Flow on Links: Yesterday, Today and Tomorrow

J D Addison
Centre for Transport Studies, UCL.

September 21, 2005

1 Introduction

This IMA conference in honour of Richard Allsop seems a timely occasion at which to both review the current state of continuum models but also to both look back to see how we got here and look to the future and some of the problems and challenges still to be addressed.

The continued growth in traffic demand makes the effective management of the road system increasingly important. This requires a good understanding both of the overall behaviour of traffic on a network and also, especially on motorways, the behaviour of traffic on a single link. While properties of the whole network can often be adequately modelled by knowing only the influxes and effluxes management of a single stretch of motorway requires models that can describe the behaviour and distribution of traffic along the whole stretch.

Good models of traffic flow along motorways are needed to support effective traffic management. From a theoretical aspect a good model gives insights into the origins of traffic behaviour, but for effective day to day management the model must also identify measurable properties of the traffic state that allow accurate predictions of the future state to be made. To be practically useful a model must also be computationally amenable to allow timely detection or prediction of developing problems.

Modelling of traffic flow covers a range from detailed microscopic models based in individual driver behaviour through less detailed microscopic models where drivers are treated more uniformly through various cell transmission models to continuum models dealing with aggregate conditions.

My original intention was to give a wide ranging survey of traffic models. However two thing made this impossible, lack of time of my part and the great range of current activity both theoretical, computational and empirical. One only has to look at the range of papers presented to the 16^{th} ISTTT(Mahmasani 2005) to get some idea of the range of activities. One side effect of this recent flurry of activity is that there are a number of articles that offer good surveys of many aspects of the historical developments. (Nagel and Nelson 2005), (Zhang

2001). Zhang (2001) in particular gives a very comprehensive survey of the state of continuum models. Rather than attempt a wider ranging paper that failed to do justice to a lot of interesting work I decide to focus on a few topics.

One particular strand has the merit of occurring over the whole period of traffic models, the story of the kinematic wave model. I will focus on the the evolution of the theoretical understanding of the model rather than its application although the two go hand in hand. While the model could be explicitly solved in a few simple but useful cases, practical computation was difficult. A decade ago new insights by Gordon Newell((Newell 1993a),(Newell 1993b),(Newell 1993c)) opened up the possibility of more extensive analytic application and more detailed computation. The kinematic wave model continues to attract interest and development(Jin and Zhang 2003),(Daganzo 2004a),(Nelson 2000),(Lebacque and Khoshyaran 2005). More recently the insights of Daganzo in particular offer further exciting possibilities new analytic applications and of easier computation.

I will touch briefly on the issue of higher order models. Do we need them? And if so what should they look like?

Finally I discuss some of the considerable theoretical and practical challenges that should be addressed. in particular he problem of finding good theoretical justifications for current and emerging models.

2 Homogeneous Models: Why and When?

While researching this paper I began to think about the nature of models and in particular continuous or continuum traffic models and their validity. A model should serve at least one of two purposes:

- It should give us insights into properties and behaviour of the system being modelled thus improving our understanding of the system ;

- It should allow the prediction of the future state a given system.

In what follows I am primarily concerned with the models meeting the first criterion. A model meeting the first criterion should at least in theory meet the second given enough ingenuity and computational effort. When seeking either to understand or to manage some aspect of a traffic system we must try to select model suitable for purpose. The model must be sufficiently tractable so that we can work efficiently with it but at the same time it must give a reasonable description of the phenomena we are interested in. In interpreting results we must always be aware of the limits of the model to ensure that results and phenomena predicted by the theory are properly interpreted, and that we do not have unrealistic expectations of what the model can do.

The standard continuous models of traffic model traffic as a fluid with the state of the fluid being given by the flow q and the density or concentration, k

. It is important to be clear about what these quantities are. They are average quantities take over time periods usually of the order of 100 seconds or more for flow and of 100 metres are more for density. This will influence the scale of what we expect to see in the model. It is thus unrealistic to expect a model describing the behaviour of such quantities to describe in detail phenomena that happen over much sorter distances and time scales, although it may still reflect the bulk net effect of such changes. The speed and density discontinuities at shockwaves that arise in the kinematic wave models are a good example. The model tells us that there is an change marking for example the back of a queue. For many modelling and traffic management purposes this is the important information. In practice under normal conditions when approaching slow moving traffic cars slow smoothly in a short distance and in a short time but this period of declaration may well be smaller than the intervals over which average quantities are measured. It is here that we come up against the limit of the trying to follow developments in fluid dynamics too closely. in a real fluid although the region of the shock may be small compared to the dimensions of the fluid be studied it is still very large compared to the dimensions of molecules making up the fluid. This may not the case for traffic. However detail of the how slowing occurred may not be important; although it could be from, for example, a safety aspect. If a more detailed understanding of what happens at the shock wave is needed then a car following model may be more appropriate. It is also worth bearing in mind that a traffic stream does not truly represent a car, but rather the conditions that a smeared out "average" car would experience.

3 Kinematic Wave Models

We will take as our beginning the work of Lighthill and Whitham (1955)), and Richards (1956)). They introduced a model that describes the evolution of traffic flow, q , and traffic density, k , along a motorway or arterial road. The model gives useful insights into some traffic phenomena such as the creation and propagation of shock waves.

While the kinematic wave model is the most widely known of the continuum flow models there has been steady interest in the development of other models.((Kerner, Konhaüser and Shilke 1996), (Bui, Nelson and Sopasakis 1996)) There has been recently been a flurry of interest and number of new models proposed.((Zhang 2001),(Zhang 1999),(Jiang, Wu and Zhu 2002)).

The kinematic wave model makes two assumptions:

- The conservation of traffic. Following Whitham (1999), on any section of road $x_1 < x < x_2$ we have the conservation equation

$$q(x_2, t) - q(x_1, t) + \frac{\mathrm{d}}{\mathrm{d}t} \int_{x_1}^{x_2} k(x, t) dx = 0. \qquad (1)$$

For appropriate assumptions of differentiability this leads to the more familiar local equation of continuity

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0. \tag{2}$$

This is the assumption made by Lighthill and Whitham (1955) and Richards (1956) in their original work. However the existence of shockwaves shows that the assumption is not universally valid.

- The equilibrium assumption. The traffic flow q is a function of the traffic density k .

$$q = Q(k). \tag{3}$$

Since there is a maximum density of traffic, associated with the "jam" density of stationary traffic we must have $Q(0) = 0$ and $Q(k_j) = 0$. The graph of $Q(k)$ in the k-q plane is the Fundamental Diagram.

Empirical observations suggest that Q should be convex downward; lines joining a pair of points on the fundamental diagram must lie on or below the curve.(See Zhang (2001) for a brief discussion of the effects of non-convex fundamental diagrams.) The convexity assumption ensures that there is a unique local maximum, giving the maximum flow. It is usually assumed that this maximum occurs at a single value k_c. From the definition of traffic speed, $v = q/k$, we obtain the equivalent assumption that the equilibrium traffic speed is a function of density.

Assuming sufficient differentiability we then obtain the first order wave equation

$$\frac{\partial k}{\partial x} + Q'(k)\frac{\partial k}{\partial x} = 0. \tag{4}$$

The characteristics of this equation are the curves of constant density and are straight lines. They represent waves of constant density propagating with velocity $Q'(k)$.

The theory of kinematic waves successfully models, at least qualitatively, two important traffic phenomena:

- Traffic moves faster than traffic conditions: $v <= \omega$.
- The occurance of shock waves.

Although conceptually simple and useful for qualitative investigation the kinematic wave model proved difficult to work with practically. Given an initial distribution at the start of the link extracting information downstream was not easy because of the occurance of shockwaves. In particular the extraction of the travel time information required the solution of the traffic trajectory

$$\frac{\mathrm{d}v}{\mathrm{d}t} = Q(k)/k. \tag{5}$$

3.1 The Accumulated Flow Formulation

The next step forward was made by (Newell (1993a)). Newell showed that the theory could be formulated in terms of the accumulated flow A(t,x) , the total amount of traffic that has passed point x up to time t. The accumulated flow A(t,x) is assumed to be continuous and piecewise differentiable. The flow and density are then obtained from the time and space derivatives respectively.

$$q(x,t) = \frac{\partial \mathrm{A}(t,x)}{\partial t},\ k = -\frac{\partial \mathrm{A}(t,x)}{\partial x}. \tag{6}$$

In his formulation Newell notes that the link need not be homogeneous. The fundamental diagram is allowed to dependent on location.

$$q = Q(k,x). \tag{7}$$

This is an aspect of the kinematic wave model that has not yet been adequately explored.

In fact there is no theoretical reason why Q should not also be function of time.(See also (Nagel and Nelson 2005), (Daganzo 2004a) There do not however seem to be compelling reasons why this this is needed. It seems more likely that a small number of fundamental diagrams may be appropriate for day and night or wet and dry conditions.

Along a characteristic of the wave equation (a wave of constant density in the homogeneous case) we have that

$$\frac{\mathrm{dA}}{\mathrm{d}t} = q - \omega k. \tag{8}$$

Thus the accumulated flow can be found by integration, analytically in simple cases and by quadrature in other cases. For the homogeneous case $q - \omega k$ is constant and results is A(t,x) being a ruled surface. Owing to the occurance of shocks the solutions of the the differential equation (8) may result in a multi-valued solution. The solution surface $\mathrm{A}(t,x)$ is then taken to be the minimum envelope.

These insights into the kinematic wave model made it possible to perform useful calculations in traffic management investigations using the kinematic wave model as the flow model. (Heydecker and Addison 1996).(Heydecker and Addison 2005).

3.2 The Variational Formulation

More recently Daganzo (2004a), in an important development, has shown that the accumulated flow A(t,x) function is obtained as the solution of a minimisation problem given a boundary D on which $\mathrm{A}(t,x)$ is known.

The fundamental diagram is taken to be neither stationary nor homogeneous.

$$q = Q(k,t,x). \tag{9}$$

The function is required to be convex in k .

Let $P = (t_P, x_P)$ be the point for which we want to know $\mathrm{A}(t, x)$; $\mathbf{P}_P$ be the set of all valid paths in the t-x plane from D to P. A path $p(s) = (s, x(s))$ in the plane is called valid if it is continuous, piecewise differentiable and $\frac{\mathrm{d}x}{\mathrm{d}s}$ is within the range of possible wave speeds.(I have restricted to paths where the ordinate t is normalised, Daganzo only requires that the path be forward pointing in time.) In what follows all paths are required to be valid. For a valid path p Daganzo defines a functional $\Delta(p)$ that has the property that if the path is a characteristic of the underlying kinematic wave model, then $\Delta(p)$ is the change in A(t,x) along the path.

The variation $\Delta(p)$ along a path is given by

$$\Delta(p) = \int_{T_B}^{t_P} R(x'(t), t, x(t))dt. \tag{10}$$

The function $R(u, t, x)$ is derived from ((8)), which gives the rate at which A(t,x) changes along a wave. Define r by

$$r = Q(k, t, x) - uk\mathrm{where}u = Q_k(k, t, x). \tag{11}$$

Assuming Q is strictly convex in k we can invert to obtain k as a function of u. Substituting in ((11)) gives r as a function $R(u, t, x)$ of u, t and x. This function is defined for an for any valid path and hence so is the variation $\Delta(p)$. The essential result is

$$\mathrm{A}(t_P,\, x_P) = \min_{p \in \mathbf{P}_P} \{B_p + \Delta(p)\} \tag{12}$$

where B_p is the value of $\mathrm{A}(t, x)$ at its starting point on the boundary D.

In a companion paper (Daganzo 2004b) Daganzo shows how by superimposing a discrete lattice over the $t - x$-plane approximate solutions can be found with controlled errors. The application of this new approach is being vigorously explored by Daganzo and his coworkers, see for example (Daganzo and Mennendex 2005).

4 Higher Order Models

Higher order models had been introduced primarily to correct perceived defects of the original kinematic wave model. There seem to two main sources of perceived defects; dissatisfaction with the existence of shocks(Somewhat ironic given the title of Richard's paper) and "spontaneous" flow break down and non-equilibrium traffic conditions. The first models introduced were those of Payne((Kerner et al. 1996),(Zhang 2001)) and Whitham (Whitham 1999). The motivation was a better understanding of shock waves. (It is not clear how seriously Whitham was offering his model as it seems to a large extent a pedagogical devise it illustrate the need for additional theories to understand shocks in hyperbolic systems.) As is now well known these higher order models suffer from a number of deficiencies. (Daganzo 1995), (del Castillo, Pintado and Benitex 2005). The two primary deficiencies sighted are that they given rise to forward moving waves and more seriously to negative flows. The problem existence of forward moving waves need not in itself be fatal provided that they decay sufficiently quickly; drivers to after all have rear view mirrors and so may respond to events behind them and information may be propagated forward by vehicle overtaking, i.e. vehicles that are travelling faster than the equilibrium speed. If such waves exist observing them in practice will prove a formidable challenge. Only by analysing video footage on a second by second basis may one be able to detect them. Despite the critical assault of (Daganzo 1995) there have been recent efforts to produce higher order models that avoid the difficulties of the older models. (Aw and Rascle 2000) achieve this by doing away with the relaxation and diffusion terms that characterise the Payne-Whitham family of models. Instead they introduce a convective derivative. Another more concerted attempt to produce a model for non-equilibrium traffic is the work of Zhang (Zhang 1998), (Li and Zhang 2001).

The question still arises as to whether we need “higher order” models and for what purpose. For many purposes of traffic modelling and management the discontinuities of shockwaves do not matter provided they are properly interpreted. It makes little difference if the there is a discontinuous change in speed or speed changes continuously over a small distance, or in a short time. If detailed knowledge of the behaviour in this region is needed then car following models should give better and more useful insights.

The issue of ’flow break’ down is also potentially resolvable in other ways, for example the use of inhomogeneous fundamental diagrams. In a recent paper Nagel and Nelson (2005) have critically compared the kinematic wave model with observational data. This excellent paper shows the difficulties inherent in deciding that the kinematic wave model is inappropriate. The clarity and careful considerations that inform the paper seems to be byproduct of the very different attitudes and expectations of the two authors. They draw attention to a number of important issues that need to be addressed. One of the most important especially in terms of justifying “higher order” models is the question of whether flow break down is a truly emergent characteristic of traffic or whether it can be explained by external factors.

Regardless of the origins of “flow breakdown” there can be little doubt that traffic exhibits “phase” transitions between flow patterns with differing characteristics. It is in understanding the change from one “phase” to another, particular the sudden onset of a transition, that higher order models may play their part. In developing such models we must bare in mind that a traffic element in such a model does not represent an individual vehicle any more than a fluid element represents as molecule. The useful non-equilibrium models will give us an insight as to how traffic conditions evolve when conditions are away from the equilibrium.

5 Tomorrow

One of the aims of this paper is to identify topics for future investigation. There are of course many problems that require the application of traffic theory models. However I shall focus on problems that seem to me to relate directly to the further development of the traffic flow theory. The problems range from the possible to the near impossible.

- Understand properly the range of phenomena that can be explained by the kinematic wave model. In a particular what phenomena does the model predict when an inhomogeneous fundamental diagram is used? One would expect that the behaviour would be very sensitive to the changes in the fundamental diagram when free flow is near to the maximum capacity. This may give rise to a form of flow break down as a transition from free flow to congested flow occurs. This should be amenable to analysis.

- Can "spontaneous" flow break down be shown to be an intrinsic property of traffic flow by deriving it from car following models? And under what conditions will it occur?

- A coherent theory of phase transition. The need for a theory of phase transition is illustrated by the empirical so called reverse λ fundamental diagram. One interpretation of the diagram is that traffic can be come "super-saturated" during free flow giving rise to higher flows than are predicted by the equilibrium fundamental diagram. At some stage this super saturated state will need to relax back into the equilibrium state.

  More generally an understanding based on driver behaviour and car following models needs to be coherently developed. See (Zhang 2001) and (Daganzo 2002) for some recent developments and discussion.

  The development of continuum models has often being guided by developments in fluid dynamics. However in fluid dynamics there is kinematic theory and statistical mechanics to guide development and understanding. In an analogous way and most challengingly is to develop a continuum model theory based on car following/driver behaviour models.

- Is it possible derive a "fundamental diagram"? What sort of fundamental diagram arises?

- Based on ensembles use a car following/driver behaviour theory to derive a (continuum) theory based on average properties of the ensembles. What are the evolution equations that arise.

- Derive bounds giving the likely deviation of observed conditions from the means. The robustness of the results derived from statistical mechanics depend on various central limit theorems.(See the exposition of (Khinchin 1949).) We cannot rely on the central limit theorems to the same degree.

References

Aw, A. and Rascle, M. (2000). Resurrection of "second order" models of traffic flow, *Siam Journal of Applied Mathematics* **60**(3): 916–938.

Bui, D. D., Nelson, P. and Sopasakis, A. (1996). The generalized bimodal traffic stream model and two regime flow, *in* J.-B. Lesort (ed.), *Transportation and Traffic Theory*.

Daganzo, C. F. (1995). A variational formulation of kinematic waves: Solution methods, *Transportation Research B* **29B**(4): 277–286.

Daganzo, C. F. (2002). "a behavioral theory of multi-lane traffic flow part i: Long homogeneous freeway sections",, *Transportation Research B* **131-158**(2): 131–158.

Daganzo, C. F. (2004a). A variational formulation of kinematic waves: basic theory and complex boundary conditions, *Transportation Research B* **39B**(2): 187–196.

Daganzo, C. F. (2004b). A variational formulation of kinematic waves: Solution methods, *Transportation Research B* **39B**(10): 934–950.

Daganzo, C. F. and Mennendex, M. (2005). A variational formulation of kinematic waves: Bottleneck properties and examples, *in* Mahmasani (2005), pp. 345–364.

del Castillo, J. M., Pintado, P. and Benitex, F. G. (2005). A formulation for the reaction time of traffic flow models, *in* C. F. Daganzo (ed.), *Transportation and Traffic Theory*, Elsevier, pp. 387–405.

Heydecker, B. and Addison, J. (1996). An exact expression of dynamic traffic equilibrium, *in* J.-B. Lesort (ed.), *Transportaion and Traffic Theory*, Pergamon, pp. 359–384. Proceedings of the 13^{th} ISTTT.

Heydecker, B. G. and Addison, J. D. (2005). Analysis of dynamic traffic equilibrium with departure time choice, *Transportaion Science* **36**(1): 39–57.

Jiang, R., Wu, Q.-S. and Zhu, Z.-J. (2002). A new continuum model for traffic flow and numerical tests, *Transportation Research* **36B**(5): 405–19.

Jin, W.-L. and Zhang, H. M. (2003). The inhomogeneous kinematic wave traffic flow model as a resonant nonlinear system, *TranSci* **37**(3): 294–311.

Kerner, B. S., Konhaüser, P. and Shilke, M. (1996). A new approach to problems in traffic flow theory, *in* J.-B. Lesort (ed.), *Transportation and Traffic Theory*, pp. 119–146. Proceedings of the 13^{th} ISTTT.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*, Courier Dover Publications.

Lebacque, J.-P. and Khoshyaran, M. M. (2005). First-order macroscopic traffic flow models: Intersection modeling, network modeling, *in* Mahmasani (2005), pp. 365–386.

Li, T. and Zhang, H. M. (2001). The mathematical theory of an enhanced nonequilibrium traffic flow model, *Network and Spatial Economic* **1**(1): 167–178.

Lighthill, M. and Whitham, G. (1955). On Kinematic Waves,II. A Theory of Traffic Flow on Long Crowded Roads., *Proc. Royal Soc.* **229A**: 317–345.

Mahmasani, H. S. (ed.) (2005). $16^t h$ *International Symposium on Transportation and Traffic Theory*, Elsevier.

Nagel, K. and Nelson, P. (2005). A critical comparison of the kinematic-wave model with observational data, *in* Mahmasani (2005), pp. 145–163.

Nelson, P. (2000). Synchronised traffic flow from a modified lighthill-whitham model, *Physics Review* **61E**(6): R6052–5.

Newell, G. F. (1993a). A simplified theory of kinematic waves in highway traffic,Part I: General theory, *Transportation Res-B*.

Newell, G. F. (1993b). A simplified theory of kinematic waves in highway traffic,Part II: Queueing at freeway bottlenecks, *Transportation Res-B* **27B**(4): 289–303.

Newell, G. F. (1993c). A simplified theory of kinematic waves in highway traffic,Part III: Multi-destination flows, *Transportation Res-B*.

Richards, P. I. (1956). Shockwaves on the highway, *Operations Research* **4**: 42–51.

Whitham, G. B. (1999). *Linear and Nonlinear Waves*, Pure and Applied Mathematics, John Wiley and Sons, Inc.

Zhang, H. M. (1998). A theory of nonequilbrium traffic flow, *Transportaion Research B* **32**(7): 485–498.

Zhang, H. M. (1999). Analysis of the stability and wave properties of a new continuum traffic theory, *Transportation Research* **33B**(6): 399–415.

Zhang, H. M. (2001). New perspectives on continuum traffic flow models, *Network and Spatial Economic* **1**(1): 9–34.