

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Statistical modelling and inference for traffic networks



**MASSEY
UNIVERSITY**

Katharina Parry
Institute of Fundamental Sciences
Massey University

A thesis submitted for the degree of
Doctor of Philosophy

2012

This thesis is my own account of my research and contains, as its main content, work that has not been previously submitted for a degree at any university.

Katharina Parry
September 2012

Acknowledgements

The world needs more people like Martin Hazelton. Many many thanks for the patience, guidance, encouragement, more encouragement and generally excellent supervision. I would like to say thank-you to Olly, who has been a great companion throughout the bad, not as bad and sometimes downright ugly stages of doing this PhD and for believing in me even when I did not. Many cheers to Leo, who knows how to makes things happen when I cannot get them to work, as well as moral support. Overall, I appreciate that nobody who was involved with me and this thesis in these past years has lost their sense of humour. And finally, I express my gratitude for the Marsden Fund scholarship as it provided the opportunity for me to do this research.

Abstract

There are two facets that are important in providing reliable forecasts from observed traffic data. The first is that the model used should describe and represent as many characteristics of the system as possible. The second is that the estimates of the model parameters need to be accurate. We begin with improved methods of statistical inference for various types of models and using various types of data; and then move onto the development of new models that describe the day-to-day dynamics of traffic systems.

Calibration of transport models for traffic systems gives rise to a variety of statistical inference problems, such as estimation of travel demand parameters. Once the ways in which vehicles move through the network are known, statistical inference becomes straightforward, however, at present, the data available are predominantly vehicle counts from a set of links in the network. The fundamental problem is that these vehicle counts do not uniquely determine the route flows, as there are a large number of possible route flows that could have led to a given set of observed link counts.

A solution to this problem is to simulate the latent route flows conditional on the observed link counts in a Markov Chain Monte Carlo sampling algorithm. This is challenging because the set of feasible route flows will typically be far too large to enumerate in practice, meaning that we must simulate from a set that we cannot fully specify. An innovative piece of work here was the extension of an existing sampling methodology that works only for linear networks to be applicable for tree networks. In simulation studies where we use the sample to estimate average route flows, we show that our method provides more reliable estimates than generalised least squares methods. This is to be expected given that our method exploits information available via second order properties of the link counts.

We provide another demonstration of how this generalised sampler can be applied whenever the need to sample from the set of latent route flows is pivotal for making statistical inference. We use the sampler to estimate travel demand parameters for day-to-day dynamic process models, an important class of model where the data has been collected on successive days and hence allows for inference using the evolution of the traffic flows over time.

A new type of data, route flows from tracked vehicles, is becoming increasingly available through emerging technologies. Our contribution was to develop a statistical likelihood model that incorporates this routing information into currently used link-count data only models. We derive some tractable normal approximations thereof and perform likelihood-based inference for these normal models under the assumption that the probability of vehicle tracking is known.

In our analysis we find that the likelihood shows irregular behaviour due to boundary effects, and provide conditions under which such behaviour will be observed. For regular cases we outline connections with existing generalised least squares methods. The theoretical analysis are complemented by simulation studies where we consider the tracking probability to be unknown and the effects on the accuracy in estimation of origin-destination matrices under estimated and/or misspecified models for this parameter.

Real link flow count data observed on a sequence of days can exhibit considerable day-to-day variability. A better understanding of such variability has increasing policy-relevance in the context of network reliability assessment and the design of intelligent transport systems. Conventional day-to-day dynamic traffic assignment models are limited in terms of the extent to which non-stationary changes in traffic flows can be represented.

In this thesis we introduce and develop an advanced class of models by replacing a subset of the fixed parameters in currently used traffic models with random processes. These resulting models are analogous to Cox process models. They are conditionally non-stationary given any realisation of the parameter processes. Numerical examples demonstrate that this new class of doubly stochastic day-to-day traffic assignment models is able to reproduce features such as the heteroscedasticity of traffic flows observed in real-life settings.

Publications arising from thesis

Parry, K. and Hazelton, M.L. (2012), 'Estimation of origin-destination matrices from link counts and sporadic routing data', *Transportation Research Part B* **46**,175-188.

Parry, K. and Hazelton, M.L. (2013), 'Bayesian inference for day-to-day dynamic traffic models'. Accepted for publication in *Transportation Research Part B*.

Parry, K., Watling, D.P. and Hazelton, M.L. (2012), 'A new class of doubly stochastic day-to-day dynamic traffic assignment models'. Submitted for publication.

Contents

Contents	vii
List of Figures	xi
List of Tables	xv
1 Traffic network models	1
1.1 Introduction	1
1.2 Traffic assignment models	6
1.2.1 Equilibrium models	6
1.2.2 Deterministic/Stochastic	7
1.2.3 Static/Dynamic	9
1.3 Statistical methods of OD matrix estimation	10
1.3.1 Maximum Entropy methods	10
1.3.2 Generalised least squares estimation	11
1.3.3 Maximum likelihood estimation	13
1.3.4 Bayesian approach	15
1.4 Structure of thesis	17
2 Improved MCMC sampling for tree networks	19
2.1 Introduction	19
2.2 Markov Chain Monte Carlo sampling	23
2.2.1 Metropolis-Hastings algorithm	25
2.2.2 Gibbs algorithm	29
2.2.3 MH-within-Gibbs algorithm	30
2.3 Methods for sampling route flows	31
2.3.1 General solution	31
2.3.2 Improvement for linear transit networks	34

Contents

2.3.3	Extension to tree networks	39
2.4	Simulation study	43
2.4.1	Topology A	46
2.4.2	Topology B	49
2.4.3	Topology C	50
2.5	Conclusion	52
3	Inference for Day-to-Day Traffic Assignment Models	55
3.1	Introduction	55
3.2	The day-to-day model likelihood	57
3.3	Modification of MCMC algorithm	59
3.3.1	Conditional Sampling of Route Flow Vectors	63
3.4	Numerical studies	64
3.4.1	Artificial example	65
3.4.1.1	Reconstruction of route flows \mathbf{y}	66
3.4.1.2	Estimation of demand parameters β and γ	69
3.4.1.3	Estimation of sensitivity ψ	71
3.4.2	State Highway 16	72
3.5	Conclusion	74
4	Incorporating Partial Routing Information	77
4.1	Introduction	77
4.2	Likelihood-based inference	80
4.3	Analysis of likelihoods	82
4.3.1	Simplified link count model	84
4.3.2	Normal link count model	86
4.3.3	Simplified updated model	87
4.3.4	Normal updated model	91
4.4	Inference for p	93
4.5	Simulation results	94
4.5.1	Experiments with Known Constant Tracking Probabilities	95
4.5.2	Experiments with Known but Varying Tracking Probabilities	97
4.5.3	Experiments with Misspecified Tracking Probabilities	99
4.5.4	Application to real network	100
4.6	Conclusion	103

5	A New Class of Day-to-Day Traffic Assignment Model	107
5.1	Introduction	107
5.2	Case studies	109
5.3	Day-to-day Markov traffic assignment models	112
5.4	An Extended Class of Models	115
5.5	Empirical study of Modelling Non-stationarity	117
5.5.1	Temporal variability	117
5.5.2	Temporal-spatial correlation structure	121
5.6	Conclusion	124
6	Conclusions	127
6.1	Inference using Bayesian methods	127
6.2	Inference using frequentist methods	128
6.3	Statistical modelling	129
6.4	Future avenues for research	130
Notation		131
Bibliography		143

Contents

List of Figures

1.1	Illustrative network, with three nodes and four links.	2
1.2	Network with property of unique routing: a linear network.	3
2.1	Simple independence chain. True parameter value indicated by dotted line and green cross represents initial candidate value.	27
2.2	Random walk chain. True parameter value indicated by dotted line and green cross represents initial candidate value.	28
2.3	Transit network with four stops.	32
2.4	Generated route flows using Tebaldi and West algorithm.	34
2.5	Lack of movement of Tebaldi and West sampler in case of very low demand.	34
2.6	General linear network.	35
2.7	Example of a tree network.	40
2.8	Simple tree network.	41
2.9	Illustration of the steps in the construction of a candidate route flow \mathbf{y}	42
2.10	Hypergeometric probabilities of associated candidate route flow vectors. Green dashed line represents probability of generated $\mathbf{y}^\dagger = (39, 7, 20, 14)$	43
2.11	Overview of the eight possible scenarios for each network, e.g. A.	44
2.12	Histogram of estimated traffic flows on each of the four routes for high demand over 100 days. The dashed line represents the $\text{Gam}(0.1, 0.1)$ prior.	46
2.13	Contour and perspective plots of the bivariate histogram of λ_{13} and λ_{14}	47
2.14	Traceplot for λ_{14} in network A, where we assume high demand and an observational period of 100 days.	48
2.15	Network B with eight nodes connected by seven links.	49
2.16	Section of State Highway 16, Auckland, New Zealand.	51
3.1	Artificial linear network with four nodes.	65
3.2	Simulated data from over a period of 10 days.	66

List of Figures

3.3	Histograms of route flows in four-node network, for days 2 to 4.	67
3.4	Histograms of route flows in four-node network, for days 5 to 7.	68
3.5	Histograms of route flows in four-node network, for days 8 to 10.	69
3.6	Posterior and prior density of demand parameter β	70
3.7	Posterior and prior density of demand parameter γ	70
3.8	Posterior and prior density of demand parameter ψ for both methods in case of simple four-node network	72
3.9	Section of State Highway 16, Auckland	73
3.10	Posterior and prior density of demand parameter ψ for both methods in case of SH16 network.	74
4.1	Linear network with only three OD pairs.	83
4.2	Contour and perspective plots of the profile log-likelihood for simplified model using link count data only.	85
4.3	Contour and perspective plots of the profile log-likelihood for the normal model using link count data only.	87
4.4	Contour and perspective plots of the profile log-likelihood for the simplified model using link counts and available routing data.	88
4.5	Contour and perspective plots of the profile log-likelihood for the normal model using link counts and available routing data.	92
4.6	Test network.	95
4.7	Part of the road network in the English city of Leicester.	101
4.8	Maximum likelihood estimates (dots) and associated 95% confidence intervals for mean route flows in the Leicester network for both models. Red crosses mark the true parameter values.	102
5.1	Locations of road sections in New York. Reproduced courtesy of 2012 Google Map data.	110
5.2	Time series and associated autocorrelation plot for traffic counts north-bound on Broadway (route 9) in the Bronx, NY.	110
5.3	Time series and associated autocorrelation plot for traffic counts north-bound on Loudon Road (route 9) in Albany, NY.	111
5.4	Time series and associated autocorrelation plot for traffic counts north-bound on Major Deegan Expressway (route 87) in the Bronx, NY.	112
5.5	Simulated time series and associated autocorrelation plot for traffic counts on two-node network with memory length $d=2$, a low learning weight $\delta = 0.01$ and fixed sensitivity level $\psi = 1$	118

5.6	Simulated time series and associated autocorrelation plot for traffic counts on two-node network with memory length $d=2$, a low learning weight $\delta = 0.01$ and random sensitivity levels where $\mathbf{E}(\psi) = 1$	119
5.7	Simulated time series and associated autocorrelation plot for traffic counts on two-node network with memory length $d=2$, a high learning weight $\delta = 0.8$ and fixed sensitivity level $\psi = 1$	120
5.8	Simulated time series and associated autocorrelation plot for traffic counts on two-node network with memory length $d=2$, a high learning weight $\delta = 0.8$ and random sensitivity levels where $\mathbf{E}(\psi) = 1$	120
5.9	Larger sample network.	121
5.10	Simulated Time series for traffic counts on Leicester network.	123
5.11	Spatial correlations in Leicester network for both fixed and random ψ	124

List of Figures

List of Tables

2.1	Network topology of four-node tree network seen in Figure 2.8.	41
2.2	Table of binary choices possible for observed link counts $\mathbf{x} = (46, 59, 21)$	42
2.3	Mean values and associated Monte Carlo standard errors (in brackets) of MCMC estimation errors ϵ_{MCMC_A}	48
2.4	Mean values and associated Monte Carlo standard errors (in brackets) of GLS estimation errors ϵ_{GLS_A}	49
2.5	Mean values and associated Monte Carlo standard errors (in brackets) of MCMC estimation errors ϵ_{MCMC_B}	50
2.6	Mean values and associated Monte Carlo standard errors (in brackets) of GLS estimation errors ϵ_{GLS_B}	50
2.7	Mean values and associated Monte Carlo standard errors (in brackets) of MCMC estimation errors ϵ_{MCMC_C}	51
2.8	Mean values and associated Monte Carlo standard errors (in brackets) of GLS estimation errors ϵ_{GLS_C}	52
4.1	Scaled root mean squared errors of maximum likelihood estimators using NOR (normal model), SIM (simplified model), and GPS (tracked vehicle data only). The results are computed for varying levels of demand and global tracking probability.	96
4.2	Scaled root mean squared errors of maximum likelihood estimators using NOR (normal model), SIM (simplified model), and GPS (tracked vehicle data only). The results are computed for varying levels of demand and marginal tracking probability p_0 . The actual tracking probabilities are $\frac{1}{2}p_0$ for vehicles originating at node 1, and $\frac{3}{2}p_0$ for those originating at node 2.	98

List of Tables

4.3	Scaled root mean squared errors of maximum likelihood estimators using NOR (normal model), SIM (simplified model), and GPS (tracked vehicle data only). The results are computed for varying combinations of level of demand, mean tracking probability, and level of variability in tracking probabilities between routes.	100
-----	---	-----