Belief Propagation for Traffic forecasting

Cyril Furtlehner (INRIA Saclay - Tao team)

context : Travesti project http ://travesti.gforge.inria.fr/)

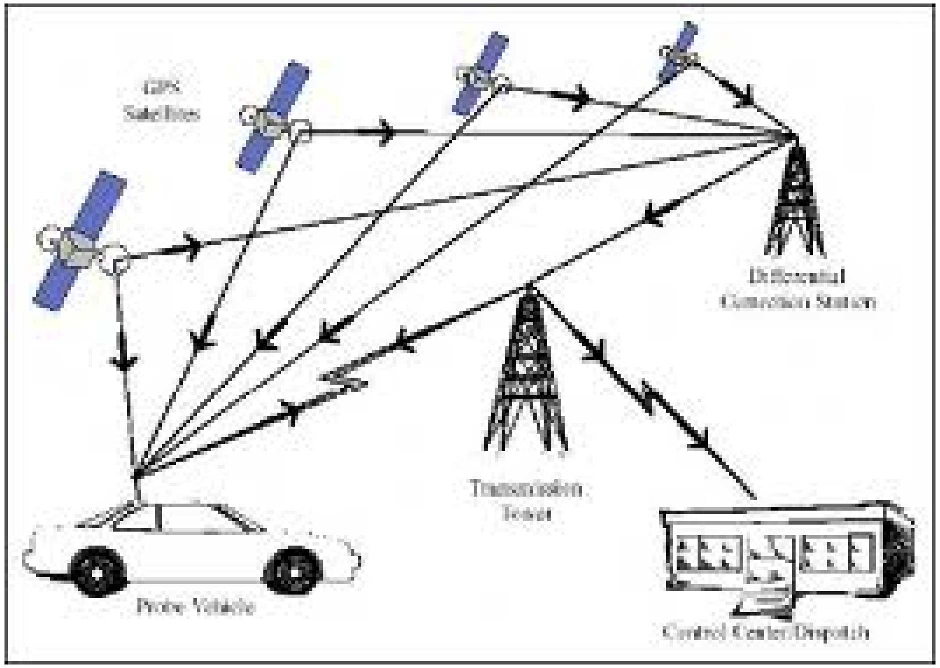
- Anne Auger (INRIA Saclay)
- Dimo Brockhoff (INRIA Lille)
- Yufei Han (CAOR Armines and INRIA Rocquencourt)
- Jean-Marc Lasgouttes (INRIA Rocquencourt)
- Fabrice Marchal (formerly at CNRS LET)
- Victorin Martin (INRIA)
- Fabien Moutarde (CAOR Armines)
- Maxim Samsonov (formerly at INRIA Saclay)

Problem at hand

Goal : reconstruct and predict road traffic on secondary network

Existing solutions based on data coming from static sensors (magnetic loops) on main roads. Too expensive to be installed in all streets.

Solution : more and more vehicles are equipped with GPS and able to exchanging data (through cellular phone connections) : Floating Car Data (FCD).



$\Longrightarrow$ we want to build an inference schema adapted to these FCD

Project guidelines

(Traffic Volume Estimation by Spatio-Temporal Inference)
(http ://travesti.gforge.inria.fr/)

Goal 1 : is to find a macroscopic description of a large scale traffic network
$\longrightarrow$ machine learning approach, clustering & dimensional reduction

At microscopic level we consider travel time along each segment (FCD observations)

Hypothesis : large scale behaviour is dominated by various dynamical congestion patterns.

Multi-modal joint measure of travel time mixing macroscopic & microscopic variables

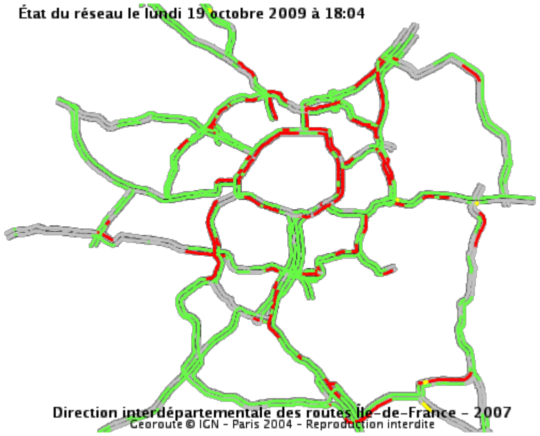
Goal 2 : Setting of a MRF to encode dependencies between degrees of freedom. (dynamics at the macroscopic level)

Real time inference : search for a model compatible with the belief propagation algorithm

Related project : FUI Pumas (http ://pumas.inria.fr/), concrete implementation in the Rouen agglomeration (1000 probe vehicles)

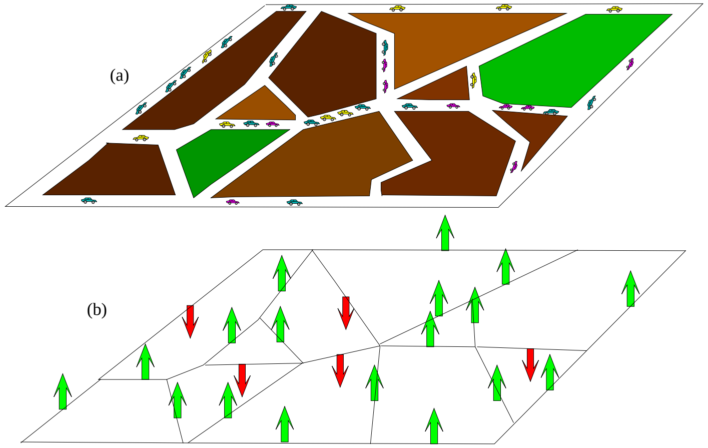
An Ising model for traffic ?

Is it possible to encode traffic data on the basis of a binary latent state $s_{i,t} \in \{-1, 1\}$ (Ising) corresponding to congested/non-congested.



Input data $x_{i,t} \in \mathbb{R}$:

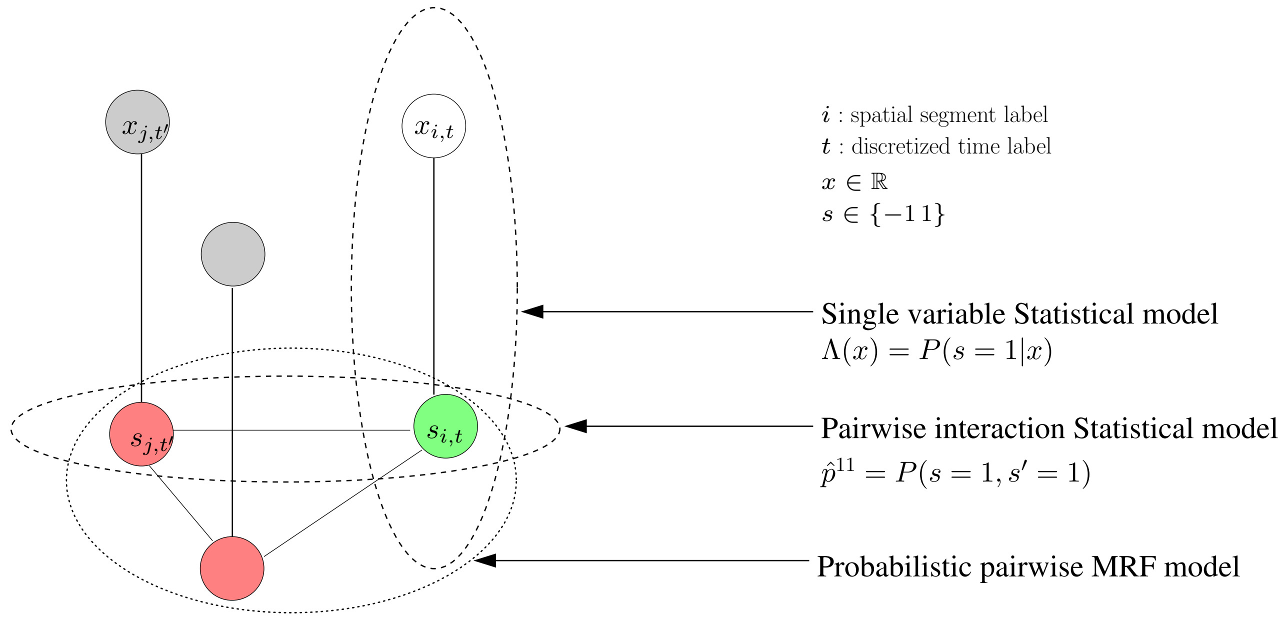
- static sensor : density and speed.
- floating car data : speed and travel times.



2 questions :

- Assuming an historical dataset of $\{\hat{x}_{i,t}\}$: which probabilistic model $P(\{x_{i,t} \in \Omega\})$?
- Given actual observations $\{x^*_{i,t}, (i,t) \in \Omega^*\}$: how to infer $\{x_{i,t}, (i,t) \in \Omega \backslash \Omega^*\}$?

Ising based proposed solution



4 components of the model :

- single variable statistical model translating real-valued observations into binary latent states
- pairwise statistical model of the dependency between latent states
- MRF model to encode the network of dependencies
- Belief propagation algorithm to decode a partially observed network $\{x^*_{i,t}\}$

Building and testing the model

Input : a learning set $\{x_i^k, i \in \{1 \ldots N\}, k = 1 \ldots M\}$.

Step 1 : define a mapping $\Lambda(x) = P(s = 1|x)$ for each link. This fixes the set of $\hat{p}_i(s_i)$

Step 2 : build (EM) set of pairwise marginals $\hat{p}_{ij}(s_i, s_j)$ based on the model hypothesis :

$$P(x_i, x_j, s_i, s_j) = \hat{p}_{ij}(s_i, s_j)P(x_i|s_i)P(x_j|s_j).$$

Step 3 : from the set $\{\hat{p}_i, \hat{p}_{ij}\}$ find a MRF
$\longrightarrow$ inverse Ising model

Experimental Test : from a test set $\{y_i^k, i = 1 \ldots N, k = 1 \ldots Q\}$, for each k, reveal one by one the variables at random and infer the other hidden one with BP.
$\longrightarrow$ plot the prediction error as a function ρ of the fraction of observed variables.

Inverse Ising Problem

Inverse Problem : from observations to an underlying Markov Random field.

Given a set of M joint observations $\{\hat{s}_i^k, i \in \{1 \ldots N\}, k = 1 \ldots M\}$, look for a model

$$P_{Ising}(\mathbf{s}) = \frac{1}{Z(\mathbf{h},\mathbf{J})} \exp\Big(\sum_i h_i s_i + \sum_{i,j} J_{ij} s_i s_j\Big)$$

Find the set $\mathbf{h}, \mathbf{J}$ of parameters s.t. the Log Likelihood is maximal

$$\mathcal{L}(\mathbf{h},\mathbf{J}) = D_{KL}(\hat{P}|P_{Ising}) = \log(Z(\mathbf{h},\mathbf{J})) - \sum_i h_i \hat{\mathbb{E}}(s_i) - \sum_{ij} J_{ij} \hat{\mathbb{E}}(s_i s_j).$$

$\mathcal{L}(\mathbf{h},\mathbf{J})$ is concave and optimal when the moment matching constraints are satisfied :

$$\frac{\partial \log Z}{\partial h_i}[\mathbf{h},\mathbf{J}] = \hat{\mathbb{E}}(s_i) \qquad\qquad \frac{\partial \log Z}{\partial J_{ij}}[\mathbf{h},\mathbf{J}] = \hat{\mathbb{E}}(s_i s_j).$$

Problem : $Z(\mathbf{h},\mathbf{J})$ is difficult to evaluate in general (exponential cost).

Inverse Ising Problem

Generic question encountered in biology, image processing...

Statistical physics based methods :

- Boltzmann machines (Hinton, Sejnowski '83)
- Linear response theory, Plefka expansion (Welling Teh '03)
- Susceptibility-Propagation (Mézard-Mora '07)
- Advanced linear response theory (Yasuda,Tanaka '09)
- Adaptative cluster expansion for neural coupling determination (Cocco et. al '09)

Sparse optimization methods in Machine learning :

- L_1 optimization with belief propagation (Lee,Ganapathi,Koller 2006)
- L_1 regularization for graph selection (Wainwright,Ravikumar,Lafferty 2008)
- L_1 penalized pseudo-likelihood (Höfling,Tibshirani 2009)

Mean-Field methods

Input are the magnetization $\hat{m}_i = \hat{\mathbb{E}}(s_i)$ and susceptibilities $\hat{\chi}_{ij} = Cov(s_i, s_j)$.

Small interaction expansion :

- Naive Mean Field :

$$J_{ij}^{MF} = \frac{\hat{\chi}_{ij}}{(1-\hat{m}_i^2)(1-\hat{m}_j^2)} \approx [\hat{\chi}^{-1}]_{ij}.$$

- Thouless-Anderson-Palmer (TAP, '77) J_{ij}^{TAP}

$$J_{ij}^{TAP} = -\frac{2[\hat{\chi}^{-1}]_{ij}}{1+\sqrt{1-8\hat{m}_i\hat{m}_j[\hat{\chi}^{-1}]_{ij}}},$$

Tree-like interactions graph : Bethe approximation

$$P_{Bethe}(\mathbf{s}) = \prod_{(i,j)\in\mathcal{T}} \frac{\hat{p}_{ij}(s_i, s_j)}{\hat{p}_i(s_i)\hat{p}_j(s_j)} \prod_i \hat{p}_i(s_i).$$

with

$$\hat{p}_{ij}(s_i, s_j) = \frac{1}{4}\big(1 + \hat{m}_i s_i + \hat{m}_j s_j + (\hat{\chi}_{ij} + \hat{m}_i\hat{m}_j)s_i s_j\big)$$

Mean-Field methods

Direct identification leads to J_{ij}^{B1}

$$J_{ij}^{B1} = \frac{1}{4}\log\Big(\frac{\hat{p}_{ij}(1,1)\;\hat{p}_{ij}(-1,-1}{\hat{p}_{ij}(-1,1)\;\hat{p}_{ij}(1,-1)}\Big).$$

Remark : $Z_{Bethe}(\mathbf{h},\mathbf{J})$ explicit function of $\hat{p}_i$ and $\hat{p}_{ij}$.

Inverse susceptibility (Nguyen, Berg 2012)

$$[\hat{\chi}^{-1}]_{ij} = \Big[\frac{1-d_i}{1-m_i^2} + \sum_{k\in\partial i}\frac{1-m_k^2}{(1-m_i^2)(1-m_k^2)-\hat{\chi}_{ik}^2}\Big]\delta_{ij} - \frac{\hat{\chi}_{ij}}{(1-m_i^2)(1-m_j^2)-\hat{\chi}_{ij}^2}\,\delta_{j\in\partial i}.$$

leads to J_{ij}^{B2} (equivalent to susceptibility propagation)

$$J_{ij}^{B2} = -\frac{1}{2}\,\mathrm{atanh}\Big(\frac{2[\hat{\chi}^{-1}]_{ij}}{\sqrt{1+4(1-\hat{m}_i^2)(1-\hat{m}_j^2)[\hat{\chi}^{-1}]_{ij}^2} - 2\hat{m}_i\hat{m}_j[\hat{\chi}^{-1}]_{ij}}\Big).$$

Two-parameters Bethe Model

Additional constraint : compatibility with belief propagation

In practice we consider 2 calibration parameters :

- a global rescaling factor α of the interactions ("inverse temperature") :
  $J_{ij} \longrightarrow \alpha J_{ij}$ for all (i,j)
- a mean connectivity K of the graph
  $\longrightarrow$ various pruning procedures

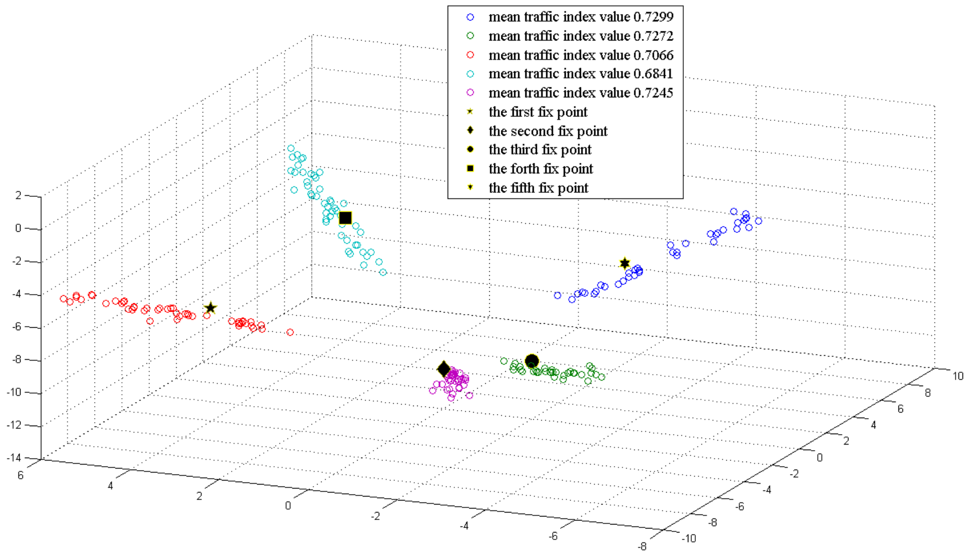
Bethe model 1

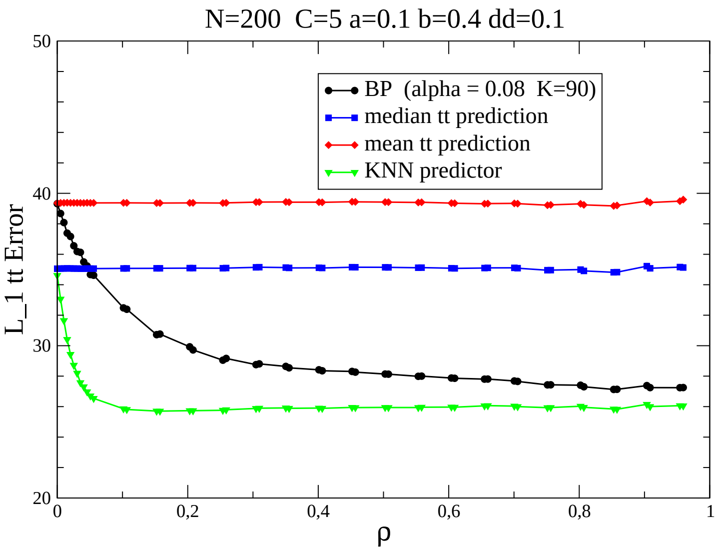
$$P^{\alpha}_{Bethe}(\mathbf{s}) = \prod_{(i,j)\in\mathcal{G}} \Bigl(\frac{\hat{p}_{ij}(s_i,s_j)}{\hat{p}_i(s_i)\hat{p}_j(s_j)}\Bigr)^{\alpha} \prod_i \hat{p}_i(s_i).$$

Various pruning procedure : e.g. Max spanning tree + thresholding on other links.

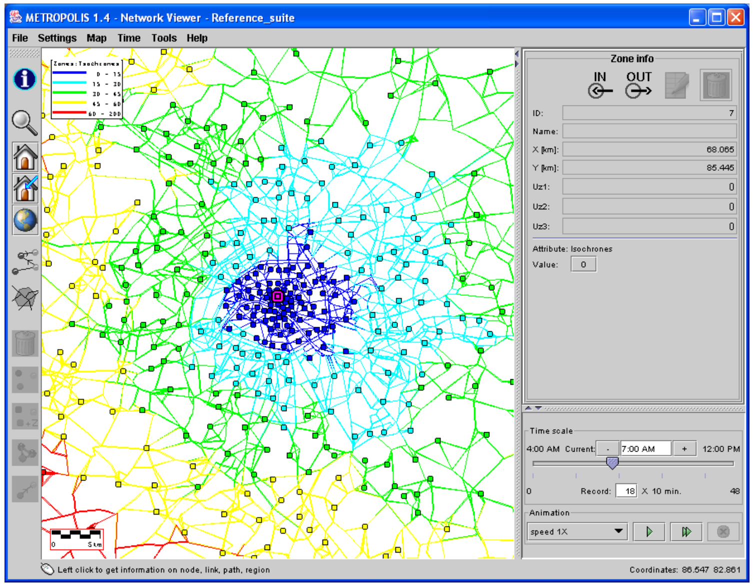
BP fixed points corresponds to congestion patterns

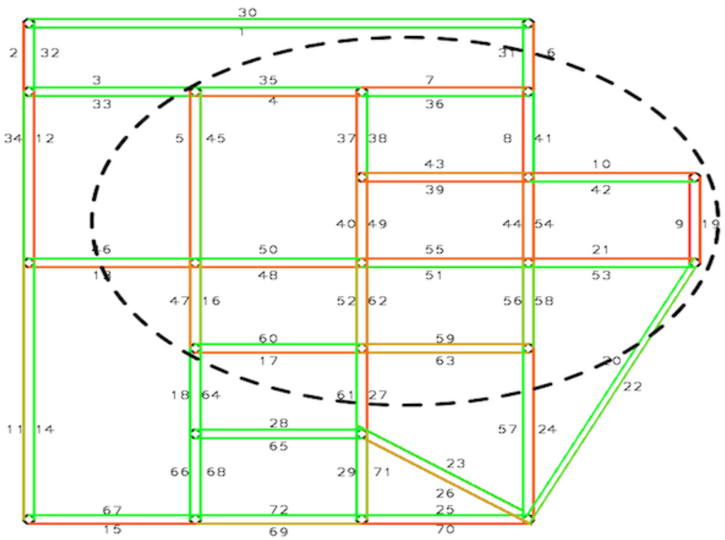
Real-valued synthetic multimodal data : combining the approximate MRF with the mapping Λ.





Simulation data : Sioux Falls Network

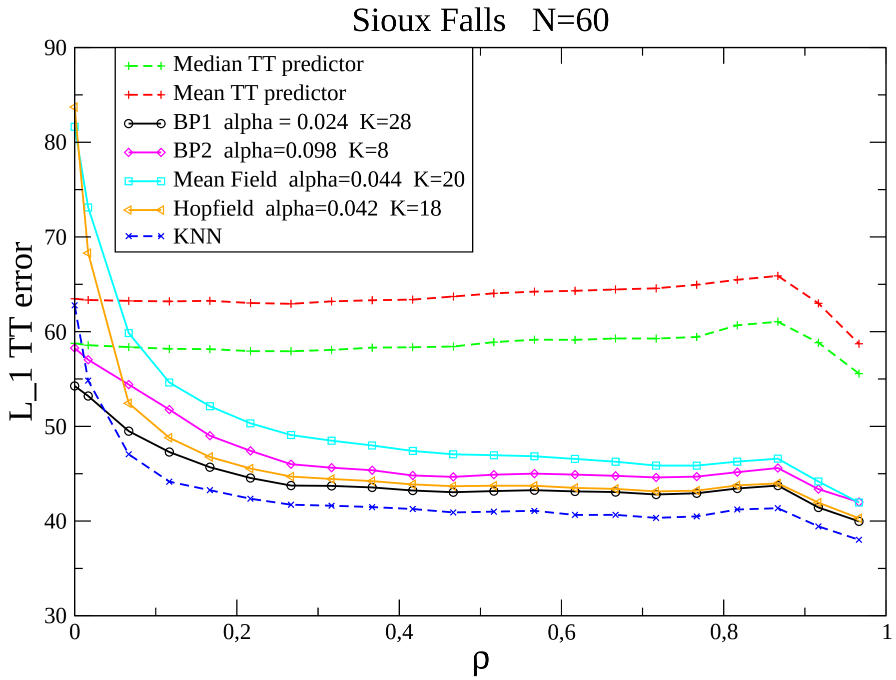


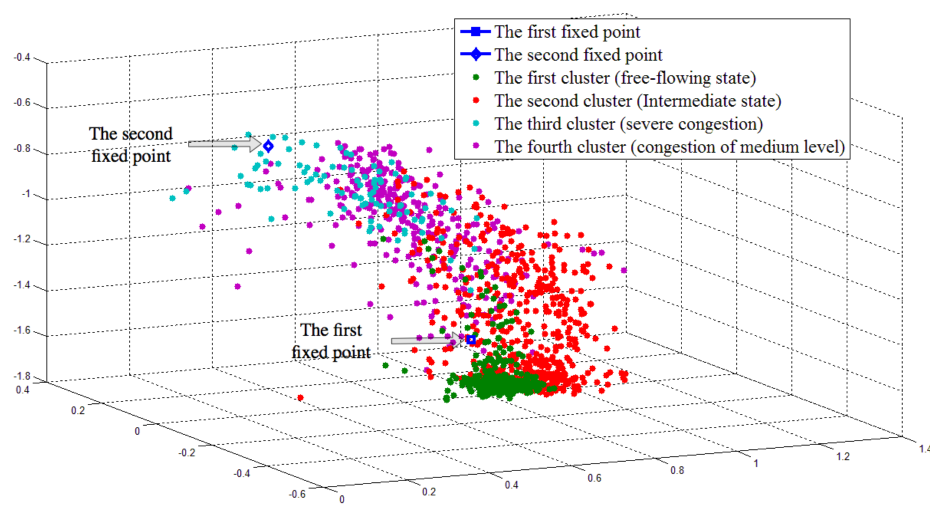


Data generated by the “METROPOLIS” traffic simulator written by Fabrice Marchal (formerly at LET, CNRS-Lyon) and André de Palma (ENS Cachan)

Experiments I

Variable are revealed one by one in a random order.

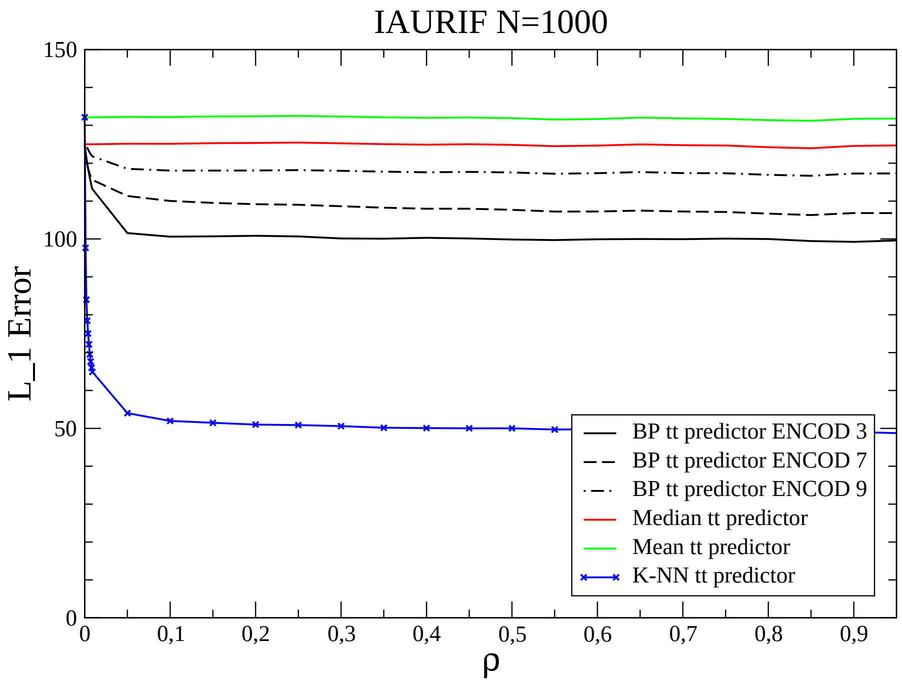


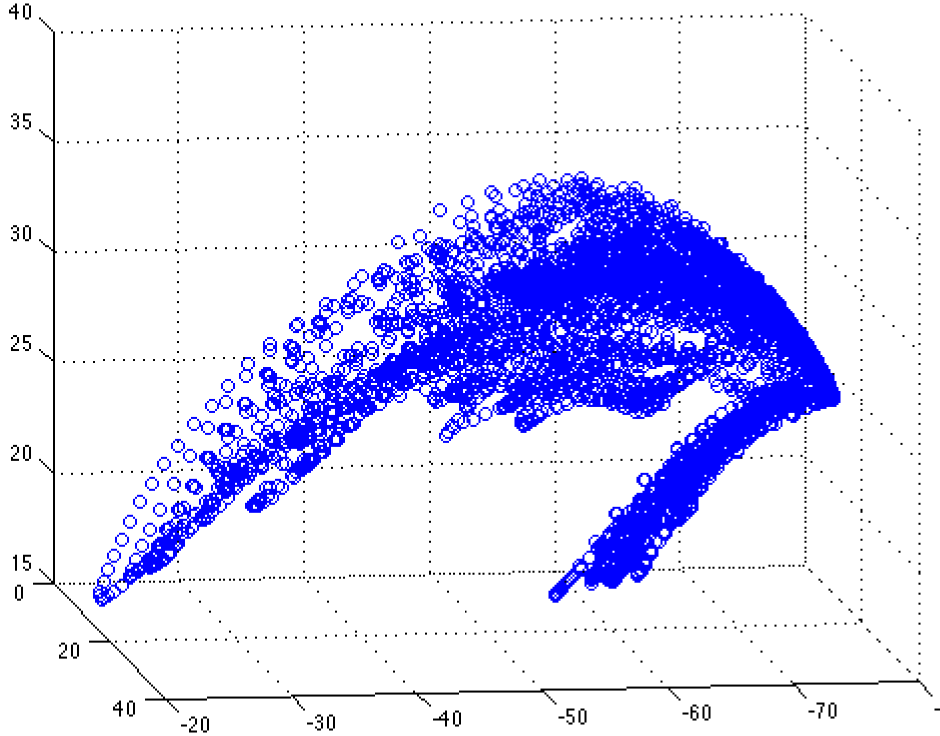


(ρ = fraction of revealed variables)

Experiments II

Variable are revealed one by one in a random order.





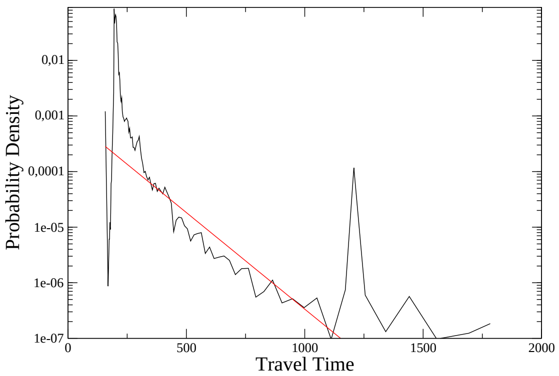
(ρ = fraction of revealed variables)

Traffic index for real data

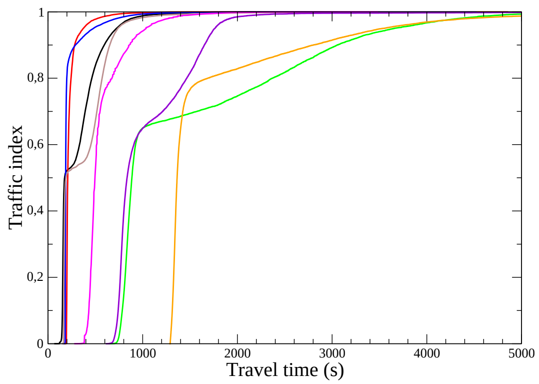
For a given travel time tt the index is defined by :

$$x_i = f(tt) \stackrel{\text{def}}{=} P(tt_i < tt),$$

using a weighted cumulative based on the segmentation

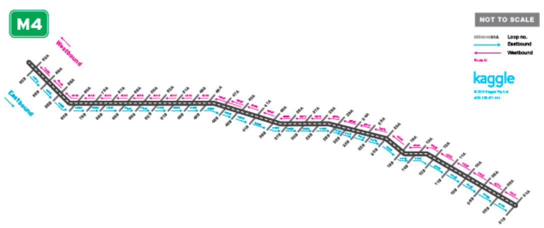


travel time distribution

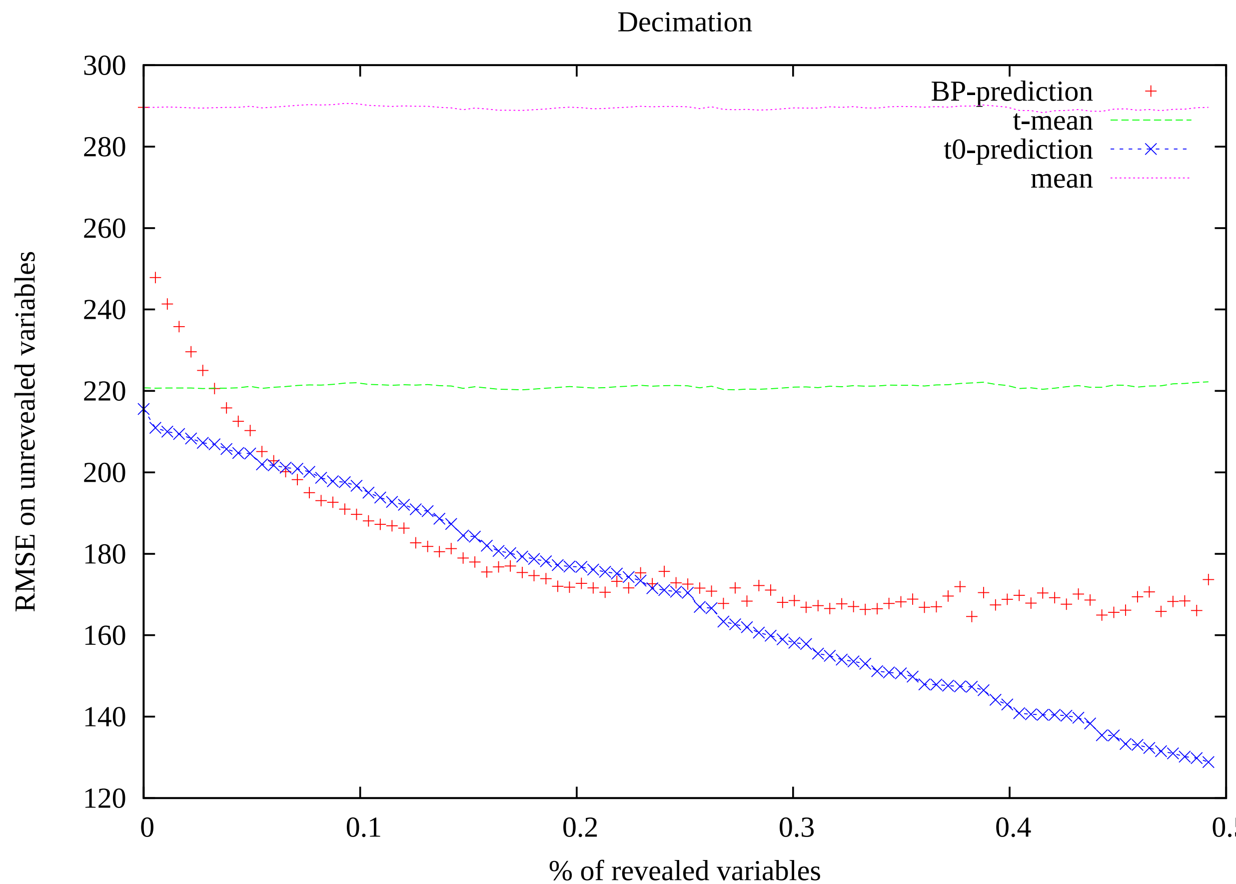


Various index maps

Highway data



- 62 segments, $\Delta t = 3$ minutes, strong correlations
- Extraction of a spanning tree with maximal weight
- the weight is the mutual information
- 3 temporal layers



Conclusion

- Ising model can be used as a building block
- Macroscopic variables have to be incorporated in the model
- Real data needed to check various hypothesis