A graphical dynamic approach to forecasting road traffic flow networks

Catriona M. Queen  Casper J. Albers

Department of Mathematics and Statistics
The Open University
Milton Keynes U.K.

One-day Conference on Traffic Modelling
The Open University, March 2009
Introduction
Want to forecast hourly traffic flows

Model needs:
- multivariate,
- work in real time,
- accommodate changes, and
- preferably interpretable.

Focus on 2 UK networks.
London
M25 / A2 / A296 Junction

Pictures taken from Google Maps
Manchester
M60 / M62 / M602 Junctions

Pictures taken from Google Maps
London
M25 / A2 / A296 Junction

Diagram

Properties

- 17 data sites
- 21 weeks of data
- Hourly counts
Manchester
M60 / M62 / M602 Junction

Diagram

Properties
- 32 data sites
- Over a year’s worth of data
- Aggregate hourly counts (by minute)
The Model
Why multivariate?

Counts at upstream sites informative about downstream sites.

Suppose:
- Minute data
- Site 1 → Site 2 takes 1 min
  ⇒ use $Y_{t-1}(1)$ to forecast $Y_t(2)$.
- We have hourly counts ⇒ use $Y_t(1)$ to forecast $Y_t(2)$. 
Represent the network by a graph

- Represent network by a directed acyclic graph (DAG).

Example road with DAG

- DAG defines set of conditional distributions: child | parents
- → breaks multivariate problem into univariate conditionals:
  \[ Y_t(1) \quad Y_t(2) | Y_t(1) \quad Y_t(3) | Y_t(2). \]
Our model is called the **Linear Multiregression Dynamic Model (LMDM)**.

**LMDMs:**

- Represent time series by a DAG.
- Model breaks into separate (conditional) univariate models (DLMs).
- Each univariate model uses parents as regressors.
- Model computations are fast and relatively straightforward no matter how complex the network.
- Interventions are easy.
- Graphical structure $\Rightarrow$ easy to handle changes in network.
Example LMDM

DAG

Observation equations:

\[
\begin{align*}
Y_t(1) &= \theta_t(1) + v_t(1), & v_t(1) &\sim N(0, V_t(1)), \\
Y_t(2) &= y_t(1)\theta_t(2) + v_t(2), & v_t(2) &\sim N(0, V_t(2)), \\
Y_t(3) &= y_t(2)\theta_t(3) + v_t(3), & v_t(3) &\sim N(0, V_t(3)).
\end{align*}
\]

Parameters evolves through system equation:

\[
\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t).
\]

Model is Bayesian so prior specified for \(\theta_0\).
LMDM assumes linear relationship between parent and child.

Plot of parent (1431A) vs child (1437A)
Do linear relationships depend on day of week?

Plot of parent (1431A) vs child (1437A) — different colour for each day.
Do linear relationships depend on hour of day?

Plot of parent (1431A) vs child (1437A) — different colour for each hour
Linear relationships: conclusion

- Linear relationships between parent and child is realistic assumption...
- ... where the linear relationship depends on hour of day.
- ⇒ Use different regression parameter for each hour.
Creating a DAG for a traffic network
Fork and DAG

Constructing the DAG

- $Y_t(1) \sim \text{Poi}()$
- $Y_t(1) \sim \mu_t + \nu_t(1), \nu_t(1) \sim N$
- $Y_t(2) | Y_t(1) \sim \text{Bin}(Y_t(1), \alpha_t)$
  - with $\alpha_t$ the proportion $1 \rightarrow 2$
  - $Y_t(2) \sim \alpha_t Y_t(1) + \nu_t(2), \nu_t(2) \sim N$
- $Y_t(3) = Y_t(1) - Y_t(2)$
  - so $Y_t(3)$ is a logical variable
- Alternative model possible
  - Equivalent under assumption
    - $Y_t(1) = Y_t(2) + Y_t(3)$
DAG for a join

**Join and DAG**

1. TRAFFIC DIRECTION

2. 3

**Constructing the DAG**

- $Y_t(1) = \mu_t(1) + \nu_t(1), \, \nu_t(1) \sim N$
- $Y_t(2) = \mu_t(2) + \nu_t(3), \, \nu_t(2) \sim N$
- $Y_t(3) = Y_t(1) + Y_t(2)$

so $Y_t(3)$ is a logical variable.
All types of road lay-out can be modelled through a combination of forks and joins.

Example: a junction
All types of road lay-out can be modelled through a combination of forks and joins.

Example: a junction
London
M25 / A2 / A296 Junction

Because of 6 missing nodes, the network splits into 2 separate DAGs and 2 isolated nodes.
Manchester
M60 / M62 / M602 Junction
Analyses and results
Forecasts

Forecast for three days, M25 network

Y(163) forecast

Y(163) error

- Y(163) forecast
- observ. / forecast
- ± 2sd

510 520 530 540 550 560 570

510 520 530 540 550 560 570

C.M. Queen, C.J. Albers — The Open University
Forecasting road traffic flow networks
Intervention required

Forecast for three days at site 167

Y(167) forecast

Y(167) error

observ. / forecast ± 2sd
Intervention required

- Same problem seen in $Y_t(167)$’s descendants.
- But not seen in $Y_t(167)$’s non-descendants.

### Example

![Diagram of a DAG for traffic network](image)

$Y_t(170B)$ without intervention

![Graph showing traffic flow](image)
Same problem seen in $Y_t(167)$’s descendants.

But not seen in $Y_t(167)$’s non-descendants.
The intervention

- Intervene for $Y_t(167)$ only.
- Large negative forecast error at time 560.
  Large positive forecast error at time 561.
- Consistent with hold-up at time 560.
- Treat $Y_{560}(167)$ as outlier.
- Intervene for $Y_t(167)$ at time $t = 561$:
  Increase forecast mean for $Y_{561}(167)$ — add on forecast error at previous time.
  Increase forecast variance for $Y_{561}(167)$ — allows for more uncertainty.
The intervention for $Y_{561}(167)$:
- improves forecast for $Y_t(167)$ at time $t = 561$,
- also improves forecasts for $Y_t(167)$'s descendants at time $t = 561$,
- does not affect the forecasts of non-descendants.
The intervention for $Y_{561}(167)$:
- improves forecast for $Y_t(167)$ at time $t = 561$,
- also improves forecasts for $Y_t(167)$’s descendants at time $t = 561$,
- does not affect the forecasts of non-descendants.

Example

$Y_t(167)$

- $Y_t(167)$ without intervention
- $Y_t(167)$ with intervention

$Y_t(169)$

- $Y_t(169)$ without intervention
- $Y_t(169)$ with intervention

C.M. Queen, C.J. Albers — The Open University
Forecasting road traffic flow networks
Conclusion
Conclusion

LMDM’s are

- Computationally fast.
- Flexible to adapt and easy to intervene.
- Promising for forecasting traffic flows.
- Easily interpretable by non-statisticians.
Future work:
- Develop an on-line monitor.
- Methods of identifying and estimating missing data.
- Automated methods for eliciting the DAG.
- Allowing feedback due to queues.
- ….

Acknowledgements:
- Kent County Council, provided London data
- The Highways Agency, provided Manchester data
References

Queen, C.M. and Albers, C.J.
Intervention and causality in a dynamic Bayesian network.

Queen, C.M. and Albers, C.J.

Queen, C.M., Wright, B.J. and Albers, C.J.
Forecast covariances in the linear multiregression dynamic model.

Queen, C.M., Wright, B.J. and Albers, C.J.
Eliciting a Directed Acyclic Graph for a Multivariate Time Series of Vehicle Counts in a Traffic Network.