# IDENTIFICATION OF REGRESSION MODELS APPLICATION IN TRAFFIC

**Pavel Dohnal**[1,2]

[1] *Department of Adaptive Systems*
*Institute of Information Theory and Automation*
*Academy of Sciences of the Czech Republic*
*Pod vodárenskou věží 4, 18208 Prague 8, Czech Republic*
[2] *Department of Control Engineering*
*Faculty of Electrical Engineering*
*Czech Technical University in Prague*
*Technická 2, 166 27 Prague 6, Czech Republic*
*E-mail:* `dohnalp@utia.cas.cz`

Abstract:
Cross-roads in cities are controlled by setting a proper green proportion and cycle length of the signal lights. Mostly, these signals are set according to a working plan whose changes during the day are given by a fixed schedule. Automatic feedback in selection of these plans can improve quality of the traffic significantly.
For the control, knowledge of traffic flow state incoming into the micro-region in the near future is necessary. Reliable and possibly multi-step prediction of the traffic flow can decide about practical success of such a feedback control.

Keywords: regression model; model order; intensity of traffic flow; prediction

## 1. INTRODUCTION

Majority of cities has problem with high density of transportation and roads are congested. This is the reason why the advanced traffic control is being applied. It uses state-space or internal models for description of reality. Ideally, they should be build as physical models. However, some features have no physical representation or they are too complex and must be modeled in another way. This is a good opportunity to use black-box input-output mixture model approach. It needs a lot of data, but we have them. Knowledge of the internal model structure is very important for a correct performance. If it is unknown, it has to be estimated.

Some of advanced traffic flow control systems use regression models to describe the traffic reality, in the longer horizon for example [1, 2]. The typical fixed daily, weekly and seasonal courses of traffic variables, like intensities, can be certainly observed and easily determined but the typical values can often radically change due to actual traffic conditions. Then, the modeling relying on the typical course of the variables can not be accurate.

## 2. PROBLEM DESCRIPTION

The control design for complex traffic systems has to be build by solving partial, mutually harmonized, subtasks. Among them, the prediction of the traffic state is of primary importance. Here, we demonstrate appropriateness of predicting traffic quantities based on a mixture model.

## 3. SOLUTIONS

### *3.1 Radical solution*

One of possible solutions of problem with high density of transportation is very radical, because it consists in rebuilding traffic area. This solution is not the best for typical Europe city. The reason is, that a lot of buildings cannot be destroyed for better traffic area.

### *3.2 Real feedback control*

Different solution is to use real feedback control [4]. micro-regions with lights are controlled by local controllers witch are used values from detectors (intensity and occupancy). micro-regions are coordinated by peoples on supervisor centers. In general the task is non trivial and practically.

**Prediction of urban traffic state:** Solution of urban transportation problems via reconstruction of the street network is expensive and very limited as it has to respect the existing urban conditions. Thus, the capacity of the network has to be efficiently exploited. This makes feedback traffic control utilizing the available traffic lights extremely important. Its design has to be able to predict future traffic state for a given regime of traffic lights. Only with such a model, the regime can be varied to increase the permeability of city cross-roads network. For it, mixture modeling seems to be suitable. Changes of daily and seasonal traffic indicate it clearly.

## 4. NOTATIONS, TERMS AND QUANTITIES

Controlled networks are split into micro-regions. They are logically self-contained transportation areas of several cross-roads with their adjoining roads. Their modeling and feedback control exploit data measured by detectors based on inductive electric coils placed under the road surface. Presence of a huge metallic object above the coil changes its magnetic properties and thus individual cars are detected. Each detector signalizes presence or absence of a car above it. From this signal, basic transportation quantities are evaluated:

- i.e. Occupancy $o$, which is defined as the portion (in %) of the time when the inspected place is occupied by cars.
- i.e. Intensity $q$ expressing the number of cars per hour.
- i.e. Density $\rho$, which is defined as the number of cars per kilometer of the traffic flow. Specifying an average length of cars, it is computed as a ratio of occupancy and average car length.

Intensity and density describe the traffic state at the detector position.

**Behavior and action of such micro-regions:** Behavior consists of all possible realizations, i.e. values of all quantities considered by the decision maker within the time span determined by the horizon of interest and related to the system.

Behavior Q = all measured, considered and chosen quantities.

Traffic intensity, road occupancy - measured.

Vehicle and flow speed - calculated and estimated.

Queue length - estimated.

Phase of crossroad control, length of light cycle, green ratio of cycle length - chosen $\Leftrightarrow$ $control\ actions$

**Daily periodicity:** The daily periodicity may be separated to daily phases. At night, the traffic intensity is very low. In the morning, the demands rapidly rise due to the cars of commuters and increased commercial transport. The intensity of the traffic reaches soon its maximum and a slight fall is observed around noon. Then, it rises again and the saturation lasts till the evening when the intensity starts to fall gradually to zero. The weekly periodicity connected with alternating work days and weekends is also strongly reflected in data.

**Quality marker:** Quality of models used as predictors is judged using the relative standard deviation $R_i$ of the prediction error $\hat{e}_{i;t} \equiv d_{i;t} - \hat{d}_{i;t}$ of $i$-th data entry $d_{i;t}$, $i$-th channel,

$$R_i \equiv \sqrt{\frac{\sum_t \hat{e}_{i;t}^2}{\sum_t (d_{i;t} - \bar{d}_i)^2}}, \quad i = 1, \ldots, d = 20, \ \bar{d}_i \equiv \text{sample mean of } d_{i;t}\ t. \tag{1}$$

In (**??**), the point prediction $\hat{d}_{i;t}$ is an approximate conditional expectation of $d_{i;t}$.

**Preprocessing of data:** Experience shows that detectors are fault-prone and their reparation is far from being easy. Therefore the filtration of their data is necessary. Here, outlier filtration and normalization to zero mean and unit standard deviation were applied on the raw data.

**Model for prediction:** The traffic system switches among several very different states. A mixture model has been chosen as a proper tool for modeling of this effect.

The model is built in the learning off line phase and used for prediction with another part of data. During the learning, the joint pdf of the modeled data record is estimated. Then, the pdf predicting selected channel(s) conditioned on other measured data is created. It is used for point or interval predictions of the channels of interest.

**Estimation of mixture models:** Both static and dynamic normal mixture models were used.

Experiments with the static case inspected the contribution of multivariate modeling and of the switching among various components.

Auto-regressions of the first order used in the dynamic case were found adequate for modeling and predicting 20 dimensional data records. Mixtures with components having higher order brought no improvements.

The initialization based on hierarchical factor splitting and implementing the normal version of Algorithm was used. It determined a finer structure of components and their number. Up to six components were found. In all cases, at least three of these components had non-negligible weights. Some components had almost zero weights but it is interesting that their omission decreased the prediction quality visibly. The rarely populated components described outlying data and consequently improved parameter estimates of the remaining components.

## 5. DECISION STEPS

### *5.1 Step 1: Learning*

**Algorithmic conversion of expert knowledge:** Car preservation law, projects of traffic networks, simulative models, traffic empiric rules.

**Data preprocessing:** Outliers missing measurements, dimension reduction, unification of differently measured data.

**Decreasing sets of descriptions f(Q):** Simultaneous structure estimation, parameters and internal quantities for normal and uniform models.

**Model validity verification**

### *5.2 Step 2: Design*

**Algorithmic conversion of objectives and constraints within ${}^{I}$f(Q):** Maximum throughput: weighted lengths of queue $\rightarrow{}^{I}$f(Q)

Multitop ${}^{I}$f(Q) $\Leftrightarrow Multipleobjectives.$

Uniform f $\Leftrightarrow constraints.$

**Optimal strategy design:** Uniform f, ${}^{I}$f : non-standard dynamical linear programming.

**Validity verification of design:** Theoretical and simulative analysis, comparison of reality and objectives even constraints.

**Implementation**

## 6. RESULTS

The mixture models predict better, even though the signal lights are preset into the mode preventing congestions inside the Strahov tunnel and thus limiting other traffic modes. The mixture model also utilize better the information from the additional entries of data records.

**Static models:** The static Auto-regression model is able to predict only the mean value of the whole data sample. The projected mixture, due to the data dependent weights of individual components, is able to follow even the ascending and descending trends of the data. With the increasing length of prediction the ability to follow data declines but the switching among components is clearly visible.

**Dynamic models:** They describe and thus predict data of this kind much better. For short-term predictions, projected high dimensional model does not dominate over the low dimensional one. For longer-term predictions, advantage of the high dimensional model is clearly visible as it includes more information about the traffic state.

## 7. CONCLUSIONS

In the Institute of Information Theory and Automation of the Academy of Science of the Czech Republic an algorithm [3,5] is developed which estimates parameters of the model, an optimal order of the regression, etc. The algorithm is tested on real traffic data samples and the overall model estimation is done.

## ACKNOWLEDGEMENTS

The research was partially supported by research center DAR, MŠMT 1M0572 and GA ČR 102/03/0049.

## REFERENCES

[1] J. Kratochvílová, I. Nagy: *Bibliographic Search for Optimization Methods of Signal Traffic Control*, Tech. Rep. 2081, ÚTIA AV ČR, Prague, 2003.

[2] J. Kratochvílová, I. Nagy: *Traffic control of microregion*, in *IFAC World Congress, Preprints*, IFAC, Ed. IFAC, Prague, 2005, accepted.

[3] L. Tesař, P. Nedoma, M. Novák: *Mixture learning script Jobcontrol (Program)*, ÚTIA AV ČR, Prague, 2004.

[4] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma, L. Tesař: *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.

[5] P. Nedoma, M. Kárný, I. Nagy, M. Valečková: *Mixtools - MATLAB Toolbox for Mixtures*, ÚTIA AV ČR, Prague, 2000.