FitFun: Improved noise models for Fundamental Diagrams

D.M. Bramich[a,*], Mónica Menéndez[a], Lukas Ambühl[b]

[a] Division of Engineering, New York University Abu Dhabi, P.O. Box 129188, Saadiyat Island, Abu Dhabi, UAE
[b] Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zurich, Switzerland

Keywords: Fundamental diagram of road traffic, Flow-density relationship, Loop detector data, Statistical modelling, GAMLSS

1. Introduction

At a macroscopic level, the fundamental equation of traffic flow:

$$q(x, t) = k(x, t)\, v(x, t) \tag{1}$$

provides a relationship between the average properties of vehicular traffic along a road at a specific location x and time t. The traffic density $k(x, t)$ is the number of vehicles per unit length at time t (veh km^{-1}), the traffic flow $q(x, t)$ is the number of vehicles passing location x per unit time (veh h^{-1}), and the space-mean speed $v(x, t)$ is the arithmetic mean of the instantaneous speeds of the vehicles over some distance (km h^{-1}; Wardrop 1952; Edie 1965). By adopting a speed-density model such that speed is a function of density only (Lighthill & Whitham 1955; Daganzo 1997), then one may write:

$$q(k) = k\, v(k) \tag{2}$$

The flow-density diagram and the relationship therein was named the "Fundamental Diagram of Road Traffic" by Haight (1963).

The fundamental diagram (FD) is central to traffic modelling (e.g. Edie 1965; Daganzo 1994; Muñoz & Daganzo 2002) and traffic management (e.g. Papageorgiou 1983; Lebacque 2005). Empirically, it has been observed to exist for different road types (i.e. highways, freeways, urban roads) in a large number of studies (e.g. Greenshields 1935; Van Aerde & Rakha 1995; Wu, Liu & Geroliminis 2011; He et al. 2018). The functional form of the FD relationship can vary considerably between individual roads, even when they are of the same type, especially since the form also depends on the location of the detector along the road (e.g. Haight 1963; Hall, Hurdle & Banks 1992; Courbon & Leclercq 2011; Ambühl et al. 2019). Furthermore, empirical FD data are inherently noisy due to the stochastic nature of traffic flow (e.g. driver behaviour/types, weather, traffic composition, transient/non-stationary traffic states, hysteresis effects, etc.) and the effects of traffic control (e.g. signal timings), combined in some measure with the uncertainty introduced during data collection and/or processing (Coifman 2014). A salient feature of empirical FDs that has been previously observed, but not yet properly considered in attempts at modelling FDs, is the change in the variance (i.e. scale) and shape (i.e. deviations from Gaussianity) of the flow (or speed) distribution as a function of density (e.g. Helbing 1997a; Helbing 1997b; Knospe et al. 2002; Guan & He 2008; Aguiléra & Tordeux 2014; Maghrour Zefreh & Török 2020). Consequently, modelling empirical FDs is a tricky business.

Starting with Greenshields (1935), many papers have been published proposing a plethora of different models for the functional form of the FD that can be used to fit empirically observed data (e.g. Gazis, Herman & Rothery 1961; Van Aerde 1995; Wang et al. 2011; Sun, Pan & Gu 2014, etc.). Specifically for the flow-density FD, previously proposed models for the functional form $q(k)$ have been comprehensively collated and reviewed by

*Corresponding author
Email address: dan.bramich@hotmail.co.uk (D.M. Bramich)

Bramich, Menéndez & Ambühl (2021) (hereafter B2021). Unfortunately, the fitting method that has been used in the vast majority of these studies is the method of least squares, which is equivalent to adopting a Gaussian noise model for the flow (or speed) with a variance that is independent of density, and employing maximum likelihood estimation. Clearly, based on our previous discussion, the conventional fitting method of least squares and the equivalent noise model are a poor choice for fitting/modelling empirical FD data. Not accounting properly for the noise limits the explanatory power of any model and introduces unwanted bias in the fit. B2021 recommend that the research community should now shift its focus away from developing increasingly complicated models for the functional form of the FD and towards exploring alternative and more appropriate noise models.

In this extended abstract, we follow up on this recommendation by describing and demonstrating a general framework (FitFun) for fitting/modelling any empirical FD that not only allows for the specification of a model component for the functional form of the FD relationship, but that also incorporates a model component for the observed noise. FitFun is general enough to encompass the vast majority of FD models that have already been proposed in the literature, and it includes model performance metrics so that model comparisons are easily facilitated.

2. Methodology: The FitFun modelling framework

FitFun allows for the specification of any model for an FD from the family of Generalised Additive Models for Location, Scale, and Shape (GAMLSS; Rigby & Stasinopoulos 2005; Stasinopoulos et al. 2017). A GAMLSS model that is specific to the flow-density FD for a set of flow-density measurement pairs $\{(q_i, k_i)\}$, where $i = 1, 2, \ldots, N_{\text{dat}}$, may be written as:

$$Q_i \overset{\text{ind}}{\sim} \mathcal{D}\,(\,\mu\,(k_i)\,, \sigma\,(k_i)\,, \nu\,(k_i)\,, \tau\,(k_i)\,) \tag{3}$$

where Q_i is a univariate random variable for flow corresponding to the ith flow measurement, $\overset{\text{ind}}{\sim}$ stipulates that the random variables Q_i have independent probability distributions, $\mathcal{D}$ is a general probability distribution with a parametric form, and μ, σ, ν, and τ are the location, scale, and shape parameters of $\mathcal{D}$. The distribution parameters are modelled as linear, non-linear, or non-parametric smoothing functions (or some combination thereof) of traffic density k.

The GAMLSS family of statistical models is very general in the sense that any parametric probability distribution can be assumed (e.g. Gaussian, gamma, Weibull, Poisson, etc.). In order to specify a GAMLSS model, one must select the probability distribution $\mathcal{D}$, and define the functions $\mu\,(k)$, $\sigma\,(k)$, $\nu\,(k)$, and $\tau\,(k)$. Fitting a GAMLSS model to data is done using maximum likelihood (ML) estimation or maximum penalised likelihood (MPL) estimation as appropriate (Rigby & Stasinopoulos 2005). Various algorithms for fitting a GAMLSS model are implemented in the `gamlss` software[1] in the R[2] programming language.

An important aspect of scientific modelling is model comparison, where one faces the task of selecting an optimal/best model for a data sample from a set of candidate models. In this context, "optimality" refers both to the Principle of Parsimony, in that the best model should constitute the simplest model that provides a good fit to the data without under- or over-fitting, and to appropriate/relevant model performance measure(s). Information criteria are used as a way to evaluate models with different numbers of parameters. FitFun uses the Akaike information criterion (AIC; Akaike 1974) and the Bayesian information criterion (BIC; Schwarz 1978) to compare model fits:

$$\text{AIC} = -2 \ln \mathcal{L}(\hat{\boldsymbol{\theta}}) \;+\; 2N_{\text{par}} \tag{4}$$

$$\text{BIC} = -2 \ln \mathcal{L}(\hat{\boldsymbol{\theta}}) \;+\; N_{\text{par}} \ln N_{\text{dat}} \tag{5}$$

where $\mathcal{L}\,(\boldsymbol{\theta})$ is the likelihood function, $\hat{\boldsymbol{\theta}}$ is a vector of ML estimators for the model parameters $\boldsymbol{\theta}$, and N_{par} is the total number of free parameters in the model. Model selection with the AIC or BIC is performed by minimising $-2 \ln \mathcal{L}(\boldsymbol{\theta})$ for each model, and then minimising the AIC or BIC over the full set of models under consideration.

[1] https://www.gamlss.com
[2] https://www.r-project.org

Literally, hundreds of models have been previously proposed for the FD relationship (see Section 3 in B2021). Fortunately, the vast majority of them can be written as GAMLSS models. FitFun therefore constitutes a modelling framework that unifies the efforts of previous researchers, and that is general enough to provide plenty of scope for the specification, exploration, and comparison of more sophisticated FD models.

3. Results

We aim to test FitFun on an unprecedently large and diverse set of loop detector (LD) data consisting of 10,150 empirical flow-density FDs drawn from all road types across 25 cities worldwide. This data set was used by B2021 to compare fits of 50 conventional flow-density FD models, and it is a cleaned and augmented version of Urban Traffic Data 2019 (UTD19) originally curated by Loder et al. 2019. The properties of the data set are given in Table 1.

We briefly demonstrate the power of the FitFun modelling framework with example fits to a couple of the empirical FDs. We consider two noise model components, the first being the conventional noise model (GaussSigCon) given by:

$$Q_i \overset{\text{ind}}{\sim} \mathcal{N}\left(q(k_i), \sigma_{\text{con}} \right) \qquad \text{for} \;\; i = 1, 2, \ldots, N_{\text{dat}} \tag{6}$$

where $\mathcal{N}(\mu, \sigma)$ represents a Gaussian distribution with mean μ and standard deviation σ, and $q(k_i)$ is the predicted flow at k_i given the adopted model component for the functional form $q(k)$. The second (more sophisticated) noise model (SN2SigNS5pNuNS3p) is given by:

$$Q_i \overset{\text{ind}}{\sim} \mathcal{SN}_2\left(q(k_i), \exp\left(S_{\sigma,\text{nat}}(k_i)\right), \exp\left(S_{\nu,\text{nat}}(k_i)\right) \right) \qquad \text{for} \;\; i = 1, 2, \ldots, N_{\text{dat}} \tag{7}$$

where $\mathcal{SN}_2(\mu, \sigma, \nu)$ represents a Skew Normal Type II distribution[3] with mode μ, scale parameter $\sigma > 0$, and skewness parameter $\nu > 0$, and $S_{\sigma,\text{nat}}(k)$ and $S_{\nu,\text{nat}}(k)$ are natural cubic splines with five and three effective free parameters, respectively.

In each row of Figure 1, we plot fits of two different models to the same empirical FD data. In the left column, the noise model component is always GaussSigCon, while in the right column, it is always SN2SigNS5pNuNS3p. For the top row, the model component for the functional form is from Van Aerde (1995), while for the bottom row, it is from Drake, Schofer & May (1966). Both the AIC and BIC indicate that the models with the SN2SigNS5pNuNS3p noise model component are $>10^{52}$ times more likely than their GaussSigCon counterparts. The median curve and the quantile regions follow the data better in the right-hand plots, although clearly there is still plenty of room for further improvement. The poorer fits for GaussSigCon seem to give less reliable estimates of physical quantities such as free-flow speed, capacity, critical occupancy, and jam occupancy.

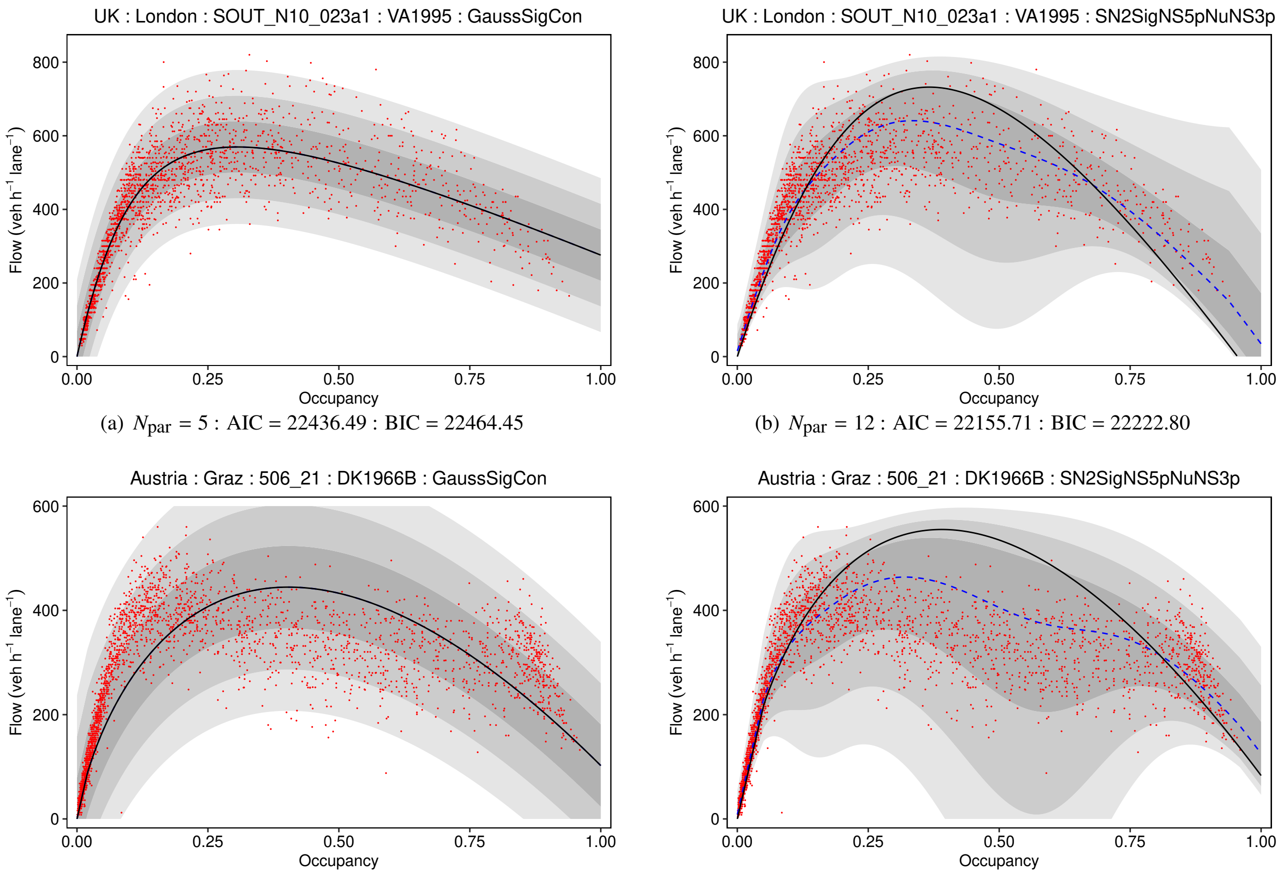
4. Conclusions

We have introduced a general framework called FitFun for fitting/modelling any empirical FD that not only allows for the specification of a model component for the functional form of the FD relationship, but that also incorporates a model component for the observed noise. We aim to test and validate FitFun by performing fits of a new set of FD models to an unprecedently large and diverse data set of empirical FDs. FitFun will help researchers move away from the restrictive conventional noise model and to focus on more appropriate noise models. This will further our understanding of traffic flow-density-speed inter-relationships at the road (or link) level by properly accounting for both their functional forms and the observed noise.

References

Aguiléra V. & Tordeux A., 2014, *A new kind of fundamental diagram with an application to road traffic emission modeling*, Journal of Advanced Transportation, 48, 165-184, doi:10.1002/atr.1248

[3] Also known as a two-piece Normal distribution. The Gaussian distribution is a special case when $\nu = 1$.



(a) $N_{\text{par}} = 5$: AIC = 22436.49 : BIC = 22464.45

(b) $N_{\text{par}} = 12$: AIC = 22155.71 : BIC = 22222.80

(c) $N_{\text{par}} = 4$: AIC = 32879.71 : BIC = 32903.52

(d) $N_{\text{par}} = 11$: AIC = 29625.69 : BIC = 29691.15

Figure 1: Plots of empirical FDs (red dots). Both plots on a single row show the same empirical FD data but with different fitted models, where the difference between the models is in the adopted noise model component. In each plot, the fitted model is represented by a black continuous curve for $q(k)$ and sequentially lighter grey regions for the $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ quantile regions. The median curve is represented by the blue dashed curve. The plot titles are all of the same format listing the country, city, loop detector ID, functional form model component (code from B2021), and noise model component.

Table 1: Properties of the B2021 data sample.

Country	City	No. Of LDs	No. Of Measurements	No. Of Good Measurements	Duration Of Time Interval (min)	Start Date	End Date	No. Of Days With Measurements
Austria	Graz	223	632,160	549,281	5	2016-04-04	2016-09-23	10
Canada	Toronto	70	398,304	351,442	15	2016-09-01	2017-01-31	61
France	Bordeaux	195	392,832	347,707	5	2016-11-21	2016-11-27	7
France	Marseille	60	864,000	759,673	3	2017-06-01	2017-07-01	31
France	Toulouse	268	900,480	730,195	3	2008-05-16	2008-06-27	7
Germany	Augsburg	355	1,833,843	1,309,142	5	2017-05-06	2017-05-25	20
Germany	Bremen	419	2,790,010	2,445,416	3	2016-09-19	2016-10-02	14
Germany	Constance	84	169,344	159,217	5	2017-02-13	2017-02-19	7
Germany	Darmstadt	189	2,701,000	1,966,707	3	2015-09-21	2016-05-06	39
Germany	Essen	29	235,888	234,446	5	2017-03-27	2017-09-30	35
Germany	Hamburg	294	14,554,197	10,099,450	3	2016-08-27	2016-12-09	105
Germany	Kassel	358	400,880	383,637	5	2016-08-28	2016-09-02	6
Germany	Speyer	134	899,392	781,421	3	2016-09-19	2016-10-02	14
Germany	Stuttgart	173	385,632	352,732	5	2016-03-21	2016-07-22	8
Germany	Wolfsburg	106	712,245	606,292	3	2016-09-19	2016-10-02	14
Italy	Cagliari	58	1,353,196	1,313,489	3	2016-05-16	2016-07-29	50
Japan	Tokyo	1,295	23,123,520	19,557,780	2.5	2017-07-01	2017-07-31	31
Spain	Madrid	315	1,814,400	1,086,628	5	2016-08-29	2017-11-11	20
Spain	Santander	60	190,080	115,283	5	2016-06-17	2016-12-02	11
Switzerland	Basel	55	110,880	105,258	5	2016-10-24	2016-10-31	8
Switzerland	Bern	342	691,314	587,992	5	2016-10-24	2016-10-31	8
Switzerland	Luzern	127	21,785,925	16,913,775	3	2015-01-01	2015-12-31	365
Switzerland	Zurich	761	65,910,654	65,826,865	3	2015-10-26	2018-06-30	372
Taiwan	Taipeh	84	490,783	444,306	3	2017-09-18	2017-10-01	14
United Kingdom	London	4,096	24,842,716	20,006,512	5	2015-05-15	2016-05-22	23
Total	-	10,150	168,183,675	147,034,646	-	-	-	-

Akaike H., 1974, *A New Look at the Statistical Model Identification*, IEEE Transactions on Automatic Control, 19, 716-723, doi:10.1109/TAC.1974.1100705

Ambühl L. et al., 2019, *Approximative Network Partitioning for MFDs from Stationary Sensor Data*, Transportation Research Record: Journal of the Transportation Research Board, No. 2673, TRB, National Research Council, Washington D.C., 94-103, doi:10.1177/0361198119843264

Bramich D.M., Menéndez M. & Ambühl L., 2021, *Fitting Fundamental Diagrams of road traffic: A comprehensive review and comparison of models using an extensive data set*, Submitted (B2021)

Coifman B., 2014, *Jam Occupancy and Other Lingering Problems with Empirical Fundamental Relationships*, Transportation Research Record: Journal of the Transportation Research Board, No. 2422, TRB, National Research Council, Washington D.C., 104-112, doi:10.3141/2422-12

Courbon T. & Leclercq L., 2011, *Cross-comparison of Macroscopic Fundamental Diagram Estimation Methods*, Procedia - Social and Behavioral Sciences, 20, 417-426, doi:10.1016/j.sbspro.2011.08.048

Daganzo C.F., 1994, *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*, Transportation Research Part B, 28, 269-287, doi:10.1016/0191-2615(94)90002-7

Daganzo C.F., 1997, *Fundamentals of Transportation and Traffic Operations*, Elsevier Science Inc., New York, Chapter 4, 66-160

Drake J.S., Schofer J.L. & May A.D., 1966, *A Statistical Analysis of Speed Density Hypotheses*, Highway Research Record, No. 154, HRB, National Research Council, Washington D.C., 53-87

Edie L.C., 1965, *Discussion of traffic stream measurements and definitions*, In “Proceedings of the Second International Symposium on the Theory of Traffic Flow”, Paris, France, 139-154

Gazis D.C., Herman R. & Rothery R.W., 1961, *Nonlinear Follow-the-Leader Models of Traffic Flow*, Operations Research, Vol. 9, No. 4, 545-567, doi:10.1287/opre.9.4.545

Greenshields B.D., 1935, *A study of traffic capacity*, Highway Research Board Proceedings, 14, 448-477

Guan W. & He S., 2008, *Statistical features of traffic flow on urban freeways*, Physica A, 387, 944-954, doi:10.1016/j.physa.2007.09.036

Haight F.A., 1963, *Mathematical Theories of Traffic Flow*, Academic Press Inc., New York, Volume 7, Chapter 3

Hall F.L., Hurdle V.F. & Banks J.H., 1992, *Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways*, Transportation Research Record, No. 1365, TRB, National Research Council, Washington D.C., 12-18

He Z., Chen M., Wang L. & Li M., 2018, *Empirical Analysis of Urban Arterial Fundamental Diagram*, Proceedings of the 37th Chinese Control Conference, Wuhan, 7719-7723, doi:10.23919/ChiCC.2018.8483985

Helbing D., 1997a, *Empirical traffic data and their implications for traffic modeling*, Physical Review E, 55, R25-R28, doi:10.1103/PhysRevE.55.R25

Helbing D., 1997b, *Fundamentals of traffic flow*, Physical Review E, 55, 3735-3738, doi:10.1103/PhysRevE.55.3735

Knospe W. et al., 2002, *Single-vehicle data of highway traffic: Microscopic description of traffic phases*, Physical Review E, 65, 056133, doi:10.1103/PhysRevE.65.056133

Lebacque J.P., 2005, *Intersection Modeling, Application to Macroscopic Network Traffic Flow Models and Traffic Management*, In “Traffic and

Granular Flow 2003”, Springer Berlin, 261-278, doi:10.1007/3-540-28091-X_26
Lighthill M.J. & Whitham G.B., 1955, *On kinematic waves II. A theory of traffic flow on long crowded roads*, Proceedings Of The Royal Society A, 229, 317-345, doi:10.1098/rspa.1955.0089
Loder A., Ambühl L., Menéndez M. & Axhausen K.W., 2019, *Understanding traffic capacity of urban networks*, Scientific Reports, 9, 16283, doi:10.1038/s41598-019-51539-5 (L2019)
Maghrour Zefreh M. & Török Á., 2020, *Distribution of traffic speed in different traffic conditions: An empirical study in Budapest*, Transport, 35, 68-86, doi:10.3846/transport.2019.11725
Muñoz J.C. & Daganzo C.F., 2002, *Moving Bottlenecks: A Theory Grounded on Experimental Observation*, Transportation and Traffic Theory in the 21st Century, Emerald Group Publishing Limited, 441-461, doi:10.1108/9780585474601-022
Papageorgiou M., 1983, *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*, Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, ISBN: 978-3-540-12237-1
Rigby R.A. & Stasinopoulos D.M., 2005, *Generalized additive models for location, scale and shape*, Applied Statistics Series C, 54, 507-554, doi:10.1111/j.1467-9876.2005.00510.x
Schwarz G., 1978, *Estimating the Dimension of a Model*, The Annals of Statistics, 6, 461-464, doi:10.1214/aos/1176344136
Stasinopoulos M.D. et al., 2017, *Flexible Regression and Smoothing: Using GAMLSS in R*, CRC Press LLC, Boca Raton
Sun L., Pan Y. & Gu W., 2014, *Data mining using regularized adaptive B-splines regression with penalization for multi-regime traffic stream models*, Journal of Advanced Transportation, 48, 876-890, doi:10.1002/atr.1232
Van Aerde M., 1995, *A Single Regime Speed-Flow-Density Relationship for Freeways and Arterials*, 74th Annual Meeting of the Transportation Research Board, Paper No. 950802, Washington D.C.
Van Aerde M. & Rakha H., 1995, *Multivariate Calibration of Single-Regime Speed-Flow-Density Relationships*, Vehicle Navigation and Information Systems Conference (VNIS), 334-341
Wang H., Li J., Chen Q.-Y. & Ni D., 2011, *Logistic modeling of the equilibrium speed-density relationship*, Transportation Research Part A, 45, 554-566, doi:10.1016/j.tra.2011.03.010
Wardrop J.G., 1952, *Some Theoretical Aspects of Road Traffic Research*, Proc. Instn. Civ. Engrs., Part II, 1, 325-378
Wu X., Liu H.X. & Geroliminis N., 2011, *An empirical analysis on the arterial fundamental diagram*, Transportation Research Part B, 45, 255-266, doi:10.1016/j.trb.2010.06.003