Data, modelling and inference in road traffic networks

Richard Gibbens

Computer Laboratory
University of Cambridge

Traffic Modelling Conference
The Open University
31 March 2009

(Joint work with Yunus Saatci)
Preamble

With thanks to ...

- Department for Transport
- Highways Agency
- EPSRC: TIME project
  http://www.cl.cam.ac.uk/research/time/
- PeMS
  Freeway performance measurement system
  http://pems.eecs.berkeley.edu/

“The intent of this project is to collect historical and real-time freeway data from freeways in the State of California in order to compute freeway performance measures.”
MIDAS
Motorway incident detection and automatic signalling

Designed for real-time closed loop control of speed limits. Presently, covers about 30% of the Highways Agency’s strategic road network. Earliest data recorded in 1995.
M25 motorway
M25 speeds

Speeds (mph) on M25 (clockwise) Mon 6 Jan 2003
Speed/flow relationship and flow breakdown

Flow (vph) vs. Speed (mph) graph showing:
- Free-flow regime
- Congested regime

Time intervals:
- 5am
- 6am
- 7am
- 8am
- 9am
- 10am
- 11am

Legend:
- 5–6
- 6–7
- 7–8
- 8–9
- 9–10
- 10–11
M25 speeds

M25 clockwise: Av. speed (mph) 6 Jan 2003
Journey time prediction

M25 clockwise: Av. speed (mph) 6 Jan 2003
Journey time prediction

M25 clockwise: Av. speed (mph) 6 Jan 2003
Real-time measurements

M25 clockwise: Av. speed (mph) 6 Jan 2003
Linear regression model with varying lags, $\delta$

Example: Mondays only

- Frozen field predictor, $T^*$, (min)
- Exact journey time, $T$, (min)

![Graph showing linear regression and historical mean](image-url)
Some notation and definitions: $T_d(t)$, $T_d^*(t)$, $\bar{T}(t)$

Let $T_d(t)$ be the journey time starting at time $t$ on day $d \in D$
Let the speeds measured at loops/sites $\ell \in \{1, \ldots, L\}$ be $S_d(\ell, t)$ and let the distance between consecutive loops be $r$.
The frozen field travel time, $T_d^*(t)$, is given by

$$T_d^*(t) = \sum_{\ell=1}^{L-1} \frac{2r}{S_d(\ell, t) + S_d(\ell + 1, t)}.$$

The historical mean travel time, $\bar{T}(t)$, for a journey starting at time of day $t$ is given by

$$\bar{T}(t) = \frac{1}{|D|} \sum_{d \in D} T(d, t).$$
Rice and van Zwet (2004) studied a varying coefficients regression model of the form

\[ T_d(t + \delta) = \alpha(t, \delta) + \beta(t, \delta) T_d^*(t) + \epsilon \]

where \( \epsilon \) is a zero mean random variable modelling the random fluctuations and measurement errors.

Model fitting with smoothed parameters

Smoothed parameters, \((\hat{\alpha}(t, \delta), \hat{\beta}(t, \delta))\), may be obtained through a weighted linear regression so as to minimize

\[
\sum_{d \in D, s} \left( T_d(s) - \alpha(t, \delta) - \beta(t, \delta) T_d^*(t) \right)^2 K(t + \delta - s)
\]

where \(K(\cdot)\) denotes a Gaussian density with mean zero and some specified variance \(\sigma^2\) and \(s\) is a general time of day value.

A further predictor

**k-Nearest Neighbour**

This predictor for $T_d(t + \delta)$ is given in terms of the $k$ closest (past) days $d_1, d_2, \ldots, d_k$ to $d$ in the sense of the distance metric (other metrics are also plausible)

$$m(d, d') = \sqrt{\sum_{t-w \leq s \leq t} \left[ T_d^*(s) - T_{d'}^*(s) \right]^2}.$$ 

The predictor for $T_d(t + \delta)$ is then

$$T_d^{kNN}(t + \delta) = \sum_{i=1}^{k} w_i T_{d_i}(t + \delta)$$

with weights $w_i$ inversely proportional to distances.

The parameter $k$ and the windowing parameter $w$ help tradeoff the accuracy with the computational overhead.
RMS prediction errors
Simple leave-one-out approach

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Mon</th>
<th>Tue/Wed/Thu</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k−Nearest neighbour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time of day
RMS prediction error (min)

<table>
<thead>
<tr>
<th>Lag (min)</th>
<th>05:00</th>
<th>10:00</th>
<th>15:00</th>
<th>05:00</th>
<th>10:00</th>
<th>15:00</th>
<th>05:00</th>
<th>10:00</th>
<th>15:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag:0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag:60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag:90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag:120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time of day

Linear regression model
Prediction intervals, given Normality assumptions (Mondays only for $t = 8\text{am}$ and $\delta = 60\text{mins}$)
Discussion

References:

