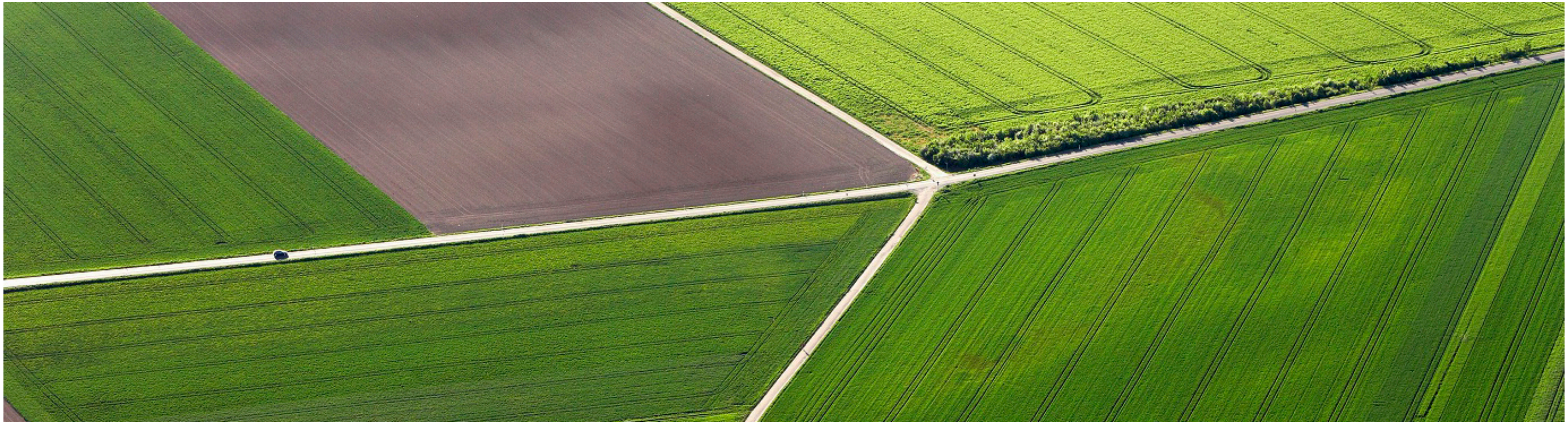
Statistical Models for Road Traffic Forecasting

Mediamobile & Insitut Mathématique de Toulouse

Collaboration with: Guillaume Allain, Thibault Espinasse, Fabrice Gamboa, Philippe Goudal, Jean-Michel Loubes

17 November 2015



Mediamobile & V-Traffic Services

Premium Traffic Information

Confidential

Mediamobile

European Leader in Traffic Broadcast

- Specialists in Traffic and Mobility since 1996
- Experts in RDS-TMC, DAB-TPEG and connected services
- Part of the TDF group, European Leader in Terrestrial Broadcast
- Present in France, Germany, Finland, Sweden, Norway, Denmark, and Poland
- Our mobility solutions are called V traffic

Our Mission

« Increase the road efficiency and the motorist safety in providing the best real time traffic and mobility information in Europe »

V-Traffic – Key dates

1996
- Creation of Mediamobile by TDF, Renault and Cofiroute
- Manufacturer and distributor of Visionaute, first PND to calculate travel time using the RDS-TMC technology.

1999
- 1st traffic information service for mobile phones (France Telecom). Among the first users of the WAP technology.

2000
- First contracts with car manufacturers to feed their in-dash navigation systems with real time traffic information
- Switch from B2C to B2B, pioneer of one-off fee model in Europe

2003
- Traffic information services marketed under the V-Trafic brand

2006
- Integration of Mediamobile into TDF's Multimedia business unit

2008
- Partnership with Orange, investigating possibilities to produce traffic information via the Floating Mobile Data tech

2009
- Joint venture with Infoblu (Italy) and ITIS Holdings (UK) to provide connected Pan-European offering to car manufacturers.

2010
- Acquisition of Destia Traffic assets in June 2010, renamed Mediamobile Nordic

2011
- BMW connected service launched, 1st TPEG connected service in France.

2012
- Toyota TPEG DAB broadcast service launched in Belgium in partnership with BeMobile.

2013
- Launch of the first TPEG DAB commercial service in Germany.
- First RDS-TMC Pan European Broadcast contract (Volvo), in partnership with Be Mobile, Infoblu, Traffic Master, TrafficNav

2014
- DAB service launched in Norway with Garmin.

Automotive manufacturers

ASTON MARTIN | Audi | BMW | CITROËN | FCA FIAT CHRYSLER AUTOMOBILES | Ford | HONDA

HYUNDAI | JAGUAR | KIA | LAND ROVER | LEXUS | McLaren

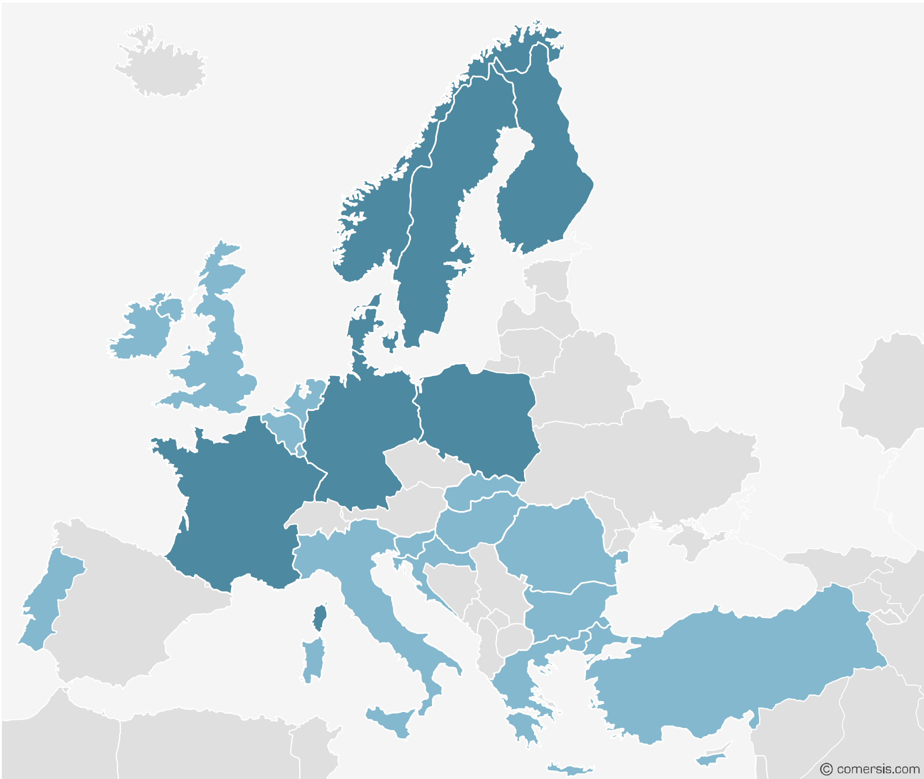
Mercedes-Benz | MINI | MITSUBISHI | NISSAN | OPEL | PEUGEOT | RENAULT

SEAT | ŠKODA | SUBARU | SUZUKI | TOYOTA | VOLVO | VW

Navigation solutions providers

ALPINE	AMARYLLO	appello.	BOSCH	clarion	Continental
DENSO	EB	KENWOOD	GARMIN	HARMAN/BECKER AUTOMOTIVE SYSTEMS	LOGICOM
MAGNETI MARELLI	mappy	MITSUBISHI ELECTRIC	Mio	NNG	UNITED NAVIGATION
SNOOPER	Sygic	Panasonic	Pioneer	ZENEC	

Mediamobile & partners
Traffic services available today (all technologies)



Mediamobile offers traffic information in over 20 European countries directly, through our partner network including ***BeMobile, Infoblu Trafficmaster*** and ***TrafficNav*** …

Mediamobile

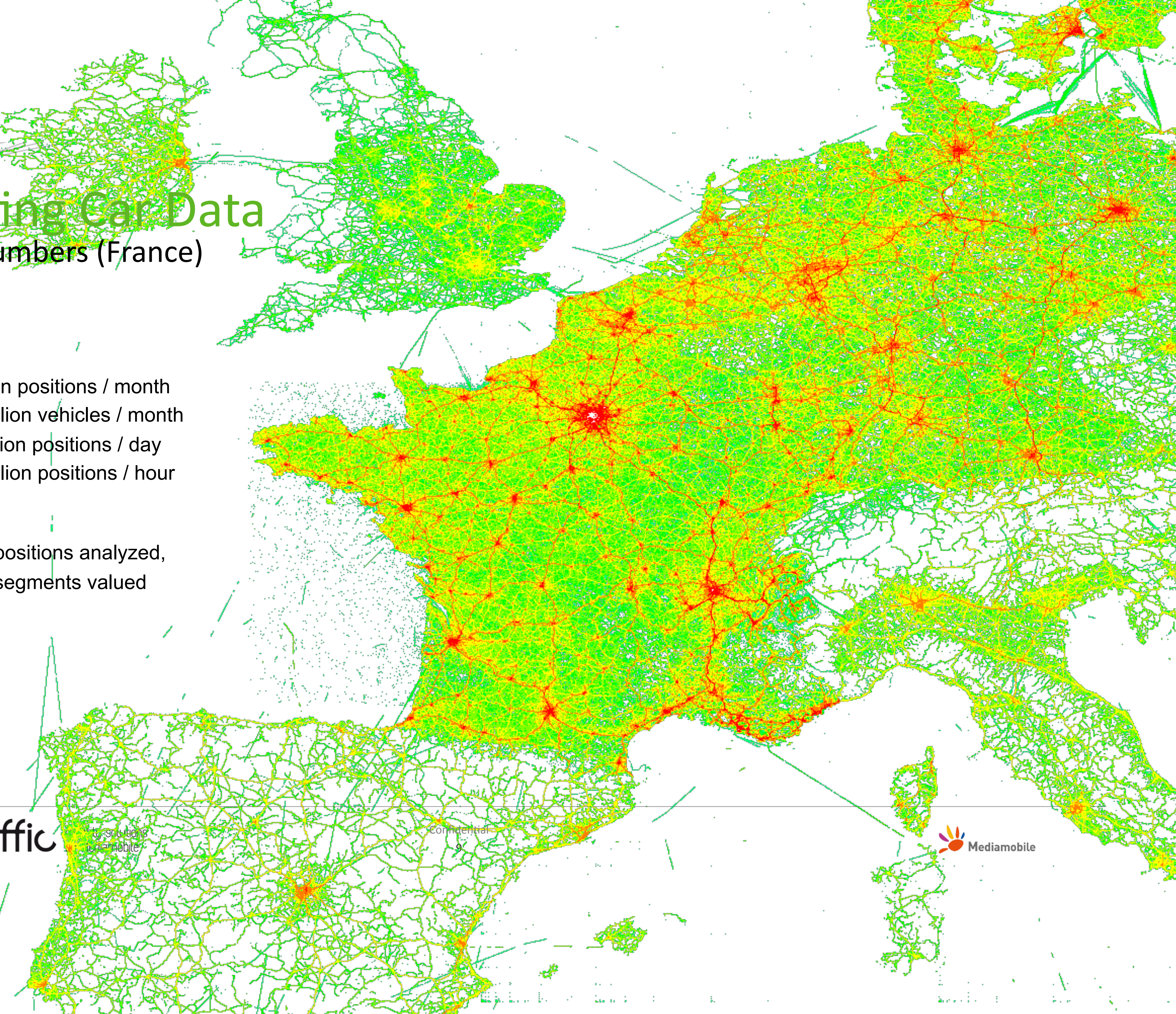
Trafficmaster, Bemobile, InfoBlue TrafficNav

Mediamobile
France
Germany
Sweden
Finland
Norway
Poland
Denmark
Trafficmaster
UK
Infoblu
Italy
Be-mobile
Belgium
Greece
Luxembourg
Portugal
Romania
Netherlands
Turkey
TrafficNav
Bulgaria
Croatia
Hungary
Ireland
Slovakia
Slovenia

V-traffic – mastering the value chain

Floating Car Data

Some numbers (France)

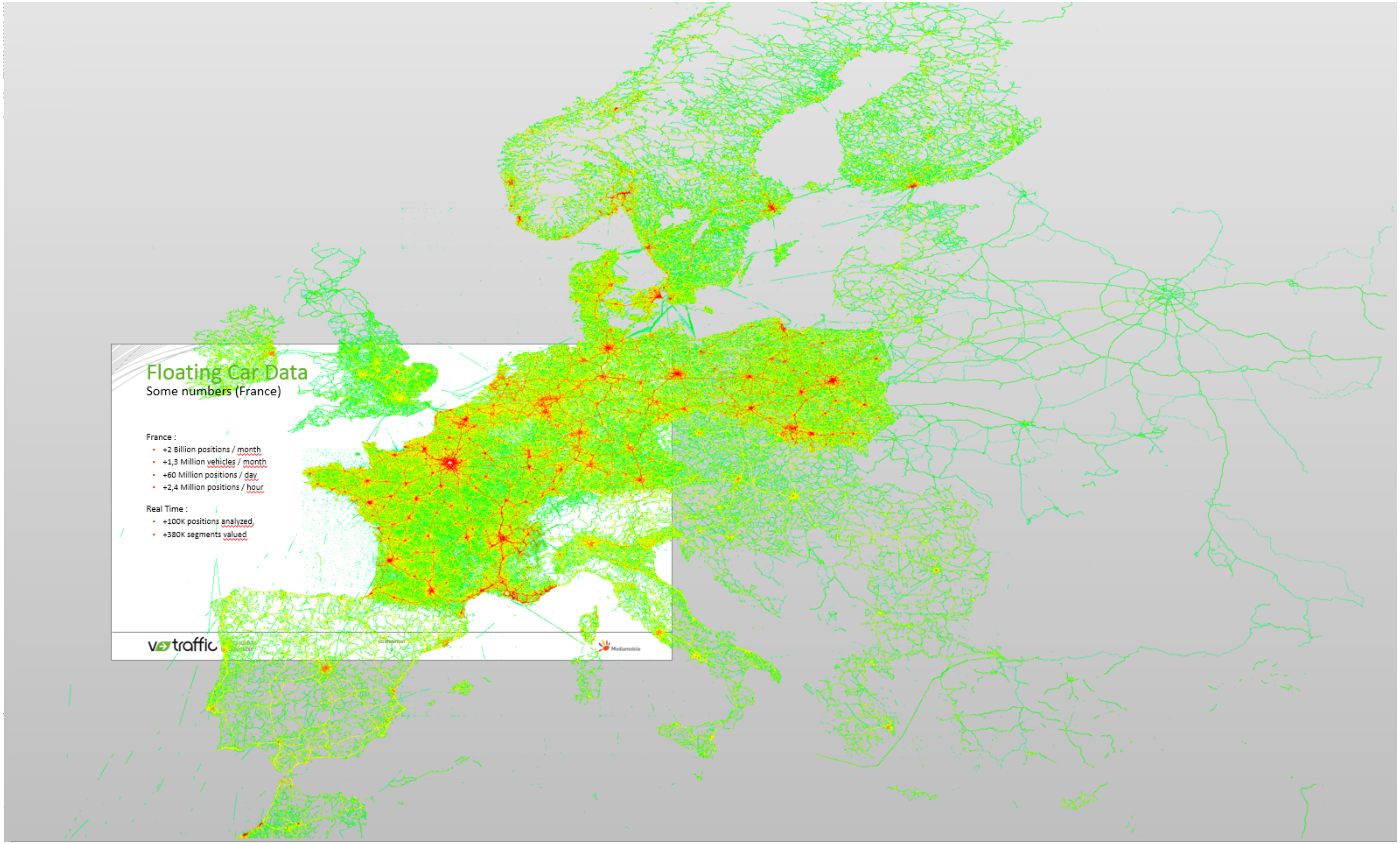


France :

- +2 Billion positions / month
- +1,3 Million vehicles / month
- +60 Million positions / day
- +2,4 Million positions / hour

Real Time :

- +100K positions analyzed,
- +380K segments valued



Floating Car Data
Some numbers (France)
France :
+2 Billion positions / month
+1,3 Million vehicles / month
+60 Million positions / day
+2,4 Million positions / hour
Real Time :
+100K positions analyzed,
+380K segments valued
traffic
Mediamobile

Overview

Objectifs de l'Etude

- **Gathering** raw traffic information
- **Processing** and agregating
- **Broadcasting** (radio, www, mobile device...)

$\Rightarrow$ Fancy new services : **forecasting** and **dynamic routine**

Industrial constraints :

- **coverage** $\begin{cases} \text{each road of the network} \\ \text{from real time to long run} \end{cases}$
- **quality/accuracy** $\begin{cases} \text{controlled speed prediction error} \\ \text{controlled jam prediction error} \end{cases}$
- **user friendly** $\begin{cases} \text{automatable} \\ \text{adaptative} \\ \text{easy to update} \end{cases}$

Road traffic data-Road network

What is a road network ?

- Graph composed of a set of pair (**edges**,**vertices**)
- **Complexity** of the graph $\rightarrow$ *Functional Road Classes (FRC)*
- **FRC** $\rightarrow$ road type classification (arterial, collector, local road...)

FRC	Number of edges	$\sum \mathrm{L[km]}$
0	46 175	22 580
1	232 572	42 793
2	462 907	75 453
3	998 808	175 790
{0,1,2,3}	1 740 462	316 616

Tab : Number of edges by FRC

- Network coverage depends on the FRC

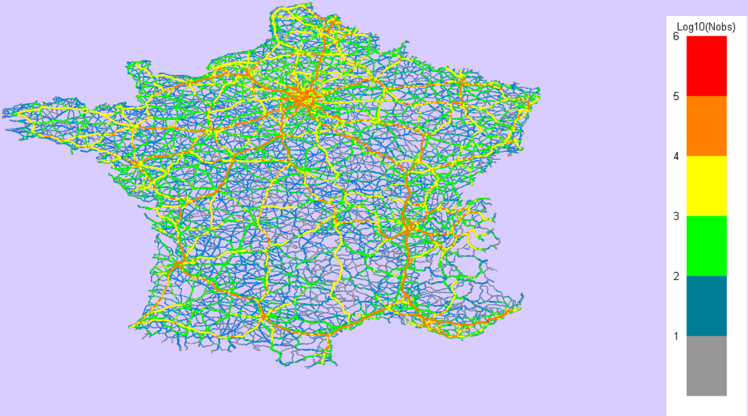
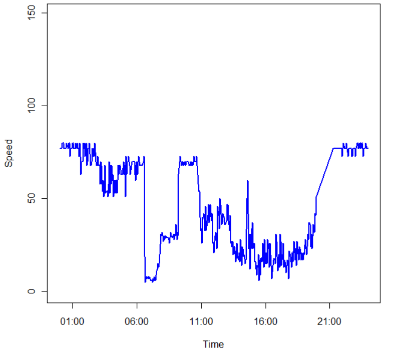


Fig : Network coverage by all FRC {0,1,2,3} from 03/01/2009 to 05/31/2009

Speed data

What is a speed data ?
Loop sensor

- speed calculated from flow and density (conservation law)



Pros

- More accurate
- 3min constant frequency

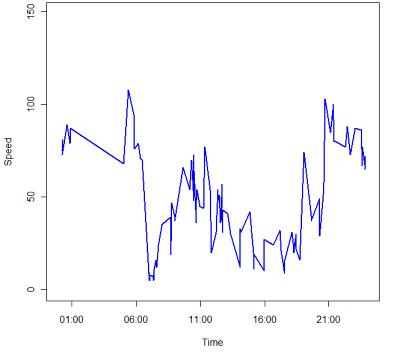
Cons

- Located only in main roads
- Thresholded at national speed limits

Speed data

GPS sensor : Floating Car Data

- positions are mapped on a graph → building speeds



Pros

- Can potentially cover all the graph
- Raw source of data

Cons

- Less accurate → GPS and map-matching error
- More variable → outlier emergence
- Random frequency → user feedback

Observations : a big data framework

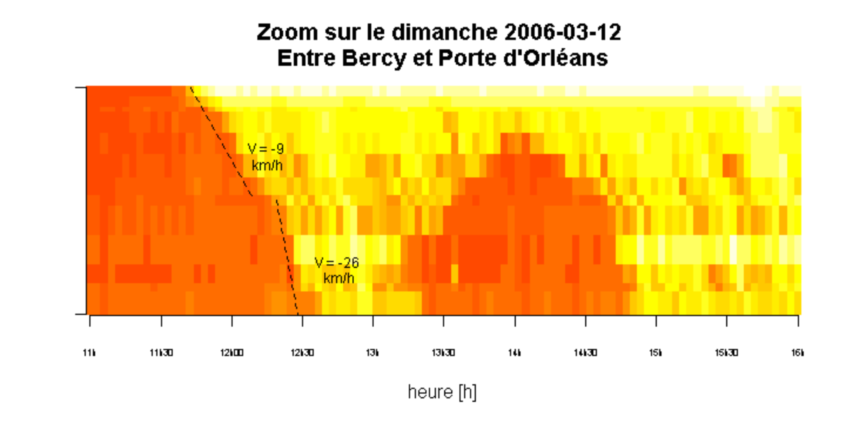
$$(x, t) \mapsto \begin{cases} V(x, t) \\ Y(x, t) \end{cases}$$

- V : Field of vehicle speeds observed at random time and space locations, observed partially on edges x of a graph (roads of the network) and observed when time goes by $t \mapsto V(x, t)$.
- spatio-temporal correlations (physics of traffic) and rupture of stationnarity
- Y : variables such as traffic events, weather conditions

large scale, online data

Overview

Local stationnarity enables to learn



Learning Features of the road traffic

Our Goal

- Appoach the road traffic dynamic with local statistical models

$$V(x_q, t_{p+h}) = \Phi(V(x,t), Y(x,t)) \rightarrow V_{q,p+h} = \Phi_{q,p,h}(\underbrace{\{V_{i,k}; i \in G, k \in T\}}_{X})$$

Problems

- High dimension of X
- All $V_{i,k}$ not influent

Solution

- Dimension Reduction
- Promoting sparse representations
- Using methodologies from machine learning

Modeling traffic dynamic with significative effects **only**

Sparse local regression in high dimension

$$V_{q,p+h} = \Phi_{q,p,h}(V_{i,k}) \rightarrow V_{q,p+h} = \sum_{i \in G, k \in T} \beta_{i,k}.V_{i,k}$$

$$\text{Estimation} \quad \widehat{\beta_{i,k}} = \underbrace{K((i,k),(q,p+h))}_{\textbf{Kernel}}$$

Kernel selection : fit road traffic dynamic

- learning a sparse set of influent parameters

$$\hat{\beta} = \arg\min_{\beta} \left(\|V_{q,p+h} - \sum_{i \in G, k \in T} \beta_{i,k}.V_{i,k}\|^2 + \lambda \sum |\beta_{i,k}| \right)$$

Functional Clustering

Dimension reduction using Functional mixture model

- speed curve V is represented with m **finite number of patterns**

$$f_1, \dots, f_i, \dots, f_m \text{ avec } f_i \in \mathbb{R}^K$$

$$V = \sum_{i=1}^{m} \mathbb{1}_{E=i}\, f_i + \epsilon_i \text{ et } f^\star = f_E$$

$$\left.\begin{array}{l} E \in \{1, \dots, m\} \text{ i.i.d. hidden R.V.} \\ \epsilon_i \in \mathbb{R}^K\,,\ \epsilon_i \sim \mathcal{N}(0, \Sigma_i \in \mathcal{M}_{K,K}) \end{array}\right\} \mathbb{E}[V|E=i] = f_i\,,\ \mathrm{Var}[V|E=i] = \Sigma_i$$

- **Classification** of E then prediction of V by $f^\star$:

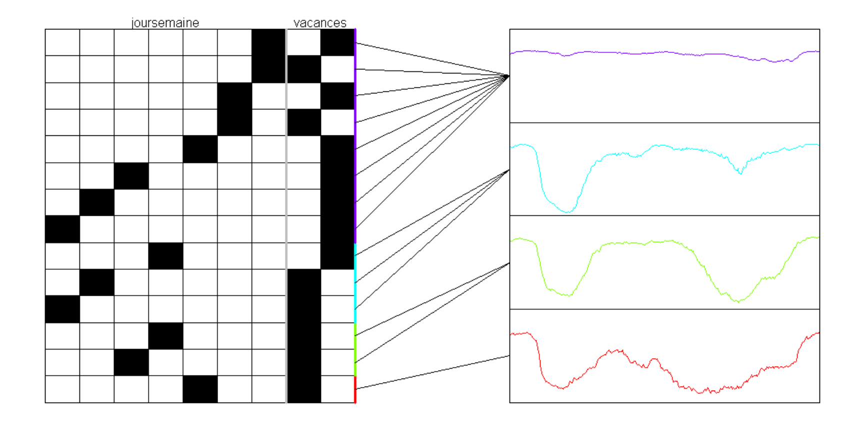
$$\begin{tikzcd} X \arrow[rr, "\text{regression}"] \arrow[dr, "\text{classification}"'] & & \hat{V} \in \mathbb{R}^K \\ & \hat{E} \arrow[ur, dashed] & \end{tikzcd}$$

The clustering methodology

Several ways to cluster functional data :

- Using latent variables : time, days, weather conditions
- Learning locally the active sets of variables using ℓ^1 penalty or (separate or group LASSO)
- Using clustering algorithms such as k-means, kernel k-means, DBSCAN ...
- Using a better representation of the data using low rank decomposition (NMF : non negative matrix factorization) or tensor factorization

An example of clustering model



Allocation Rules [Patent IMT Mediamobile]

Frame :

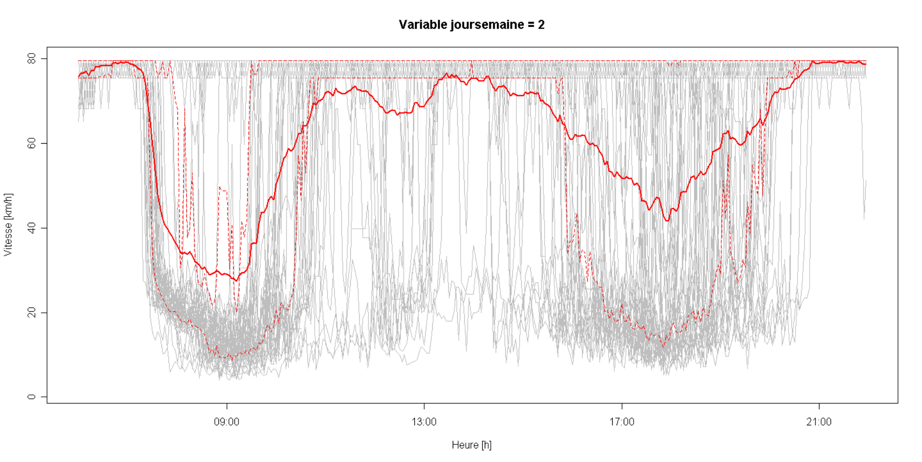
$\rightarrow$ Prediction **in the day** D
$\rightarrow$ Speeds V^p are **known**
$\Big\}$ p fixed, $X = (V^p, C)$

- X Time series
- How many patterns ?
    - h big et p small :
      $\Rightarrow m$ small
    - $\rightarrow$ h small and p big :
      $\Rightarrow m$ **big**
- Need for efficient clustering algorithm but off-line to get a **collection of identifiable features**

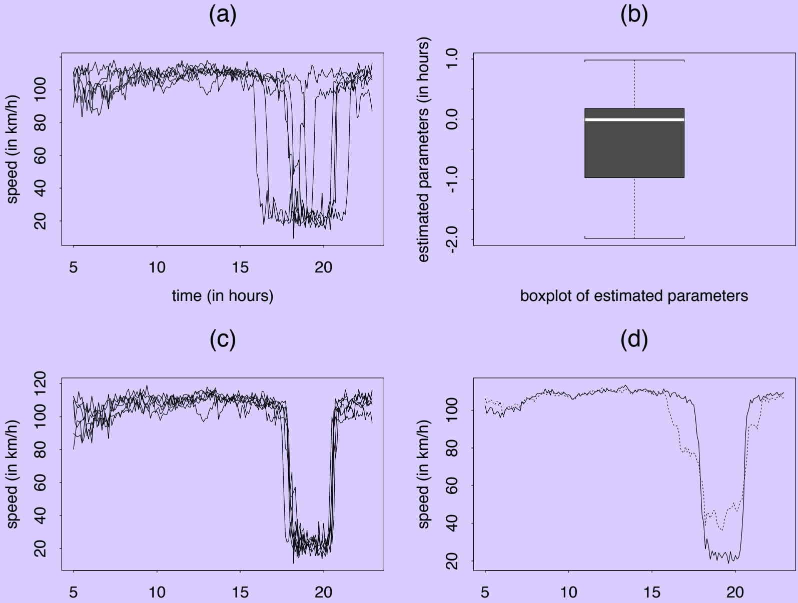
Key point : finding features that respects the structure of road trafficking .

Overview

Shift on traffic jams



Shift on traffic jams



Registration of Shape invariant model (SIM)

From a practical question to a theoretical model : **how to extract a feature from curves with deformations** ?

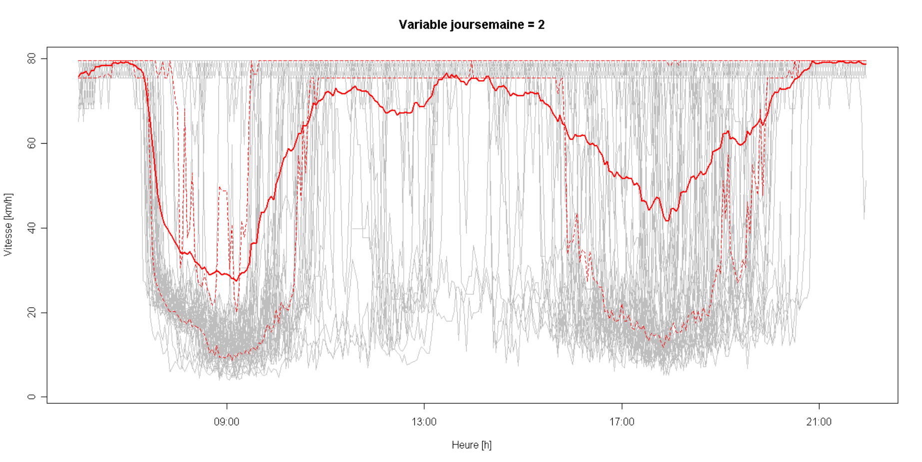
$$Y_{ij} = f_j^*(x_i) + \varepsilon_{ij} \quad i = 1 \ldots n_j,\ j = 1 \ldots J.$$

- **there exists** $f^\star : \mathbb{R} \to \mathbb{R}$ **with**

$$f_j^*(\cdot) = a_j^* f^\star(\cdot - \theta_j^*) + \upsilon_j^* \quad (\theta_j^*, a_j^*, \upsilon_j^*) \in \mathbb{R}^3,\ \forall j = 1 \ldots J.$$

  $f^\star$ is the **feature** that conveys the structure of traffic data.
  More than 12 research papers, 4 Phd inspired by the paper by Gamboa, Loubes, Maza [2007] with several distances, online methods and other modifications

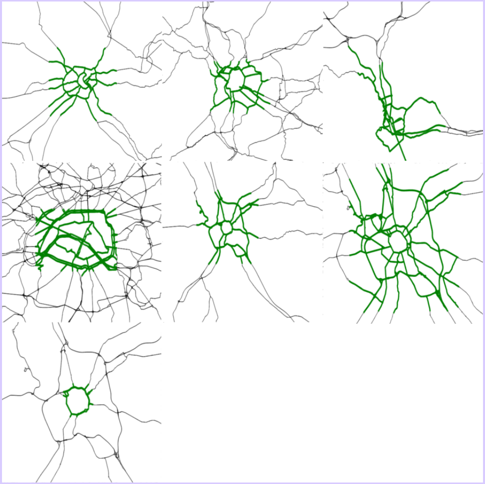
Shift on traffic jams



Overview

Graph of roads network

Modeling : Random process $(X_i^{(n)})_{n\in\mathbb{Z}, i\in G}$



Graph of roads network

Modeling : Random process $(X_i^{(n)})_{n\in\mathbb{Z}, i\in G}$

- Indexed by (discrete) time $\mathbb{Z}$ and the **graph** G of the road traffic network

Graph of roads network

Modeling : Random process $(X_i^{(n)})_{n\in\mathbb{Z}, i\in G}$

- Indexed by (discrete) time $\mathbb{Z}$ and the **graph** G of the road traffic network

Objective

Use spatial information to predict : build a model for covariance operators of X indexed by a graph

Gaussian Process on Graph : Origin of the Problem

Trafic : Predict the speed of the vehicles with missing values

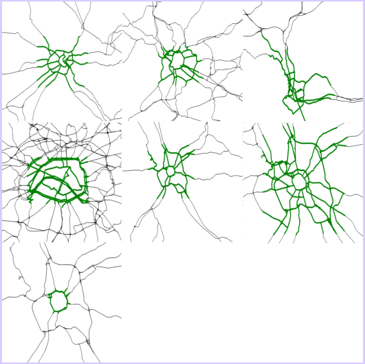
Until now : Spatial dependency is not well exploited

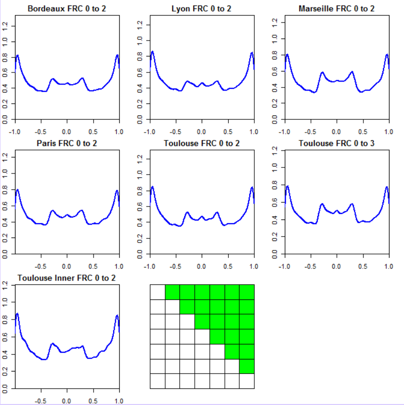
Aims

- Give a model that uses spatial dependency
- Estimate the spatial correlation
- Spatial filtering

Methodology : use the **spectral representation of the graph** eigenvalues and eigenvectors of the graph. The covariance of the process is a function of the spectrum of the graph.

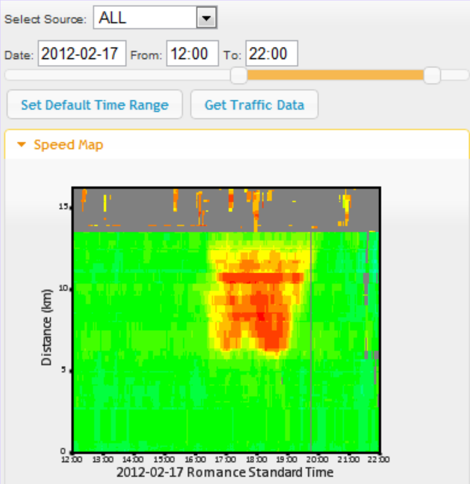
Spectrum of the road network



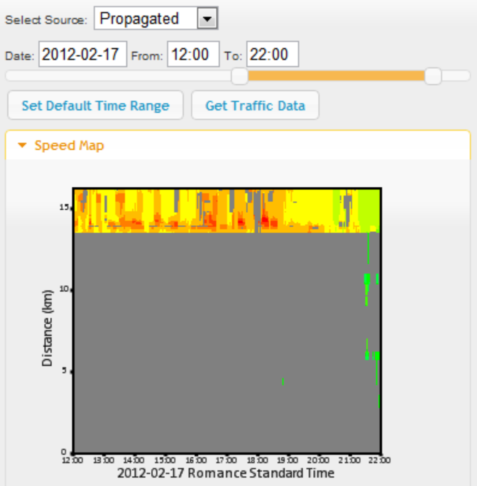


The concrete problem

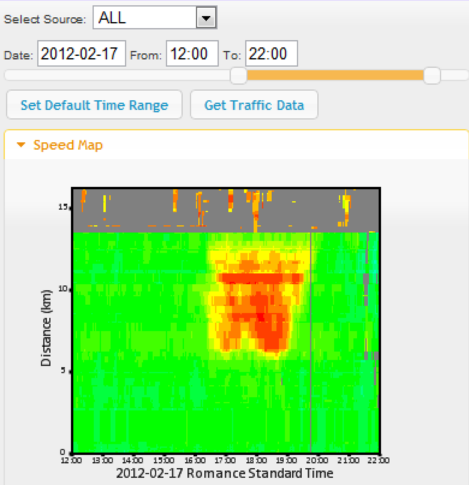
The concrete problem

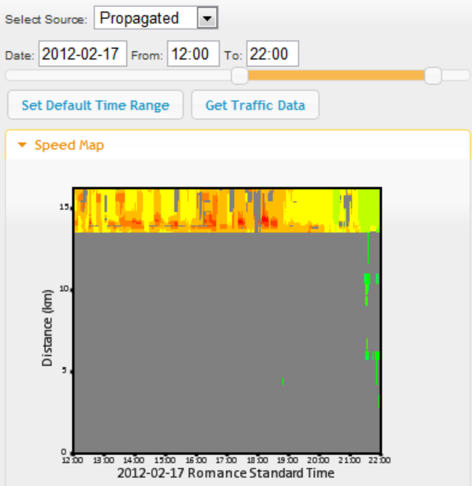


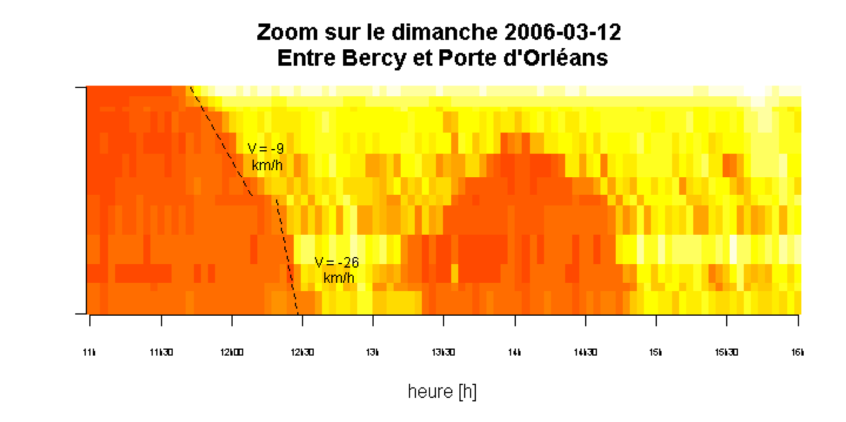
A solution ?


Select Source: Propagated
Date: 2012-02-17 From: 12:00 To: 22:00
Set Default Time Range
Get Traffic Data
Speed Map
Distance (km)
0
5
10
15
12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00
2012-02-17 Romance Standard Time

Let's compare






Zoom sur le dimanche 2006-03-12
Entre Bercy et Porte d'Orléans
V = -9
km/h
V = -26
km/h
11h
11h30
12h00
12h30
13h
13h30
14h
14h30
15h
15h30
16h
heure [h]

Thank you for your Attention