PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2008) **366**, 2017–2032
doi:10.1098/rsta.2008.0018
Published online 6 March 2008

Mechanisms for spatio-temporal pattern formation in highway traffic models

By R. Eddie Wilson*

Department of Engineering Mathematics, University of Bristol, Queen's Building, University Walk, Bristol BS8 1TR, UK

A key qualitative requirement for highway traffic models is the ability to replicate a type of traffic jam popularly referred to as a *phantom jam*, *shock wave* or *stop-and-go wave*. Despite over 50 years of modelling, the precise mechanisms for the generation and propagation of stop-and-go waves and the associated spatio-temporal patterns are in dispute. However, the increasing availability of empirical datasets, such as those collected from motorway incident detection and automatic signalling system (MIDAS) inductance loops in the UK or the next-generation simulation trajectory data (NGSIM) project in the USA, means that we can expect to resolve these questions definitively in the next few years. This paper will survey the essence of the competing explanations of highway traffic pattern formation and introduce and analyse a new mechanism, based on dynamical systems theory and bistability, which can help resolve the conflict.

Keywords: nonlinear dynamics; highway traffic modelling; stop-and-go waves

1. Introduction

Recent reports estimate that delays due to road traffic congestion cost UK businesses up to £20 billion annually (Confederation of British Industry 2003). As economies grow, so will road traffic: the UK forecast is 30% growth in the period 2000–2015 (House of Commons Select Committee on Transport 2005). Hence there is an intense international effort in *Intelligent Transport Systems* (ITS) in which information and communication technologies are used to manage traffic in order to alleviate congestion.

On the English motorways and trunk roads, known collectively as the strategic road network, the Highways Agency has employed schemes such as Controlled Motorways (automatically reduced mandatory speed limits, such as on London's M25 orbital motorway), ramp metering (traffic lights on on-ramps, which release just a few vehicles at a time) and most recently Active Traffic Management (hard shoulder running on Birmingham's M42 motorway); see http://www.highways.gov.uk. These schemes activate automatically in peak times in an attempt to stabilize flow and hence reduce congestion and accidents. The investment in telematics infrastructure has been significant—approximately £100 million for Active Traffic Management alone.

*re.wilson@bristol.ac.uk

One contribution of 16 to a Discussion Meeting Issue 'Networks: modelling and control'.

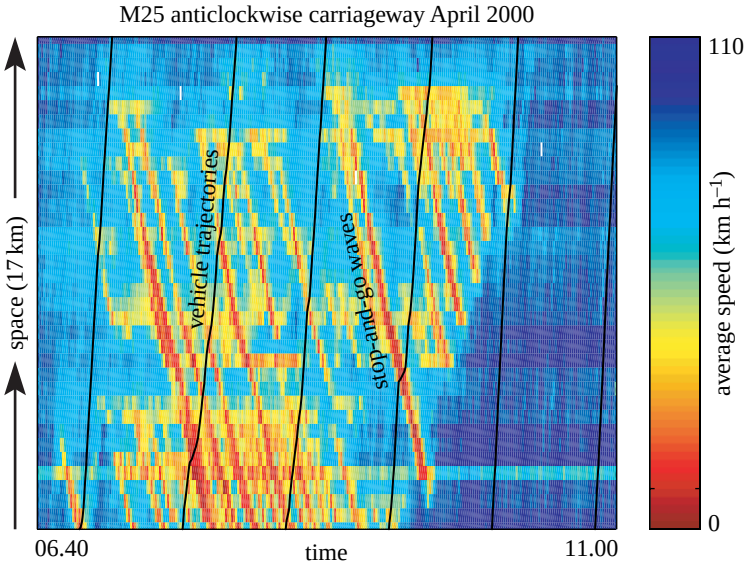

Figure 1. A typical sample of spatio-temporal speed data captured from the inductance loop system on London's orbital motorway, with some estimated average vehicle trajectories superimposed. This paper is concerned with the mechanisms for the formation of stop-and-go congestion waves, which are a common feature of highways around the world.

The context of this paper is the fundamental traffic models and control algorithms that will in future form the kernel of ITS. Traffic flow models operate at a range of different scales, from (i) whole-link models, which output travel time without modelling within-link traffic structure, through (ii) macroscopic models, which are generally formulated as partial differential equations (PDEs) and which regard traffic as a fluid-like continuum, to (iii) car-following models which consider individual vehicle dynamics. Here we discuss classes (ii) and (iii). Note that it is generally accepted that there will never be a single definitive model for highway traffic flow and moreover, despite the very wide range of traffic models available, there has until recently been insufficient data for a detailed evaluation and verification or optimization of competing models. Consequently, the academic modelling literature has grown and branched but has usually failed to connect to the real applications. However, there is now substantial mileage in using data routinely captured from ITS hardware to refine the fundamental traffic models themselves.

Figure 1 displays a small subset of spatio-temporal data captured from the Highways Agency's motorway incident detection and automatic signalling system (MIDAS) hardware, whose detection system consists of sets of inductance loops buried in the road surface and spaced typically at 500 m intervals around the motorway network. These loops, which are typical of highways in many Western countries, are equipped with signal processing electronics that measure the time and lane number of passing vehicles and estimate their speeds and lengths. In normal operation, a roadside outstation bundles these data into 1 min averages which are then sent to a control centre.

In everyday terms, the pattern shown in figure 1 is commonly referred to as a *phantom jam* or a *shock wave* although in the scientific literature the terms *stop-and-go wave* or *wide moving jam* are preferred, since the structure, which propagates upstream against the flow of traffic, consists of two sharp interfaces (one at which vehicles brake and the other at which vehicles accelerate) bounding a plateau of slow-moving traffic. In fact, this pattern is just one member of more complicated classifications developed by Kerner & Rehborn (1997) and Treiber *et al.* (2000).

The first (and now famous) mathematical explanation of traffic jams and their propagation was attempted with the hydrodynamic LWR model due to Lighthill & Whitham (1955) and Richards (1956), which describes traffic via continuous density $\rho(x,t)$ and velocity $v(x,t)$ variables that satisfy the continuity equation

$$\rho_t + (\rho v)_x = 0, \tag{1.1}$$

supplemented by the speed–density relation

$$v = \hat{V}(\rho), \tag{1.2}$$

where $\hat{V}$ is a prescribed decreasing function that models the fact that sparse traffic tends to drive quickly, whereas dense traffic drives more slowly for safety reasons. From here one obtains the *fundamental diagram*

$$Q := \rho V(\rho) \tag{1.3}$$

for traffic flux. The choice $\hat{V} = v_{\max}(1-\rho/\rho_{\max})$ (known as Greenshield's model) is typical in that it yields a quadratic unimodal Q, and consequently the result that a highway's maximum traffic flow is attained at intermediate densities and speeds.

Moreover, the theory of characteristics may be used to analyse the wave types of (1.1) and (1.2), and the analysis depends qualitatively on the shape of Q. In particular, for strictly convex Q (such as for Greenshield's model), the LWR model captures the upstream (decelerating traffic) interface of a stop-and-go wave as a classical shock; however, at the downstream (accelerating traffic) interface, a rarefaction fan is predicted (whereas the inductance loop data indicate that the downstream interface remains sharp). In §4 we will return to this point.

The focus of this paper, however, is on car-following models that consider vehicles to be discrete entities moving in continuous time and space (see Helbing 2001 for a review). Such models involve ordinary or delay differential equations that describe each driver's acceleration response to the vehicle(s) immediately ahead. As shown in figure 2, it is typical for such models to possess an instability in some parameter regimes, which leads to the folding up of traffic into structures that resemble stop-and-go waves—an idea which dates back to Herman *et al.* (1959) who analysed linear *follow-the-leader* models in which a driver's acceleration is proportional to the relative velocity of the vehicle ahead. However, the nonlinear rejuvenation of this area is much more recent—beginning with the *optimal velocity* model proposed by Bando *et al.* (1995). Despite an intensive international effort in this area over the last 10 years, there are, in my view, significant gaps in our mathematical understanding, which now need to be addressed.

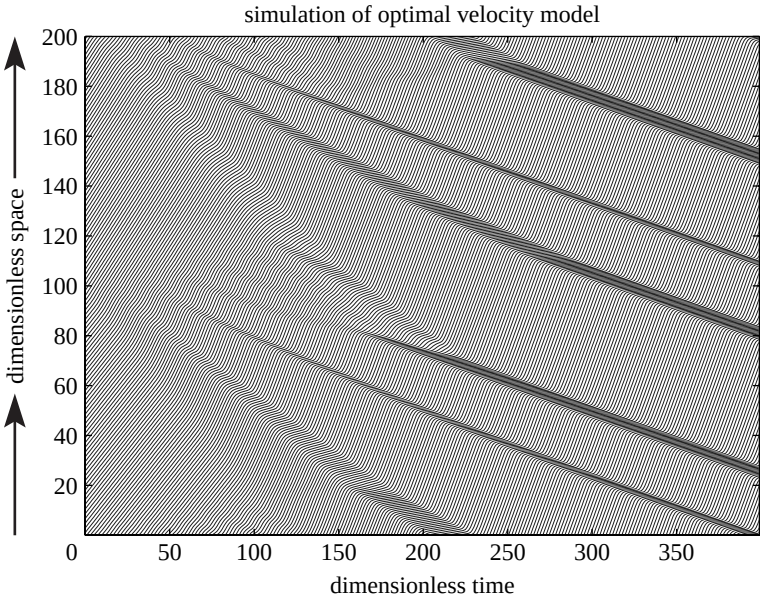

Figure 2. Trajectory plots for a typical car-following model solved on a ring road. A small perturbation is added to uniform flow (parallel trajectory) initial data, which magnifies to produce large-scale wave features. Here we simulate 100 vehicles with the optimal velocity model (2.7) and unstable parameters $h_* = 2$, $\alpha = 1.5$.

The paper is laid out as follows. First in §§2 and 3, we present a formulation of the standard linear instability analysis of car-following models, which is new in that it is couched in almost entirely general terms. In particular, under very mild constraints on the car-following model under consideration, we are able to show that if linear instability occurs, then its onset has the same mechanism, namely via a dispersion relation whose real part is quadratic in small wavenumber and whose curvature changes sign as the bifurcation condition is crossed.

Since the onset of instability in car-following models occurs at long wavelengths, it is my view that it should be captured by macroscopic PDE models, and in §4 we discuss the limitations of existing PDE theories in this regard. It should be said that there is a substantial community who believe that the stability issues are not central to pattern formation and we outline their arguments.

Next in §5, we give an account of recent developments in the *nonlinear* stability analysis of car-following models. In particular, we describe how it is possible to construct models in which smooth (so-called *uniform*) flows are linearly stable for all parameters, and yet for which there are patterns like those in figure 2. In my opinion, such a mechanism may go a long way to help resolve the conflict in views between the different traffic modelling communities. Finally, in §6 we present conclusions.

2. Car-following model framework

Our starting point is the standard situation depicted in figure 3. We consider a single lane of traffic labelled 1, 2, etc., in the upstream direction. Displacements and velocities are denoted $x_n(t)$ and $v_n(t) \geq 0$, respectively, and our models shall

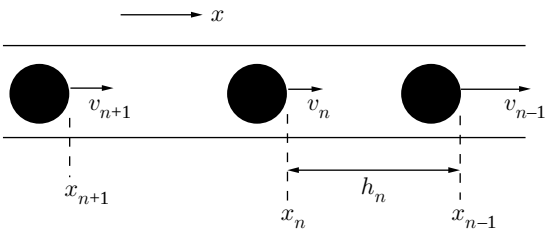

Figure 3. General scheme and notation for car-following models.

also involve the front-to-front spacing $h_n(t) := x_{n-1}(t) - x_n(t) > 0$ of consecutive vehicles, commonly referred to as the headway. Note that overtaking is neglected in our framework in return for analytical tractability. In fact, next-generation simulation project (NGSIM) data indicate that lane changing at congested merges is key in triggering stop-and-go waves. Therefore, our approach is to view lane changing as an external perturbation to a single-lane model whose stability should then be analysed.

In their simplest form, car-following models consist of a set of coupled differential equations for the trajectory of each vehicle, which typically supplements the kinematic relations $\dot{x}_n = v_n$ with a behavioural model

$$\dot{v}_n = f(h_n, \dot{h}_n, v_n), \tag{2.1}$$

which describes how drivers accelerate/decelerate in response to the motion of the vehicle in front, and their own velocity.

Here, motivated by data, we focus on models where there is a monotone increasing *optimal velocity* function V such that

$$f(h_*, 0, V(h_*)) = 0 \quad \text{for all} \quad h_* > 0 \tag{2.2}$$

and consequently a one-parameter family of steady driving solutions known as uniform flows. In this case, the connection with the LWR model (1.1), (1.2) is the obvious one given by $V(h_*) \equiv \hat{V}(\rho)$ where $\rho = 1/h_*$.

We now consider small perturbations to the equilibria by setting $h_n = h_* + \tilde{h}_n(t)$ and $v_n = V(h_*) + \tilde{v}_n(t)$, where $\tilde{h}_n$ and $\tilde{v}_n$ are small. Assuming f is sufficiently smooth, this linearization yields

$$\dot{\tilde{v}}_n = (\mathrm{D}_h f)\tilde{h}_n + (\mathrm{D}_{\dot{h}} f)\dot{\tilde{h}}_n + (\mathrm{D}_v f)\tilde{v}_n, \tag{2.3}$$

where the partial derivatives $\mathrm{D}f$ are evaluated at the constant equilibrium arguments $(h_*, 0, V(h_*))$, and necessary constraints for rational driver behaviour are

$$\mathrm{D_h} f, \mathrm{D}_{\dot{h}} f \geq 0 \quad \text{and} \quad \mathrm{D}_v f \leq 0. \tag{2.4}$$

Note that equation (2.3) may be re-expressed in the form

$$\dot{v}_n = (\mathrm{D}_h f)(h_n - h_*) + (\mathrm{D}_{\dot{h}} f)\,\dot{h}_n + (\mathrm{D}_v f)(v_n - V(h_*)). \tag{2.5}$$

A broad class of car-following models may then be obtained by a dynamic relaxation where h_* and $V(h_*)$ are replaced by the time-varying quantities $H(v_n)$ and $V(h_n)$,

respectively. Here $H := V^{-1}$ is the *optimal headway* function and we obtain

$$\dot{v}_n = \alpha(V(h_n) - v_n) + \beta\dot{h}_n + \gamma(h_n - H(v_n)), \tag{2.6}$$

with $\alpha, \beta, \gamma \geq 0$, consisting of a blend of optimal velocity, relative velocity and optimal headway terms. This class is a strict subset of the general model (2.1), but has the advantage that the equilibrium and near-equilibrium structures are clearly exhibited. Moreover, well-known models in the literature are captured as special cases: for example, $\alpha > 0$, $\beta = \gamma = 0$ give the classic optimal velocity model (Bando *et al.* 1995)

$$\left.\begin{aligned} \dot{v}_n &= \alpha(V(h_n) - v_n), \\ V(h) &= \tanh(h-2) + \tanh 2, \end{aligned}\right\} \tag{2.7}$$

with which we illustrate this paper. Here we have adopted Bando's original non-dimensional choice for $V(h)$ which has (what is believed to be) the correct *sigmoidal* shape.

In addition to model (2.1), we wish to consider mildly non-local stimuli to driver behaviour, where the spacings of several vehicles in front are also considered. This multi-anticipative generalization takes the form

$$\dot{v}_n = f\left(h_n^{(1)}, h_n^{(2)}, \ldots, h_n^{(m_h)}, \dot{h}_n^{(1)}, \dot{h}_n^{(2)}, \ldots, \dot{h}_n^{(m_{\dot{h}})}, v_n\right), \tag{2.8}$$

where m_h and $m_{\dot{h}}$ are the numbers of vehicles ahead that are considered in headway and headway rate terms, respectively, and multiple headways $h_n^{(k)}$ are defined by

$$h_n^{(k)} = h_n + h_{n-1} + \cdots + h_{n-k+1}. \tag{2.9}$$

In particular $h_n^{(1)} = h_n$. As before, we assume the existence of an equilibrium function V so that

$$f(h_*, 2h_*, \ldots, m_h h_*, 0, 0, \ldots, 0, V(h_*)) = 0, \tag{2.10}$$

for any h_*. As a specific example, we may consider

$$\dot{v}_n = \sum_{k=1}^{m_\alpha} \alpha_k \left\{ V\left(\frac{h_n^{(k)}}{k}\right) - v_n \right\} + \sum_{k=1}^{m_\beta} \beta_k \dot{h}_n^{(k)} + \sum_{k=1}^{m_\gamma} \gamma_k \left\{ \frac{h_n^{(k)}}{k} - H(v_n) \right\} \tag{2.11}$$

as a direct generalization of equation (2.6) (cf. Lenz *et al.* 1999; Wilson *et al.* 2004).

3. Linear stability analysis

Now the goal is to show that the linear stability analysis of uniform flow situations works out in a very similar way under quite unrestrictive assumptions on the detail of the model. First, we eliminate the velocity variable and write small amplitude dynamics in terms of headway variables alone. This is achieved by noting that $\dot{h}_n = v_{n-1} - v_n$ and consequently $\ddot{h}_n = \dot{v}_{n-1} - \dot{v}_n$, so that (2.3) yields

$$\ddot{\tilde{h}}_n = (\mathrm{D}_h f)(\tilde{h}_{n-1} - \tilde{h}_n) + (\mathrm{D}_{\dot{h}} f)(\dot{\tilde{h}}_{n-1} - \dot{\tilde{h}}_n) + (\mathrm{D}_v f)\dot{\tilde{h}}_n. \tag{3.1}$$

The exponential ansatz

$$\tilde{h}_n = \mathrm{Re}\,(c\mathrm{e}^{\mathrm{i}n\theta}\,\mathrm{e}^{\lambda t}) \tag{3.2}$$

then yields the quadratic

$$\lambda^2 + \{(\mathrm{D}_{\dot{h}} f)(1 - \mathrm{e}^{-\mathrm{i}\theta}) - (\mathrm{D}_v f)\}\lambda + (\mathrm{D}_h f)(1 - \mathrm{e}^{-\mathrm{i}\theta}) = 0, \tag{3.3}$$

to solve for the (generally complex) growth rate λ in terms of the discrete wavenumber θ, $0 < \theta \leq \pi$.

(*a*) *Case of short-wavelength perturbations:* $\theta = \pi$

Note that $\theta = \pi$ gives the shortest possible perturbation wavelength, corresponding to a fluctuation of period two cars. Equation (3.3) then yields

$$\lambda^2 + \{2(\mathrm{D}_{\dot{h}} f) - (\mathrm{D}_v f)\}\lambda + 2\mathrm{D}_h f = 0, \tag{3.4}$$

for which all coefficients are positive. Consequently, there are two real roots with negative real parts. The conclusion is that models of type (2.1) cannot in general propagate short-wavelength instabilities, provided the sensible sign conventions (2.4) are maintained.

(*b*) *Case of long-wavelength perturbations:* $\theta = 0+$

We now consider the case of long-wavelength perturbations for which θ is small and positive. Note that $\lambda = 0$, $\theta = 0$ always solve (3.3) because the uniform flow under consideration is just one member of a continuous family of such solutions. The strategy is thus to seek small solutions λ in terms of a regular perturbation expansion $\lambda = \lambda_1\theta + \lambda_2\theta^2 + \cdots$ and determine the direction in which the dispersion relation bends at $\theta = 0$. Equating the first two powers of θ yields

$$O(\theta)\colon\ -(\mathrm{D}_v f)\lambda_1 + (\mathrm{D}_h f)\mathrm{i} = 0, \tag{3.5}$$

$$O(\theta^2)\colon\ \lambda_1^2 + (\mathrm{D}_{\dot{h}} f)\mathrm{i}\lambda_1 - (\mathrm{D}_v f)\lambda_2 + \frac{1}{2}(\mathrm{D}_h f) = 0. \tag{3.6}$$

Therefore $\lambda_1 = \mathrm{i}(\mathrm{D}_h f)/(\mathrm{D}_v f)$ is purely imaginary and the growth is neutral at leading order. The $O(\theta^2)$ relation then yields the real expression

$$\lambda_2 = \frac{(\mathrm{D}_h f)}{(\mathrm{D}_v f)^3}\left\{\frac{1}{2}(\mathrm{D}_v f)^2 - (\mathrm{D}_h f) - (\mathrm{D}_{\dot{h}} f)(\mathrm{D}_v f)\right\}, \tag{3.7}$$

whose bracket consists of a balance of terms of different signs, allowing the possibility of changes in stability as either the model or the parameters are changed. For instability of arbitrarily large wavelength perturbations, we require

$$\frac{1}{2}(\mathrm{D}_v f)^2 - (\mathrm{D}_h f) - (\mathrm{D}_{\dot{h}} f)(\mathrm{D}_v f) < 0. \tag{3.8}$$

For the optimal velocity model (2.7), this gives the standard instability condition

$$\alpha < 2\,V'(h_*), \tag{3.9}$$

and the stability limit is thus $\alpha = 2$ since $V'_{\max} = 1$.

(c) General condition for instability

Now that we have considered the extreme cases of short and long wavelengths, we perform a marginal stability analysis for the general wavelength, by seeking the neutral stability curve. If we set $\lambda = \mathrm{i}\omega$ (ω real) in (3.3) and equate real and imaginary parts, we obtain

$$-\omega^2 + \{(\mathrm{D}_{\dot{h}} f)\sin\theta\}\omega + (\mathrm{D}_h f)(1-\cos\theta) = 0, \tag{3.10}$$

$$\{(\mathrm{D}_{\dot{h}} f)(1-\cos\theta) - (\mathrm{D}_v f)\}\omega + (\mathrm{D}_h f)\sin\theta = 0. \tag{3.11}$$

Here we use the second equation to eliminate ω in the first, and then apply the half-angle formulae $\sin\theta = 2SC$ and $1-\cos\theta = 2S^2$, where $S = \sin(\theta/2)$ and $C = \cos(\theta/2)$. Simplification and division by the non-zero factor $4S^2(\mathrm{D}_h f)$ then yields

$$\frac{1}{2}(\mathrm{D}_v f)^2 - C^2\{(\mathrm{D}_h f) + (\mathrm{D}_{\dot{h}} f)(\mathrm{D}_v f)\} = -2S^2\{(\mathrm{D}_{\dot{h}} f)^2 S^2 - (1+C^2)(\mathrm{D}_{\dot{h}} f)(\mathrm{D}_v f)\}, \tag{3.12}$$

whose r.h.s. is negative (recall the sign conventions (2.4)). Since $0 < C^2 \leq 1$, we may compare the l.h.s. with (3.8) and conclude that the general model is linearly unstable to arbitrarily long-wavelength perturbations at the marginal stability point of any other mode.

(d) Linear stability for the multi-anticipative model

Linearization of (2.8) about the uniform flow state gives

$$\dot{\tilde{v}}_n = \sum_{k=1}^{m_h} (\mathrm{D}_{h^{(k)}} f)\tilde{h}_n^{(k)} + \sum_{k=1}^{m_{\dot{h}}} (\mathrm{D}_{\dot{h}^{(k)}} f)\dot{\tilde{h}}_n^{(k)} + (\mathrm{D}_v f)\tilde{v}_n, \tag{3.13}$$

and consequently by using (2.9) and thus $\tilde{h}_{n-1}^{(k)} - \tilde{h}_n^{(k)} = \tilde{h}_{n-k} - \tilde{h}_n$, we obtain

$$\ddot{\tilde{h}}_n = (\mathrm{D}_h f)^\dagger \left(\sum_{k=1}^{m_h} z_k \tilde{h}_{n-k} - \tilde{h}_n \right) + (\mathrm{D}_{\dot{h}} f)^\dagger \left(\sum_{k=1}^{m_{\dot{h}}} w_k \dot{\tilde{h}}_{n-k} - \dot{\tilde{h}}_n \right) + (\mathrm{D}_v f)\dot{\tilde{h}}_n, \tag{3.14}$$

where $(\mathrm{D}_h f)^\dagger := \sum_{k=1}^{m_h} (\mathrm{D}_{h^{(k)}} f)$ and $(\mathrm{D}_{\dot{h}} f)^\dagger := \sum_{k=1}^{m_{\dot{h}}} (\mathrm{D}_{\dot{h}^{(k)}} f)$ are analogous to the total derivatives $(\mathrm{D}_h f)$ and $(\mathrm{D}_{\dot{h}} f)$ in the simple case without multi-anticipation, and $z_k := (\mathrm{D}_{h^{(k)}} f)/(\mathrm{D}_h f)^\dagger$ and $w_k := (\mathrm{D}_{\dot{h}^{(k)}} f)/(\mathrm{D}_{\dot{h}} f)^\dagger$ are non-negative weights with $\sum w_k = \sum z_k = 1$. The exponential ansatz (3.2) thus yields the quadratic equation (cf. equation (3.3))

$$\lambda^2 + \left\{ (\mathrm{D}_{\dot{h}} f)^\dagger \left(1 - \sum_{k=1}^{m_{\dot{h}}} w_k \, \mathrm{e}^{-\mathrm{i}k\theta} \right) - (\mathrm{D}_v f) \right\} \lambda + (\mathrm{D}_h f)^\dagger \left(1 - \sum_{k=1}^{m_h} z_k \, \mathrm{e}^{-\mathrm{i}k\theta} \right) = 0. \tag{3.15}$$

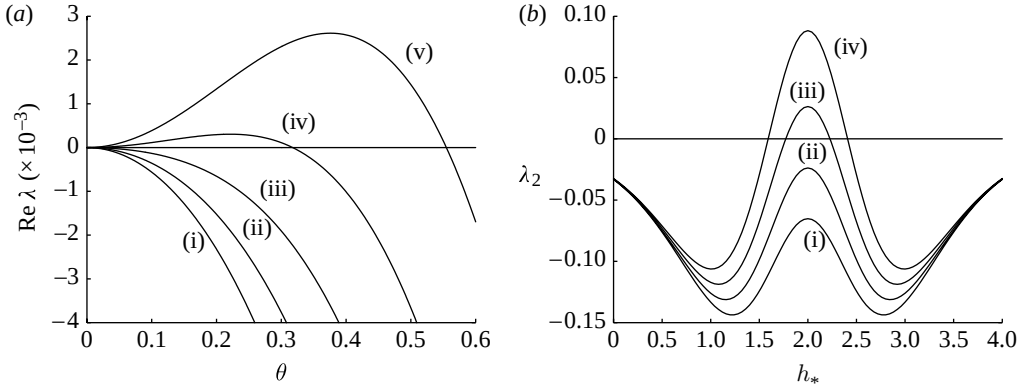

Figure 4. Usual form for the onset of instability in car-following models, illustrated here for the optimal velocity model (2.7), for which the onset is at $\alpha = 2$ and $h_* = 2$, see condition (3.9). (*a*) Dispersion relation plotting Re λ against discrete wavenumber θ, which shows the generic onset of instability at small wavenumber and hence at large wavelength. This plot is for $h_* = 2$ and (i) $\alpha = 2.25$, (ii) $\alpha = 2.15$, (iii) $\alpha = 2.05$, (iv) $\alpha = 1.95$ and (v) $\alpha = 1.85$. (*b*) Plot of λ_2 against headway showing the usual situation where the onset of long-wavelength instability occurs at mid-range values of the headway and hence velocity. Here (i) $\alpha = 2.3$, (ii) $\alpha = 2.1$, (iii) $\alpha = 1.9$ and (iv) $\alpha = 1.7$.

The analysis of small θ solutions proceeds similarly to before to yield $\lambda_1 = \mathrm{i}(\sum kz_k)(\mathrm{D}_h f)^\dagger/(\mathrm{D}_v f)$ and (cf. equation (3.7))

$$\lambda_2 = \frac{(\mathrm{D}_h f)^\dagger}{(\mathrm{D}_v f)^3}\left\{\frac{1}{2}\left(\sum k^2 z_k\right)(\mathrm{D}_v f)^2 - \left(\sum kz_k\right)^2 (\mathrm{D}_h f)^\dagger - \left(\sum kw_k\right)\left(\sum kz_k\right)(\mathrm{D}_{\dot{h}} f)^\dagger (\mathrm{D}_v f)\right\}. \tag{3.16}$$

So far our numerical investigations indicate that the onset of instability occurs at infinite wavelength in the same way as models without multi-anticipation, and hence the stability boundary is given by a sign change in equation (3.16). However, a proof may not be achieved so simply since equation (3.15) is parametrized by two independent complex numbers.

4. Discussion

The general situation that we have now established is described by the dispersion relation plot in figure 4*a*. In summary, if a model is unstable to a mode of any one wavelength, then it is also unstable to all longer wavelengths. Moreover, as a parameter is varied, the onset of instability occurs at infinite wavelength (zero wavenumber) via a change in sign of the second derivative of the growth rate Re λ. Thus in marginally unstable situations, only the longest wavelengths are magnified, which may explain how models that are entirely local in terms of their interactions give rise to structures (i.e. stop-and-go waves) whose wavelength is many times the vehicle spacing.

Our next step is to identify the mean headway h_* as a special parameter and then analyse the onset of instability for different h_* as a separate external parameter, such as the sensitivity α in the optimal velocity model (2.7), is varied.

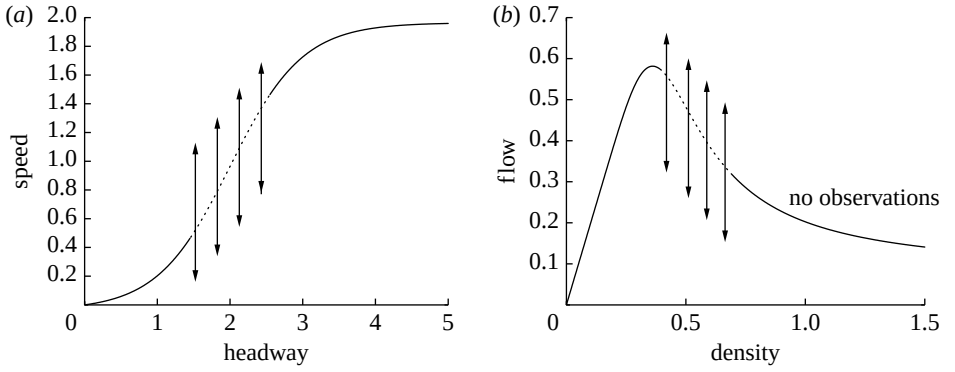

Figure 5. Plots of (*a*) speed against headway and (*b*) flow against density illustrated for uniform flows corresponding to the non-dimensional optimal velocity function from (2.7). The dotted section corresponds to mid-range headway where the onset of linear stability typically occurs and periodic behaviour results. Consequently, numerical observations derived from simulations in this regime are widely spread and this result is in agreement with the empirical flow–density data captured from inductance loop systems.

From our remarks above concerning the dominance of long-wavelength effects, it is sufficient to detect sign changes in the coefficient λ_2, which govern the component of Re λ that is quadratic in small wavenumbers (figure 4*b*). There is in fact no general result, but for many models the onset of instability occurs at mid-range values of h_*. Such cases lead to the interpretation presented in figure 5. In particular, since there is a range of densities for which uniform flow is linearly unstable, in which one would expect periodic oscillations, one should see periodic variations in flow. In fact, large fluctuations are found in empirical flow data to the right of the fundamental diagram's maximum, thus supporting the theory that we outline here. Finally, we should say that there is an open question concerning whether uniform flow restabilizes at large densities (small headways), since inductance loop systems, which are fixed in space and calculate time averages, do not properly capture the details of very dense (and hence almost stationary) traffic.

In order to explain the central conflict in the highway traffic modelling community, we now turn our attention to PDE models. Our discussion begins with the model of Kerner & Konhäuser (1993), which supplements the continuity equation (1.1) with an equation for convective acceleration

$$v_t + vv_x = \alpha(\hat{V}(\rho) - v) - \beta\frac{\rho_x}{\rho} + \mu\frac{v_{xx}}{\rho}, \tag{4.1}$$

whose r.h.s. terms involve relaxation to an optimal velocity $\hat{V}(\rho)$, in addition to pressure (to model drivers' anticipation) and diffusion. It may be shown that the uniform flow equilibria in this model become linearly unstable and thus give rise to solutions that resemble stop-and-go waves, in much the same way as we have described for car-following models.

Unfortunately, it is known (Daganzo 1995) that in such *second-order* models (by which we mean that there is a dynamic equation for velocity v), the pressure gradient term can cause unphysical effects such as solution modes that propagate

downstream faster than traffic, or even backward flowing traffic when density gradients are extreme. Daganzo's observation led to a sequence of papers beginning with Aw & Rascle (2000) and Zhang (2002), which rectified these issues by evaluating the pressure gradient in a Lagrangian frame that moves with drivers. The state of the art in this theory is described by Lebacque *et al.* (2007), where (1.1) is supplemented by

$$\boldsymbol{I}_t + v\boldsymbol{I}_x = -\boldsymbol{f}(\boldsymbol{I}), \tag{4.2}$$

where $\boldsymbol{I} = \boldsymbol{I}(\rho, v)$ is a vector of so-called Lagrangian markers that are either conserved on vehicle trajectories or relaxed according to the dynamics of $\boldsymbol{f}$. In this setting, Aw & Rascle (2000) use a scalar marker $I = v - \hat{V}(\rho)$ with $f(I) = \alpha I$.

Model framework (1.1), (4.2) has two particular problems in my view.

(i) First, in common with the standard LWR model (equations (1.1) and (1.2)), a discontinuity in initial data at which traffic accelerates will collapse via a rarefaction fan when the fundamental diagram (1.3) is strictly convex. Consequently, this model framework does not propagate the downstream interface of a stop-and-go wave in the manner that data indicate it should. Those who adhere to these models thus propose a piecewise-linear triangular construction for the fundamental diagram, which has coincidental advantages from the point of view of tractability (Daganzo (1994) and many subsequent papers). In my view, this is a contrived solution which is not supported by empirical flow-density data.

(ii) Second, the dispersion relation is incompatible with that of the car-following theory that we have outlined. Small perturbations to uniform flow are propagated at the characteristic speeds of the hyperbolic system, either unchanged or damped via the dynamics of the source term $\boldsymbol{f}$. In particular, it is not possible for uniform flow to be linearly unstable unless the source term itself is excitable or breaks the strict formulation of the rules presented here (see Greenberg 2004; Siebel & Mauser 2006). In either case, it is not clear how to inherit the small wavenumber scaling $\mathrm{Re}\,\lambda \sim \lambda_2\theta^2$ of the car-following dispersion relation.

Highway traffic modellers are thus fractured into several communities. First, we have 'one-phase' modellers who use strictly hyperbolic PDEs together with a triangular fundamental diagram where necessary and who oppose the idea that traffic flow is unstable at any density. Rather, these researchers propose that random large-amplitude events at the microscale cause traffic jams (Daganzo *et al.* 1999). Second, we have 'two-phase' modellers from a theoretical physics background who believe that instability is at the heart of stop-and-go waves, and who typically use car-following models since they have yet to establish a PDE theory with good global existence properties. This latter community is further fractured owing to disputes concerning the classification of spatio-temporal patterns (see Schönhof & Helbing 2007 for a comprehensive discussion based on empirical data). It seems to me that it is necessary to try and reconcile the efforts of these diverse modelling schools.

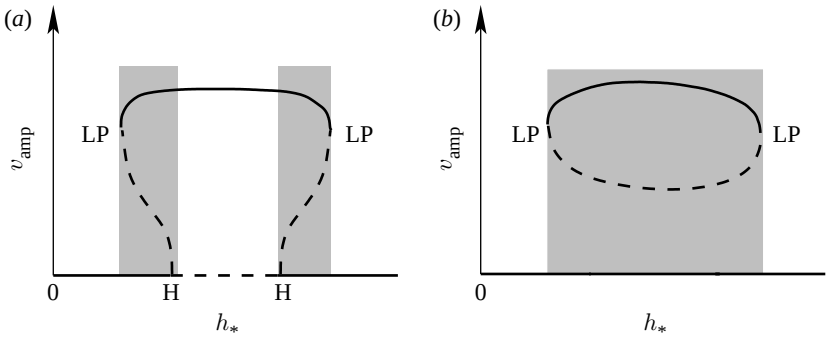

Figure 6. Schematic of branches of periodic solutions on a ring road. The semi-norm $v_{\mathrm{amp}} := \max_t v_n(t) - \min_t v_n(t)$ is plotted as the bifurcation parameter h_* is varied. The abscissa thus corresponds to uniform flow solutions where $v_n(t) \equiv V(h_*)$. Solid lines denote linearly stable solutions whereas dashed lines denote linearly unstable ones, and shaded regions denote bistability in that there are two coexisting stable solutions. (*a*) Standard situation where uniform flow is unstable for mid-range headways. Loss of stability is at subcritical Hopf bifurcations labelled H. The bifurcating branches of unstable periodic solutions subsequently turn back at cyclic folds (also known as limit points, labelled LP), to yield a branch of stable large-amplitude solutions that correspond to stop-and-go waves. (*b*) A situation in which bistability is possible without uniform flow ever being unstable.

5. New pattern mechanisms based on nonlinear instability

Our discussion now returns to the stability analysis of car-following models and in particular the simplified situation where there is a large number N of identical vehicles driving on a single-lane ring road of length Nh_*. As we identified in §3, it is a typical property of such models that the uniform flow solutions lose stability to small-amplitude, long-wavelength perturbations. Moreover, as parameters change, stability is usually lost first for mid-range values of h_*. A recent sequence of papers (Gasser *et al.* 2004; Orosz *et al.* 2004, 2005; Orosz & Stepan 2006) has used ideas from dynamical systems theory to analyse what type of time-varying solution is generated at the loss of linear stability.

The principal tools have been normal form analysis (see Kuznetsov 1995 for an introduction), in which the car-following model is expanded to cubic order at the bifurcation so as to analyse the curvature (i.e. sub- or supercritical) of bifurcating solution branches, and numerical continuation packages such as AUTO (Doedel *et al.* 1997) and DDE-BIFTOOL (Engelborghs *et al.* 2002), which deal with ordinary and delay differential equation models, respectively. These packages are able to follow branches of equilibria and periodic solutions of differential equation systems as parameters are varied. Moreover, stability information is computed along solution branches and codimension-one bifurcation points where the stability of solutions changes are detected automatically.

The typical overall form of the bifurcation diagram for the optimal velocity model is shown in figure 6*a* (see Gasser *et al.* (2004) and Orosz *et al.* (2004, 2005) for numerical computations). This sketch indicates that as headway is varied, the loss of stability is subcritical, so that the bifurcating branch of periodic solutions is itself unstable. Moreover, as the amplitude of this branch increases, there is

subsequently a cyclic fold at which the branch of periodic solutions turns back and regains stability. The interpretation of this solution structure is that there are ranges of headway values for which uniform flow is linearly stable and yet there coexist large-amplitude stop-and-go waves that are themselves linearly stable. Consequently, real-world traffic may flow apparently smoothly and stably in normal circumstances, yet an exceptionally large-amplitude perturbation (for example caused by a truck overtaking a truck) may cause the flow to jump on to the large-amplitude stop-and-go wave branch.

The bistability property described here may help resolve the conflict that we described earlier between the two-phase community, where spontaneous flow breakdown is accepted, and the one-phase hyperbolic PDE/traffic engineering community, where it is not—since instability at the linear level is no longer required for pattern formation.

So far we require uniform flow to be linearly unstable over some range of headways to enable bistability. However, one might design models in which the bifurcation diagram took the form shown in figure 6*b*: here uniform flow is linearly stable at all headway values and yet there is a disconnected branch (known as an isola) of periodic solutions that correspond to stop-and-go waves.

To this end, we have recently constructed a model for which bistability is possible with uniform flow stable at all headways. This new model takes the form

$$f(h_n, \dot{h}_n, v_n) = \alpha(h_n, \Delta v_n)\Delta v_n, \tag{5.1}$$

where $\Delta v_n = V(h_n) - v_n$ defines the velocity difference from the standard optimal velocity model (2.7), and we take

$$\alpha(h_n, \Delta v_n) = \begin{cases} 2.5 & \text{if } |\Delta v_n| < 0.05 \text{ and } 1.4 < h_n < 2.6, \\ 1.5 & \text{otherwise.} \end{cases} \tag{5.2}$$

In essence, this construction does a 'cut-and-paste' operation on the dynamics of the standard optimal velocity model by varying the sensitivity parameter α. The low-α region of phase space admits the large-amplitude (stable) stop-and-go wave solution, whereas the high-α setting guarantees linear stability of uniform flow through mid-range headway values (figure 7). However, it is not yet known whether bistability without instability is possible in models in which the dynamics has not been contrived in this way.

6. Conclusion

Our over-arching aim is the development of a model framework for forecasting highway traffic flow. This is a grand challenge in mathematical modelling since at its core lies the complexity and unpredictability of human driver behaviour. Moreover, the challenge is multiscale, since the spatio-temporal range of a forecast should depend on the application in question: 'queue ahead' warning systems operate over scales of several kilometres and several minutes; travel time forecasts may operate over a scale of hours; whereas strategic planning forecasts are concerned with the performance of the whole national network over a time scale of years.

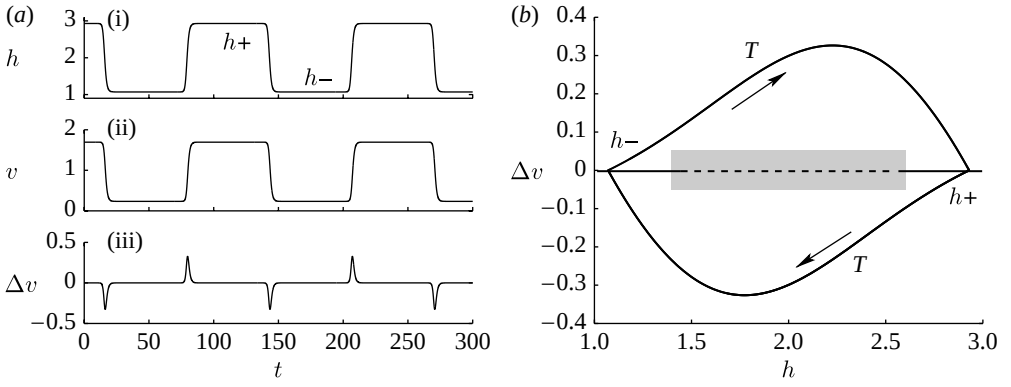

Figure 7. Stop-and-go wave structure as a periodic solution on a ring road, illustrated here for the optimal velocity model (2.7) with sensitivity $\alpha = 1.5$, mean headway $h_* = 2$ and $N = 100$ vehicles. (a(i)) Headway h, (ii) velocity v and (iii) $\Delta v := V(h) - v$ against time for a single vehicle, showing the plateau structure. Note that each vehicle's trajectory is identical to its predecessor up to a simple time shift, i.e. $h_n(t) = h_{n-1}(t-\tau)$ and $v_n(t) = v_{n-1}(t-\tau)$. (b) Trajectory data T from (a) re-mapped into $(h, \Delta v)$ phase space, with the uniform flow equilibria added as a horizontal axis and the dashed line denoting the linearly unstable mid-range of headways. By increasing the sensitivity α in the shaded region, we may stabilize uniform flow at all headway values yet the large-amplitude stop-and-go wave solution T persists unchanged.

In principle, at least, the MIDAS system and similar datasets from overseas contain sufficient detail for making and evaluating macroscopic forecasts whose scale is coarser than the 1 min × 500 m resolution. However, there is still vigorous debate over what form mathematical models should take and indeed even in the fundamental mechanisms for pattern formation.

In particular, the form of the onset of linear instability in car-following models is entirely generic, occurring via a change in the curvature of the dispersion relation at zero wavenumber. However, this mechanism is absent from the models proposed by the hyperbolic PDE community.

In this respect, it seems that the difference between linear and nonlinear stabilities must be recognized when classifying models. In particular, this distinction may help resolve conflict since, as we have shown, it is possible for a car-following model to exhibit bistability where uniform flow is linearly stable and yet linearly stable stop-and-go waves coexist. In this setting, the one-phase community is (partially) correct, since uniform flow is linearly stable, yet the two-phase community is also (partially) correct, since instability, albeit nonlinear instability, is at the heart of pattern formation. Schönhof & Helbing (2007) have recently supported this resolution by performing a detailed evaluation of empirical flow patterns using simulations with bistability properties.

There is now scope for a much more serious quantitative examination of data and a fitting of models, which needs to take place at the macroscopic PDE level using standard loop data and at the microscopic level using novel data sources that have only recently become available, for example camera trajectory data or unaveraged inductance loop data. In this way, we can expect over the next few years to definitively resolve the conflict between the various traffic modelling schools.

The author acknowledges the support of an EPSRC Advanced Research Fellowship (grant no. EP/E055567/1) and access to the MIDAS data system granted by the English Highways Agency.

References

Aw, A. & Rascle, M. 2000 Resurrection of "second order" models of traffic flow. *SIAM J. Appl. Math.* **60**, 916–938. (doi:10.1137/S0036139997332099)

Bando, M., Hasebe, K., Nakayama, A., Shibata, A. & Sugiyama, Y. 1995 Dynamical model of traffic congestion and numerical simulation. *Phys. Rev. E* **51**, 1035–1042. (doi:10.1103/PhysRevE.51.1035)

Confederation of British Industry 2003 *The UK as a place to do business: is transport holding the UK back?* London, UK: CBI.

Daganzo, C. F. 1994 The cell transmission model: a simple dynamic representation of highway traffic. *Trans. Res. B* **28**, 269–287. (doi:10.1016/0191-2615(94)90002-7)

Daganzo, C. F. 1995 Requiem for second-order fluid approximations of traffic flow. *Trans. Res. B* **29**, 277–286. (doi:10.1016/0191-2615(95)00007-Z)

Daganzo, C. F., Cassidy, M. J. & Bertini, R. L. 1999 Possible explanations of phase transitions in highway traffic. *Trans. Res. A* **33**, 365–379. (doi:10.1016/S0965-8564(98)00034-2)

Doedel, E. J., Champneys, A. R., Fairgrieve, T. F., Kuznetsov, Y. A., Sandstede, B. & Wang, X. 1997 Continuation and bifurcation software for ordinary differential equations. Technical report, Department of Computer Science, Concordia University. See http://indy.cs.concordia.ca/auto.

Engelborghs, K., Luzyanina, T. & Roose, D. 2002 Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL. *ACM T. Math. Softw.* **28**, 1–21. (doi:10.1145/513001.513002)

Gasser, I., Sirito, G. & Werner, B. 2004 Bifurcation analysis of a class of 'car-following' traffic models. *Physica D* **197**, 222–241. (doi:10.1016/j.physd.2004.07.008)

Greenberg, J. M. 2004 Congestion redux. *SIAM J. Appl. Math.* **64**, 1175–1185. (doi:10.1137/S0036139903431737)

Helbing, D. 2001 Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067–1141. (doi:10.1103/RevModPhys.73.1067)

Herman, R., Montroll, E. W., Potts, R. B. & Rothery, R. W. 1959 Traffic dynamics—analysis of stability in car following. *Oper. Res.* **7**, 86–106.

House of Commons Select Committee on Transport 2005 Report RP21A. Written evidence by the secretary of state for transport.

Kerner, B. S. & Konhäuser, P. 1993 Cluster effect in initially homogeneous traffic flow. *Phys. Rev. E* **48**, 2335–2338. (doi:10.1103/PhysRevE.48.R2335)

Kerner, B. S. & Rehborn, H. 1997 Experimental properties of phase transitions in traffic flow. *Phys. Rev. Lett.* **79**, 4030–4033. (doi:10.1103/PhysRevLett.79.4030)

Kuznetsov, Y. A. 1995 *Elements of applied bifurcation theory.* Applied Mathematical Sciences, vol. 112. Berlin, Germany: Springer.

Lebacque, J.-P., Mammar, S. & Salem, H. H. 2007 Generic second order traffic flow modelling. In *Transportation and traffic theory 2007* (eds R. E. Allsop, M. G. H. Bell & B. G. Heydecker), pp. 755–776. London, UK: Elsevier.

Lenz, H., Wagner, C. K. & Sollacher, R. 1999 Multi-anticipative car-following model. *Eur. Phys. J. B* **7**, 331–335. (doi:10.1007/s100510050618)

Lighthill, M. J. & Whitham, G. B. 1955 On kinematic waves II: a theory of traffic flow on long crowded roads. *Proc. R. Soc. A* **229**, 317–345. (doi:10.1098/rspa.1955.0089)

Orosz, G. & Stepan, G. 2006 Subcritical Hopf bifurcations in a car-following model with reaction-time delay. *Proc. R. Soc. A* **462**, 2643–2670. (doi:10.1098/rspa.2006.1660)

Orosz, G., Wilson, R. E. & Krauskopf, B. 2004 Global bifurcation investigation of an optimal velocity traffic model with driver reaction time. *Phys. Rev. E* **70**, 026 207. (doi:10.1103/PhysRevE.70.026207)

Orosz, G., Krauskopf, B. & Wilson, R. E. 2005 Bifurcations and multiple traffic jams in a car-following model with reaction-time delay. *Physica D* **211**, 277–293. (doi:10.1016/j.physd.2005.09.004)

Richards, P. I. 1956 Shockwaves on the highway. *Oper. Res.* **4**, 42–51.

Schönhof, M. & Helbing, D. 2007 Empirical features of congested traffic states and their implications for traffic modeling. *Transport. Sci.* **41**, 135–166. (doi:10.1287/trsc.1070.0192)

Siebel, F. & Mauser, W. 2006 On the fundamental diagram of traffic flow. *SIAM J. Appl. Math.* **66**, 1150–1162. (doi:10.1137/050627113)

Treiber, M., Hennecke, A. & Helbing, D. 2000 Congested traffic states in empirical observations and microscopic simualtions. *Phys. Rev. E* **62**, 1805–1824. (doi:10.1103/PhysRevE.62.1805)

Wilson, R. E., Berg, P., Hooper, S. & Lunt, G. 2004 Many-neighbour interaction and non-locality in traffic models. *Eur. Phys. J. B* **39**, 397–408. (doi:10.1140/epjb/e2004-00205-y)

Zhang, H. M. 2002 A non-equilibrium traffic model devoid of gas-like behaviour. *Trans. Res. B* **36**, 275–290. (doi:10.1016/S0191-2615(00)00050-3)