=1}^{50} a_{ij}\alpha_i \left[1 + \left(\frac{y_i}{\beta_i}\right)^2\right]$$

where $a_{ij} = 1$ if link i is part of route j, and $a_{ij} = 0$ otherwise. The flow on link i is denoted y_i, and α_i and β_i are link specific parameters, with the latter representing the link capacity. The vector of link flows can be computed from route flows by $\boldsymbol{y} = A\boldsymbol{x}$, where $A = (a_{ij})$ is the link-path incidence matrix. Disutilities are modelled using (1) with $\beta = 0.05$, and route choice probabilities are computed using the logit model (19) with parameter $\theta = 0.1$.

In our first test with this network, we simulate flows for a sequence of 50 days. On day 15 we impose a 50% reduction in the capacity of link 7; the capacity returns to normal the next day. This could represent the effects of minor road works or an accident, for example. We will now focus on travel between node 1 and node 20, which corresponds to journeys from the centre of the city to the University of Leicester. There are several plausible routes for this journey, of which two carry a significant amount of traffic. These are labelled routes 12 and 16 (from the total of 123 routes). Link 7 forms part of route 12, but is not part of route 16.

We conducted 1000 simulations of the system using the Markov model. The results are plotted in the standard manner for routes 12 and 16 in Figure 5. As expected, the reduction in capacity of link 7 results in a temporary shift of travellers from route 12 to route 16. Clearly

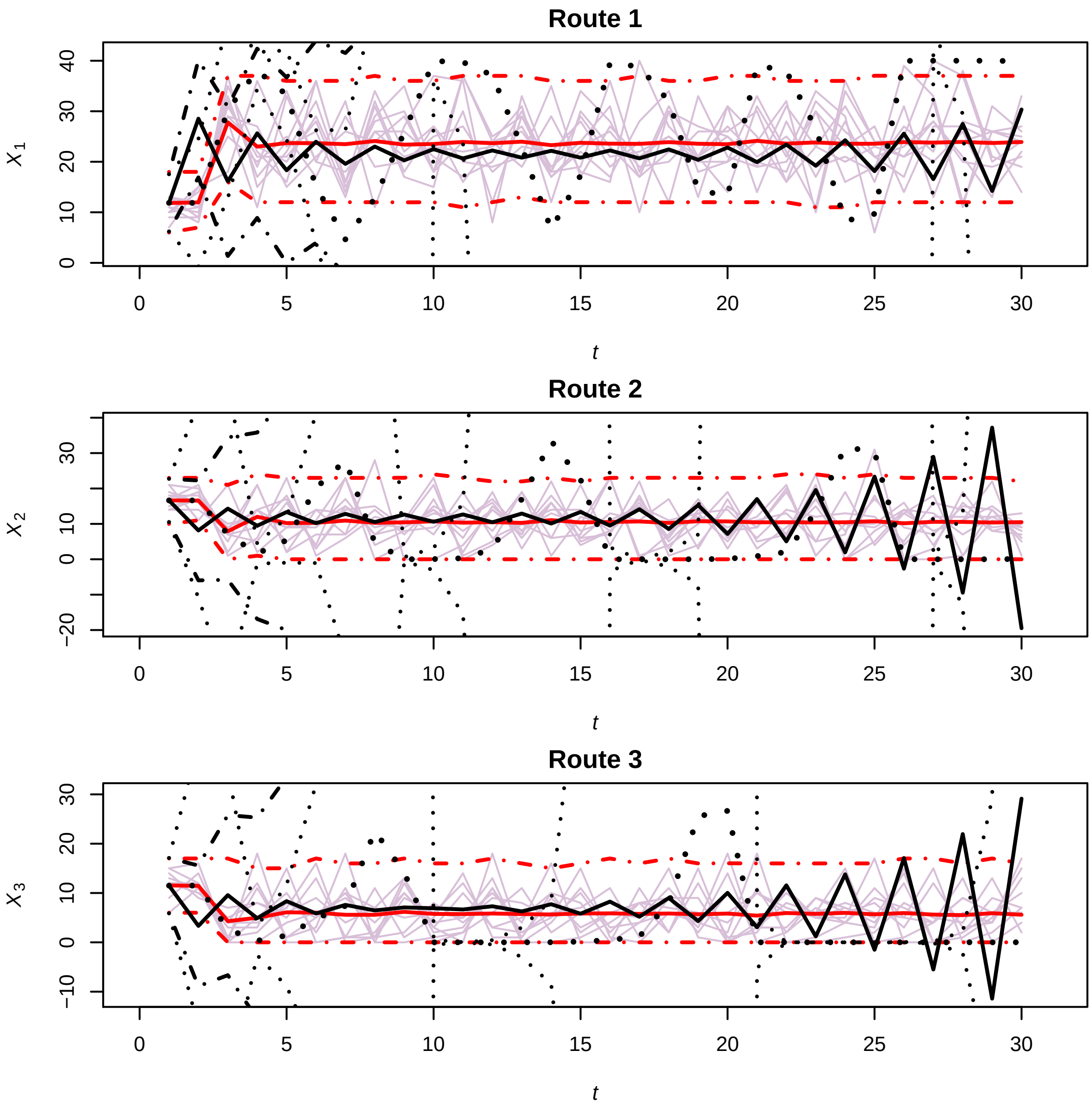

Figure 3: Three-route example with logit parameter $\theta = 1.1$ and greatly reduced link capacities. The unbroken lines depict the true mean flow (red line) and our linear approximation thereof (black line). The dotted central line is the mean approximation using Davis and Nihan's (1993) nonlinear methodology. The outer lines indicate limits of 95% prediction intervals for flows, matched to the mean by plotting symbol and/or colour. The jagged light purple lines show the realized time plots of traffic flows from 10 simulations of the model.

our linear approximation methodology has provided an accurate representation of the transient behaviour of the system following the disruption. Naturally Davis and Nihan's (1993) nonlinear approximation is also excellent. However, we note that it took more than ten times as long to run as our linear approximation (40.6 CPU seconds versus 3.8 CPU seconds).

For our second test with the Leicester network, we simulate the system for 730 days (i.e. two years). On day 366 we reduce the capacity on link 7 by 20%, and then hold it therefore there for the entirety of the second year. Again we focus on the results for flows on routes 12 and 16, which are plotted in the standard manner in Figure 6. We observe that Davis and Nihan's (1993) nonlinear approximation is extremely accurate throughout the period. Our linear approximation for the evolution of the mean flows is also very accurate, but there is a

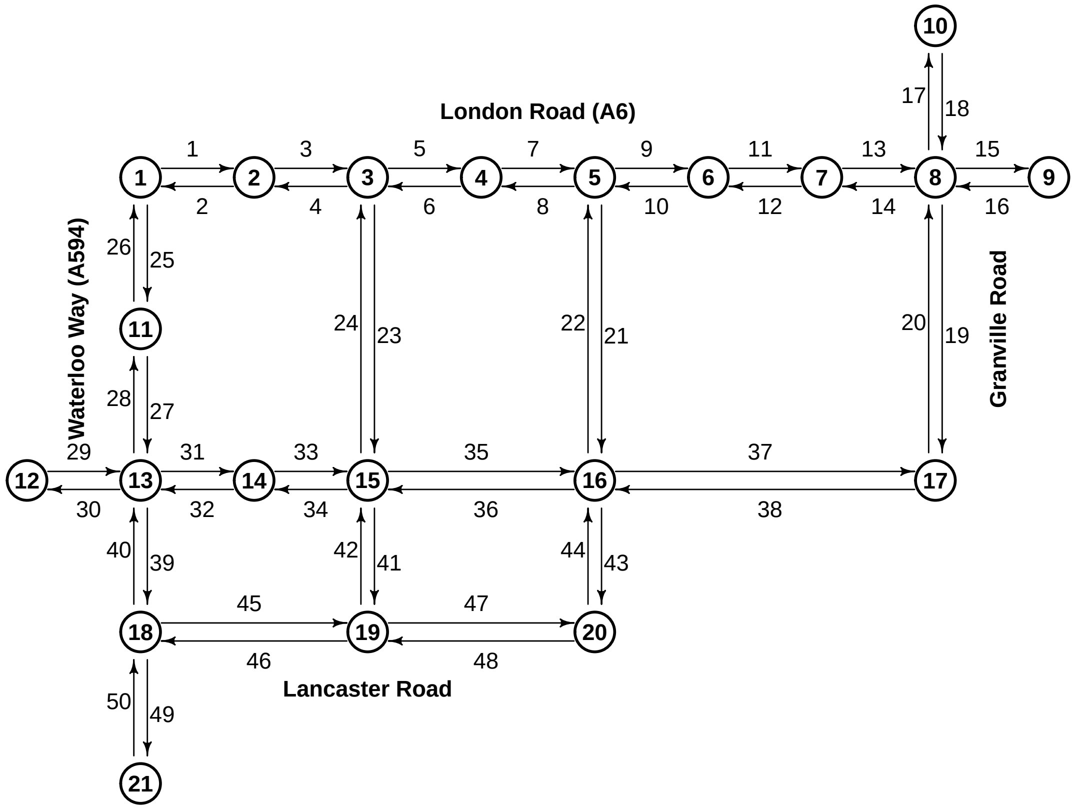

Figure 4: Abstraction of part of the road network in the English city of Leicester.

small inflation in the range of the prediction intervals. By the end of the two years the system has settled down to a new stationary distribution. It is interesting to note how well our linear methodology manages to describe this new distribution, given that the approximation is based on the initial SUE flow pattern. Clearly the means of the two stationary distributions are sufficiently close for assumption (A4) to be largely applicable.

For this longer simulation experiment, our linear approximation took 12.4 CPU seconds to run. By comparison, Davis and Nihan's (1993) nonlinear approximation required 575.8 CPU seconds.

5. Conclusions

Stochastic process models of transportation networks provide a rich description of both transient dynamics and random variation within a self-consistent framework. They therefore seem highly suited to many contemporary opportunities and challenges, such as understanding the disruptive impacts of planned road maintenance or network changes, or quantifying the impacts of policies and changes in demand on network (un)reliability. They have already been shown to be particularly effective in assessing the effectiveness of measures designed to mitigate the impacts of unexpected variation, such as through pricing (Liu et al. 2017), information (Zhao et al. 2018), or traffic control (Liu et al. 2006). However, conventional Monte Carlo-based methods of estimating such processes suffer from several difficulties, not only high computational

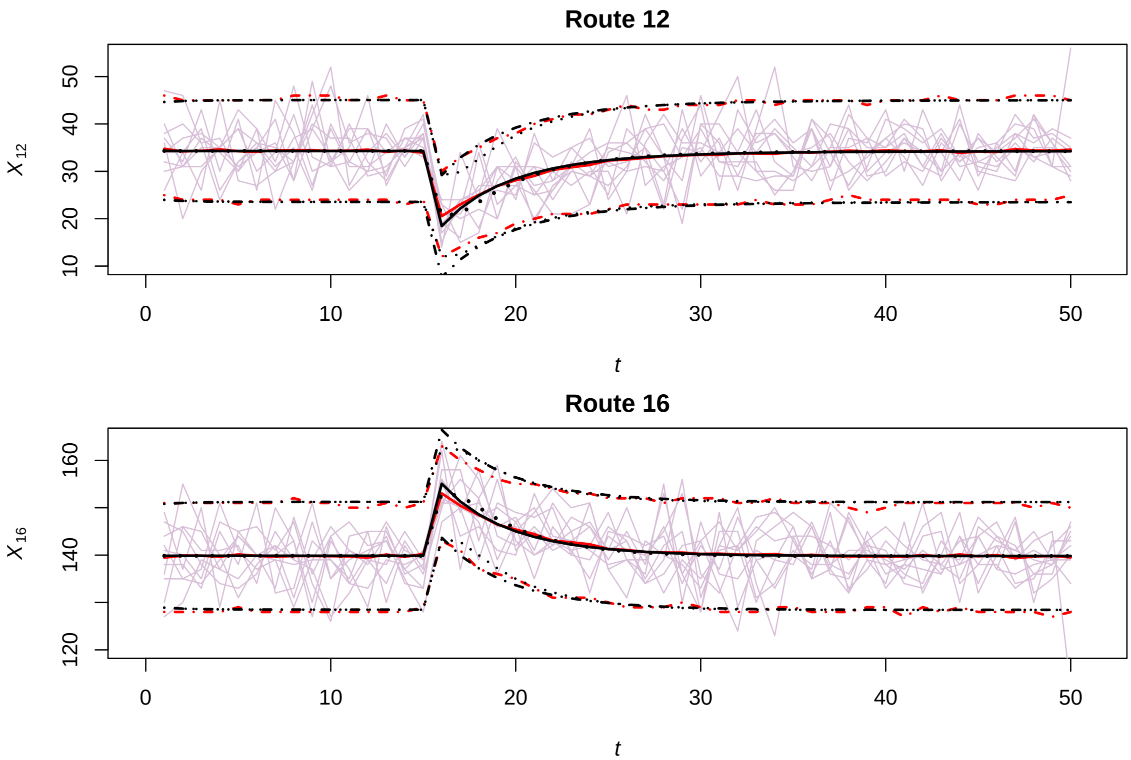

Figure 5: Routes 12 and 16 for the Leicester network, with an intervention applied at day $t = 15$ only. The unbroken lines depict the true mean flow (red line) and our linear approximation thereof (black line). The dotted central line is the mean approximation using Davis and Nihan's (1993) nonlinear methodology. The outer lines indicate limits of 95% prediction intervals for flows, matched to the mean by plotting symbol and/or colour. The jagged light purple lines show the realized time plots of traffic flows from 10 simulations of the model.

demands but also the difficulty in interpreting model outputs, a key challenge being to separate systematic change in these outputs from random variation.

In the present paper we have derived theoretical results and associated approximations that allow such problems to be circumvented, by allowing the transient evolution of the first two moments of the state variables to be modelled without the need for simulation, and with only knowledge of an equilibrium state. In numerical experiments, we have demonstrated how our theoretical results allow us to anticipate the circumstances in which such an approximation may work well, and where it may break down. For practical application with standard types of smooth, monotonic cost and probability functions, the user need only check assumptions (A2) and (A4). The former simply requires calculation of the eigenvalues of the matrix M (evaluated at SUE). Assumption (A4) concerns the distance of the initial flow pattern from SUE (or the stationary mean), and is more difficult to assess. Our numerical studies suggest that linear approximation works well for mean flows of order 100 when the initial flow pattern is up to 3 standard deviations distant from SUE. More extreme initial states can be accommodated for higher levels of travel demand (i.e. larger values of ζ), based on the discussion at the end of Section 2.

Our methods, and the earlier ones of Davis and Nihan (1993), are restricted to Markovian models, in which the state of the system at day $t+1$ is independent of the state at day $t-1$ given the state at day t. This is a classical assumption in the literature, which is far less restrictive than it might at first appear. As we saw in Section 3, by defining the system state

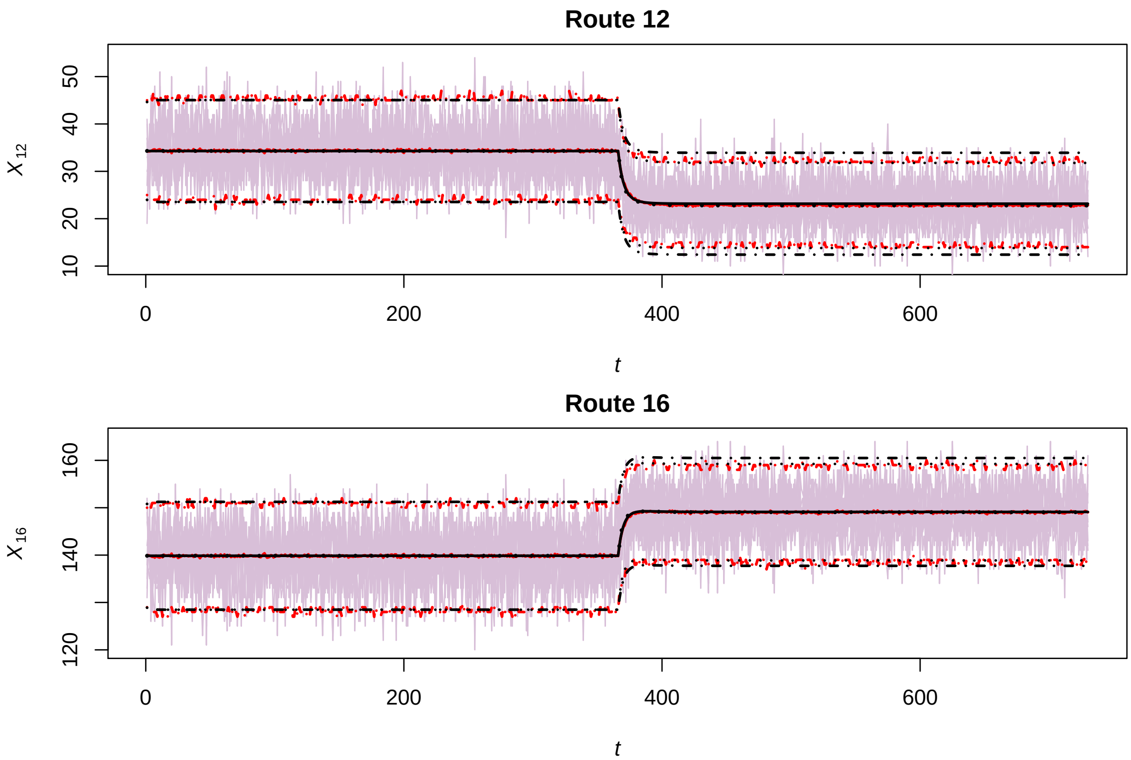

Figure 6: Routes 12 and 16 for the Leicester network, with an intervention applied continuously from day 366. The unbroken lines depict the true mean flow (red line) and our linear approximation thereof (black line). The dotted central line is the mean approximation using Davis and Nihan's (1993) nonlinear methodology. The outer lines indicate limits of 95% prediction intervals for flows, matched to the mean by plotting symbol and/or colour. The jagged light purple lines show the realized time plots of traffic flows from 10 simulations of the model.

in terms of the pattern of route flows over the previous m days, we retain the Markov structure even though travellers now have a longer memory on which to base decisions on choice of route. Moreover, inclusion of both route flows and disutilities in the state vector gives rise to a very rich class of models. Note that for the model in Section 2, the route flow process $\{\boldsymbol{x}^t\}$ is not itself Markovian because it depends on the initial disutility $\boldsymbol{u}^0$ courtesy of equation (1), but $\{\boldsymbol{s}^t\}$ is a Markov process. In general, it is difficult to envisage a day-to-day traffic model based on a classical cost-updating and minimization framework which cannot be represented as a Markov process through suitable choice of state vector (cf. Watling and Cantarella 2015). Nevertheless, we acknowledge that the practicability of Davis and Nihan's (1993) approximation in particular will lessen as the size of the state vector increases.

There exist many natural, further applications of the work reported, whether the focus is on understanding disruption and resilience, or on evaluating or designing robust control/pricing/information measures in a kind of stochastic process counterpart to the work of Cromvik and Patriksson (2010). A computationally-efficient approximation such as the one derived in the present paper might also be deployed in the same spirit that a metamodel has been shown to be useful in approximating complex model constraints in optimal control or parameter estimation problems (Osorio and Chong 2015).

There is also significant potential in seeking future generalisations and extensions of the results presented in the present paper through relaxation and refinement of the model assumption. For example, in our present study we presumed quite conventional assumptions concerning risk-neutral travel behaviour, yet the framework presented (in representing endogenous sources of variation) is clearly high suited to representing various kinds of risk-averse behaviour, as for example in Praksash and Srinivasan (2017). As an alternative direction to explore, it is notable that in our traffic model we have supposed steady state conditions as modelled by explicit functions between travel time and flow, and clearly this is rather conducive to an analysis based on Jacobians. However, analytical analyses of stochastic process models have been shown to be feasible with simple within-day dynamic network loading models (Balijepalli and Watling 2005), and we can draw confidence that these might be extended in the future, given the advances that have now been made in calculus for more complex dynamic network loading models (Shen et al. 2007, Osorio et al. 2011, Rinaldi et al. 2016, Song et al. 2018).

Acknowledgement

The second author acknowledges financial support from Royal Society of New Zealand Marsden Grant 14-MAU-017. The authors thank three anonymous referees for their helpful comments, which led to significant improvements to this paper.

References

N.C. Balijepalli, D.P. Watling, 2005. Doubly Dynamic Equilibrium Distribution Approximation Model for Dynamic Traffic Assignment, In: *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction - Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, Mahmassani HS (Ed), Emerald Group Publishing Ltd, 741-760.

M.G.H. Bell and C. Cassir, 2000. *Reliability of Transport Networks*, Research Studies Press, Baldock, UK.

S. Boyer, S. Blandin, L. Wynter, 2015. Stability of transportation networks under adaptive routing policies. *Transportation Research Part B* 81, 886903.

G.E. Cantarella GE, and D.P. Watling DP, 2016. Modelling road traffic assignment as a day-to-day dynamic, deterministic process: a unified approach to discrete- and continuous-time models. *EURO Journal on Transportation and Logistics*, 5, 69–98.

E. Cascetta, 1989. A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research* 23B, 1–17.

G.C. Cantarella, and E. Cascetta, 1995. Dynamic processes and equilibrium in transportation networks: Towards a unifying theory. *Transportation Science* 29, 305–329.

Castillo, Calviño, Nogal and Lo, 2014. On the probabilistic and physical consistency of traffic random variables and models. *Computer Aided Civil and Infrastructure Engineering* 29, 496–517.

A. Chen, Z. Zhou, P. Chootinan, S. Ryu, C. Yang, S.C. Wong, 2011. Transport network design problem under uncertainty: a review and new developments. *Transport Reviews* 31, 743–768.

C. Cromvik, M. Patriksson, 2010. On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, Part II: Applications. *Journal of Optimization Theory and Applications*, 144, 479500.

C.F. Daganzo and Y. Sheffi, 1977. On stochastic models of traffic assignment. *Transportation Science* 11, 253–274.

G.A. Davis and N.L. Nihan, 1993. Large population approximations of a general stochastic traffic assignment model. *Operations Research* 41, 169–178.

G. Flötteröd and C. Osorio, 2017. Stochastic network link transmission model. *Transportation Research Part B* 102, 180209.

M.L. Hazelton, 2002. Day-to-day variation in Markovian traffic assignment models. *Transportation Research* 36B, 61–72.

M.L. Hazelton, 2015. Network tomography for integer-valued traffic. *Annals of Applied Statistics* 9(1), 474–506.

M.L. Hazelton, D.P. Watling, 2004. Computation of equilibrium distributions of Markov traffic-assignment models. *Transportation Science*, 38, 331–342.

H. Huang, T. Liu, H. Yang, 2010. Modeling the evolutions of day-to-day route choice and year-to-year ATIS adoption with Stochastic User Equilibrium. *Journal of Advanced Transportation* 42(2), 111-127.

S.E. Jabari and H.X. Liu, 2013. A stochastic model of traffic flow: Gaussian approximation and estimation. *Transportation Research Part B* 47, 1541.

W.H.K. Lam, H.K. Lo and S.C. Wong, 2014. Advances in equilibrium models for analyzing transportation network reliability. *Transportation Research Part B: Methodological* 66, 1–162.

R. Liu, D. Van Vliet, D. Watling, 2006. Microsimulation models incorporating both demand and supply dynamics. Transportation Research Part A, 40, 125150.

Z. Liu, S. Wang, B. Zhou, Q. Cheng, 2017. Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics. *Transportation Research Part C: Emerging Technologies*, 79, 58-72.

S.Nakayama, and D.P. Watling, 2014. Consistent formulation of network equilibrium with stochastic flows. *Transportation Research Part B: Methodological* 66,50–69.

C. Osorio, G. Fltterd, M. Bierlaire, 2011. Dynamic network loading: a stochastic differentiable model that derives link state distributions. *Transportation Research Part B: Methodological*, 45, 14101423.

C. Osorio, L. Chong, 2015. A Computationally Efficient Simulation-Based Optimization Algorithm for Large-Scale Urban Transportation Problems. *Transportation Science*, 49, 623-636.

C. Osorio, and J. Yamani, 2017. Analytical and scalable analysis of transient tandem Markovian finite capacity queueing networks. *Transportation Science* 51(3), 823-840.

K. Parry, D.P. Watling, M. Hazelton, 2016. A new class of doubly stochastic day-to-day dynamic traffic assignment models. *EURO Journal of Transportation and Logistics*, 5, 5–23.

S. Peeta and A.K. Ziliaskopoulos, 2001. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics* 1, 233–265.

A.A. Prakash, K.K. Srinivasan, 2017. Finding the Most Reliable Strategy on Stochastic and Time-Dependent Transportation Networks: A Hypergraph Based Formulation. *Networks and Spatial Economics*,17, 809-840.

R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`.

M. Rinaldi, W. Hempe, C.M.J. Tampere, 2016. A sensitivity-based approach for adaptive decomposition of anticipatory network traffic control. *Transportation Research Part C: Emerging Technologies*, 66, 150-175.

Y. Sheffi, 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall. Englewood Cliffs, NJ.

W. Shen, Y. Nie, H.M. Zhang, 2007. On Path Marginal Cost Analysis and its Relation to Dynamic System-Optimal Traffic Assignment. In: *Transportation & Traffic Theory 2007*, (eds. R. Allsop, M. Bell, B. Heydecker), Elsevier, Oxford, UK, 327-360.

M.J. Smith and D.P. Watling, 2016. A route-swapping dynamical system and Lyapunov function for stochastic user equilibrium. *Transportation Research Part B* 85, 132141.

W. Song, K. Han, Y. Wang, T.L. Friesz, E. del Castillo, 2018. Statistical metamodeling of dynamic network loading. *Transportation Research Part B: Methodological*, in press.

W.Y. Szeto and S.C. Wong, 2012. Dynamic traffic assignment: model classifications and recent advances in travel choice principles. *Central European Journal of Engineering* 2, 1–18.

D.P. Watling, 1999. Stability of the stochastic assignment problem: a dynamical systems approach. *Transportation Research* 33B, 281–312.

D.P. Watling, D.S. Milne, S.D. Clark, 2012. Network impacts of a road capacity reduction: Empirical analysis and model predictions. *Transportation Research Part A*, 46, 167–189.

D.P. Watling, and G.E. Cantarella, 2013. Modelling sources of variation in transportation systems: Theoretical foundations of day-to-day dynamic models. *Transportmetrica B*, 1, 3–32.

D.P. Watling, and G.E. Cantarella, 2015. Model representation and decision-making in an ever-changing world: The role of stochastic process models of transportation systems. *Networks and Spatial Economics* 15, 843–882.

W. Zhang, W. Guan, J. Ma, J. Tian, 2015. A nonlinear pairwise swapping dynamics to model the selfish rerouting evolutionary game. *Networks and Spatial Economics* 15(4), 1075-1092.

X. Zhao, C. Wan, J. Bi, 2018. Day-to-Day Assignment Models and Traffic Dynamics Under Information Provision. *Networks and Spatial Economics*, in press.

S. Zhu, D. Levinson, H.X. Liu, K. Harder, 2010. The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. *Transportation Research Part A*, 44, 771-784.