A General Stochastic Process for Day-to-Day Dynamic Traffic Assignment: Formulation, Asymptotic Behaviour, and Stability Analysis

Giulio E. Cantarella, University of Salerno
David P. Watling, University of Leeds

Abstract – This paper presents a general modelling approach to day-to-day dynamic assignment to a congested network through discrete-time stochastic and deterministic process models including an explicit modelling of users' habit as a part of route choice behaviour, through an exponential smoothing filter, and of their memory of network conditions on past days, through a moving average or an exponentially smoothing filter. An asymptotic analysis of the mean process is carried out to provide a better insight. Results of such analyses are also used for deriving conditions, about values of the system parameters, assuring that the mean process is dissipative and/or converges to some kind of attractor. Numerical small examples are also provided in order to illustrate the theoretical results obtained.

Keywords: day-to-day dynamics, stochastic process models, mean process, deterministic process models, stability analysis

1. Introduction

The development, since the 1970s, of efficient computational methods for implementing network equilibrium models has arguably had one of the most significant impacts of academic research on transport planning practice. In many countries, such methods are an embedded element of procedures for cost-benefit analysis of proposed schemes, and are used widely for operational planning of traffic measures. With this class of approach now extended to consider multiple classes of users, within-day dynamic traffic interactions, unreliability and heterogeneity/mis-perception of users, their potential applicability is wider than it has ever been. Such facts are important to appreciate when proposing any approach that may be viewed as an *alternative* to the network equilibrium philosophy. Many large transport investments have been justified on the basis of equilibrium predictions, and so there is a political 'price' in practitioners moving to any "new" method. Academic researchers can help considerably in this process by better understanding the linkages between what might appear to be apparently diverse methods, and in particular by understanding the connection of any alternative approaches to network equilibrium. The objective, for example, could be to better understand the cases in which network equilibrium may be justified as an approximation to some real-world situation, and those cases in which it may potentially give misleading results. The present paper is motivated by exactly this desire to better understand the connections between approaches, and to understand where network equilibrium is a useful notion in this context. This includes the possibility, in some cases, that we calculate equilibrium in exactly the same way as we do at present, but the *meaning* or conceptualisation of the computed state is different, and suggests additional or alternative ways to utilise the computed state.

The focus of the present paper will be on what have become known as day-to-day dynamic models of route choice, focusing on (among other elements) how users adapt their route decisions over repeated trips. The term 'day-to-day' dynamic is useful to distinguish these approaches from 'within-day dynamic' models, these latter focusing on issues such as time-dependent OD demand rates, the spatial and temporal interactions of traffic flows, the influence on users' time-dependent choice of route and possibly departure time, and the possibility for users to make *en route* diversions during a journey. In order to focus our discussion, we do not consider within-day dynamic issues in the present paper, though we note that there are several papers that consider the combination of day-to-day and within-day dynamics (Cascetta & Cantarella, 1991; Balijepalli & Watling, 2005; Liu et al, 2006; Friesz et al, 2011), and note that it is possible to transfer many of the arguments of the kind used here (admittedly at the price of far greater complexity) to the combined case. We note that the term 'day-to-day dynamic' is intended, therefore, to be indicative of the kind of process being considered, but it need not be that these models are representing a real, continuous sequence of complete days. In this respect it is good to have in mind the following suggestion of an 'epoch-to-epoch dynamic' model, where:

> *'... 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'. (Cascetta, 1989)*

There exist two clear classes of model of day-to-day dynamic route choice, namely Deterministic Processes (DPs) and Stochastic Processes (SPs). DPs are more naturally associated with traditional equilibrium models of transport systems, in the sense that point equilibria may emerge, under certain assumptions, as the convergent limits of such processes under some long-term steady conditions. DPs also allow transitions to be examined, especially when some 'shock' or designed change occurs (He & Liu, 2012). A recent review and synthesis of DP models in discrete and continuous time may be found in Cantarella & Watling (2015); in this paper no comparison is carried out with DPs in continuous time and/or based on Wardrop approach to route choice behaviour (for a recent paper see Guo et al., 2015), since results obtained with these models can hardly transferred to the kind of models discussed in this paper.

SPs are more naturally associated with modelling the variability that is seen to occur in real-life systems, even under relatively stable operating conditions; they are thus able to represent both dynamic transitions and steady-state fluctuations. A review of SP models is provided in Watling & Cantarella (2014). The two types of approach draw on quite different mathematical disciplines, DPs emerging from non-linear dynamical system theory (typically interested in mappings over continuous state spaces), whereas SPs arising originally from the study of probability theory and Markov chains (over discrete state space).

Although some numerical evidence relating DPs and SPs exists (e.g. Cantarella & Cascetta, 1995; Watling, 1996), relatively little general, theoretical evidence exists concerning their relation for general traffic networks. The exceptions to this are the works of Davis & Nihan (1993) and Hazelton & Watling (2004), both of whom developed asymptotic approximation results for SPs, as demands and capacities grow in tandem. In the present paper we develop a 'natural' relation between DPs and SPs, which emerges from viewing DPs as a joint process in the statistical moments of the corresponding SP. This work is inspired by the general (asymptotic) theory mentioned above, and the series of two-link examples recently studied in Watling & Cantarella (2013). We shall here extend the work presented in Watling & Cantarella (2013) in several ways, particularly focusing on the development of the mean of a SP as a DP, as well as other results in literature.

The models presented and discussed in this paper extend our previous theoretical work on discrete-time stochastic and deterministic process models into a general modelling approach to day-to-day dynamic assignment so as to: (a) relate to general traffic networks (not just two-link networks); (b) include an inertia/habit effect modelled through an exponential smoothing filter; and (c) incorporate learning models with finite or infinite memory, bridging moving average and exponentially smoothed approaches.

The theoretical approximation of an SP model is first derived as a DP in the vector of flow means. Analysing the resulting DP, conditions are established to ensure uniqueness of the equilibrium, and to ensure its (local) asymptotic stability, conditions for the system being dissipative are also stated. Numerical examples are provided in order to motivate the work, to illustrate the theoretical results obtained, and to explore the generality of the asymptotic (large demand/large capacity) approximation, even in cases where demand and capacity are not "large".

The paper is organised as follows. Section 2 presents basic notations and briefly reviews SUE models; then section 3 discusses some simple but effective approaches to modelling dynamic learning and choice behaviour and analyses resulting Deterministic Process models. Section 4 describes the proposed SP model and some solution approaches as well as an asymptotic approximation to the mean of this process. Finally, in sections 5 and 6 we discuss the main findings and identify several potential future research directions.

2. Basic notations, definitions and equations in SUE models

In this section we will briefly review the basic notations and definitions adopted, as well as fixed-point models for stochastic user equilibrium assignment (Cantarella, 1997).

Our starting is that demand is segregated into multiple *classes*, each class containing users moving on the same origin-destination (OD) movement and in the same user category (i.e. with the same behavioural parameters)[1]. Let

n_{CL} be the number of user classes;
n_i be the number of acyclic (or elementary) routes available for users of class i;
$n = \sum_i n_i$ be the total number of routes available across all user classes[2];
$d_i > 0$ be the demand flow for user class i, assumed integer and greater than zero;
$\mathbf{d}$ be the demand flow vector of dimension n_{CL};
$\mathbf{D}_{[i]} = d_i\, \mathbf{I}_{n_i}$ be a diagonal matrix of dimensions $n_i \times n_i$, with entries on the main diagonal equal t o d_i;
$\mathbf{p}_{[i]} \geq \mathbf{0}$ and $\mathbf{1}^{\mathrm{T}}\, \mathbf{p}_{[i]} = 1$ be the route fraction vector of dimension n_i for user class i;
$\mathbf{x}_{[i]} = d_i\, \mathbf{p}_{[i]} \geq \mathbf{0}$ and $\mathbf{1}^{\mathrm{T}}\, \mathbf{x}_{[i]} = d_i$ be the route flow vector of dimension n_i for user class i;
$\mathbf{D}$ be a $n \times n$ block diagonal matrix, with each block given by $\mathbf{D}_{[i]}$;
$\mathbf{p}$ be the route fraction block vector of dimension n, with each block given by $\mathbf{p}_{[i]}$;
$\mathbf{x} = \mathbf{D}\,\mathbf{p} \geq \mathbf{0}$ be the route flow block vector of dimension n, with each block given by $\mathbf{x}_{[i]}$;
$\mathbf{w}_{[i]}$ be the route cost vector of dimension n_i for user class i;
$\mathbf{w}$ be the route cost block vector of dimension n, with each block given by $\mathbf{w}_{[i]}$;
n_{LINK} be the number of links;
$\mathbf{c}$ be the link cost vector of dimension n_{LINK};
$\mathbf{f}$ be the link flow vector of dimension n_{LINK};
$\mathbf{f_b}$ be the link base flow vector of dimension n_{LINK}, link flows not depending from modelled user route choice behaviour;
$\mathbf{B}$ be the $n_{LINK} \times n$ link-route incidence matrix, with entries equal to 1 or 0 depending on whether a link is part of the given route. Each class is assumed connected by at least two routes, since the demand flow of any user class with only one available route induces further link flows that can be directly added to the base link flow vector $\mathbf{f_b}$,.

The link flows are given by:

---

[1] For readers unfamiliar with such notation, it is suggested that on a first reading it makes sense to suppose there is a single user category (and so classes refer only to origin-destination pairs), and then on a second reading to consider the generalisation to multiple user categories, since conceptually there is little difference.
[2] It is worth noting that according to the above notations the collection of acyclic routes available for travel, across all classes, are indexed 1, 2, ..., n, in such a way that the n_i routes corresponding to class i have indices $\{1 + \sum_{j=1}^{i-1} n_j\,, 2 + \sum_{j=1}^{i-1} n_j\,, \dots, n_i + \sum_{j=1}^{i-1} n_j\}$, for i = 1, 2, ..., n_{CL}.

$$\mathbf{f} = \mathbf{B}\,\mathbf{x} + \mathbf{f_b}$$

We shall suppose the route travel costs **w** are sum of a term linear-additive in the (generic) link travel costs:

$$\mathbf{w} = \mathbf{B}^{\mathrm{T}}\,\mathbf{c} + \mathbf{w_o}$$

where possibly of another term including specific or non-additive route costs, $\mathbf{w_o}$.

Congestion on the n_{LINK} links of the network is modelled through travel cost functions:

$$\mathbf{c} = \mathbf{c}(\mathbf{f})$$

In most cases the link travel cost actually depends on the flow capacity ratio, rather than on the flow value itself.

Then the *link* travel cost-flow functions imply corresponding *route* travel cost-flow functions such that the route travel costs when the route flows are **x** are given by:

$$\mathbf{w}(\mathbf{x}) = \mathbf{B}^{\mathrm{T}}\,\mathbf{c}(\mathbf{B}\,\mathbf{x} + \mathbf{f_b}) + \mathbf{w_o}$$

In our subsequent analysis this implied relationship above between route costs and route flows, will often be used rather than the underlying relationship between link costs and link flows.

The route choice fractions **p** result from the user route choice behaviour and can be expressed as a function of the route disutilities by applying any model derived from the Random Utility Theory, such as Logit, C-Logit, Probit, Gammit, … :

$$\mathbf{p}_{[i]} = \mathbf{p}_{[i]}(\mathbf{z}_{[i]}) \qquad i = 1, 2, \ldots, n_{CL}$$
$$\mathbf{p} = \mathbf{p}(\mathbf{z})$$

where in this case:
$\mathbf{z}_{[i]} = \mathbf{w}_{[i]}(\mathbf{x})$ is the route disutility vector of dimension n_i for user class i;
$\mathbf{z} = \mathbf{w}(\mathbf{x})$ is the route disutility block vector of dimension n, with each block given by $\mathbf{z}_{[i]}$.

The *stochastic user equilibrium* assignment searches for mutually consistent flows and costs, assuming that a RUM is used to described the route choice behaviour. It can be expressed by fixed-point models with respect to route (or arc) flows (or costs), such as:

$$\mathbf{x}_{\mathrm{SUE}} = \mathbf{D}\,\mathbf{p}(\mathbf{w}(\mathbf{x}_{\mathrm{SUE}})) \qquad (2.1)$$

Existence of solution is guaranteed if both the cost function and the route choice model are continuous (and the network is connected). Uniqueness is guaranteed under the commonly adopted conditions of positive definiteness of the Jacobian of cost function, $\mathbf{J_w} = \nabla_{\mathbf{x}}\,\mathbf{w}(\mathbf{x})$, and the negative semi-definiteness of the choice probability function Jacobian, $\mathbf{J_p} = \nabla_{\mathbf{z}}\,\mathbf{p}(\mathbf{z})$, the latter holding under mild assumptions (Cantarella, 1997). Furthermore, as discussed in Bifulco et al (2013) and in therein quoted references, the invertibility of the matrix $\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_w}$ is a weaker (sufficient) condition to guarantee uniqueness of the SUE solution.

The models and results presented in the following of this paper are stated with respect to route costs, disutilities and flows, it seems worth noting that they hold as well with respect to link variables. Generally this is not the case for day-to-day dynamic models based on Wardrop route choice behaviour.

3. Modelling learning and choice behaviour in DP models

The specification of a day-to-day dynamic process models for assignment requires the explicit modelling of

- user habit: how users make a choice today, possibly repeating yesterday choice to avoid the effort needed to take a decision, or reconsidering it according to the forecasted level of service,

- user learning and forecasting process: how users forecast the level of service that they will experience today, from experience and other sources of information.

In this section we will describe some simple approaches to address the two above issues (sub-sections 3.1 and 3.2), which will allow us to specify and analyse a Stochastic Process (SP) model in the following section 4. In this section a Deterministic Process (DP) model is also analysed (sub-section 3.3) to support comparison between two approaches to user learning and forecasting process and to approximate the mean process of the SP model.

We suppose a learning process for users whereby the disutility $z_j^{(t-1)}$ of each route j forecasted at the end of travelling on day t – 1 is used when making decisions for the following day t. This forecasted disutility is assumed to be the accumulated knowledge up to the end of day t – 1, so generalising the notion of disutility introduced in the above section 2. Let

$\mathbf{x}^{(t)}$ be the route flow block vector on day t;
$\mathbf{z}^{(t)}$ be the forecasted route disutility block vector on day t.

3.1 Modelling the dynamic choice process

In the following, we specify how users make decisions based on learnt experiences (modelled as in the following sub-section 3.2). Specifically, we assume:

- A fixe proportion α ($0 < \alpha \leq 1$) of users reconsider their previous day's choice, and those that do decide to reconsider then make choices in proportions according to a random utility model (possibly then repeating the previous day's choice); and
- The remaining users choose between the available routes in proportions equal to the fraction of users that actually chose those routes on the previous day.

Under such a behavioural model, users of class i are now assumed to have two reasons for choosing any route j available for them: either they choose it out of habit, which a proportion $x_j^{(t-1)}/d_i$ of them do (where $x_j^{(t-1)}$ is the number of users that actually chose route j yesterday, and d_i is the class demand flow for class i), or their choice behaviour can be modelled through a RUM, such as Logit, C-Logit, Probit, … . The proportion of users choosing for the first reason is $1-\alpha$ and for the second reason is α. For those that *do* decide to reconsider their choice, then conditionally on the vector of disutilities $\mathbf{z}^{(t-1)}$ at the end of day t – 1, each user of class i chooses a route independently of one another, with choice proportions given by a random utility model, with the proportions of choosing a route j available for that class given by $p_j(\mathbf{z}^{(t-1)})$. Collecting the relevant proportions together for class i, we then denote the vector function for each user class i as: $\mathbf{p}_{[i]}(\mathbf{z}_{[i]}^{(t-1)})$ and the collection of all such functions across the all the classes as: $\mathbf{p}(\mathbf{z}^{(t\,-1)})$. Then we get an exponential smoothing filter ES(α):

$$\mathbf{x}^{(t)} = \alpha\, \mathbf{D}\, \mathbf{p}(\mathbf{z}^{(t-1)}) + (1 - \alpha)\, \mathbf{x}^{(t-1)} \qquad (t = 2, 3, 4, \ldots) \qquad (3.1)$$

or $\mathbf{x}^{(t)} = \mathbf{D}\, (\alpha\, \mathbf{p}(\mathbf{z}^{(t-1)}) + (1 - \alpha)\, \mathbf{D}^{-1}\, \mathbf{x}^{(t-1)}) \quad (t = 2, 3, 4, \ldots) \qquad (3.1')$

Equation (3.1) tries to model in simple but effective way user inertia to change and how much users are prone to review their habit; this simple model also allows us to develop the consideration about convergence and stability in sub-section 3.3. In more general approaches the proportion α may itself be a function of some disutility reliability variables. For instance in approaches regarding ATIS equipped users α may depend on the ATIS aggregate reliability, and thus may change over time. This issue is addressed by Bifulco and Simonelli (2005), Bifulco et al. (2007, 2009, 2011) through a modelling approach consistent with this paper. However, embedding this approach in a (complete) multi-user framework allowing for the kind of stability analysis carried out in sub-section 3.3 is still an open issue (see also section 6).

In disaggregate approaches, a proportion α is defined for each route separately depending on the difference between experienced and forecasted (or ATIS provided) costs. The use of probabilistic thresholds leads to route choice switching models. This approach is rather effective when only two routes are available between each O-D pair, since there is no need of any route choice model. Indeed, when more than two routes are available, a conditional route choice function should be applied to model route choice behaviour of users who decide to reconsider their yesterday choice.

3.2 Modelling the learning behaviour process

We have seen (Cantarella, 2013, for further details, and Bifulco et al, 2014, for further models) that in the case of DP models, an especially convenient form of specification of learning model is one in which the forecasted disutility at the end of a day (say yesterday) is a convex combination of the previous day's forecasted disutility and the present day's actual travel cost:

$$\mathbf{z}^{(t-1)} = \beta\, \mathbf{w}(\mathbf{x}^{(t-1)}) + (1 - \beta)\, \mathbf{z}^{(t-1)} \qquad (t = 1, 2, 3, \ldots) \qquad (3.2)$$

where the memory weighting parameter satisfies $0 < \beta \leq 1$, and where at $t = 1$, we suppose:

$\mathbf{z}^{(0)} = \mathbf{w}(\mathbf{x}^{(0)})$ for a given route flow vector $\mathbf{x}^{(0)}$.

Under such a model, we may recursively apply the expressions above in order to relate the forecasted disutility on any day t to the complete history of travel costs (and hence flows) since the beginning of the process, leading to an exponential smoothing filter ES(β):

$$\mathbf{z}^{(t-1)} = \beta\, \mathbf{w}(\mathbf{x}^{(t-1)}) + \beta\, \Sigma_{k = 2,\ldots,t-1}\, (1 - \beta)^{k-1}\, \mathbf{w}(\mathbf{x}^{(t-k)}) + (1 - \beta)^{t-1}\, \mathbf{w}(\mathbf{x}^{(0)}) \qquad (t = 1, 2, 3, \ldots)\,.$$

Clearly, then, such a process would not be Markovian if we defined state variables in terms of the flows, since users never entirely 'forget' an initial experience (except in the special case of $\beta = 1$). This can be resolved by instead considering the couple ($\mathbf{x}^{(t)}$, $\mathbf{z}^{(t)}$) as state variables, depending (with the assumptions we shall subsequently make) only on the couple ($\mathbf{x}^{(t-1)}$, $\mathbf{z}^{(t-1)}$). Such a model would be entirely amenable to analysis by standard numerical methods.

We note that from the viewpoint of representing real-world systems, the assumption of an 'infinite learning' model adopted in several previous studies of the deterministic process model was made due to the considerable mathematical simplicity it affords, being an approximation of real-world user behaviour (see below for some numerical examples), rather than a belief that users never forget an experience, however old. It is therefore quite easy to justify an assumption of finite memory, and we shall adopt this in our stochastic process analysis in section 4. We shall, however, specify this model in order to capture the possibility (at one extreme) of mimicking as closely as possibly the 'infinite memory' assumption commonly adopted in studies of deterministic process models.

Therefore, we shall use the relative weighting of actual travel costs as suggested by the β-model above, but will truncate the past memory by only considering some pre-specified fixed number of previous days m. This means that at day t, we only start the summation at day $t - m + 1$. In order that the truncated process retains the property of being a convex combination of the actual costs, we apply a scaling factor to ensure that the non-neglected weights sum to 1, leading to a convex moving average filter with (normalised) decreasing weights defined by one-parameter MA(β, m) and for $\beta < 1$ and $m > 1$ given by:

$$\eta_k = \beta\,(1-\beta)^{k-1}/(1-(1-\beta)^m) \geq 0 \quad (k = 1, 2, \ldots, m) \qquad \text{with } \Sigma\, \eta_k = 1 \text{ and } \beta \in \,]0,1]$$

$$\text{or} \quad \eta_k = \eta_{k-1}\,(1-\beta) \quad (k = 2, \ldots, m) \qquad \text{with } \eta_1 = \beta\,/\,(1-(1-\beta)^m)$$

It is worth noting that condition $\beta = 1$ and $m > 1$ yields to $\eta_1 = 1$ and $\eta_{k+1} = 0$ ($k = 2, \ldots, m$), moreover for $m = 1$ it is assumed $\eta_1 = 1$. Since all these conditions lead to a ES(1) filter, a proper MA filter is only obtained with $m > 1$ and $\beta < 1$ that is $\eta_1 < 1$. If the weights η_k are assumed strictly decreasing with respect to k, then $\eta_1 > 1/m$.

In this case after the initialization step (see below), the forecasted disutilities are given by:

$$\mathbf{z}^{(t\text{-}1)} = \beta\,/\,(1-(1-\beta)^m)\,\mathbf{w}(\mathbf{x}^{(t\text{-}1)}) + \Sigma_{k=2,\ldots,m}\, \eta_k\, \mathbf{w}(\mathbf{x}^{(t-k)})$$

$$\text{or } \mathbf{z}^{(t\text{-}1)} = \Sigma_{k=1,\ldots,m}\, \eta_k\, \mathbf{w}(\mathbf{x}^{(t-k)}) \qquad (t = m+1, m+2, \ldots) \qquad (3.3)$$

$$\text{or } \mathbf{z}^{(t\text{-}1)} = \mathbf{CM}^{(t\text{-}1)} \cdot \boldsymbol{\eta}$$

where $\mathbf{CM}^{(t\text{-}1)}$ is the $m \times n$ memory matrix of costs with m columns given by the costs in the m previous days, $\mathbf{w}(\mathbf{x}^{(t-k)})$ $k = 1, 2, \ldots, m$.

At the end of each day (t-1) the current cost memory matrix $\mathbf{CM}^{(t\text{-}2)}$ is updated by dropping last column, moving all others columns rightwards and putting $\mathbf{w}(\mathbf{x}^{(t\text{-}1)})$ as first column in $\mathbf{CM}^{(t\text{-}2)}$. Initialization of $\mathbf{CM}^{(t)}$, say specification of $\mathbf{CM}^{(0)}$, may be carried out assuming that:

- all the m columns of $\mathbf{CM}^{(0)}$ are equal to $\mathbf{w}(\mathbf{x}^{(0)})$ for any given route flow vector $\mathbf{x}^{(0)}$;
- the ES filter (3.2) is applied for m days to fill the m columns of matrix $\mathbf{CM}^{(0)}$.

It is worth noting that as m goes to infinite we get the exponential smoothing filter described above by equation (3.2) with either of the above initialization approaches. Moreover as m goes to infinite,

- $\eta_1 = \beta\,/\,(1-(1-\beta)^m) > \beta$ goes to β,
- $\eta_m = \beta\,(1-\beta)^{m-1}/(1-(1-\beta)^m) > 0$ goes to 0.

For example of convergence for $\beta \geq 0.40$ and m > 10 both differences are less than 0.01.

Moreover, looking at equation (3.3), we can see that the weights given to actual travel experiences in the past depend only on the relative distance in time they are away from the present, i.e. the model MA(β, *m*) is time-homogeneous.

3.3 Stability analysis through Deterministic process models

This subsection analyses whether the system described by the deterministic process model DP (3.1, 3.3) based on MA filter is dissipative with respect to the memory depth *m* and the memory weighting parameter β. A brief review of results concerning DP (3.1, 3.2) based on ES is also reported below for comparison's sake. In all DP models discussed below fixed-points states given by $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)} = \mathbf{x}^*$ are equal to the SUE flow pattern $\mathbf{x}_{SUE}$.

Three cases will be discussed below depending on values of memory depth *m* and of memory weighting parameter β, whichever is the value of habit parameter α (α = 1 leading to particular cases). [Appendix 1 briefly review deterministic process models, say discrete-time time-homogeneous Markovian non-linear systems.]

• We will first assume that the memory is large enough that other past days are considered beside yesterday but not all past days, say $1 < m < \infty$, and $\beta < 1$, thus $\eta_1 > \beta$ and $\eta_m > 0$. In this case it is necessary to re-formulate DP 1(3.1, 3.3) to obtain a Markovian DP.

The system state at day *t* is described by a *m*-block vector, one block ${}^h\mathbf{x}^{(t)}$ $h = 1, ..., m - 1$ for each of the *m* days to be kept in memory. Thus: the first block ${}^1\mathbf{x}^{(t)}$ contains today's route flows, as in vector $\mathbf{x}^{(t)}$ already introduced; the second block ${}^2\mathbf{x}^{(t)}$ contains yesterday's route flows, say $\mathbf{x}^{(t-1)}$; and so on. Therefore, on each day *t* today's flows—contained in the first block ${}^1\mathbf{x}^{(t)}$ —are updated according to equation (3.1), then each of all the other blocks are used to keep a memory of the previous days' flows, whilst the *m*-th day in the past is no longer recorded. According to this state definition today's state only depends on yesterday's, leading to the following Markovian DP:

$$ {}^1\mathbf{x}^{(t)} = (1-\alpha)\, {}^1\mathbf{x}^{(t-1)} + \alpha\, \mathbf{D}\, \mathbf{p}\left(\Sigma_{k=1,\dots,m}\, \eta_k\, \mathbf{w}({}^k\mathbf{x}^{(t-1)})\right) \qquad (3.4) $$

$$ {}^{h+1}\mathbf{x}^{(t)} = {}^h\mathbf{x}^{(t-1)} \qquad h = 1, \dots, m-1 \qquad (3.5) $$

From the analysis reported in Appendix 2 it turns out that the system may not be dissipative, especially for high values of β and low values of *m*. However, whichever is the value of β, there always exists a large enough memory depth m^* such that for any memory deeper than this value m^* the system is *dissipative* from any starting state.

Figure 1 shows the minimum memory depth value m^* against β needed to obtained a value of η_m less than: 0.1, 0.05, 0.01, 0.005, 0.001 (from top to bottom).

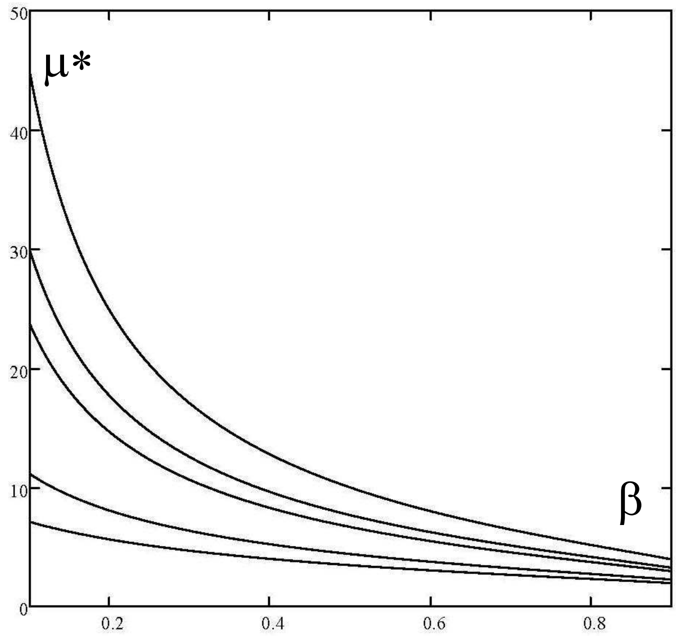

Figure 1. Value m against β need to obtain a value of η_m less than: 0.1, 0.05, 0.01, 0.005, 0.001 (from top to bottom).*

This brief analysis shows that the DP systems based on MA filter for modelling learning process may not be dissipative, that is may not converge to any kind of attractor (from some starting states at least) as time goes to infinite. On the other hand, if the memory depth is large enough the system is always dissipative. A full stability and bifurcation analysis of this case is still an open issue, worth of further research effort.

• On the other hand, if $\beta < 1$ and the memory depth is so large that it may be considered infinite, no day is ever forgotten (even though just a small weight is given to days far in the past), say $m \to \infty$, then $\eta_1 \to \beta$ and $\eta_m \to 0$. In this case MA filter (3.3) tends to the ES (3.2) filter whichever is the value of β, and we get the general $\alpha\beta$ Markovian DP model recently discussed in detail in Cantarella (2013). The determinant of the Jacobian of such a model is $(1-\alpha)^n\ (1-\beta)^n$ always in the range $]-1, 1[$, thus the system is always dissipative, that is it always converges to an attractor (not necessarily a fixed-point) from any starting state. (In that paper an in-depth fixed-point stability and bifurcation analysis is carried out and further earlier references are also reported.)

• A special case is obtained if memory refers to yesterday only, $m = 1$, or $\beta = 1$ then for MA filter $\eta_1 = \eta_m = 1$, thus the MA(β, m = 1), the MA(β = 1, m) and the ES(β = 1) filters give the same DP model:

$$\mathbf{x}^{(t)} = (1-\alpha)\, \mathbf{x}^{(t-1)} + \alpha\, \mathbf{D}\, \mathbf{p}(\mathbf{w}(\mathbf{x}^{(t-1)})) \tag{3.6}$$

DP (3.6) is Markovian with Jacobian matrix **J** given by:

$$\mathbf{J} = (1-\alpha)\, \mathbf{I} + \alpha\, \mathbf{G} \qquad \text{where } \mathbf{G} = \mathbf{D}\, \mathbf{J_p}\, \mathbf{J_w}$$

Since the determinant of **J** may be out of the range]−1, 1[, the system may be not *dissipative*, that is it may not converge to any kind of attractor. In this case even if there is a unique fixed-point $\mathbf{x}^*$, it may be an attractor from some starting states only but not from all, or it is not an attractor at all. An in-depth fixed-point stability and bifurcation analysis can easily be carried out noting that for each eigenvalue γ of matrix **G** an eigenvalue λ of matrix **J** is given by $(1-\alpha) + \alpha\gamma$, being a special case of the case briefly discussed above.

Figure 2 shows the evolution of a route flow from day 75 to day 90 (basic data are in appendix 3) applying DP with MA with memory depth $m = 2, \ldots, 6$ or with ES. DP with MA and $m = 6$ is almost undistinguishable of DP with ES both reach a fixed-point attractor equal to the unique SUE. It can easily be seen that a short memory leads the system towards other kind of attractors than the unique fixed-point, which is not (locally) stable.

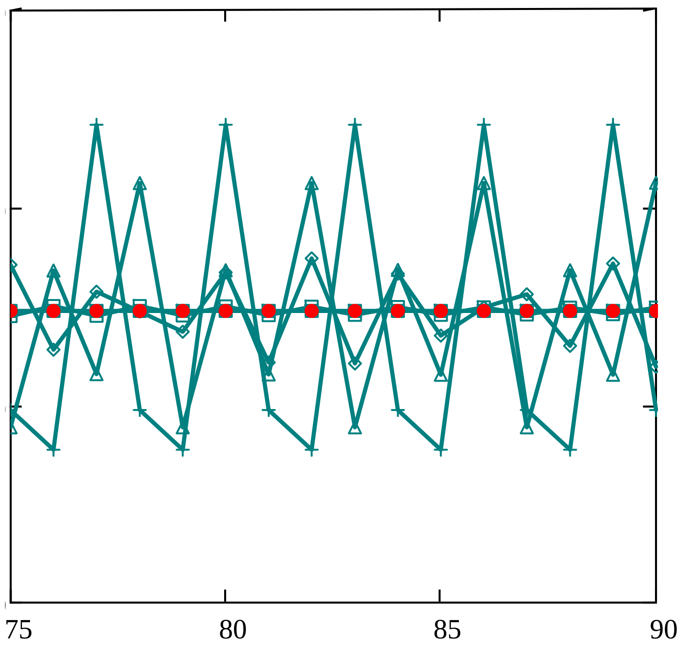

Figure 2. A route flow against day given by DP with MA m = 2(+), 3(△), 4(◇), 5(□), 6(■), and with ES (●) (see also the text above.)

It seems worth noting that all results presented above still hold if the DP model (3.1, 3.3) or (3.1, 3.2) are formulated with respect to link variables; this is always the case when route choice user behaviour is modelled through random utility models. On the other hand, if this behaviour is modeled according to Wardrop I principle, link- and route- based models are generally different (see for instance Guo et al., 2015).

4. A Stochastic Process model for day-to-day assignment

In conventional SUE models we are familiar with the idea of modelling randomness/heterogeneity in users' *perceptions* of travel costs, and we also include here such a feature. Additionally we shall here suppose that the *actual* travel costs experienced are also randomly distributed. In the present paper we suppose that the only source of randomness in the actual travel costs is the randomness in flows. This is an extreme and unnecessarily restrictive assumption, and in practice there are likely to be many other unobserved sources of variation in the actual travel costs, e.g. due to weather, incidents, vehicle-mix. The model defined could be extended to represent such variations, either through postulating a probability distribution of elements of the parameters of the cost functions (e.g. the capacities), and/or by assuming additional additive variation on the distribution of travel costs generated by variable flows and/or variable parameters (i.e. this would be in addition to the flow-based variation captured in the postulated model). These are important factors to consider, yet in line with the rest of the paper we neglect them here in order to focus on the main thrust of the paper. For a discussion of some additional sources of variation that might be modelled using such processes, the interested reader is referred to Watling & Cantarella (2013, 2014).

Due to the several sources of uncertainty above mentioned we suppose that the number of user travelling on route j on day t as well as the corresponding route disutility are modelled as random variables, $X_j^{(t)}$ and $Z_j^{(t)}$ respectively, whose realisations are denoted by $x_j^{(t)}$ and $z_j^{(t)}$. Thus, let

$\mathbf{Z}_{[i]}$ be the route disutility vector of dimension n_i for user class i;
$\mathbf{Z}$ be the route disutility block vector of dimension n, with each block given by $\mathbf{Z}_{[i]}$;
$\mathbf{X}_{[i]} = d_i\, \mathbf{p}_{[i]} \geq \mathbf{0}$ and $\mathbf{1}^{\mathrm{T}}\, \mathbf{X}_{[i]} = d_i$ be the route flow vector of dimension n_i for user class i;
$\mathbf{X}$ be the route flow block vector of dimension n, with each block given by $\mathbf{X}_{[i]}$.

4.1 The overall SP model

The above assumptions combined with the dynamic choice process (3.1) and the MA filter (3.3) (or the ES filter (3.2)), lead to an m-dependent Markov process in discrete state space, whereby the conditional probability distribution of the state on any day t, as represented through the vector random variable $\mathbf{X}^{(t)}$, is fully determined by the previously-realised values of the states $\{\mathbf{X}^{(t-k)} : k = 1, 2, \ldots, m\}$. The assumptions may be summarised as:

$$\mathbf{X}_{[i]}^{(t)} \mid \{\mathbf{X}^{(t-k)} : k = 1, 2, \ldots, m\} \sim \text{Multinomial}(d_i, (1-\alpha)\,(\mathbf{X}_{[i]}^{(t-1)}/d_i) + \alpha\ \mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)}))$$

independently for each $i = 1, 2, \ldots, n_{CL}$

for some vector of cost functions $\mathbf{w}(.)$, choice model $\mathbf{p}(.)$, demand vector $\mathbf{d}$, memory length $m \geq 1$, normalised learning weights $\{\eta_1, \eta_2, \ldots, \eta_m\}$, reconsideration probability $0 < \alpha \leq 1$. Actually, it will be more convenient, below, to capture this model by writing it (entirely equivalently) as:

$$\mathbf{Z}^{(t-1)} = \textstyle\sum_{k=1,\ldots,m} \eta_k\, \mathbf{w}(\mathbf{X}^{(t-k)}) \qquad (4.1)$$

$$\mathbf{Y}_{[i]}^{(t)} = (1-\alpha)\,(\mathbf{X}_{[i]}^{(t-1)}/d_i) + \alpha\ \mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)}) \qquad (4.2)$$

$$\mathbf{X}_{[i]}^{(t)} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}^{(t-1)} \sim \text{Multinomial}(d_i, \mathbf{Y}_{[i]}^{(t)}) \qquad (4.3)$$

independently for each $i = 1, 2, \ldots, n_{CL}$.

where $Y^{(t)}_{[i]}$ is the vector of the *composite route choice probabilities*, including both habit and choice process, for user class *i*, with entries y_i.

Some remarks about model (4.1-4.3):

- the composite route choice probability for route *i*, y_i, is a random variable since it is a function of random variables;
- the composite route choice probability for route *i*, y_i, depends on route costs or disutilities, even though probability α does not (in more advanced models probability α may also depend on costs and disutilities, (see section 6 for further comments);
- according to equations (4.2) and (4.3) the choice behaviour of any two users of the same class or of different classes are assumed independent conditional on the remembered past states; still in the *unconditional* distribution all users' route choices may affect all the others through congestion, say the cost function introduced in the sub-section 2.1;
- equation (4.1) needs to be properly initialized as described in the sub-section 3.2

As established first by Cascetta (1989), if the random utility model **p**(.) is such that a non-zero probability is assigned to all feasible alternatives (as satisfied by regular random utility models defined on an infinite support), then the process above has a unique stationary probability distribution to which it converges, regardless of the initial conditions, that is it is *regular*.

Model (4.1-4.3) may be solved through Monte Carlo techniques. At this aim it is useful noting that a Multinomial random variable is obtained by independently repeating *n* times a Categorical (also called "generalized Bernoulli") random variable (in the very same way that a Binomial is obtained by independently repeating *n* times a Bernoulli r. v.).

On each day *t* first disutilities **Z** are updated through equation (4.1), and choice probabilities **Y** through equation (4.2). Then for each user class *i* the inverse distribution function method is applied d_i times to the categorical distribution defined by the choice probabilities, using a sample of d_i pseudo-random numbers uniformly distributed over [0,1]. This way, an unbiased estimate of the mean of the route flows **X** is obtained; the same approach allows us to estimate any other moment, such as variance, and the unique stationary distribution. This solution approach may be applied to real cases, provided that routes are explicitly enumerated. Solution methods not requiring such enumeration are still an open issue.

From standard properties of the Multinomial distribution the corresponding mean process is given by:

$$E[\mathbf{X}^{(t)}_{[i]} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}^{(t-1)}] = d_i\,((1-\alpha)(\mathbf{X}^{(t-1)}_{[i]}/d_i) + \alpha\,\mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)})) =$$
$$= E[\mathbf{X}^{(t)}_{[i]} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}^{(t-1)}] = (1-\alpha)\,\mathbf{X}^{(t-1)}_{[i]} + \alpha\, d_i\,\mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)})) \quad (i = 1, 2, \ldots, n_{CL}). \qquad (4.4)$$

It should be noted that $E[\mathbf{X}^{(t)}_{[i]} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}^{(t-1)}]$ in the above equation is a random variable since it is a function of random variables $\mathbf{Z}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$. Collecting equations (4.4) together across all classes, and using the notation for writing the demands introduced in section 2.1, it then follows that:

$$E[\mathbf{X}^{(t)} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}^{(t-1)}] = (1-\alpha)\,\mathbf{X}^{(t-1)} + \alpha\,\mathbf{D}\,\mathbf{p}(\mathbf{Z}^{(t-1)}). \qquad (4.4')$$

Applying a statistical identity to equation (4.4') above then yields an expression for the unconditional mean process:

$$\begin{aligned} E[\mathbf{X}^{(t)}] &= E[E[\mathbf{X}^{(t)}|\mathbf{Z}^{(t-1)}, \mathbf{X}^{(t-1)}]] = \\ &= (1-\alpha)\, E[\mathbf{X}^{(t-1)}] + \alpha\,\mathbf{D}\, E[\mathbf{p}(\mathbf{Z}^{(t-1)})] \end{aligned} \quad (4.5)$$

A stability analysis of the mean process (4.5) requires that it is put in a Markovian form to apply results from deterministic process theory (see appendix 2 for details). The DP model described in section 3.3 is an approximation to the mean process (4.5). An analysis of the variance will be the topic of a future paper.

If the above assumptions are combined with the dynamic choice process (3.1) and the ES filter (3.2) we get a 1-dependent Markov process but in continuous state space of flows and forecasted costs, as discussed in Cantarella and Cascetta (1995).

4.2 Asymptotic analysis

An asymptotic analysis of the mean process is carried-out in the next sub-sections, by further developing and extending the asymptotic analyses presented in Watling & Cantarella (2013), which drew on earlier work of Davis & Nihan (1993) and Hazelton & Watling (2004). It may be worth stressing that the purpose of this analysis is exploiting relationships with DP models not providing a solution method for real cases.

In order to make some progress in *analytically* capturing the evolution of this process, the analysis is based on an asymptotic analysis whereby we examine the behaviour of the process as the demands are scaled by $\varsigma > 0$, and so the scaled demands are denoted by $\varsigma \boldsymbol{d}$, and in particular the behaviour of the process as ς becomes large, but in a special sense.

Since simply scaling the demand alone would clearly change the nature of the demand/network being modelled, and so not give any meaningful results, what we analyse is what happens when the demand is 'scaled' for the purposes of modelling route choice, but the scaling is reversed when it is substituted in the congestion relationships. We might think of this process, intuitively, as one in which demands and link capacities are scaled *in tandem*, if we are adopting travel cost functions whose actual argument is the ratio of flow to capacity. Thus if the vector $\mathbf{x}_\varsigma = (x_{\varsigma 1}, x_{\varsigma 2}, \ldots, x_{\varsigma n})$ denotes the flows under a demand scaling of ς on the n routes of the network as above, then $w_{j\varsigma}(\mathbf{x}_\varsigma)$ denotes the travel cost on route j when the route flows are $\mathbf{x}_\varsigma$ for $j = 1, 2, \ldots, n$. Noting that reversing the scaling the route flow vector would be $\varsigma^{-1}\mathbf{x}_\varsigma$ we are thus motivated to consider functions of the form:

$$w_{j\varsigma}(\mathbf{x}_\varsigma) = w_j(\varsigma^{-1}\mathbf{x}_\varsigma)$$

where $w_i(.)$ is a function independent of ς (that is the underlying true route cost functions, as defined in section 2.1). We use $\mathbf{w}_\varsigma(\mathbf{x}_\varsigma) = (w_{1\varsigma}(\mathbf{x}_\varsigma), w_{2\varsigma}(\mathbf{x}_\varsigma), \ldots, w_{n\varsigma}(\mathbf{x}_\varsigma))^{\mathrm{T}}$ and $\mathbf{w}(\varsigma^{-1}\mathbf{x}_\varsigma) = (w_1(\varsigma^{-1}\mathbf{x}_\varsigma), w_2(\varsigma^{-1}\mathbf{x}_\varsigma), \ldots, w_n(\varsigma^{-1}\mathbf{x}_\varsigma))^{\mathrm{T}}$ to denote the corresponding vector mappings.

All the above presented equations (4.1-4.5) can easily be re-written taking into account the scaling factor as:

$$\mathbf{Z}^{(t-1)} = \textstyle\sum_{k=1,\ldots,m} \eta_k\, \mathbf{w}_\varsigma(\mathbf{X}_\varsigma^{(t-k)}) \quad (4.6)$$

$$\mathbf{Y}_{[i]}^{(t)} = (1-\alpha)\, (\mathbf{X}_{\varsigma[i]}^{(t-1)}/(\varsigma\, d_i)) + \alpha\ \mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)}) \quad (4.7)$$

$$\mathbf{X}^{(t)}_{\varsigma[i]} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}_{\varsigma}^{(t-1)} \sim \text{Multinomial}(\zeta\, d_i, \mathbf{Y}^{(t)}_{[i]}) \tag{4.8}$$

independently for each $i = 1, 2, \ldots, n_{CL}$

where notation $\mathbf{X}^{(t)}_{\varsigma[i]}$ highlights that the left hand side of equation (4.8) is slightly different from $\mathbf{X}^{(t)}_{[i]}$ in equation (4.3) since it depends on scale parameter ζ.
The corresponding mean process is given by:

$$\text{E}[\mathbf{X}^{(t)}_{\varsigma[i]} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}_{\varsigma}^{(t-1)}] = (1-\alpha)\, \mathbf{X}^{(t-1)}_{\varsigma[i]} + \alpha\, \varsigma\, d_i\, \mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)})) \quad (i = 1, 2, \ldots, n_{CL}). \tag{4.9}$$

$$\text{E}[\mathbf{X}_{\varsigma}^{(t)} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}_{\varsigma}^{(t-1)}] = (1-\alpha)\, \mathbf{X}_{\varsigma}^{(t-1)} + \alpha\, \varsigma\, \mathbf{D}\, \mathbf{p}(\mathbf{Z}^{(t-1)}). \tag{4.9'}$$

$$\text{E}[\mathbf{X}_{\varsigma}^{(t)}] = \text{E}[\text{E}[\mathbf{X}_{\varsigma}^{(t)} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}_{\varsigma}^{(t-1)}]] = $$
$$= (1-\alpha)\text{E}[\mathbf{X}_{\varsigma}^{(t-1)}] + \alpha\, \varsigma\, \mathbf{D}\, \text{E}[\mathbf{p}(\mathbf{Z}^{(t-1)})]. \tag{4.10}$$

After Cascetta (1989) the SP model (4.6, 7, 8) is regular. Furthermore, Davis & Nihan (1993) studied a wide class of stochastic process models, and showed that, as demand $\varsigma\mathbf{D} \to \infty$ in tandem with the network capacities, so the stationary distribution converges to a multivariate normal distribution with mean equal to the conventional Stochastic User Equilibrium (SUE) solution.

The process we shall analyse is an extension of that considered by Hazelton & Watling (2004); their process is exactly ours for the case $\alpha = 1$. Like them, we use Davis & Nihan's result to develop an asymptotic approximation to moments of the stationary distribution of the process, based only on knowledge of the SUE solution and other input data to the traffic assignment process. Watling & Cantarella (2013) further extended this work for the case of uncongested two-route networks and for congested two-route networks with $\alpha = 1$, deriving expressions to describe the dynamics of the process only in terms of its means, variances and covariances. The body of work above has been the motivation for our present analysis. In particular we shall aim to extend the analysis of Watling & Cantarella (2013) to the case of general networks for a general value of α ($0 < \alpha \leq 1$). However, differently from the goals of these works, we shall focus on a process in which only the first moment, the mean, is used to approximate the evolution of the process, with a particular goal to explore the relationship to 'equivalent' deterministic process models which neglect variability.

In order to do so we make the following distributional approximations, following Hazelton & Watling (2004), where assuming $\mathbf{w}_{\varsigma}(.)$ and $\mathbf{p}(.)$ to be continuously differentiable:

$$\mathbf{w}_{\varsigma}(\mathbf{X}) = \mathbf{w}_{\varsigma}(\mathbf{x}_{\text{SUE}}) + \varsigma^{-1}\, \mathbf{J}_{\mathbf{w}}\, (\mathbf{X} - \mathbf{x}_{\text{SUE}}) + \text{O}_{\text{p}}(\varsigma^{-0.5})$$

$$\mathbf{p}(\mathbf{Z}) = \mathbf{p}(\mathbf{w}_{\varsigma}(\mathbf{x}_{\text{SUE}})) + \mathbf{J}_{\mathbf{p}}\, (\mathbf{Z} - \mathbf{w}_{\varsigma}(\mathbf{x}_{\text{SUE}})) + \text{O}_{\text{p}}(\varsigma^{-0.5})$$

where

$\mathbf{x}_{\text{SUE}}$ is the (assumed) unique SUE solution satisfying: $\mathbf{x}_{\text{SUE}} = \zeta\, \mathbf{D}\, \mathbf{p}(\mathbf{w}_{\varsigma}(\mathbf{x}_{\text{SUE}}))$, consistent with defintion in section 2;

$\mathbf{J}_{\mathbf{w}} = \nabla_{\mathbf{x}}\, \mathbf{w}_{\varsigma}(\mathbf{x} = \mathbf{x}_{\text{SUE}})$ and $\mathbf{J}_{\mathbf{p}} = \nabla_{\mathbf{z}}\, \mathbf{p}(\mathbf{z} = \mathbf{w}_{\varsigma}(\mathbf{x}_{\text{SUE}}))$ are respectively the Jacobian matrix of $\mathbf{w}_{\varsigma}(.)$ evaluated at $\mathbf{x}_{\text{SUE}}$ and the Jacobian matrix of $\mathbf{p}(.)$ evaluated at $\mathbf{w}_{\varsigma}(\mathbf{x}_{\text{SUE}})$.

[Note that since these are statements about relationships between random variables, then so must the order notation logically be a statement about distributions. In particular we say a random variable $A = \text{O}_{\text{p}}(\varsigma^n)$ if there exists an $a > 0$ such that $\lim_{\varsigma\to\infty} \Pr(|A/\varsigma^n| > a) = 0$.]
In simple terms, this indicates that as $\varsigma \to \infty$ then we can regard the transformation $\mathbf{w}_{\varsigma}(\mathbf{X})$

of the random variable **X** as a *linear* transformation, given by the first order Taylor series approximation about the SUE solution.

From equation (4.10) we may obtain (as proved in appendix 4) that:

$$\varsigma^{-1}\,\alpha\,(\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_w})\,\boldsymbol{\mu}^* = \varsigma^{-1}\,\alpha\,(\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_w})\,\mathbf{x}_{\mathrm{SUE}} + O(\varsigma^{-0.5}) \;.$$

where in stationarity $\boldsymbol{\mu}^* = \boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)} = \boldsymbol{\mu}^{(t-2)} = \ldots = \boldsymbol{\mu}^{(t-m)}$ with $\boldsymbol{\mu}^{(t)} = \mathrm{E}[\mathbf{X}^{(t)}]$.

Now, as discussed in Bifulco et al (2013), the invertibility of the matrix $\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_w}$ is a condition that may be adopted for assuming uniqueness of the SUE solution (it is weaker than the commonly adopted conditions of positive definiteness of the Jacobian of cost function and the negative semi-definiteness of the choice probability function Jacobian). Thus, under the assumption that $(\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_w})^{-1}$ exists, and since $\alpha > 0$, we obtain:

$$\varsigma^{-1}\,\boldsymbol{\mu}^* = \varsigma^{-1}\,\mathbf{x}_{\mathrm{SUE}} + \alpha^{-1}(\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_w})^{-1}\,O(\varsigma^{-0.5}) \;.$$

This generalises the model and the result in Hazelton & Watling (2004) to include habit modelling, say $0 < \alpha \leq 1$; indeed the model in Hazelton & Watling (2004) turns out to be a special case where no kind of habit occurs, say $\alpha = 1$.
It implies that for large ς we have a justification to approximate $\varsigma^{-1}\,\boldsymbol{\mu}^*$ by $\varsigma^{-1}\,\mathbf{x}_{\mathrm{SUE}}$, since $\boldsymbol{\mu}^*$ and $\mathbf{x}_{\mathrm{SUE}}$ both grow with ς. The DP models (3.1, 3.3) and (3.1, 3.2) discussed in section 3.3 are an approximation to the asymptotic mean process above described.

4.3 Some numerical examples

This section reports the results of some numerical examples of the asymptotic behaviour of the SP model (4.6, 7, 8) comparing it with the DP model (3.1, 3). At this aim, it is useful to restate the model (4.6, 7, 8) as the following equivalent model with a slightly different definition of **X** as highlighted by notation **X'**:

$$\mathbf{Z}^{(t-1)} = \textstyle\sum_{k=1,\ldots,m} \eta_k\, \mathbf{w}(\mathbf{X}'^{(t-k)}) \qquad (4.11)$$

$$\mathbf{Y}_{[i]}^{(t)} = (1-\alpha)\,(\,\mathbf{X}'^{(t-1)}_{[i]}/d_i) + \alpha\;\mathbf{p}_{[i]}(\mathbf{Z}^{(t-1)}) \qquad (4.12)$$

$$\mathbf{X}'^{(t)}_{[i]} \mid \mathbf{Z}^{(t-1)}, \mathbf{X}'^{(t-1)} \sim (1/\zeta)\cdot \mathrm{Multinomial}((\zeta\, d_i),\; \mathbf{Y}_{[i]}^{(t)}) \qquad (4.13)$$

independently for each $i = 1, 2, \ldots, n_{CL}$

where $r_i = \zeta\, d_i$ is the numbers of users given scale factor ζ and demand flow d_i for each user class i. It is worth noting that this way equations (4.11) and (4.12) are equal to equations (4.1) and (4.2) respectively.

Figures 3 and 4 (cfr Figure 2) show the evolution of a route flow from day 75 to day 90, for m = 3 and 5 respectively, as a results of SP model (4.11, 12, 13) with r = 10, 100, 1000, 10000 users[3] and of the corresponding DP model (3.1, 3). The SP model has been solved through the Monte Carlo techniques already described. As expected from the above asymptotic analysis, results with the SP model with 10000 users are very close to those with the DP model. As the memory depth m increases from 3 to 5 the observed fluctuations become smaller for high numbers of users. Figure 5 (cfr Figure 2) shows the evolution of a route flow from day 75 to day 90, for m = ∞, that is with ES, as a results of SP

[3] The values of demand and route flows are irrelevant since the ratios flow/capacity remain unchanged.

model (4.11, 12, 13) with r = 10, 100, 1000, 10000 users and of the corresponding DP model (3.1, 3). The reported results are consistent with those in Figures 3 and 4.

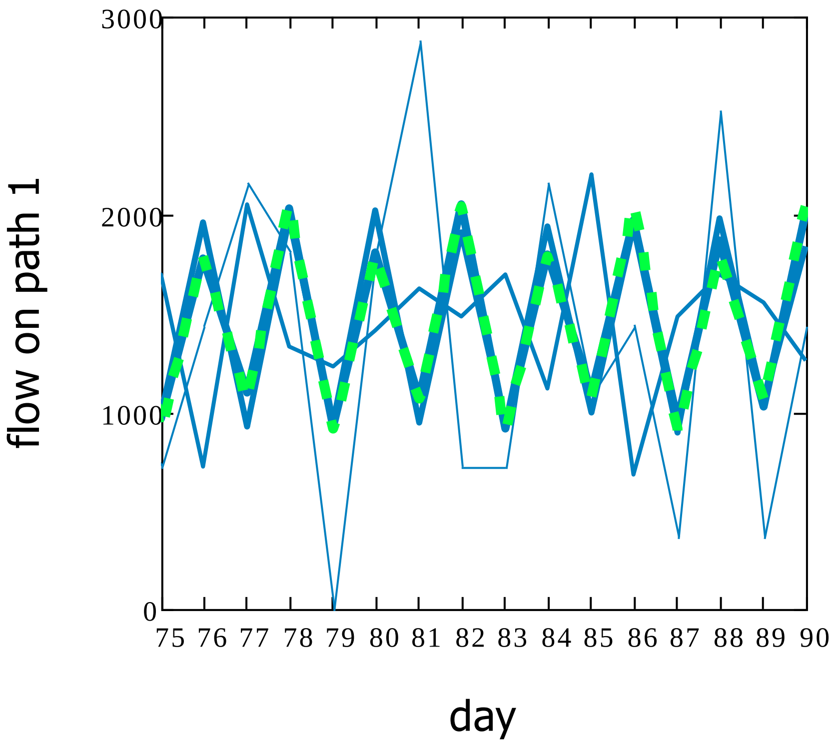

Figure 3. A route flow against day with MA and m = 3 given by SP with10, 100, 1000, 10000 users (thin to thick line), or DP (dashed line) as in Figure 2 (△).

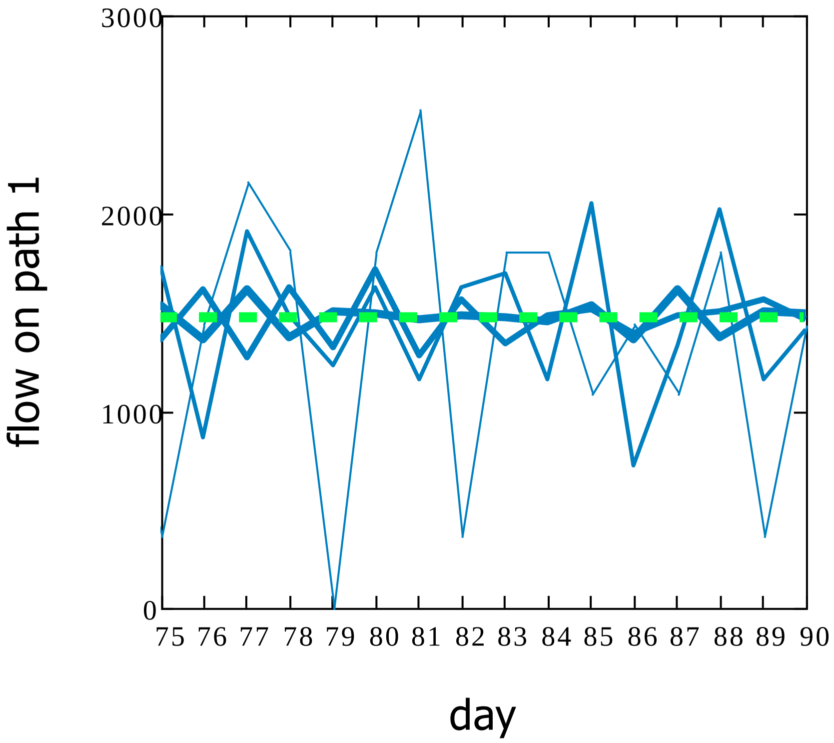

Figure 4. A route flow against day with MA and m = 5 given by SP with10, 100, 1000, 10000 users (thin to thick line), or DP (dashed line) as in Figure 2 (☐).

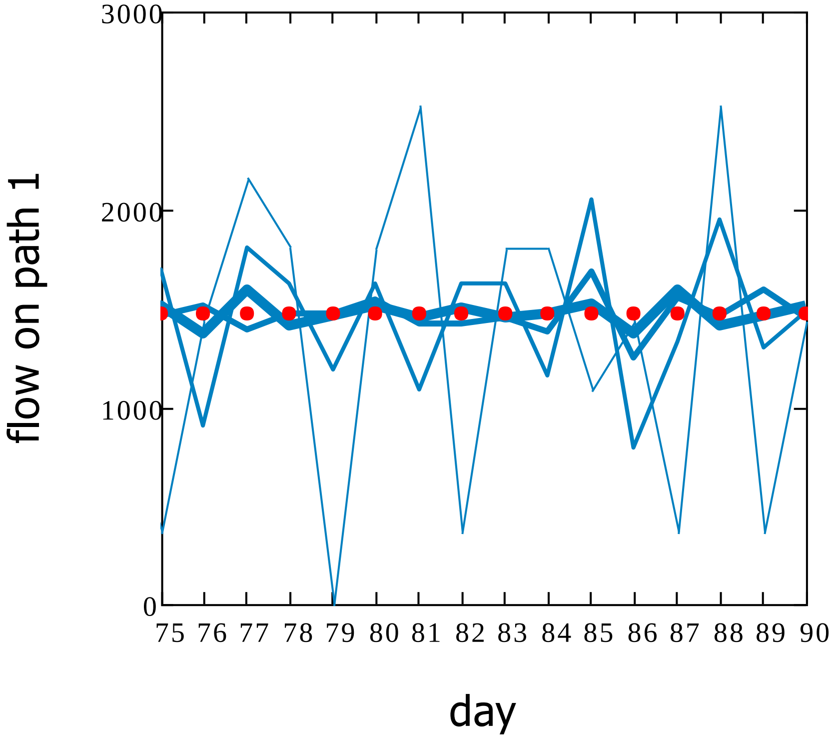

Figure 5. A route flow against day with ES given by SP with10, 100, 1000, 10000 users (thin to thick line) or DP (dotted line) as in Figure 2 (●).

5. Discussion

In the paper we have presented several technical results concerning stochastic process models, but in the midst of the technical details it can be easy to miss the key implications of the work. In this section we aim to draw out these implications.

Result 1: Asymptotic Mean Process Dynamics

(a) *Asymptotically, as demands/capacities grow, in the stochastic process model the mean flows* $\boldsymbol{\mu}^{(t)}$ *depend only on* $\{\boldsymbol{\mu}^{(t-k)} : k = 1,\ldots,m\}$, *and not on anything else in the previous history of the process (such as variances, covariances, etc. in previous days).*

(b) *Asymptotically, dependence of* $\boldsymbol{\mu}^{(t)}$ *on* $\{\boldsymbol{\mu}^{(t-k)} : k = 1,\ldots,m\}$ *is linear and can be expressed through knowledge of only: the SUE solution, the choice probability Jacobian evaluated at SUE, the cost function Jacobian evaluated at SUE, the value of* α, *the values of m and* $\{\eta_k: k = 1,2,\ldots,m\}$ *and the demand vector* $\mathbf{d}$. *Significantly, these are all input parameters, or—in the case of the SUE solution—something that can be readily derived from the input parameters without regard to the system dynamics.*

To our knowledge, these result are not explicitly stated, and not in this way, in any previous papers, and anyway no previous paper showed the technical results for a case including inertia effects (through α). It holds for a general network and for multiple classes.

It is important to appreciate that Result 1 holds for any model of learning weights based on the previous *m* days, as long as the weights sum to one. It is not necessary that the learning weights decay with time, for example. We could have a learning process, for example, where users put a weight of zero on the previous 4 days and a weight of one on the experience 5 days ago (as might occur, for example, in a model in which days are weekdays, and a user travelling on Fridays only learns from previous Fridays). It is also not necessary that the learning weights give rise to stable behaviour in the deterministic system above; still our dynamic equations hold for other types of system.

It is also important that Result 1 holds under a range of assumptions for the cost functions, cost function parameters, choice probability functions, choice probability function parameters, and value of α. This will mean that it holds to describe processes that are very different in nature, with very different kinds of trajectory.

Result 1(b) implies that even though there is a dependence of mean only on means, the nature of this dependence does depend on system parameters. Therefore if, for example, we have a problem with given cost functions and choice probability functions (so a specific, single SUE solution), and fix the learning weights $\{\eta_k: k = 1,2,\ldots,m\}$, then from some given initial conditions the mean process will yield a variety of different trajectories by varying α. These different trajectories will take a different amount of time to reach the given SUE solution (if the case) from some given starting conditions, and so Result 1(b) may be used to numerically compute (an estimate of) the amount of time that the mean process will be transient, from some given initial conditions and depending on the system parameters.

It should be remarked that Result 1 makes some 'distributional assumptions' concerning the form of the cost functions and the choice probability functions, and their relation to the demand multiplier. We also assume a unique SUE. It seems we are making some restrictions, therefore, and so it be interesting to make further exploration of Result 1(a) in terms of what happens beyond these restrictions, e.g. with asymmetric cost functions and multiple SUE (building on Watling, 1996).

Result 2: Asymptotic Result on Equilibrium of Mean Process and SUE

Asymptotically, the equilibrium of the mean process in Result 1 is an SUE, under certain conditions on the Jacobian of cost functions evaluated at SUE, the choice probability Jacobian evaluated at SUE, and the demand vector, but not on the behavioural parameters α *and* $\{\eta_k$: $k = 1,2,\ldots,m\}$.

This establishes that asymptotically, the point equilibria of the mean process are invariant to the behavioural parameters (α and $\{\eta_k$: $k = 1,2,\ldots,m\}$). This corresponds to results known for deterministic process models (Cantarella & Cascetta, 1995).

Result 3: Asymptotic Result on Equilibrium of Mean Process

The nature of the approximate mean process, namely whether it is dissipative and/or whether it converges to a stable fixed point, is determined by α, $\{\eta_k$: $k = 1,2,\ldots,m\}$*, and the Jacobian of the mean process evaluated at SUE.*

In contrast to Result 2, this result establishes that the behavioural parameters affect the nature of the approximate mean process, but that we can anticipate this nature from knowing the behavioural parameters. Again, this corresponds to what is already known for deterministic process models (Cantarella & Cascetta, 1995).

However, to balance these results we should also mention the following result:

Result 4: Asymptotically the variance of the process depends on more than the mean

It is *not* *true that in general: Asymptotically as demands/capacities grow, in the stochastic process model the* *variance* *of the process at time t depends only on* $\{\mathbf{\mu}^{(t-k)}$: $k = 1,\ldots,m\}$*, and not on anything else in the previous history of the process (such as variances, covariances etc. in previous days).*

In order to prove this negative result we may use a counter-example, for which we can refer to the analysis of two-link networks in Watling and Cantarella (2013). This negative result is important for addressing a relatively common mis-perception that modelling stochastic processes is akin to adding 'noise' to a deterministic process.

6. Conclusions and Research Perspectives

In this paper we have sought to integrate and extend various previous works concerning stability of deterministic processes and the analysis of stochastic processes. We have presented a generic stochastic process model for a general multi-class network, including notions of user habit, learning and choice, and have analysed this model theoretically by developing an asymptotic approximation for the mean process of such a model. We have shown how this approximating mean process relates to SUE. We have then used the tools of deterministic dynamical systems to analyse the mean process, and have shown how the nature of the learning process can be used to anticipate the nature of the mean process, including whether it is dissipative, converges to a stable fixed point, etc.

Even if the presented models and results are stated with respect to route costs, disutilities and flows, it seems worth noting that they hold as well with respect to link variables. Generally this is not the case for day-to-day dynamic models based on Wardrop route choice behaviour.

While the present paper has been wholly theoretical in nature, we believe that there is important future research in analysing such systems through Monte Carlo simulations of the process, as in the small numerical examples described at the end of sections 4 and 3. In doing so, insights may be obtained that would enrich and add to the theoretical insights provided here.

In particular, apart from open issues already mentioned above, linking to the Results we highlighted in section 5, we would suggest that interesting investigations would be to:

- explore the impact on the process trajectory of the cost function parameters, choice model parameters and α, and relate the findings to Result 1(a);
- explore analytically or numerically the transient time for the mean process to reach SUE from given starting conditions, as a function of α;
- explore the impact of changing parameters to increase the variance of the process and to see its consequential impact on the mean process, by changing the number of users (e.g. not large), or in cases with multiple SUE, or if the system is near-periodic, or the effect of the learning weights;
- explore the impact of parameters, such as α and β, differentiated by user class;
- re-examine Result 1(a) and extend results in sub-section 3.3 in the light of the various kinds of learning processes suggested by Horowitz (1984);
- re-examine Result 1(a) and extend results in sub-section 3.3 in the light of of habit models where α changes over time depending on (aggregate or disaggregate) difference between forecasted and experienced disutilities (as proposed in some papers on continuous-time DP models);
- illustrate and further explore Result 2 using the tests above;
- illustrate and explore both stable and unstable cases, and relate to Result 3;
- explore the strength of the dependence of the process on previous variance/ autocorrelations, and the extent to which knowledge of the mean is almost sufficient (related to Result 4).

Calibration of SP models is still open issue, see Parry & Hazelton (2013), and Shao et al. (2014) for some approaches to this problem.

Acknowledgements – Authors greatly appreciated the comments and suggestions received from the anonymous referees, which helped them to produce a much better paper; they also wish to thank the Guest Editor for his patience waiting for the revised version of the paper. In addition authors wish to thanks their institutions, University of Salerno and University of Leeds, for their support. This work was partially supported under UNISA local grant no. ORSA145550 (2014).

References

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.

Bifulco G.N., Cantarella G.E., Simonelli F., Velonà P. (2013). Equilibrium and day-to-day stability in traffic networks under ATIS. In: Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transportation Systems, Dresden, pp 503–512.

Bifulco, G.N., Cantarella, G.E., Simonelli, F. (2014). Design of signal setting and advanced traveler information systems. J. of Intelligent Transportation Systems 18(1), 30-40.

Cantarella G.E. (2013). Day-to-day dynamic models for Intelligent Transportation Systems design and appraisal. Transportation Research Part C 29, 117–130.

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. (2016). Modelling road traffic assignment as a day-to-day dynamic, deterministic process: a unified approach to discrete- and continuous-time models. In EURO J Transp Logist (2016) 5:69–98.

Cantarella G.E., Gentile G., Velonà P. (2010). Uniqueness of Stochastic User Equilibrium. In Proceedings of 5th IMA conference on Mathematics in Transportation. (London, UK, April 2010).

Cascetta E. & Cantarella G.E. (1991). A day-to-day and within-day dynamic stochastic assignment model. Transportation Research Part A 25, 277-291.

Davis G. & Nihan N. (1993). Large population approximations of a general stochastic traffic assignment model. Operations Research 41, 169-178.

Friesz, T.L., Kim, T., Kwon C., Rigdon, M.A. (2011). Approximate network loading and dual-time-scale dynamic user equilibrium. Transportation Research B 45, 176-207.

Guo Ren-Yong, Yang H., Huang Hai-Jun, Tan Z. (2015). Link-based day-to-day network traffic dynamics and equilibria. Transportation Research Part B 71, 248–260.

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.

Horowitz J.L. (1984).The stability of stochastic equilibrium in a two-link transportation network. Transportation Research Part B 18, 13-28.

Liu R., Van Vliet D. & Watling D. (2006). Microsimulation models incorporating both demand and supply side dynamics. Transportation Research Part A 40, 125-150.

Parry K. & Hazelton M. L. (2013). Bayesian inference for day-to-day dynamic traffic models. Transportation Research Part B 50, 104–115

Shao H., Lam W.H.K., Sumalee A. Chen Anthony (2014). Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts. Transportation Research Part B 68, 52–75.

Watling D.P. (1996). Asymmetric problems and stochastic process models of traffic assignment. Transportation Research Part B 30, 339-357.

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

Watling D.P. & Cantarella G.E. (2015). Model Representation & Decision-Making in an Ever-Changing World: The Role of Stochastic Process Models of Transportation Systems. In Networks and Spatial Economics, 15 (3), pp. 843-882.

Appendix 1: brief review of deterministic processes

A brief review of main definitions regarding deterministic processes is reported below. (for more details see for instance Stokey and Lucas (1989) among many others).

A **deterministic process** is a discrete-time time-homogeneous Markovian non-linear system, and may be specified as:

$\mathbf{x}^t = \varphi(\mathbf{x}^{t-1})$ $\quad (t \in \mathbb{N};\ \mathbf{x}^t \in \mathcal{S} \subseteq \Re^n$

where today *state*, $\mathbf{x}^t$, depends on yesterday state, $\mathbf{x}^{t-1}$ only (Markovian systems) through the *transition function* $\varphi(\cdot)$ from the *state space* $\mathcal{S}$ to the state space $\mathcal{S}$.

Any time-discrete system with finite memory (today state, $\mathbf{x}^t$ directly depends on a finite number of previous days states) may still be formulated as a deterministic process, with a duly specification of the system state to include (finite) memory of the past states (an example is given in sub-section 3.3). (For more details see Cantarella and Watling, 2015.)

A **deterministic process** with a differentiable transition function is called **differentiable.** Let $\mathbf{J}$ = Jac[$\varphi(\cdot)$] be the Jacobianmartix of the transition function, and λ_j be one of the n eigenvalues of $\mathbf{J}$, omitting the argument $\mathbf{x}$, then det($\mathbf{J}$) $= \prod_j \lambda_j$; moreover let

$\delta(\mathbf{J}) = |\ \det(\mathbf{J})\ | = \prod_j |\ \lambda_j\ |$ be the absolute value of the determinant of matrix $\mathbf{J}$;
$\rho = \max_j |\ \lambda_j\ |$ be the spectral radius of the determinant of matrix $\mathbf{J}$,
$\nu_\tau = |||\ \mathbf{J}\ |||$ be any matrix norm of matrix, where subscript τ highlights that there exist several different norms.

It results that: if $\nu_\tau < 1$ for some τ then $\rho < 1$, and if $\rho < 1$ then $\delta < 1$.

A **self reproducing set** (**srs**) of states is a subset S of the state space $\mathcal{S}$ having the following properties:

- has a dimension strictly less than the dimension of the state space, n;
- the system cannot evolve towards a state out of the srs starting from its interior;
- the srs is minimal, that is it does not strictly include any other srs.

An **attractor** is a srs that

- has an *attraction domain* (also called basin of attraction), which is a proper super-set of the srs such that from any initial state belonging to the domain the system converges towards the srs; the attraction domain may be a proper sub-set of the state space.

An example of the above definitions is a fixed-point state, $\mathbf{x}^* = \varphi(\mathbf{x}^*)$, which is an attractor if it has a an attraction domain, otherwise it is a repulsor, if from any other initial state the system diverges from the fixed-point state, or a saddle, if from some initial states the system converges to the fixed-point state and from others diverges from it.

More generally, there are four main **types of attractors**:

- *fixed-point attractors*: the system always takes up the same point;
- *k-periodic attractors*: the system periodically moves among k points;
- *quasi-periodic attractors*: the system moves on a toroidal surface containing infinite many points;
- *a-periodic attractors*: the system moves within a fractal set.

The basic analysis usually carried out about a deterministic process considers fixed-point states and sees if they are attractors. If they are not attractors it is necessary to verify whether the system converges and towards which attractors if it does (see for instance Cantarella, 2013, for more details).

A deterministic process is called **dissipative** if a (small enough) ball round the initial state will shrink as the system evolves; in this case the system will converge to an attractor (but not necessarily a fixed-point), possibly depending on the starting state. A sufficient condition is:

$$\delta(\mathbf{J}(\mathbf{x})) < 1 \qquad \forall \mathbf{x}$$

with $\delta \neq 0$; otherwise, $\delta = 0$, the analysis should be moved to a space with reduced dimensions where $\delta \neq 0$, through proper linear transformations. A deterministic process may be dissipative from a sub-set of the state space only.

A fixed point is called **locally stable** if it is an attractor, say if from any initial state belonging to the attraction domain the system converges towards it. A sufficient condition is:

$$\rho(\mathbf{J}(\mathbf{x})) < 1 \qquad \forall \mathbf{x}$$

A fixed point is called **globally stable** if it is an attractor and its attraction domain is the whole state space, say from any initial state belonging to the state space the system converges towards it. A sufficient conditions is:

$$\nu_\tau (\mathbf{J}(\mathbf{x})) < 1 \quad \text{for some } \tau \qquad \forall \mathbf{x}$$

This conditions actually means that $\boldsymbol{\varphi}(\cdot)$ is strictly non-expansive (an extension of uniformly non-expansive or contraction as in Banach theorem, which also guarantees uniqueness and the rate of convergence). (This feature actually assures that a Lyapunov function exists, which is a more general global stability condition.)

It can easily noted that the sufficient condition for a fixed-point being globally stable implies the sufficient condition for it being locally stable, which in turn implies the condition for dissipativeness.

References

Stokey NL, Lucas RE (1989). **Recursive methods in economic dynamics**. Harvard University Press, Cambridge.

Appendix 2: dissipativeness of DP (3.4, 3.5)

Since DP (3.4, 3.5) is Markovian its dissipativeness may be analysed looking at its Jacobian matrix $\mathbf{J}$; its structure is given below for m = 4, where $\mathbf{G}_k = \mathbf{D}\ \mathbf{J_p}\ \mathbf{J_c}(^k\mathbf{x}^{(t-1)})$ and its entries depend on system parameters, such as demand flows, link capacities.

	block 1 (*day t–1*)	block 2 (*day t–2*)	block 3 (*day t–3*)	block 4 (*day t–4*)
block 1 *(day t)*	$\alpha\ \eta_1\ \mathbf{G}_1 + (1-\alpha)\ \mathbf{I}$	$\alpha\ \eta_2\ \mathbf{G}_2$	$\alpha\ \eta_3\ \mathbf{G}_3$	$\alpha\ \eta_4\ \mathbf{G}_4$
block 2 *(day t–1)*	**I**	**0**	**0**	**0**
block 3 *(day t–2)*	**0**	**I**	**0**	**0**
block 4 *(day t–3)*	**0**	**0**	**I**	**0**

From matrix algebra the absolute value of the determinant of the above Jacobian matrix $\mathbf{J}$ $|\det(\mathbf{J})|$ is equal to absolute value of the determinant of the following matrix $\mathbf{J'}$ obtained through properly interchanging some columns

	block 1 (*day t–1*)	block 2 (*day t–2*)	block 3 (*day t–3*)	block 4 (*day t–4*)
block 1 *(day t)*	$\alpha\ \eta_4\ \mathbf{G}_4$	$\alpha\ \eta_1\ \mathbf{G}_1 + (1-\alpha)\ \mathbf{I}$	$\alpha\ \eta_2\ \mathbf{G}_2$	$\alpha\ \eta_3\ \mathbf{G}_3$
block 2 *(day t–1)*	**0**	**I**	**0**	**0**
block 3 *(day t–2)*	**0**	**0**	**I**	**0**
block 4 *(day t–3)*	**0**	**0**	**0**	**I**

In addition, remembering from block-matrix analysis that

if $\mathbf{M}$ =

$\mathbf{M}_{11}$	$\mathbf{M}_{12}$
$\mathbf{M}_{21}$	$\mathbf{M}_{22}$

and $\det(\mathbf{M}_{22}) \neq 0$ then $\det(\mathbf{M}) = \det(\mathbf{M}_{22})\ \det(\mathbf{M}_{11} - \mathbf{M}_{12}\ \mathbf{M}_{22}^{-1}\ \mathbf{M}_{21})$, thus it yields that

$$|\det(\mathbf{J})| = \alpha\ \eta_m \det(\mathbf{G_m})$$

Omitting arguments for simplicity's, $\mathbf{G_m} = \mathbf{D}\ \mathbf{J_p}\ \mathbf{B}^T\ \mathbf{J_C}\ \mathbf{B}^T$ is singular, $\det(\mathbf{G_m}) = 0$, since $\mathbf{J_p}$ is singular, $\det(\mathbf{J_p}) = 0$, due to the normalization of probabilities; thus the above result, however elegant, is rather useless as such.

The DP (3.4, 3.5) can be reformulated avoiding redundant route variables. Indeed one route choice probability or flow is redundant because it may easily be obtained from the others (for each class there are at least two routes). After Cantarella *et al.* (2010), for each user class i any of the first $\tilde{n}_i = n_i - 1$ routes is called an independent route (iro), and the equations below hold with respect to iro choice probability and flow vectors, $\tilde{\mathbf{p}}_{[i]}$ and $\tilde{\mathbf{x}}_{[i]}$:

$$\tilde{\mathbf{p}}_{[i]} = \mathbf{E}_{[i]}\ \mathbf{p}_{[i]} \qquad \tilde{\mathbf{x}}_{[i]} = \mathbf{E}_{[i]}\ \mathbf{x}_{[i]} \qquad \forall i$$

$$\mathbf{p}_{[i]} = \mathbf{L}_{[i]}\ \tilde{\mathbf{p}}_{[i]} + \mathbf{e}_{[i]} \qquad \mathbf{x}_{[i]} = \mathbf{L}_{[i]}\ \tilde{\mathbf{x}}_{[i]} + \mathbf{x_{b[i]}} \qquad \forall i$$

where, given $\mathbf{I}_i$ the $(n_i \times n_i)$ identity matrix, and $\tilde{n}_i = n_i - 1$;
$\mathbf{E}_{[i]}$ is the $(\tilde{n}_i \times n_i)$ matrix obtained by dropping the last row from the identity matrix $\mathbf{I}_i$;
$\mathbf{e}_{[i]}^T = [0, 0, \ldots, 1]$ is a $(n_i \times 1)$ vector, given by the last row of the identity matrix $\mathbf{I}_i$;
$\mathbf{1}_{[i]} = [1, 1, \ldots, 1]^T$ is an $(n_i \times 1)$ vector with all entries equal to one;
$\mathbf{L}_{[i]} = (\mathbf{I}_{[i]} - \mathbf{e}_{[i]}\ \mathbf{1}_{[i]}^T)\ \mathbf{E}_{[i]}^T$ is a $(n_i \times \tilde{n}_i)$ matrix obtained from the $(\tilde{n}_i \times \tilde{n}_i)$ identity matrix by adding at the bottom one more row $(1 \times \tilde{n}_i)$ with all entries equal to -1;
$\mathbf{x_{b[i]}} = d_i\ \mathbf{e}_{[i]}$ is a $(n_i \times 1)$ vector.

Collecting the above vectors and matrices into block vectors or matrices we get:

$\tilde{\mathbf{p}} = \mathbf{E}\,\mathbf{p}$ $\qquad\qquad$ $\tilde{\mathbf{x}} = \mathbf{E}\,\mathbf{x}$

$\mathbf{p} = \mathbf{L}\,\tilde{\mathbf{p}} + \mathbf{e}$ $\qquad\qquad$ $\mathbf{x} = \mathbf{L}\,\tilde{\mathbf{x}} + \mathbf{x_b}$

where

$\tilde{\mathbf{p}}$ is the iro choice probability block vector with a $(\tilde{n}_i \times 1)$ block for each user class $\tilde{\mathbf{p}}_{[i]}$;
$\tilde{\mathbf{x}}$ is the iro flow block vector with a $(\tilde{n}_i \times 1)$ block for each user class $\tilde{\mathbf{x}}_{[i]}$;
$\mathbf{E}$ is a row-block matrix with a $(\tilde{n}_i \times n_i)$ block for each user class $\mathbf{E}_{[i]}$;
$\mathbf{e}$ is a block vector with a $(n_i \times 1)$ block for each user class $\mathbf{e}_{[i]}$;
$\mathbf{L}$ is a $(n \times \tilde{n})$ diagonal-block matrix with a $(n_i \times \tilde{n}_i)$ block for each user class, $\mathbf{L}_{[i]}$;
$\mathbf{x_b}$ is a block vector with a $(n_i \times 1)$ block for each user class $\mathbf{x_b}_{[i]}$.

Thus an equivalent formulation of the DP (3.4, 3.5) with respect to iro flows is given by:

$$^{1}\tilde{\mathbf{x}}^{(t)} = (1-\alpha)\,^{1}\tilde{\mathbf{x}}^{(t-1)} + \alpha\,\mathbf{E}\,\mathbf{D}\,\mathbf{p}(\Sigma_{k=1,\dots,m}\,\eta_k\,\mathbf{w}(\mathbf{L}\,^{k}\tilde{\mathbf{x}}^{(t-1)} + \mathbf{x_b})) \qquad (A.1)$$

$$^{h+1}\tilde{\mathbf{x}}^{(t)} = {}^{h}\tilde{\mathbf{x}}^{(t-1)} \qquad h = 1, \dots, m-1 \qquad (A.2)$$

In most cases, such as for invariant RUM's, for a user class i the route choice probabilities $\mathbf{p}_{[i]}$ do not actually depend on forecasted disutilities $\mathbf{z}_{[i]}$ but only on their differences. Let

$\tilde{\mathbf{z}}_{[i]} = \mathbf{L}_{[i]}^{\mathrm{T}}\ \mathbf{z}_{[i]}$ be the vector of iro disutility differences, with an entry for each iro given by the iro disutility minus the disutility of the last route, with

$$\tilde{\mathbf{z}}_{[i]} = \mathbf{L}_{[i]}^{\mathrm{T}}\ \mathbf{z}_{[i]} = \mathbf{L}_{[i]}^{\mathrm{T}}\,\mathbf{w}_{[i]}(\mathbf{x}) \qquad (A.3)$$

Thus the iro choice probabilities $\tilde{\mathbf{p}}_{[i]}$ may be specified as a function of the iro disutilites differences, $\tilde{\mathbf{z}}_{[i]}$, as $\tilde{\mathbf{p}}_{[i]} = \tilde{\mathbf{p}}_{[i]}(\tilde{\mathbf{z}}_{[i]})$, thus:

$$\mathbf{p}_{[i]} = \mathbf{L}_{[i]}\,\tilde{\mathbf{p}}_{[i]}(\tilde{\mathbf{z}}_{[i]}) + \mathbf{e}_{[i]} \qquad (A.4)$$

Thus combining equations (A.1), (A.2), (A.3), and (A.4) an equivalent formulation of the DP (3.4, 3.5) with respect to iro flows and disutilites is given by:

$$^{1}\tilde{\mathbf{x}}^{(t)} = (1-\alpha)\,^{1}\tilde{\mathbf{x}}^{(t-1)} + \alpha\,\tilde{\mathbf{D}}\,\tilde{\mathbf{p}}(\Sigma_{k=1,\dots,m}\,\eta_k\,\mathbf{L}^{\mathrm{T}}\mathbf{w}(\mathbf{L}\,^{k}\tilde{\mathbf{x}}^{(t-1)} + \mathbf{x_b})) \qquad (A.5)$$

$$^{h+1}\tilde{\mathbf{x}}^{(t)} = {}^{h}\tilde{\mathbf{x}}^{(t-1)} \qquad h = 1, \dots, m-1 \qquad (A.6)$$

$\tilde{\mathbf{D}} = \mathbf{E}\ \mathbf{D}\ \mathbf{L}$, being a block diagonal matrix of dimensions ($(n - n_{CL}) \times (n - n_{CL})$), with each block given by $\tilde{\mathbf{D}}_{[i]}$ being a diagonal matrix of dimensions $\tilde{n}_i \times \tilde{n}_i$, with entries on the main diagonal equal to $d_i > 0$.

DP (A.4, A.5) looks like DP (3.4, 3.5) and shares with it the Jacobian matrix structure with reference to matrices $\tilde{\mathbf{G}}_{\mathrm{k}} = \tilde{\mathbf{D}}\ \tilde{\mathbf{J}}_{\mathbf{p}}\ (\mathbf{L}^{\mathrm{T}}\,\mathbf{J_w}(\mathbf{L}\ ^{\mathbf{k}}\tilde{\mathbf{x}}^{(t-1)} + \mathbf{x_b})\ \mathbf{L})$ where $\tilde{\mathbf{J}}_{\mathbf{p}} = \nabla_{\mathbf{z}}\,\tilde{\mathbf{p}}(\tilde{\mathbf{z}})$ is the Jacobian matrix of choice function $\tilde{\mathbf{p}}(.)$ evaluated at $\tilde{\mathbf{z}} = \Sigma_{k=1,\dots,m}\ \eta_k\ \mathbf{L}^{\mathrm{T}}\ \mathbf{w}(\mathbf{L}\ ^{\mathbf{k}}\tilde{\mathbf{x}}^{(t-1)} + \mathbf{x_b})$, and $\mathbf{J_w} = \nabla_{\mathbf{x}}\,\mathbf{w}(\mathbf{x})$ is the Jacobian matrix of the route cost flow function, thus:

$$\det(\tilde{\mathbf{G}}_{\mathrm{k}}) = \det(\tilde{\mathbf{D}})\,\det(\tilde{\mathbf{J}}_{\mathbf{p}})\,\det(\mathbf{L}^{\mathrm{T}}\,\mathbf{J_w}\,\mathbf{L}) \qquad (A.7)$$

It is worth noting that in general $\det(\tilde{\mathbf{D}}) > 0$, and $\det(\tilde{\mathbf{J}}_{\mathbf{p}}) \neq 0$ under mild assumptions, say $\tilde{\mathbf{p}}_{[i]}(\tilde{\mathbf{z}}_{[i]})$ is a strictly positive invariant RUM for each user class i, as proved in the below.

Indeed, remembering from matrix algebra that if a real symmetric matrix has strictly positive diagonal entries and is strictly column diagonally dominant it is

positive semi-definite and non-singular thus positive definite, it suffices to observe (after Cantarella, 1997) that given $\tilde{v}_j = -\tilde{z}_{j]}$:
$\partial\tilde{p}_i(\tilde{v}) / \partial\tilde{v}_j = \partial\tilde{p}_j(\tilde{v}) / \partial\tilde{v}_i$ for an invariant RUM,
$\partial\tilde{p}_i(v) / \partial\tilde{v}_i > 0$ with $\partial\tilde{p}_j(v) / \partial\tilde{v}_i < 0 \quad \forall i \neq j$ for a strictly positive RUM,
$\partial\tilde{p}_i(\tilde{v}) / \partial\tilde{v}_i > \sum_{j\neq i} | \tilde{p}_j(\tilde{v}) / \partial\tilde{v}_i |$
since $\sum_j \partial\tilde{p}_j(\tilde{v}) / \partial\tilde{v}_i > 0 \Rightarrow \partial\tilde{p}_i(\tilde{v}) / \partial\tilde{v}_i > -\sum_{j\neq i} \partial\tilde{p}_j(\tilde{v}) / \partial\tilde{v}_i$
because $\sum_j \partial p_j(v) / \partial v_i = 0$, and $\partial p_{ni}(v) / \partial v_i < 0$,
where $\partial\tilde{p}_j(\tilde{v}) / \partial\tilde{v}_i < 0 \quad \forall i \neq j$. ☐

Matrix $\mathbf{L}^T \mathbf{J_w} \mathbf{L}$, say $\det(\mathbf{L}^T \mathbf{J_w} \mathbf{L})$, is further analysed below. Let

$\mathbf{B}$ be the row-block link-route incidence matrix with a block $\mathbf{B}_{[i]}$ for each class i;
$\tilde{\mathbf{B}} = \mathbf{B}\,\mathbf{L}$ be the $(n_{LINK} \times \tilde{m})$ the row-block link – independent route incidence matrix with a block $\tilde{\mathbf{B}}_{[i]}$ for each class i;

the Jacobian matrix of the iro cost flow function $\mathbf{L}_{[i]}^T\, \mathbf{w}_{[i]}(\mathbf{x})$ is given by:

$$\tilde{\mathbf{J}}_\mathbf{w} = \mathbf{L}^T \mathbf{J_w} \mathbf{L} = \tilde{\mathbf{B}}^\mathbf{T} \mathbf{J_c} \tilde{\mathbf{B}} \tag{A.8}$$

since $\mathbf{J_w} = \mathbf{B^T} \mathbf{J_c} \mathbf{B}$, with $\mathbf{J_c} = \nabla_\mathbf{f}\, \mathbf{c}(\mathbf{f})$ being the Jacobian matrix of the link cost flow function.

According to equation (A.8), equation (A.7) becomes:

$$\det(\tilde{\mathbf{G}}_k) = \det(\tilde{\mathbf{D}}) \det(\tilde{\mathbf{J}}_\mathbf{p}) \det(\tilde{\mathbf{B}}^\mathbf{T} \mathbf{J_c} \tilde{\mathbf{B}}) \tag{A.9}$$

Assuming that $\mathbf{J_c}$ is a (not necessarily symmetric) positive definite matrix with respect to real vectors, thus it is non singular, $\det(\mathbf{J_c}) \neq 0$, two cases may occur, as discussed below.

• The rank of the link-iro incidence matrix is equal to the number of iro's, $\text{rank}(\tilde{\mathbf{B}}) = r = \tilde{m}$, thus $r = \tilde{m} \leq n_{LINK}$, and $\tilde{\mathbf{B}}$ is full rank. In this case, matrix $\tilde{\mathbf{B}}^\mathbf{T} \mathbf{J_c} \tilde{\mathbf{B}}$ is a (not necessarily symmetric) positive definite matrix with respect to real vectors, thus it is non singular.

Indeed if $\mathbf{Q}$ is a $(n \times n)$ (not necessarily symmetric) positive definite matrix with respect to real vectors (but necessarily with respect to complex vectors too):

$$\mathbf{x}^T \mathbf{Q}\, \mathbf{x} > 0 \qquad \forall\, \mathbf{x} \neq 0, \mathbf{x} \in \Re^n$$

and if $\mathbf{M}$ is a $(n \times m)$ full rank matrix with $m \leq n$, then the $(m \times m)$ matrix $\mathbf{M}^T \mathbf{Q}\, \mathbf{M}$ is positive definite matrix with respect to real vectors:

$$\mathbf{y}^T \mathbf{M}^T \mathbf{Q}\, \mathbf{M}\, \mathbf{y} > 0 \qquad \forall\, \mathbf{y} \neq 0, \mathbf{y} \in \Re^m$$

since $\mathbf{M}\,\mathbf{y} > \mathbf{0}\ \forall\, \mathbf{y} \neq 0, \mathbf{y} \in \Re^m$. ☐

Thus, in this case $\det(\tilde{\mathbf{G}}_k) \neq 0$.

• The rank of the link-iro incidence matrix is less than the number of iro's, $\text{rank}(\tilde{\mathbf{B}}) = r < \tilde{m}$. In this case, matrix $\tilde{\mathbf{B}}$ may be expressed as the product of two full rank matrices both with rank r through a rank factorization: $\tilde{\mathbf{B}} = \tilde{\mathbf{B}}_\mathbf{1} \times \mathbf{R}$.

Indeed, a $(n \times m)$ matrix $\mathbf{Q}$ with rank $r \leq \min(n, m)$ contains r linearly independent columns making up the $(n \times r)$ full rank sub-matrix $\mathbf{Q_1}$ so that

$\mathbf{Q} = [\mathbf{Q_1} \mid \mathbf{Q_2}]$, where the $(n \times (m - r))$ matrix $\mathbf{Q_2}$ contains the linearly dependent columns, with $\mathbf{Q_2} = \mathbf{Q_1} \times \mathbf{A}$ for a suitable $(r \times (m - r))$ full rank matrix $\mathbf{A}$. Hence $\mathbf{Q} = [\mathbf{Q_1} \mid \mathbf{Q_2}] = \mathbf{Q_1} \times [\mathbf{I_r} \mid \mathbf{A}]$, where the $(r \times n)$ matrix $\mathbf{R} = [\mathbf{I_r} \mid \mathbf{A}]$ is full rank. ☐

In this case DP (A.5, A.6) can be reformulated in the space of the iro's corresponding to the r linearly independent columns of $\tilde{\mathbf{B}}$ through a linear transformation defined by matrix $\mathbf{R}$. In this space the reformulated DP (A.5, A.6) leads to $\det(\tilde{\mathbf{G}}_k) \neq 0$, properly redefining matrix $\tilde{\mathbf{G}}_k$ (details are not explicitly reported for brevity's sake).

Some assumptions about the link-route incidence matrix are useful to reduce the number of linearly dependent columns (or rows):

1. each link belongs to at least a route, thus no row is null,
2. each route contains at least one arc, thus no column is null,
3. no pair of routes are equal, thus no pair of columns are equal,
4. no route is properly contained in another route (a);

All the above assumptions are quite mild and/or reasonable and can easily be accepted. On the other hand two links may well have equal rows if the share all routes, as it occurs for instance for two link in series or in parallel.

From the above considerations, with reference to the Jacobian matrix $\mathbf{J}°$ of the DP (A.5, A.6), possibly re-adapted to the appropriate space, we get:

$$|\det(\mathbf{J}°)| = \alpha\, \eta_m\, |\det(\tilde{\mathbf{G}}_\mathbf{m})| \neq 0$$

If the partial derivatives in matrix $\tilde{\mathbf{G}}_\mathbf{m}$ are well-defined, say finite and continuous, the absolute value of the determinant of matrix $\tilde{\mathbf{G}}_\mathbf{m}$, say $|\det(\tilde{\mathbf{G}}_\mathbf{m})|$, is a continuous function defined over a compact set, thus $|\det(\tilde{\mathbf{G}}_\mathbf{m})|$ has an upper bound, g_{MAX}, (and a lower bound, as well), and $\alpha\, \eta_m\, |\det(\tilde{\mathbf{G}}_\mathbf{m})| \leq \alpha\, \eta_m\, g_{MAX}$, and $\alpha\, \eta_m\, g_{MAX} < 1$ implies $|\det(\mathbf{J}°)| < 1$. Value of g_{MAX} cannot easily computed, an approximation may be obtained through matrix norms; this issue will be discussed in a future paper.

It can easily demonstrated that η_m is decreasing with m, and $\lim_{m\to\infty} \eta_m = 0$ [authors wish to thanks an anonymous reviewer who raised this point providing the mathematical details], thus whichever the value of β is, there always exists a large enough memory depth m^* such that for any memory deeper than this value m^* the system is *dissipative* from any starting state. The minimum memory depth value m^* is defined by:

$$m^* = \min\{m:\ \eta_m < 1 / (\alpha\, g_{MAX})\}$$

It is worth noting that the entries of matrix $\tilde{\mathbf{G}}_\mathbf{m}$ as well as g_{MAX} have no dimension, thus the above condition is not affected by the units used to measure flows or costs.

A similar analysis can be carried out with respect to link variables.

Appendix 3: features of the toy network.

This appendix describes basic data of the toy network used to develop the examples in the main text. The network (a simple representation of North-South motorway connections from Napoli to Salerno in Italy) is described by the graph in Figure. A.1, with four nodes, {A, B, C, D}, and five arcs, {1 = (A,C), 2 = (B,C), 3 = (B,D), 4 = (A,B), 5 = (C,D)}.

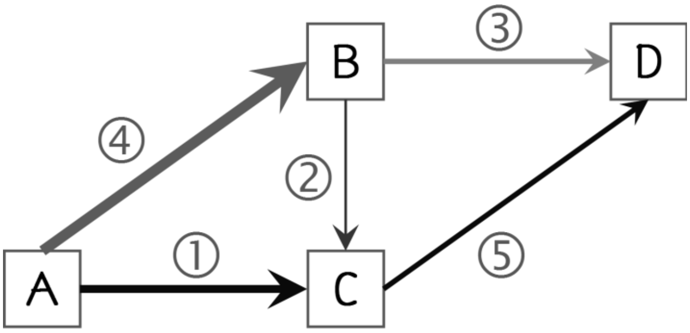

Figure A.1. *Network used in the examples.*

Davidson (hyperbolic) travel time function describes cost flow relationship for each arc *a*:

$c_a = c_{o,a}(1 + \chi_a\ f_a\ /\ (cap_a - f_a))$ for $f_a \in\ ]0, cap_a[$

where $c_{o,a}$ is the null flow cost, cap_a is the capacity of the arc, and χ_a a shape parameter. Since this function shows a vertical asymptote at capacity, a first order approximation is commonly considered when the flow to capacity ratio is greater that a pre-fixed threshold in the range [0,1[; this parameter models how congestion affects costs (a null value meaning no effect at all); a value 0.80 is used. According to these assumptions the arc cost function is continuously differentiable (thus continuous) and strictly increasing.

Only one O-D pair (A-D) is considered connected by three paths, {A-C-D, A-B-C-D, A-B-D}, with flow equal to 0.75 of the maximum flow that can traverse the network at saturation. Path choice behaviour is modelled by an invariant Logit choice model with dispersion parameter $\theta = (\sqrt{6}\ /\ \pi)\ \sigma = 0.780\ \sigma$, where σ is the standard deviation of path perceived utility; it is assumed that $\sigma = 0.30 \times 35$, where 35 is the cost of the shortest path, thus $\theta \cong 8$. Under these assumptions the path choice function is continuous, strictly positive, and increasing with symmetric (positive semi-definite) Jacobian.

The dynamic choice process is modelled through the ES(α) (3.1) with $\alpha = 0.60$. The learning behaviour process is modelled through the ES(β) (3.2) or the MA(β, *m*) (3.3) filter with $\beta = 0.40$ in both cases, different values of *m* are considered. The MA is initialized by applying the ES filter (3.2) for *m* days to fill the *m* columns of cost memory matrix.

Sufficient conditions for existence and uniqueness of one fixed-point state are satisfied by above assumptions. The DP model based on ES evolves towards the unique fixed-point attractor equal to the SUE; this may not be case for DP with MA. When convergence occurs, it requires less than 75 days; thus, in all examples in the main text the evolution over time on path 1 / arc 1 is shown from day 75 to day 90. Results shown are not affected by the initial states, as shown by some examples not reported for brevity's sake.

Appendix 4: proof of results in sub-section 4.2

Recalling equation (4.10) in the main text:

$$\mathrm{E}[\mathbf{X}_\varsigma^{(t)}] = \mathrm{E}[\mathrm{E}[\mathbf{X}_\varsigma^{(t)}|\mathbf{Z}^{(t-1)}, \mathbf{X}_\varsigma^{(t-1)}]] = $$
$$= (1-\alpha)\, \mathrm{E}[\mathbf{X}_\varsigma^{(t-1)}] + \alpha\,\varsigma\,\mathbf{D}\,\mathrm{E}[\mathbf{p}(\mathbf{Z}^{(t-1)})] \qquad (4.10)$$

and dividing it through by ς yields:

$$\varsigma^{-1}\,\mathrm{E}[\mathbf{X}^{(t)}] = \varsigma^{-1}(1-\alpha)\,\mathrm{E}[\mathbf{X}^{(t-1)}] + \alpha\,\mathbf{D}\,\mathrm{E}[\mathbf{p}(\mathbf{Z}^{(t-1)})]\,.$$

Using the distributional approximation for $\mathbf{p}(\mathbf{Z})$ in the neighbourhood of stationarity:

$$\varsigma^{-1}\,\mathrm{E}[\mathbf{X}^{(t)}] = \varsigma^{-1}(1-\alpha)\,\mathrm{E}[\mathbf{X}^{(t-1)}] + \alpha\,\mathbf{D}\,(\mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})) + \mathbf{J_p}\,(\mathrm{E}[\mathbf{Z}^{(t-1)}] - \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})) + \mathrm{O}(\varsigma^{-0.5})$$

where we say $f(\varsigma)$ is $\mathrm{O}(\varsigma^n)$ if $\lim_{\varsigma\to\infty} f(\varsigma)/\varsigma^n = u < \infty$ for some finite constant u.

Recalling that the condition for SUE is $\mathbf{x}_{\mathrm{SUE}} = \zeta\,\mathbf{D}\,\mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))$, the above may be simplified to:

$$\varsigma^{-1}\,\mathrm{E}[\mathbf{X}^{(t)}] = \varsigma^{-1}(1-\alpha)\,\mathrm{E}[\mathbf{X}^{(t-1)}] + \alpha\,\varsigma^{-1}\,\mathbf{x}_{\mathrm{SUE}} + \alpha\,\mathbf{D}\,\mathbf{J_p}\,(\mathrm{E}[\mathbf{Z}^{(t-1)}] - \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})) + \mathrm{O}(\varsigma^{-0.5})\,.$$

Now, also we have that:

$$\mathrm{E}[\mathbf{Z}^{(t-1)}] = \Sigma_{k=1,\ldots,m}\,\eta_k\,\mathrm{E}[\mathbf{c}_\varsigma(\mathbf{X}^{(t-k)})]$$
$$= \Sigma_{k=1,\ldots,m}\,\eta_k\,(\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\,\mathbf{J_c}\,(\mathrm{E}[\mathbf{X}^{(t-k)}] - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5}))$$
$$= \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\,\Sigma_{k=1,\ldots,m}\,\eta_k\,\mathbf{J_c}\,(\mathrm{E}[\mathbf{X}^{(t-k)}] - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5})$$

and so

$$\mathrm{E}[\mathbf{Z}^{(t-1)}] - \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) = \varsigma^{-1}\,\Sigma_{k=1,\ldots,m}\,\eta_k\,\mathbf{J_c}\,(\mathrm{E}[\mathbf{X}^{(t-k)}] - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5})\,.$$

Substituting into the expression above for $\varsigma^{-1}\,\mathrm{E}[\mathbf{X}^{(t)}]$, and denoting $\boldsymbol{\mu}^{(t)} = \mathrm{E}[\mathbf{X}^{(t)}]$, yields:

$$\varsigma^{-1}\,(\boldsymbol{\mu}^{(t)} - \alpha\,\mathbf{x}_{\mathrm{SUE}}) = \varsigma^{-1}(1-\alpha)\,\boldsymbol{\mu}^{(t-1)} + \varsigma^{-1}\alpha\,\Sigma_{k=1,\ldots,m}\,\eta_k\,\mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c}\,(\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5})$$

which after some slight rearrangement can be written as:

$$\varsigma^{-1}(\boldsymbol{\mu}^{(t)} - \mathbf{x}_{\mathrm{SUE}}) = \varsigma^{-1}(1-\alpha)(\boldsymbol{\mu}^{(t-1)} - \mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\alpha\,\Sigma_{k=1,\ldots,m}\,\eta_k\,\mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c}\,(\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5}).$$

Thus, asymptotically with small error relative to ς, we can relate the mean $\boldsymbol{\mu}^{(t)}$ of the process to the means $\{\boldsymbol{\mu}^{(t-k)} : k = 1,2,\ldots,m\}$ on the preceding m days, at least approximately in a neighbourhood of stationarity where Davis & Nihan's result may be assumed to approximately hold.

Note that in stationarity, $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)} = \boldsymbol{\mu}^{(t-2)} = \ldots = \boldsymbol{\mu}^{(t-m)} = \boldsymbol{\mu}^*$ (say), and the dynamic equations above give:

$$\varsigma^{-1}\,(\boldsymbol{\mu}^* - \mathbf{x}_{\mathrm{SUE}}) = \varsigma^{-1}(1-\alpha)\,(\boldsymbol{\mu}^* - \mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\alpha\,\Sigma_{k=1,\ldots,m}\,\eta_k\,\mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c}\,(\boldsymbol{\mu}^* - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5})$$
$$= \varsigma^{-1}\,((1-\alpha)\,\mathbf{I} + \alpha\,\mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c})\,(\boldsymbol{\mu}^* - \mathbf{x}_{\mathrm{SUE}})\,\Sigma_{k=1,\ldots,m}\,\eta_k + \mathrm{O}(\varsigma^{-0.5})$$
$$= \varsigma^{-1}\,((1-\alpha)\,\mathbf{I} + \alpha\,\mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c})\,(\boldsymbol{\mu}^* - \mathbf{x}_{\mathrm{SUE}}) + \mathrm{O}(\varsigma^{-0.5})$$

implying that:

$$\varsigma^{-1}\,\alpha\,(\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c})\,\boldsymbol{\mu}^* = \varsigma^{-1}\,\alpha\,(\mathbf{I} - \mathbf{D}\,\mathbf{J_p}\,\mathbf{J_c})\,\mathbf{x}_{\mathrm{SUE}} + \mathrm{O}(\varsigma^{-0.5})\,.$$