

# Learning Traffic Flow Dependencies in Highway Networks

Samitha Samaranayake, Sébastien Blandin  
Systems Engineering, UC Berkeley



# Motivation

---

- Modeling the evolution of traffic state
- Given observations, estimating traffic state
- Given the current state, forecasting the traffic state
- Routing vehicles on the highway network

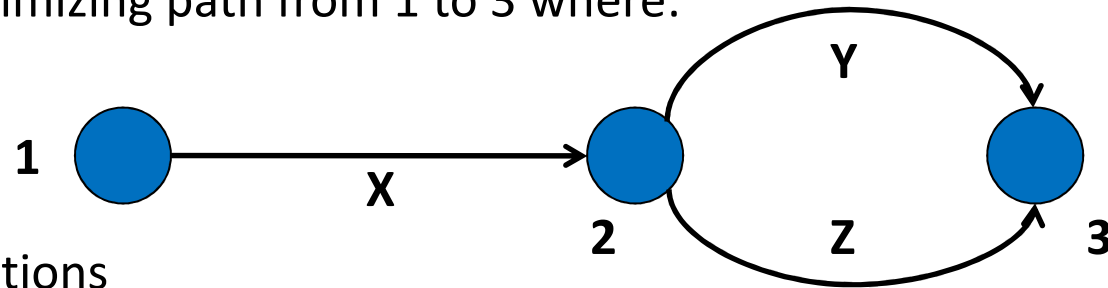


This is a complex system



# Routing example

- Finding the variance minimizing path from 1 to 3 where:  
 $X \perp Y$  and  $X \not\perp Z$



- A priori case: no observations

- The variance of path X-Y is:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

- The variance of path X-Z is:

$$\text{var}(X + Z) = \text{var}(X) + \text{var}(Z) + 2 \text{cov}(X, Z)$$

- Adaptive case: travel time on X is known

- The expected variance on Y is:

$$E(\text{var}(Y|X)) = \text{var}(Y)$$

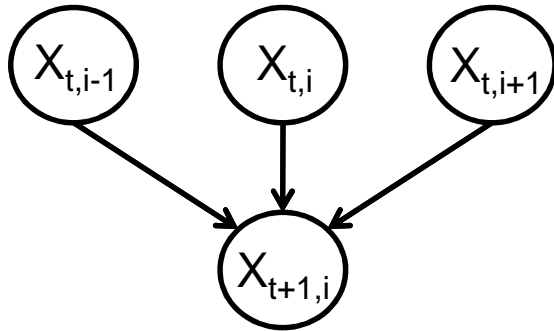
- The expected variance on Z is:

$$E(\text{var}(Z|X)) = \text{var}(Z) - \text{var}(E(Z|X))$$



## Modeling dependencies of traffic flow

- The traffic state at a given location is determined by the previous traffic states at nearby locations .



Assume that the state in cell  $X_{t+1,i}$  only depends on the state of cells  $X_{t,i-1}$ ,  $X_{t,i}$  and  $X_{t,i+1}$ .

- The state at each location is generally not known with certainty, so we will treat them as random variables

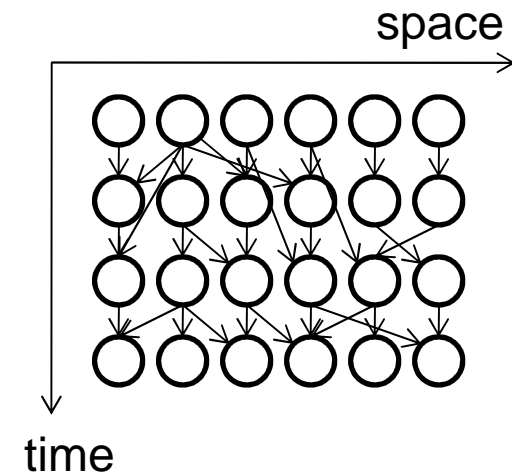
$P(X_{t+1,i} = x_{t+1,i})$  is a function of the values  $X_{t,i-1}$ ,  $X_{t,i}$  and  $X_{t,i+1}$

- These random variables are defined by a joint probability density function



# Structure learning problem

- Objective : Given a dataset find the graphical model structure which best describes the generative distribution
- Method: Find the maximum likelihood parameters and posterior probability for each structure and search over the structure space to find the optimal structure



- The posterior contains an integral that is hard to compute, so we approximate it using the following scoring function:

- *Bayesian Information Criterion* (BIC)

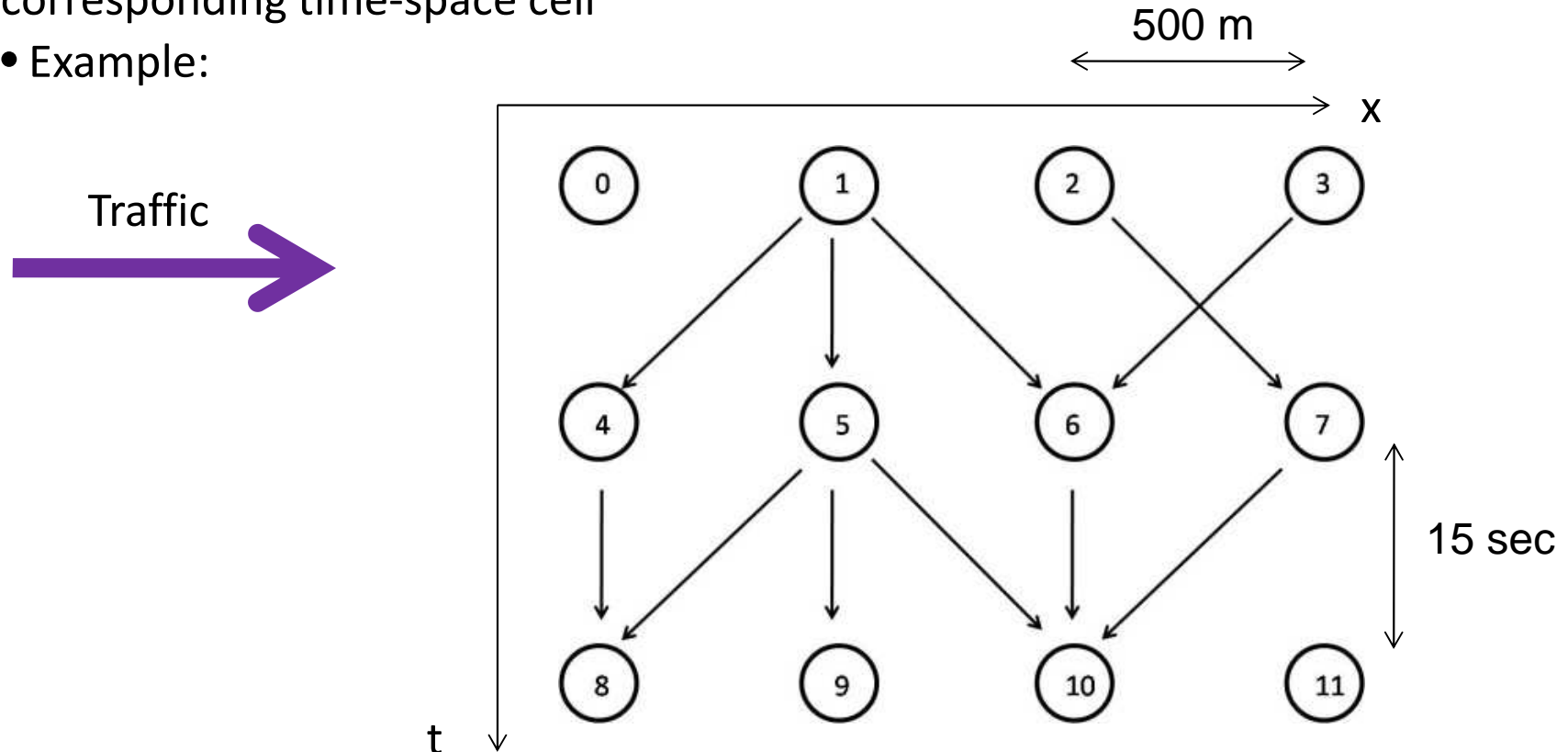
$$\text{BIC}(D, \mathcal{G}, \Theta) = \log \left( P(D|\hat{\Theta}, \mathcal{G}) \right) - \frac{d}{2} \log m + O(1)$$

- Where  $D, \mathcal{G}, \Theta$  are respectively the data, graph structure and parameters,  $d$  is the number of edges in the graph and  $m$  is the number of data points per node.



# Experiment Setting

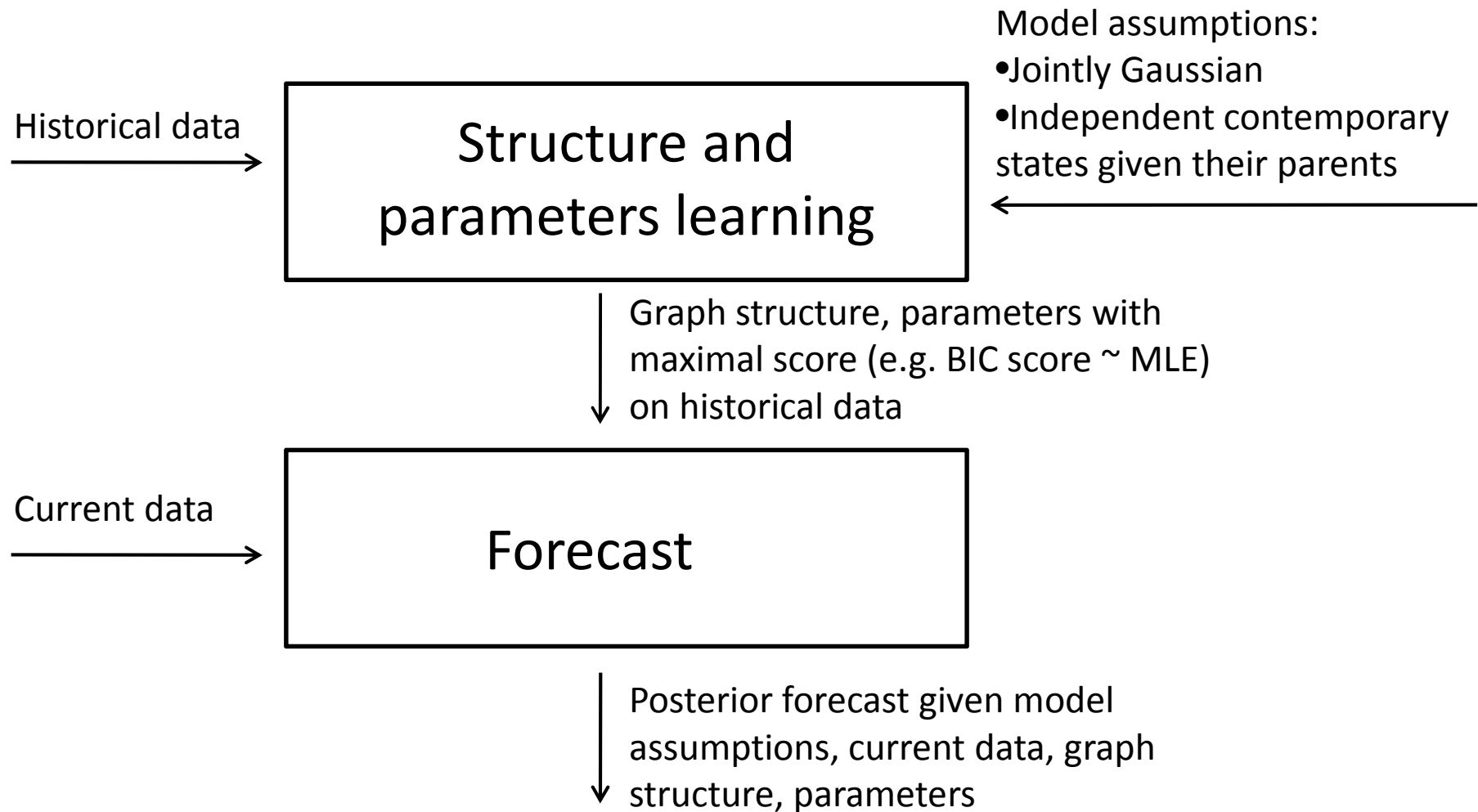
- Model the highway network with a 2 dimensional graphical model
- Each node is a random variable representing mean traffic speed over the corresponding time-space cell
- Example:



- Assumptions:
  - Distribution is jointly Gaussian
  - Dependencies are strictly forward in time

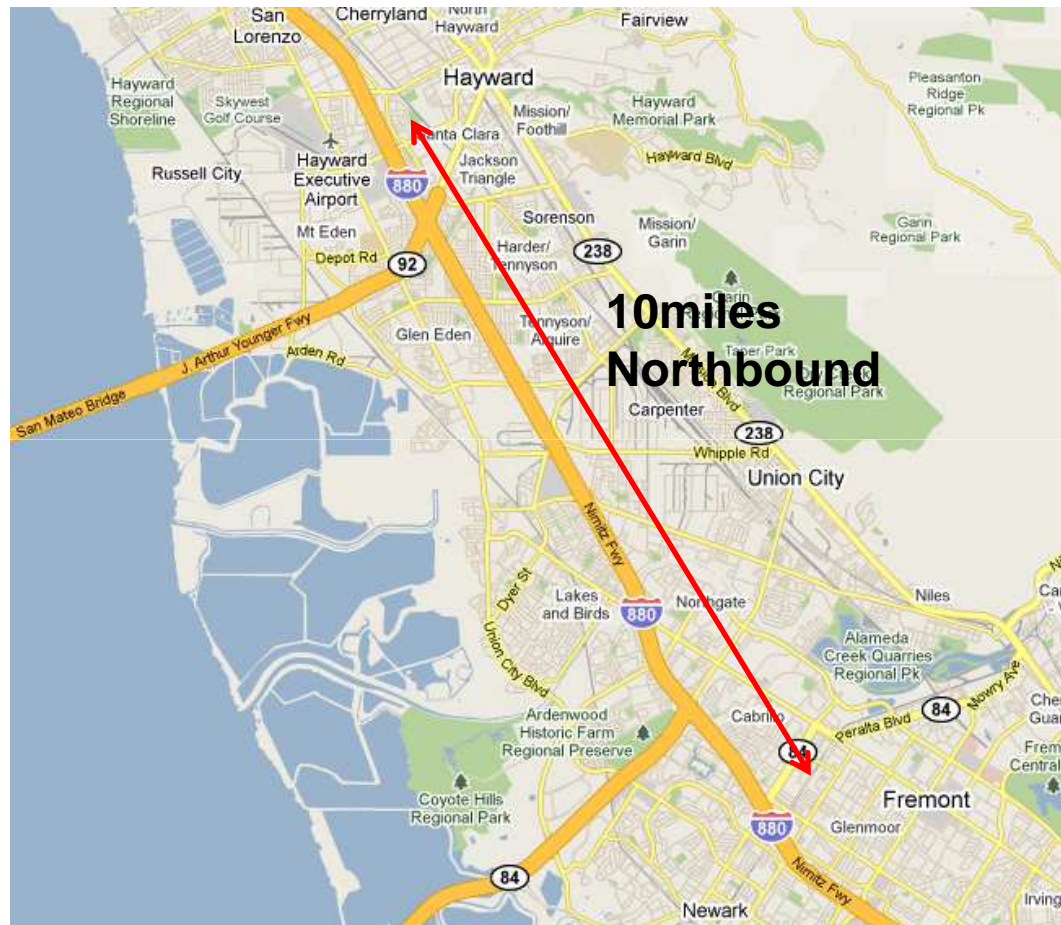


# Method Flow Chart for forecast





# Forecast experiment



- Traffic estimates from the Mobile Millennium system

- Real time estimates every 30 sec for 400 m discretization cells over the whole Bay Area highway network

- The structure is learnt on one day of data (Feb 1<sup>st</sup>, 2010)

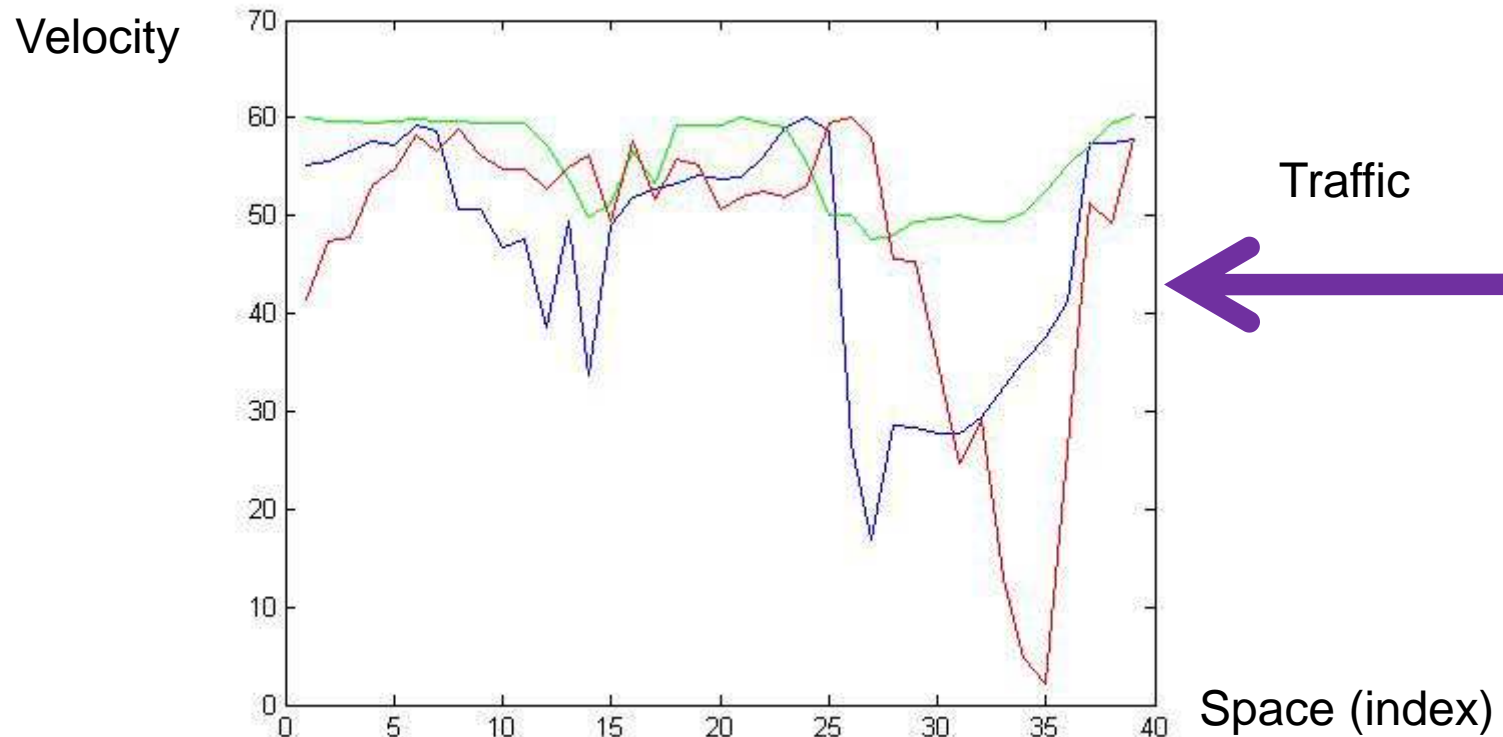
- The forecast is made for Feb 2<sup>nd</sup>, 2010





## Experimental results

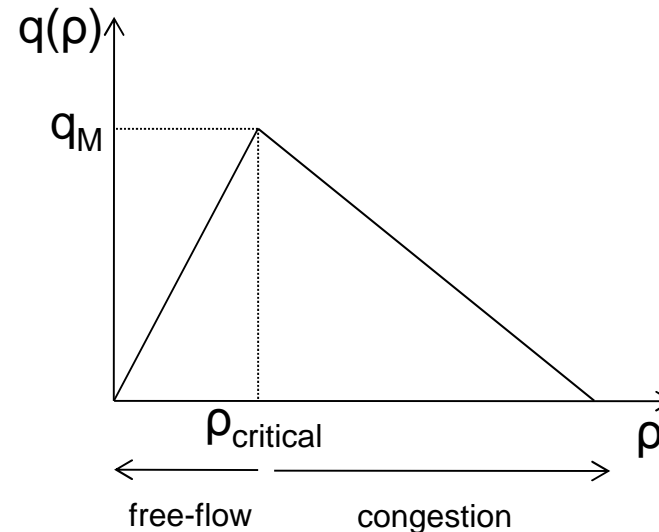
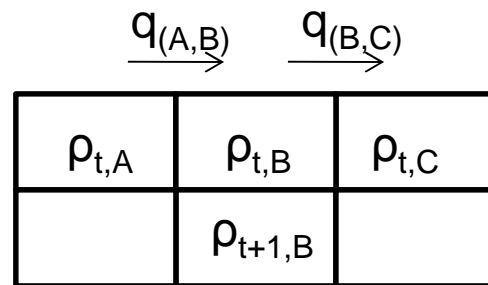
- 10 minutes forecast at the beginning of afternoon rush hour
- Structure selection allowed to consider 25 neighboring nodes from the 5 previous time steps, time discretization is 10 minutes
- True value, forecast with training on previous day, forecast with training on morning rush hour





# Limitations of the model

- We assumed that the distributions were jointly Gaussian, which restricts the model to learning linear relationships between the traffic states, traffic flow dependencies are not linear!
- Consider the discretization of the LWR PDE using the Godunov scheme



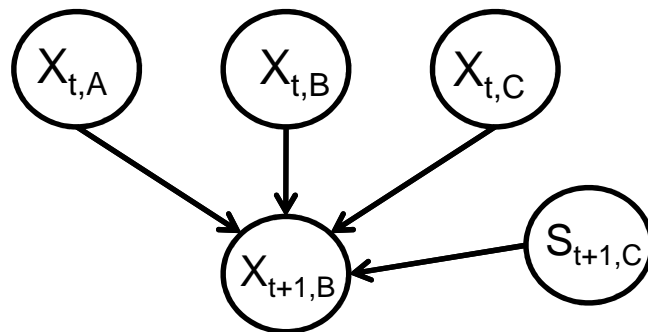
$$\rho_{t+1,B} = \rho_{t,B} + T/x (q_{t,B} - q_{t,A})$$

$$q(A, B) = \begin{cases} q(B) & \text{if } \rho_A \geq \rho_B \geq \rho_{critical} \\ q(\rho_{critical}) & \text{if } \rho_A \geq \rho_{critical} \geq \rho_B \\ q(A) & \text{if } \rho_{critical} \geq \rho_A \geq \rho_B \\ \min(q(A), q(B)) & \text{if } \rho_B \geq \rho_A \end{cases}$$



## Improving the model

- The state of node B at time  $t+1$  has a non-linear relationship with the state of nodes A, B and C at time  $t$ , but the relationship is linear if we know whether A, B and C are in free-flow or congestion at time  $t$
- Assuming that there is at most one wave propagating in the network, there are eight possible linear equations to consider for  $\rho_{t+1,B}$
- Now we reformulate the maximum likelihood function for each cell as follows:



where for each data point we now pick from eight different joint Gaussians based on the selection variable  $S$



## Closing remarks

---

- We are currently working on parameter learning for just the complete free flow and complete congestion cases. This will be extended to cover all eight modes.
- The state space explodes as we increase the number of dependent cells. Therefore, a more general formulation is required when extending this to structure learning.
- The mode is currently determined by looking at the states of the parent cells. However, in certain cases we might not know what the actual state is. Such cases will require a mixture model which makes the inference problem harder.