Framework for Traffic Pattern Identification: Required Step for Short-term Forecasting

Jordi Casas (corresponding author)
TSS - Transport Simulation Systems and Universitat de Vic
Passeig de Gràcia 12, 08007 Barcelona (Spain)
Phone +34 93 317 1693
Fax +34 93 317 1716

Email: casas@aimsun.com

Aleix Ruiz de Villa
TSS - Transport Simulation Systems
Passeig de Gràcia 12, 08007 Barcelona (Spain)
Phone +34 93 317 1693
Fax +34 93 317 1716
Email: aleix.ruizdevilla@aimsun.com

Alex Torday
TSS - Transport Simulation Systems,
Passeig de Gràcia 12, 08007 Barcelona (Spain)
Phone +34 93 317 1693
Fax +34 93 317 1716
Email: torday@aimsun.com

ABSTRACT

In the world of transport management, the term ‘anticipation’ is gradually replacing ‘reaction’. Indeed, the ability to forecast traffic evolution in a network should ideally form the basis for many traffic management strategies and multiple ITS applications. Real-time prediction capabilities are therefore becoming a concrete need for the management of networks, both for urban and interurban environments, and today’s road operator has increasingly complex and exacting requirements. Recognising temporal patterns in traffic or the manner in which sequential traffic events evolve over time have been important considerations in short-term traffic forecasting. However, little work has been conducted in the area of identifying or associating traffic pattern occurrence with prevailing traffic conditions. This paper presents a framework for detection pattern identification based on finite mixture models using the EM algorithm for parameter estimation. The computation results have been conducted taking into account the traffic data available in an urban network.

1 INTRODUCTION

Numerous techniques for traffic data forecasting using historical detector data and/or real-time measurements have been developed using statistical methods, Neural Networks, Fuzzy-logic, Support Vector Machines and Traffic Simulation. (Vlahogianni et al, 2004) summarizes different methodologies of traffic data forecasting where some examples of the different methods could be: (Smith. and. Demetsky, 1996) described a nonparametric regression method; (Williams et al. 1998) introduced seasonal autoregressive integrated moving average and exponential smoothing methods, the Time series Analysis using the ARIMA family of models (Davis et al. 1991; Hamed et al. 1995; Williams 2001, Stathopoulos and Karlaftis, 2003), the neural network applications (Smith and Demetsky 1994; Clark and al, 1998; Dia , 2001; Ishak and Alescandru 2003). Other methodologies has been analysed: the support vector machines (Chen and Grant-Muller, 2001; Zhang and Xie, 2007; Castro-Neto et al, 2009; Zhang et al, 2011), Fuzzy-logic (Stathopoulos et al, 2010) and the use of Traffic Simulation (Ruiz et al, 2007;Torday et al., 2010).

Recognising temporal patterns in traffic or the manner in which sequential traffic events evolve over time have been important considerations in short-term traffic forecasting. However, little work has been conducted in the area of identifying or associating traffic pattern occurrence with prevailing traffic conditions. The effectiveness of predictions is based on the knowledge of such patterns (Kantz and Schreiber 1997).

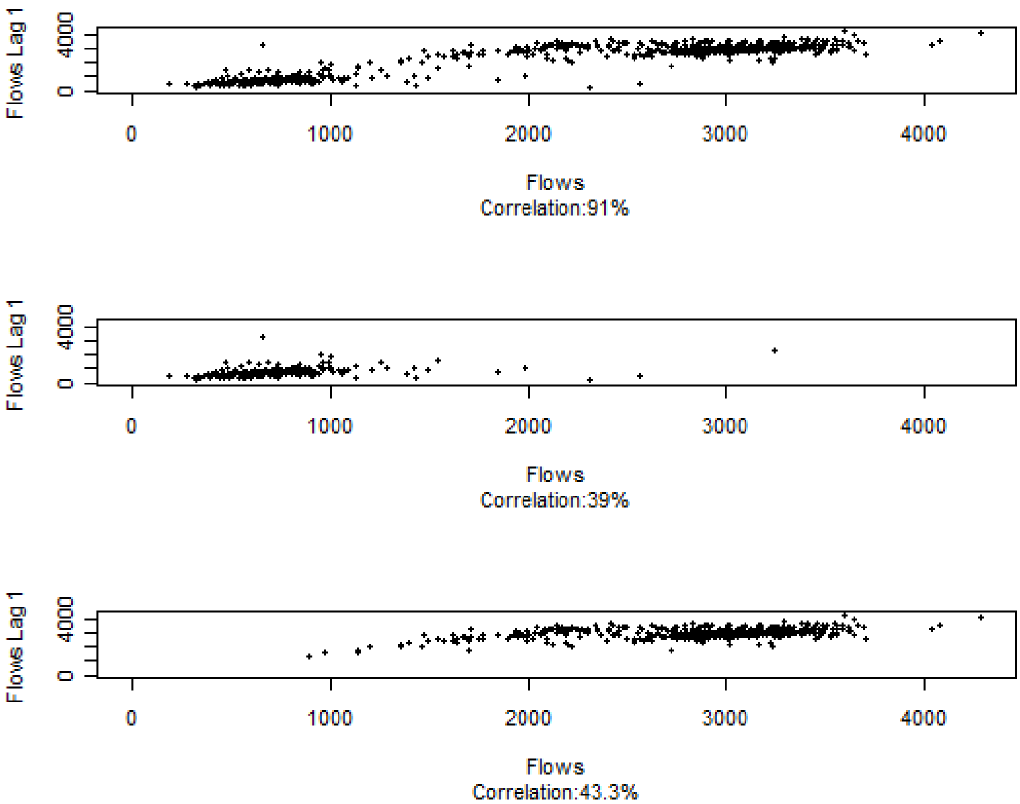
The present paper focuses on the study of the temporal patterns of traffic that arise in urban networks regarding both the statistical structure of sequential traffic events as well as the evolution of the specific structure in time. The aim is to develop a detection pattern identification framework and identify the pattern associated with real-time detection measures for short-term prediction. As a secondary consequence, this framework could be applied to incident detection in urban networks.

This paper is structured as follows: in the introduction we explain the main motivations and applications of a pattern identification procedure. In section 2, we explain the statistical methodology used to perform this task and we show that mixing patterns can lead to false conclusions. A real data example is studied in section 3 and we finally end with a conclusions section.

1.1 Detection Patterns for Time Series Analysis: The False Friend Correlation

The classic setting of time series analysis, such as ARIMA technique, is based on stationarity and linearity. Generally speaking, this means that there exists a linear relationship between time lags of a certain quantity (linearity) and this relationship is constant through time (stationarity). We will see that, in the presence of different patterns, analysing the data as a whole can lead to wrong conclusions about time lag correlations. Let us explain this through an example.

We analysed traffic flows from Madrid (Spain) for 153 days of the year 2005 (see Results section). In the first graph of FIGURE 1 we plotted traffic flows at time intant *t* versus traffic flows at instant *t-1*, where *t* runs from 7.45 am to 8.45 am, at every quarter-of-an-hour. Since the correlation coefficient is 91%, we could easily conclude that both linearity and stationarity hold. On the other hand, we applied the pattern identification framework described below, with two components, each one representing a different pattern. Thus, each day (comprising traffic flows between 7.30 am and 8.45 am) is assigned to one of the two patterns. We plotted the two sets of days separately in the second and third graphs in FIGURE 1. Observe that correlations have suffered a large decrease. Now, hypothesis of stationarity or linearity are clearly not verified by any of the two groups.

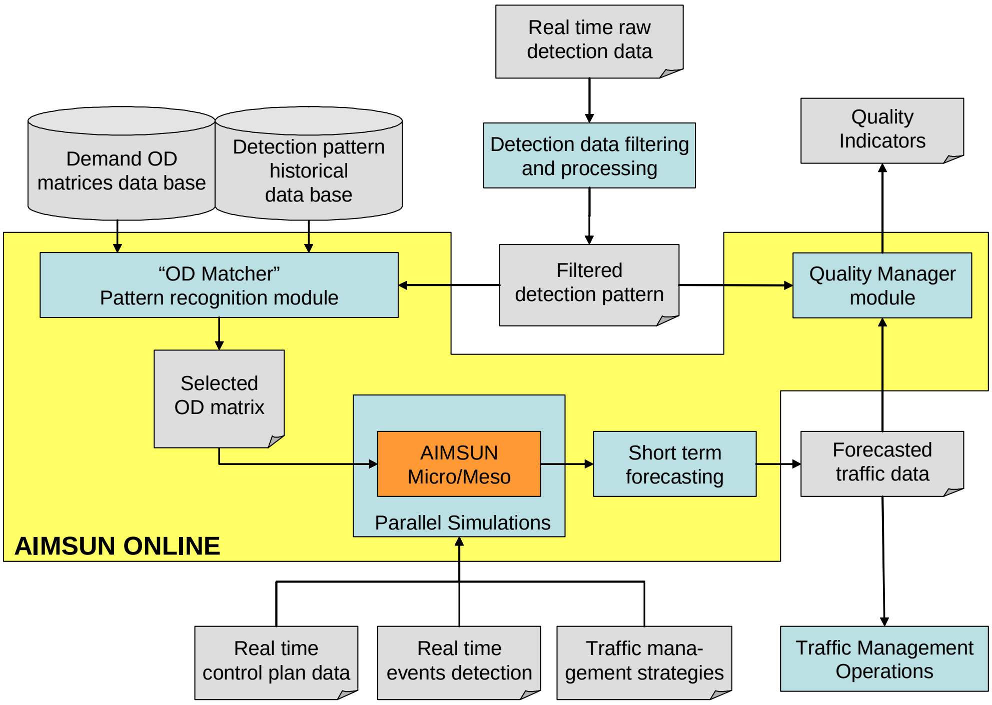


1. Example of false correlation

In section 2.4 we will study this effect in depth. We will see that it can be caused by mixing data from different patterns. Moreover, we will argue that, in this case, global correlation should be rejected and any attempt to study linearity should be done in each group separately. We conclude that pattern identification becomes a necessary step in order to prevent false conclusions in any linear setting.

1.2 Detection Patterns for Simulation

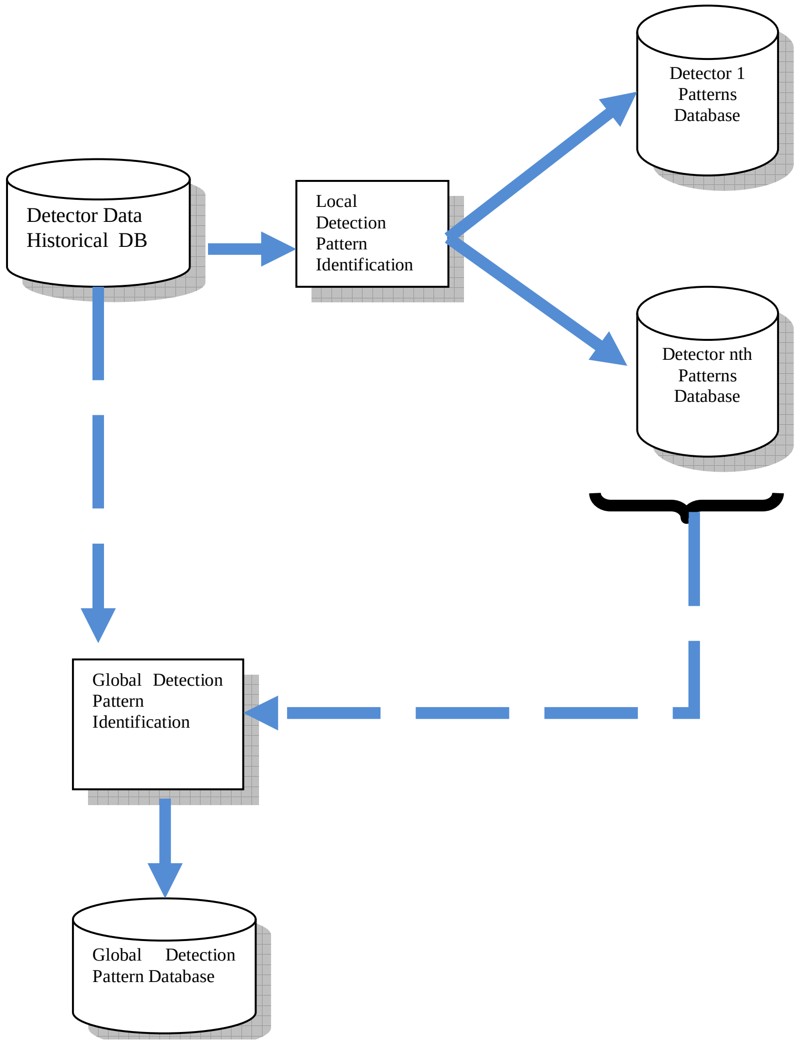
(Torday et al., 2010) propose Aimsun Online as a framework for a simulation-based decision support system for real-time traffic management.



2 General architecture of the Aimsun Online solution

In the Aimsun Online architecture (see FIGURE 2) the basic input for this application is the real-time detection data, which serves to determine the current level of demand in the network and consequently allows the loading of the corresponding OD matrix into the simulation model before it runs. This operation is carried out through the OD Matcher module, which compares the latest flow measurements from each detector (typically, the time series from the last 30 minutes) with pre-prepared daily traffic patterns. These patterns result from an offline treatment of the detection historical database (ideally containing more than one year´s data).

Considering this architecture and underlying the motivation of having detection pattern identification, the basic input required is the detection pattern definition used in different steps of the forecasting process.



3. Local and Global Pattern Identification

Integrating the detection pattern identification in this architecture, this paper proposes the application of the developed methodology. This process of Detection Pattern Identification is based on a sequential process, where the first stage is the local detection pattern identification and the second stage is the global pattern identification (see FIGURE 3).

1. Local Detection Pattern Identification (LDPI): Calculates the different patterns applying the EM algorithm for each detector, taking into account the data available for each individual detector inside the historical database.
2. Global Detection Pattern Identification (GDPI): Calculates the different global patterns taking into account all data available within the historical data and each individual detector pattern calculated in the previous step.

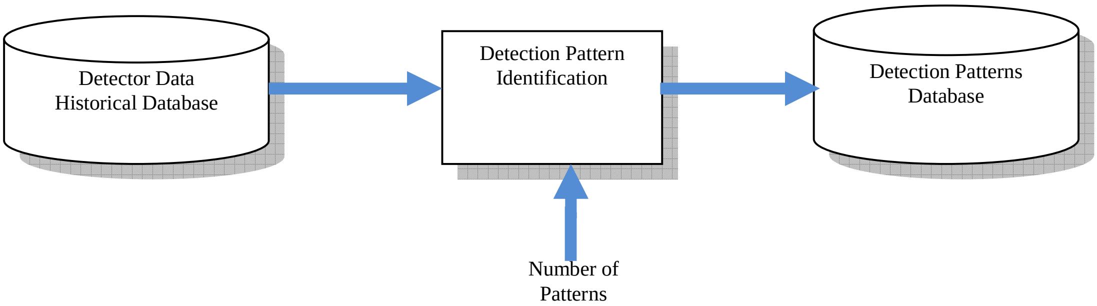
The application of this generic framework in a simulation-based decision support system for real-time traffic management is determined by the following elements:

- Traffic demand definition (in FIGURE 2 identified as “OD Matcher”): GDPI determines the global detection patterns and, when the system receives the detector measurements in real-time, classifies the real data according to the different patterns calculated and gives as output the traffic demand defined in terms of OD matrices associated with the pattern that reproduces the real-time detector measures.

- Quality Manager: LDPI determines the pattern for each individual detector, according to the real-time detector measures, and then using the simulation results combined with a prediction based on patterns establishes a rate of goodness of different types of forecasting methods.
- Incident Detection Manager (this is not represented explicitly in FIGURE 2 but this module could be added as part of the Quality Manager): LDPI determines the pattern for each individual detector according to the real-time detector measures and then establishes the presence, location and severity of possible incidents. The incident detection is based on the "Outlier Detection" methodology, explained in Section 2.

2 METHODOLOGY

The methodology proposed in this paper for detection pattern identification is based on finite mixture (FM) models using the EM model for parameter estimation. The generic framework (see FIGURE 4) receives as input the real detection data historical database and the number of patterns as a user-defined parameter, and generates as output a detection patterns database. The number of patterns a priori is a user-defined parameter but in the numerical results this has been calculated using the ICL criterion.



4. Generic framework for detection pattern identification

2.1 Finite Mixture Models

Given densities $f_1(y;\theta_1),\dots,f_g(y;\theta_g)$ where θ_i denote the density parameters and $y=(y^1,\dots,y^n)$, and a collection of proportions $\pi_1,\dots,\pi_g$ satisfying $\pi_i \geq 0$, and $\pi_1+\cdots+\pi_g=1$ we define the density

$$f(y;\ \vartheta)=\pi_1 f_1(y;\theta_1)+\cdots+\pi_g f_g(y;\theta_g), \qquad (1)$$

where $\vartheta=(\pi_1,\dots,\pi_g;\ \theta_1,\dots,\theta_g)$. The density $f(y,\vartheta)$, has a direct interpretation. Say we have g different groups, each one represented by proportion π_i of the whole population. In our case, neither the set of parameters $\theta_1,\dots,\theta_g$, nor the proportions $\pi_1,\dots,\pi_g$ (nor the number of components) are known, so they have to be estimated.

A classic specification for the functions $f_i(y;\theta_i)$ is the normal density. In this case $\theta_i=(\mu_i;\Sigma_i)$, where μ_i is the vector of expectations and Σ_i is the matrix of variances-covariances. A remarkable property is that any density function can be well approximated by a mixture of normal densities if the number of components is large enough. In fact, in the extreme case that g equals the number of observations of the sample, the normal mixture model can be regarded as a parametric density estimation method with a Gaussian kernel (see Silverman, 1986).

2.2 Parameter Estimation: EM Algorithm

The natural way to estimate the unknown parameters is the maximum likelihood estimation:

$$L=\prod_j f(y_j;\ \vartheta).$$

However, the traditional approach (to find the roots of the gradient of the log likelihood) doesn't lead to a closed-form solution. Instead, the problem is solved by means of the EM algorithm. See (Dempster et al., 1977; McLachlan and Krishnan, 2008) and (McLachlan and Peel, 2000) for a comprehensive treatment of finite mixtures.

Let $y_1,\dots,y_N$, be vectors of observations that are independent but may have dependent elements. In our study, they are flows in a particular detector over some time interval $I=(t_i)$, e.g. T=(8, 8.25, 8.5, 8.75, 9) hours and $y_1=(101.2, 87, 83, 95, 93)$ vehicles per hour registered at 23/04/2005. In the parameter setting, equal variance-covariance matrices should be taken in order to ensure global finite maximum of the likelihood, i.e., $\Sigma_1,\dots,\Sigma_g=\Sigma$. In the EM approach, unobserved variables z are considered realisations of the corresponding labeling of each observation. That is, for each observation y_j, we consider a random vector of binary elements $z_j = (z_{j,i})$, with $z_{j,i} = 1$ if the observation y_j belongs to the component i, and 0 otherwise. If we were to observe these vectors, the problem of estimating the parameters of the normal distributions would be simple. In the terminology of the EM algorithm, this method and the corresponding dataset are called 'complete'. The realised dataset is called 'incomplete'. Then the complete data likelihood (as a function of the variables y and z) is specified and the parameters are estimated using an iterative procedure consisting of two steps: (E) expectation and (M) maximisation. Given initial conditions $(\pi_1^0, \dots, \pi_g^0;\ \mu_1^0, \dots, \mu_g^0, ; \Sigma^0)$, E step in the k+1th iteration is performed by

$$\tau_{i,j}^{k+1} = \frac{\pi_i^k f_i(y_j; \mu_i^k; \Sigma^k)}{\sum_l \pi_l^k f_l(y_j; \mu_l^k; \Sigma^k)}, \tag{2}$$

for each observation j and each group i. Then, parameters are updated applying the M step,

$$\pi_i^{k+1} = \frac{1}{n} \sum_j \tau_{i,j}^{k+1},$$

$$\mu_i^{k+1} = \frac{\sum_j y_j \tau_{i,j}^{k+1}}{\sum_j \tau_{i,j}^{k+1}},$$

and

$$\Sigma^{k+1} = \frac{1}{n} \sum_i \sum_j \tau_{i,j}^{k+1} (y_j - \mu_i^{k+1})(y_j - \mu_i^{k+1})^T.$$

The algorithm stops when the the likelihood L of the new iteration differs from its value in the previous iteration by less than an *a priori* set threshold. Quantities $\tau_{i,j}$ when the algorithm has converged is the conditional probability of the observation j of belonging to group i (the conditional expectation of the elements of z_j). Then assignment to a group i' follows when the corresponding $\tau_{i',j}$ is the maximum among the others $\tau_{i,j}$.

2.3 Outlier Detection

In this paper we follow the scheme described by (Longford and D'Urso, 2011). Outliers are handled by introducing a new group, Group 0, governed by an "improper" component, that is, with constant density $f_0 = D$ (so $\int_{-\infty}^{\infty} f = \infty$), which will compete with the previous ones. The same formulas are used, replacing f_0 with D. Observe that the estimated π_0 is a non-decreasing function of D.

D cannot be estimated by statistical procedures. A desired level of outliers is sought instead. Given π_0^{des}, at each iteration k+1 we perform a Newton-Rapshon type algorithm to the function $\pi_0^{k+1} = \pi_0^{k+1}(D)$, so that $\pi_0^{k+1} = \pi_0^{des}$. Recall

$$\tau_{0,j}^{k+1} = \frac{\pi_0^k D}{\sum_{l \geq 1} \pi_l^k f_l(y_j; \mu_l^k; \Sigma^k) + \pi_0^k D},$$

$$\pi_0^{k+1} = \frac{1}{n} \sum_j \tau_{0,j}^{k+1}.$$

Then,

$$\frac{\partial \pi_0^{k+1}}{\partial D} = \frac{1}{nD} \sum_j \tau_{0,j}^{k+1} - (\tau_{0,j}^{k+1})^2,$$

and the formulae for the Newton-Raphson algorithm becomes,

$$D_{m+1} = D_m - \frac{(\pi_0^{k+1}(D) - \pi_0^{des})}{\partial \pi_0^{k+1}(D_m)/\partial D} = D_m - \frac{\frac{1}{n}\sum_j \tau_{0,j}^{k+1} - \pi_0^{des}}{\frac{1}{nD_m}\sum_j \tau_{0,j}^{k+1} - (\tau_{0,j}^{k+1})^2}.$$

Observe that if any $\tau_{0,j}^{k+1} \neq 1,0$, then $\frac{\partial \pi_0^{k+1}}{\partial D} > 0$ (because $0 \leq \tau_{0,j}^{k+1} \leq 1$ for all j), and the function is strictly increasing. Since $\pi_0^{k+1}(0) = 0$, and $\lim_{D \to \infty} \pi_0^{k+1}(D) = 1$, there is a unique root of the equation

$$\pi_0^{k+1}(D) - \pi_0^{des} = 0 \tag{3}$$

over the domain $\{D > 0\}$. It may happen that D_{m+1}<0. This implies that $\pi_0^{k+1}(D_m) - \pi_0^{des} > 0$, so we can assure that the positive root of (3) lies in the interval $(0, D_m)$. In this case we proceed applying a one-step bisection algorithm, i.e. $D_{m+1} = {}^{D_m}/_{2}$.

2.4 Finite Mixtures and False Friend Correlations

Traditionally, techniques such as ARIMA (at least its elementary versions), rely on stationarity and linearity. This means, in particular, that given a time series $\{X_t\}_t$ and k,

$$E[X_t] = \mu,$$
$$Var[X_t] = \sigma^2,$$
$$Cov[X_t, X_{t-k}] = \gamma[k] \quad (4)$$

and

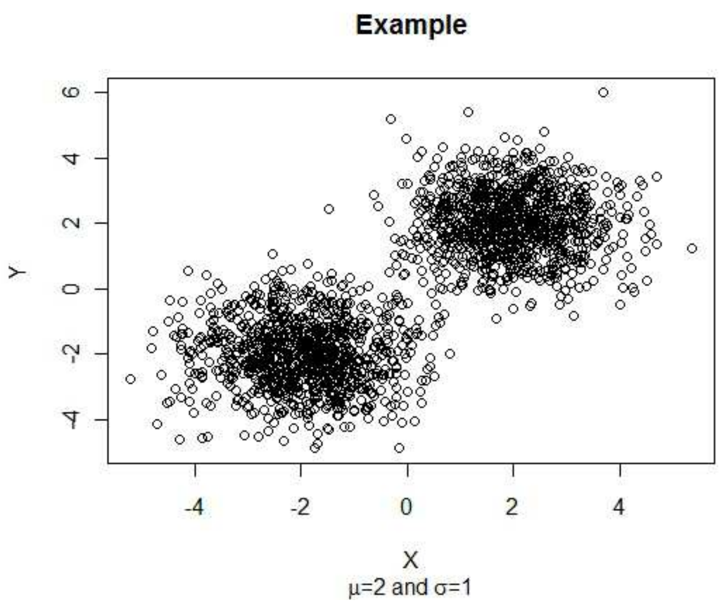
$$E[X_t | X_{t-1}, X_{t-2}, \ldots] \text{ is a linear function on } X_{t-1}, X_{t-2}, \ldots \quad (5)$$

See, for instance, Chatfield (2004). One empirical way to check (4) and (5) is by plotting the lagged pairs, say $\{(X_t, X_{t-1})\}_t$, and search for linearity and high correlation. We will see that finite mixtures can emulate a highly correlated process without having a true linear relationship. We construct an example of a mixture of two groups, none of them having internal correlation but with the whole process exhibiting high levels of correlation.

Consider two multivariate Gaussian vectors

$$(X_1, Y_1) \sim \mathcal{N}\big((\mu, \mu);\ \Sigma\big)$$
$$(X_2, Y_2) \sim \mathcal{N}\big((-\mu, -\mu);\ \Sigma\big),$$

(with densities f_1, f_2 respectively) describing two different groups, with $\mu > 0$ and $\Sigma = \sigma^2\text{Id}$. Consider a random independent sample taking values from each group with probability 1/2 for each (see FIGURE 5).



5. Plot of Synthetic Data

We denote the resulting process by (X, Y). Then, its underlying density is a mixture type density

$$f(x,y) = \frac{1}{2} f_1(x,y) + \frac{1}{2} f_2(x,y) = \frac{1}{2}\phi(x;\mu,\sigma^2)\phi(y;\mu,\sigma^2) + \frac{1}{2}\phi(x;-\mu,\sigma^2)\phi(y;-\mu,\sigma^2),$$

where

$$\phi(\,x\,;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

denotes the density of a Gaussian process with mean μ and variance σ^2. The marginal distribution of X is

$$f_X(x) = \frac{1}{2}\phi(x;\mu,\sigma^2) + \frac{1}{2}\phi(x;-\mu,\sigma^2).$$

Then,

$$E[X] = \frac{1}{2}E[X_1] + \frac{1}{2}E[X_2] = \frac{1}{2}\mu + \frac{1}{2}(-\mu) = 0,$$

$$Var[X] = \frac{1}{2}(E[X_1]^2 + Var[X_1]^2) + \frac{1}{2}(E[X_2]^2 + Var[X_2]^2) - E[X] = \frac{1}{2}(\mu^2 + \sigma^2) + \frac{1}{2}((-\mu)^2 + \sigma^2) = \mu^2 + \sigma^2.$$

Analogous results also hold for Y. Let's calculate the correlation coefficient between X and Y:

$$Cov(X,Y) = \int\int xyf(x,y)dxdy = \frac{1}{2}\int\int xyf_1(x,y)dxdy + \frac{1}{2}\int\int xyf_2(x,y)dxdy = I_1 + I_2.$$

The term I_1 can be written as

$$I_1 = \frac{1}{2}\int\int[(x-\mu)(y-\mu) + \mu x + \mu y - \mu^2]f_1(x,y)dxdy = I_{11} + I_{12} + I_{13} + I_{14}.$$

Then,

$$I_{11} = 0,$$

because is the covariance between X_1 and Y_1. We also have that,

$$I_{12} = \frac{1}{2}\mu\int\int x\phi(x;\mu,\sigma^2)\phi(y;\mu,\sigma^2)dxdy = \frac{1}{2}\mu^2 = I_{13},$$

and,

$$I_{14} = -\frac{1}{2}\mu^2.$$

Finally $I_1 = \frac{1}{2}\mu^2$. The term I_2 can be calculated analogously, so

$$Cov(X,Y) = \mu^2.$$

Thus,

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]}\sqrt{Var[Y]}} = \frac{\mu^2}{\mu^2 + \sigma^2}.$$

Taking μ big enough, since $\lim_{\mu\to\infty} Corr(X,Y) = 1$, this new process can exhibit high levels of correlation.

Let's calculate now $E[Y\,|X = x]$. The conditional density of Y given $X = x$ is

$$f_{(Y|X=x)}(y) = \frac{f(x,y)}{f_X(x)} = \frac{\phi(x;\mu,\sigma^2)\phi(y;\mu,\sigma^2) + \phi(x;-\mu,\sigma^2)\phi(y;-\mu,\sigma^2)}{\phi(x;\mu,\sigma^2) + \phi(x;-\mu,\sigma^2)}.$$

Then,

$$E[Y|X = x] = \int yf_{(Y|X=x)}(y)dy = \mu\frac{\phi(x;\mu,\sigma^2)}{\phi(x;\mu,\sigma^2) + \phi(x;-\mu,\sigma^2)} - \mu\frac{\phi(x;-\mu,\sigma^2)}{\phi(x;\mu,\sigma^2) + \phi(x;-\mu,\sigma^2)}.$$

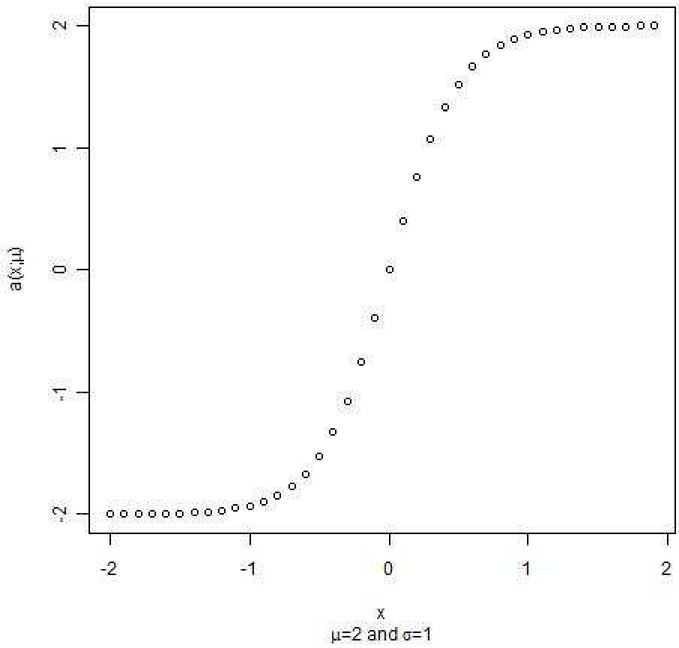
Denote

$$a(x;\mu) = \frac{\phi(x;\mu,\sigma^2)}{\phi(x;\mu,\sigma^2) + \phi(x;-\mu,\sigma^2)} = \frac{1}{1 + e^{\frac{(x-\mu)^2}{2\sigma^2} - \frac{(x+\mu)^2}{2\sigma^2}}} = \frac{1}{1 + e^{-\frac{2x\mu}{\sigma^2}}}.$$

Then $a(x;-\mu) = \frac{e^{-\frac{2x\mu}{\sigma^2}}}{1+e^{-\frac{2x\mu}{\sigma^2}}}$. Finally,

$$E[Y|X = x] = \mu\,(2a(x;\mu) - 1).$$

Observe that $\lim_{x\to\infty} E[Y|X = x] = \mu$, and $\lim_{x\to-\infty} E[Y|X = x] = -\mu$. We stress the fact that $E[Y\,|X = x]$ is not even a linear function on the variable x. Linearity of $E[Y\,|X = x]$ in x is an elementary hypothesis of any linear model (and so for ARIMA's). In this example, any assumption of linearity between the variables X and Y is clearly violated. In FIGURE 6 we show an example of the function $a(x;\mu)$ with $\mu = 2$ and $\sigma = 1$. Some linearisation could be argued in the interval $(-1,1)$, however recall that the means of both initial distributions are 2 and -2. So the "linear" interval $(-1,1)$ lies in a "low probability" zone.



6. Function a(x,µ)

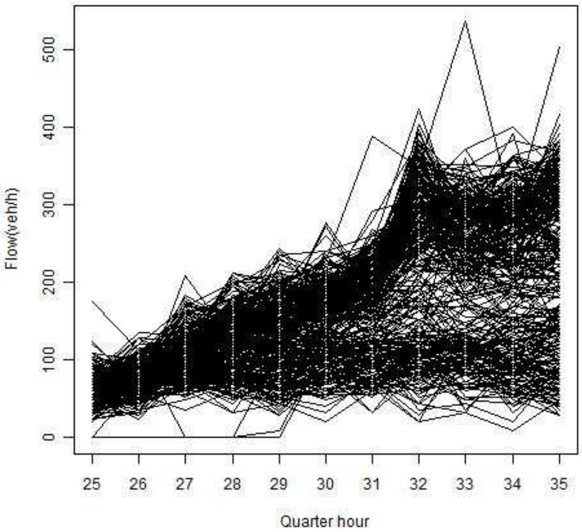
We might still think that the globality of the process brings some information about its behavior, and so the global linearity could be useful for forecasting, to take one example. The answer is that if there were a true linear relationship, it should still hold for subsets of the sample, which is not clearly the case. Moreover, knowing more information about the sample, for instance to which group the observation belongs, should bring more accuracy to our predictions but applying a linear scheme to any of the two example groups is indeed a bad decision, since variable X_i and Y_i are uncorrelated.

This shows that properties of the component distributions (normality, linear regression, independence, and others) are not retained by mixing. That is, the mixture does not have these properties, even if every component does. This implies that by mixture analysis we may discover features of the components (or even of every component) that cannot be found in the entire dataset, comprising the union of the subsets that belong to the components.

We conclude that linear relationship, if any exists, should be considered for each group independently. This analysis shows the importance of the previous cluster analysis when applying any forecasting technique.

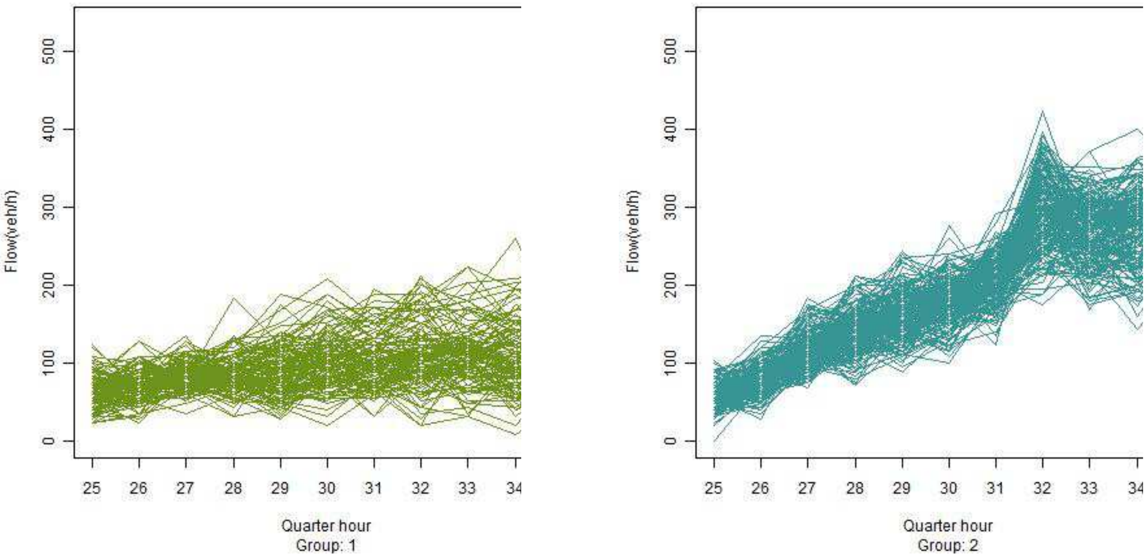
3 RESULTS

Data from 31 detectors in Madrid (Spain) registering traffic flow every quarter-of-an-hour throughout the year 2005 has been analysed. No missing data technique has been applied, so those days with any missing value among the 96 quarters-of-an-hour have been dropped.



7. Madrid detector data

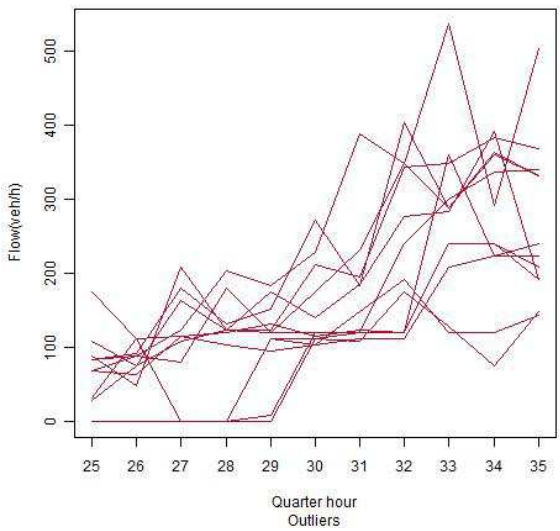
In FIGURE 7, we display the whole bunch of trajectories, each one representing a different day for the detector 40001, between 6.15 a.m.(qh=25) and 8.45 a.m.(qh=35). We chose only two groups (the simplest case) for illustration and the time interval selected is the morning rush hour. Intuition suggests that the groups should be strongly associated with the days of the week. In fact, 75% of Group 1 is composed of Saturdays, Sundays and holidays, while these days make up only 2% of Group 2(see FIGURE 8). Similar results are obtained for many (or all) the other detectors. This reinforces the intuition that forecasting of traffic or any related analysis should be conducted separately for working days and weekends and holidays. This is a straightforward conclusion from intuition. However, we could also think that Mondays or Fridays exhibit different patterns from the rest of the days of the week. Analyzing the sample proportions of the groups more deeply we also conclude that this is not the case, which is not such a direct conclusion.



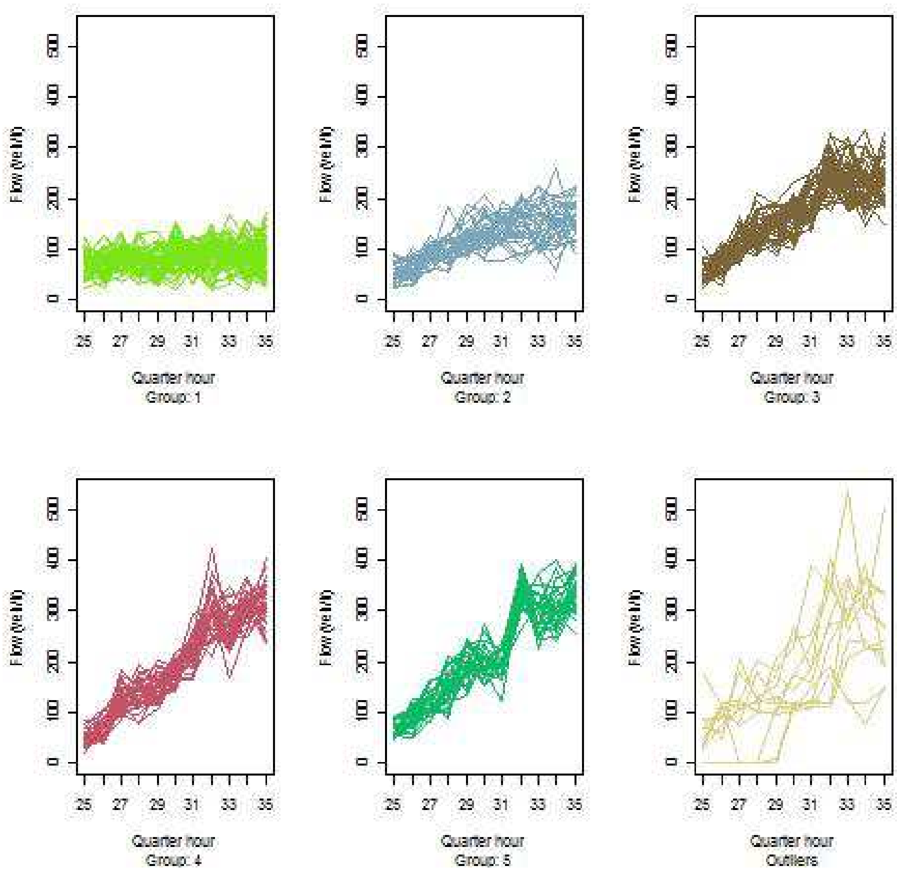
8. Groups 1 and 2

In FIGURE 9 we display the outliers set *a priori* at 5% of the sample. Some of them can be easily identified by eye from FIGURE 7, namely the observation with zeros from quarters 25-29, the outliers showing spikes at quarters 31 and 33 or the outlier with nearly 200 veh/h at quarter 25.

A more detailed analysis is based on greater numbers of groups, or on a number determined from the data. We applied the ICL criterion (see McLachlan and Peel (2000)) which is an AIC fashion criterion based on Bayesian statistics. In this example, 6 is the number of groups suggested by the ICL criterion (see FIGURE 10). More intricate patterns can be observed in the results. As in the earlier analysis, 73% of groups 1 and 2 are composed of weekends/holidays and these comprise only 2% of the rest of the groups.



9. Outlier detection



10. Patterns with 6 groups

4 CONCLUSIONS

Pattern identification has become a key step to better understanding of traffic behaviour. Here we perform this task via clustering, which is one of the main topics in the Data Mining field. We have seen that the EM algorithm adapts successfully to traffic data classification, combining a strong statistical basis with very intuitive results. We highlight its recent variation to outlier detection. The improper component technique has shown itself to be a suitable way to uncover outliers.

Section 2.3 has shown us the strength of pattern identification in forecasting. Clearly, ignoring clustering can lead us to false conclusions.

We want to emphasize how well suited mixture models are to the problem of traffic management in real time. Note that each group or pattern can be regarded as a different scenario. Knowing to which group a new observation belongs would be very useful when applying different policies. So, given a particular pattern identification (performed previously by the EM algorithm), a new vectorial observation can be easily classified using a minor modification of (2). In fact, for a new observation to be completely classified, we need the values of the observation over the whole time domain of the pattern, e.g. from quarter 25 to 35 in the Madrid example. However, this is not the case with real-time data. One option for overcoming this problem is to use the same formula (2), but only with the vector means and the variances-covariances matrix of the time intervals observed up until the moment. In this way, we obtain a probabilistic assignation of real- time data to each one of the scenarios.

We would like to mention two future research possibilities: The first is to analyse the influence of other variables such as weather or sport events to each one of the groups; the second is the use of these patterns as a primary step for forecasting.

5 ACKNOWLEDGMENTS

A.Ruiz de Villa would like to thank N. T. Longford (SNTL, Barcelona) for introducing him into the world of the EM algorithm and for all the great statistical advice.

6 REFERENCES

Castro-Neto M., Young-Seon Jeong, Myong-Kee Jeong, et al (2009). *Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions*. Expert Systems with Applications, 36, 6164-6173.

Chatfield, C. (2004) *The analysis of time series. An introduction*. Sixth edition. Chapman & Hall/CRC Texts in Statistical Science Series.

Chen, H., and S. Grant-Muller. (2001) *Use of Sequential Learning for Short-Term Traffic Flow Forecasting*. Transportation Research Part C, Vol. 9, No. 5, pp. 319–336.

Davis G. A., Niham N. L., Hamed M. M., and Jacobson L. N. (1991) *Adaptive forecasting of freeway traffic congestion*. Transportation Research Record, 1287, 29-33.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) *Maximum likelihood from incomplete data via the EM algorithm*. With discussion. J. Roy. Statist. Soc. Ser. B 39, no. 1, 1–38.

Dia, H.(2001) An Object-*Oriented Neural Network Approach to Short-Term Traffic Forecasting*. European Journal of Operation Research, Vol. 131, pp. 253–261.

Hamed M. M., H. R. Al-Masaeid and Z. M. Bani Said (1995) *Short-term prediction of traffic volume in urban arterials*, ASCE Journal of Transportation Engineering, 121(3), pp. 249–254.

Ishak S., and Alecsandru C. (2003) *Optimizing traffic prediction performance of Neural networks under various topological, input and traffic condition settings*. Transportation Research Board 82nd Annual Meeting, Washington D. C.

Kantz H., and Schreiber T. (1997) *Non-linear time series analysis*. Cambridge Non-linear Science: Series 7, Cambridge University Press.

Longford, N. T. and D'Urso, P. (2011) *Mixture models with an improper component*. J.Appl.Stat. 1360-0532.

McLachlan, G.J.and Krishnan, T. (2008) *The EM algorithm and extensions*. Second edition. Wiley Series in Probability and Statistics. Wiley-Interscience.

McLachlan, G. and Peel, D. (2000) *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience.

Ruiz N, Unnikrishnan A, Waller, T. (2007). *Integrated Traffic Simulation-Statistical Analysis Framework for Online Prediction of Freeway Travel Time*. Transportation Research Record: Journal of the Transportation Research Board. Volume 2039 / 2007 pp 24-31

Silverman, B. W. (1986) *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Smith, B. L., and M. J. Demetsky (1994) *Short-Term Traffic Flow Prediction:Neural Network Approach*. In Transportation Research Record 1453, TRB, National Research Council, Washington, D.C., 1994, pp. 98–104.

Smith, B. L. and M. J. Demetsky (1996) *Multiple-Interval Freeway Traffic Flow Forecasting*. Transport Research Record, 1554, pp. 136-141.

Stathopoulos A., and Karlaftis M. G. (2003) *A multivariate state-space approach for urban traffic flow* *modelling and prediction*. Transportation Research Part C, 11 (2), pages 121-135.

Stathopoulos A., Karlaftis M., and Dimitriou L. (2010).*Fuzzy Rule-Based System Approach to Combining Traffic Count Forecasts*. Transportation Research Record: Journal of the Transportation Research Board, No. 2183, Transportation Research Board of the National Academies, Washington,D.C., 2010, pp. 120–128.

Torday A., Salom J., Garcia D., Delgado M., Funes G. (2010) *Simulation-based Decision Support System for Real-Time Traffic Management*. Proc. 89th Annual Meeting of Transportation Research Board Annual Conference, Washington DC

Vlahogianni, E. I., J. C. Golias, and M. G. Karlaftis (2004) *Short-Term Traffic Forecasting: Overview of Objectives and Methods*. Transportation Reviews, Vol. 24, pp. 553-557.

Williams, B.M. (2001). *Multivariate vehicular traffic flow prediction-evaluation of ARIMAX Modeling*. Journal of the Transportation Research Board, No. 1776, TRB, National Research Council, Washington, D.C.,194-200.

Williams, B. M., P. K. Durvasula, and D. E. Brown (1998) *Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models.* Transportation Research Record, 1644, pp. 132-141.

Zhang Y.L., Xie Y.C. (2007). *Short-Term Freeway Traffic Volume Forecasting Using v-Support Vector Machines*. Journal of the Transportation Research Board, No. 2024, TRB, National Research Council, Washington, D.C., 92-99.

Zhang N, Zhang Y, Lu H. (2011) *Short-term Freeway Traffic Flow Prediction Combining Seasonal Autoregressive Integrated Moving Average and Support Vector Machines*. Proc. 90th annual meeting of transportation research board annual conference, Washington DC