Fitting fundamental diagrams to LWR using flow data

Jeremie Coullon, Yvo Pokern, Benjamin Heydecker, and Eugeny Buldakov

July 14, 2017

The Lighthill Whitham Richards (LWR, in [2] and [3]) model of traffic flow was developed over 50 years ago to model the propagation of shock waves on motorways. Although it has been extended in many ways to model different phenomena and mechanisms in traffic (capacity drop, vehicle acceleration, flow instability), a nonparametric statistical approach to inferring the underlying partial differential equation is so far absent from the literature.

To use the LWR model, a suitable fundamental relationship must be chosen and fitted to data. This relationship gives rise to the so-called fundamental diagram (FD). Choosing the relationship is usually done based on theoretical properties of the FD and fit to flow-density traffic data. Such theoretical properties include the concavity (or non-concavity) of the FD (see [1]) and the interpretability of its parameters.

There are two limitations to this approach of fitting the FD. Firstly, quantifying the uncertainty in the estimate of FD parameters in the flow-density plane would not include a time component. As flow-density data is generated from a process that varies in space and time, the fit or misfit of a FD should rather be evaluated by solving LWR and comparing the output to traffic data in both space and time. Secondly, density must be estimated from traffic data, which leads to different results depending on the estimation procedure used. For example, density can be estimated from flow and speed data or from occupancy data, yielding different values of flow for a given density value. Furthermore, each of these different procedures have disadvantages; for example, one needs either to estimate the average vehicle length if one uses occupancy, and one needs to convert time mean speed to space mean speed if one uses speed and flow.

To remedy the first of these limitations, we estimate the parameters in two FDs: one due to Underwood ([4]) and another due to del Castillo ([1]). We estimate these parameters by solving the inverse problem for LWR using a Bayesian framework. We elicit a prior for the FD parameters, and construct a likelihood by solving LWR and comparing it to flow data on a stretch of M25 motorway using a Poisson statistical model. As we consider traffic data in both time and space, the uncertainty quantification for the parameters is then justified, as discussed above.

The result of this inference is a joint distribution of the parameters (the posterior distribution) which describes the uncertainty in the parameters. This uncertainty arises primarily because neither are flow and density measured in continuous space and time, nor can they be. The statistical model thus bridges the gap between discrete observations and continuous models introducing uncertainty in the process. Other important aspects are features and properties of traffic that LWR is unable to reproduce such as multiple lanes, variation in flow parameters (such as capacity) or bottlenecks in the road - these are also subsumed into the statistical error model as a first approximation. We compare this fitting procedure to a straightforwards fitting of fundamental diagrams directly to flow-density data, and find that the fitted wavespeeds are different; the proposed methodology finds the correct wavespeeds in traffic.

The data is from MIDAS detectors which measures the state of traffic on a 1-minute and 500m aggregate scale. We consider a stretch of road that has no on/off ramps (as LWR cannot model on/off ramps without adding source terms), and consider the road to be single lane motorway with heterogeneous drivers. These are necessary assumptions to make to be able to use LWR (namely, without extensions such as the ability to model multiple lanes).

To address the second limitation, we would like the inference to be independent of the density estimation procedure used. However to solve LWR numerically we need density data as boundary conditions in the inlet and outlet of the stretch of road. To solve this we therefore consider these boundary conditions to be unknowns, and impute these functional nuisance parameters as part of a fully Bayesian methodology.

Multiplying the prior and the likelihood results in the posterior distribution, which is the probability distribution of the FD parameters given the flow data. As this posterior is not available analytically, we generate samples from it using a Markov chain Monte Carlo (MCMC) algorithm. We use a Gibbs sampler that samples from the conditional distributions of the FD parameters and boundary conditions, with a Metropolis-Hastings accept-reject criterion. As proposals we use normal/log normal distributions with diminishing adaptation of variance for the FD parameters, and a preconditioned random walk (RWMH) on the space of Ornstein-Uhlenbeck paths for the boundary condition.

Finally, we include an entropy based goodness of fit testing which takes predictive uncertainty into account: we test the predictive power of the model with estimated FD on unseen traffic data.

From this inference, we can also recover from the boundary condition imputation a density estimate that yields a good LWR fit. From this we can evaluate the density estimation procedures commonly used and identify the one that corresponds most closely to the imputed boundary conditions.

The next step in the research is to perform inference of a nonparametric FD. A nonparametric formulation of a function does not assume a specific parametric form, so is allowed to vary freely. This allows a flexible FD to be fitted, revealing clearly the extent of the uncertainty in this estimate (for example in the FD at high density). Using the posterior resulting from the estimation, one can also perform hypothesis testing on such theoretical properties of the FD as concavity, and provide a data-driven contribution to this debate.

In summary, this research consists of a nonparametric statistical treatment of the parameters in LWR: namely the FD and its nuisance parameters (the boundary conditions). As we obtain a posterior distribution of FDs rather than just point estimates, we gain an understanding of the uncertainty in these estimates. The outcomes of our approach thus include an assessment of model fit for different parametric forms of FD, estimates of traffic quantities such as capacity or critical density and a novel density estimation procedure each with consideration of the uncertainties involved.

References

[1] del Castillo, J. (2012). Three new models for the flow-density relationship: derivation and testing for freeway and urban data. Transportmetrica 8(6): 443-465

[2] Lighthill, M. H., and Whitham, G.B., 1955. On Kinematic Waves 2: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society of London A, 229, pp.317-345.

[3] Richards, P.I., 1956. Shock Waves on the Highway. *Operations Research*, 4 (1) pp.42-51

[4] Underwood, R.T., 1961. Speed, Volume, and Density Relationships: Quality and Theory of Traffic Flow. *Yale Bureau of Highway Traffic*, 141-188.