MODELLING OF TRAFFIC INTENSITIES

K. Dedecius, I. Nagy, R. Hofman

Department of Adaptive Systems, Institute of Information Theory
and Automation Academy of Sciences of the Czech Republic

Faculty of Transportation Sciences, Czech Technical University in Prague

Abstract

The paper concerns the Bayesian modelling of traffic intensities measured in towns and cities. It proposes the use of a normal regressive model of low order, whose parameters (regression coefficients) are estimated using partial forgetting. This method allows to track the time varying parameters and respects the different variability rate of the absolute term modelling the mean value. The general theoretical concept of Bayesian modelling and parameter estimation is complemented with the theory related to the normal model *in lieu*.

1 Introduction

Statistical modelling of traffic intensities in the urban areas becomes a significant task [6]. Increasing road traffic is accompanied by a wide range of negative factors, influencing the environment (air pollution), local economy ("opportunity costs", fuel costs, wear of vehicles and roads), health and other domains, see, e.g. [7, 13, 15]. Obviously, there appear yet many other externalities. Couple of solutions to the issue appeared in the last years, e.g.

- Junction improvements, e.g., roundabout junctions which have become popular in the Czech Republic in the past decade.
- Reversible lanes with direction switching, separate lanes for specific users like local mass transport systems, etc.
- Urban planning with respect to the local traffic needs.
- Intelligent traffic systems (ITS) like variable traffic signs, high occupancy toll lanes, cordon zones with congestion pricing, variable speed limits, collision avoidance systems, dynamic light systems etc. [14].

Many of these modern intelligent systems use statistical modelling of intensities (and other variables). However, as regards this type of modelling, we can use complex models demanding special technical equipment (computers, information channels, etc.) on one hand or to use simpler models at the cost of limitations of the overall system performance on the other hand.

Modelling of traffic intensities may be evaluated with the autoregressive models of low order, but their basic forms can fail, simply due to the time-varying mean value of the signal (cf. Fig. 1). While, at night, the roads are almost empty, in the morning the intensity rapidly grows to reach its first daily peak. The second peak occurs in the afternoon, then the intensity usually slowly decreases. To solve this non-stationarity, we employ the absolute term and model the mean value with it.

Specific notation: $'$ denotes transposition, $f(a|b)$ is a conditional probability density function in which a random variable (and its realization) a is conditioned by a random variable b (or its realization). $\mathbb{E}[\cdot]$ denotes mean value of the argument. Time $t = 1, 2, \ldots$ is discrete. x^* denotes a set of x values (hypotheses, functions...). Furthermore, let us introduce the notation

$$f(x = x_t | y = y_t, \boldsymbol{d}(\tau)) \equiv f_{t|\tau} f(x|y).$$

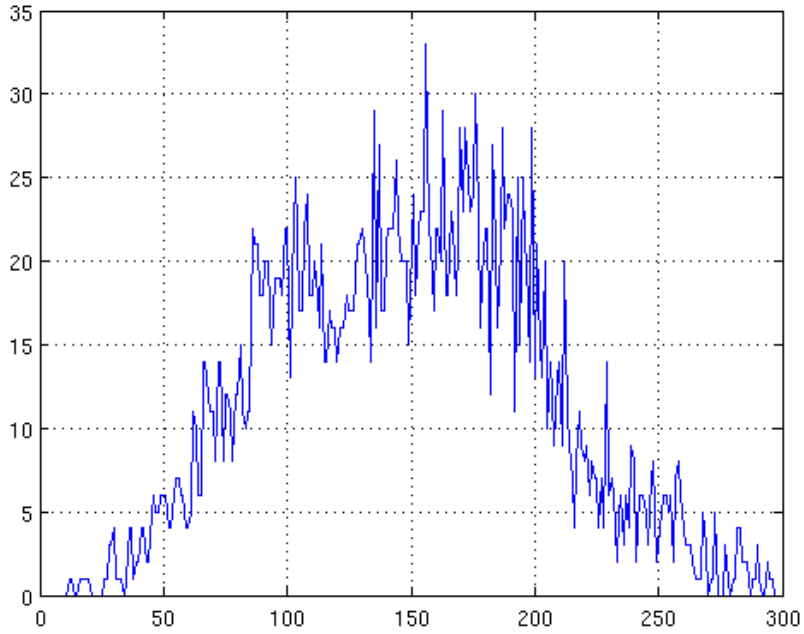

Figure 1: A sample of course of traffic intensities during 1 day

2 Normal regressive model

For the sake of convenience, we employ the n-th order autoregressive model AR(n) in the form

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + k_t + e_t, \qquad t = 1, 2, \ldots, \tag{1}$$

where y_t denotes traffic intensity measured at time instant t, a_i are regression coefficients and k_t denotes the absolute term of the model. Its purpose is to model the mean value of the signal. The term e_t stands for the normally distributed white noise with zero mean and constant variance σ^2 [12],

$$e_t \sim \mathcal{N}(0, \sigma^2). \tag{2}$$

Under the assumptions on noise whiteness, the regressive model (1) may be expressed with a probability density function (pdf) [8]

$$f_{t|t-1}(y|\boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{\psi}_t'\boldsymbol{\theta}_t, \sigma^2) \tag{3}$$

where $m = n + 2, \boldsymbol{\psi}_t \in \mathbb{R}^m$ and $\boldsymbol{\theta}_t \in \mathbb{R}^m$ denote a column regression vector and a vector of regression parameters,

$$\boldsymbol{\psi}_t = (y_t, \ldots, y_{t-n}, 1)' \quad \text{and} \quad \boldsymbol{\theta}_t = (a_{1,t}, \ldots, a_{n,t}, k_t)'.$$

The term $\boldsymbol{\Theta}_t$ is a set of model parameters, which in the case of the normal model (3) is $\boldsymbol{\Theta}_t = \{\boldsymbol{\theta}_t, \sigma^2\}$. In this work, we focus especially on the regression coefficients aggregated in $\boldsymbol{\theta}_t$. Under general conditions, it is possible to avoid using the absolute term k_t, however, it will play a fundamental role in the further reading.

2.1 Estimation

Suppose that the model (3) is known up to a set of parameters $\boldsymbol{\Theta}_t$, whose elements are to be estimated. The Bayesian paradigm, considering the parameters to be random variables, allows us to represent their distribution with a pdf

$$f_{t|t-1}(\boldsymbol{\Theta}) \equiv f(\boldsymbol{\Theta}_t|\boldsymbol{d}(t-1)). \tag{4}$$

Apparently, their distribution is conditionally dependent on the previous measurements, which are the only source of information available to the model.[1] Estimation of these parameters has two steps:

Data update: incorporating new gathered data into the distribution of parameters through the Bayes' theorem [8]:

$$f_{t|t}(\boldsymbol{\Theta}) = \frac{f_{t|t-1}(y|\boldsymbol{\Theta}) f_{t|t-1}(\boldsymbol{\Theta})}{\int_{\boldsymbol{\Theta}^*} f_{t|t-1}(y|\boldsymbol{\Theta}) f_{t|t-1}(\boldsymbol{\Theta}_t) \mathrm{d}\boldsymbol{\Theta}} \tag{5}$$

Time update: reflecting the (potential) time-variability of parameters in $\boldsymbol{\Theta}_{t+1}$ [8]:

$$f_{t+1|t}(\boldsymbol{\Theta}) = \int_{\boldsymbol{\Theta}^*} f(\boldsymbol{\Theta}_{t+1}|\boldsymbol{\Theta}_t, \boldsymbol{d}(t)) f_{t|t-1}(\boldsymbol{\Theta}) \mathrm{d}\boldsymbol{\Theta}. \tag{6}$$

Let us first analyse the time update procedure. If the parameters are constant, then the normal model $f(\boldsymbol{\Theta}_{t+1}|\boldsymbol{\Theta}_t, \boldsymbol{d}(t))$ is identical with the Dirac distribution. In this case, the integral in (6) represents an identity functional and

$$f_{t+1|t}(\boldsymbol{\Theta}) = f_{t|t}(\boldsymbol{\Theta}).$$

The consequences are obvious: (i) the time update may be omitted if the parameters are constant, and (ii) under the normality of the model and under the quadratic criterion, the constant parameters' point estimates are identical to the frequentists' ones obtained from the static linear regression.

The recursive Bayesian estimation exploits the fact, that given a conjugate prior distribution, the posterior is of the same type. The normal distribution, describing the model (3) is a member of the exponential family; it can be proved, that any member of this family, meeting certain conditions, possesses a conjugate counterpart. One of these conditions is the existence of a sufficient statistics [3], allowing to avoid working with a large set of data by their transformation into a set of smaller non-increasing dimension

$$f(a|\boldsymbol{d}(t)) = f(a|\boldsymbol{S}_t). \tag{7}$$

The single-output normal model (3) is conjugated with the Normal inverse-gamma $\mathcal{N}i\mathcal{G}(\boldsymbol{V}, \nu)$ prior. Its sufficient statistics are the number of degrees of freedom $\nu \in \mathbb{R}$, sometimes referred to as the counter, and the extended information matrix $\boldsymbol{V} \in \mathbb{R}^{m \times m}$. The data update rules (5) for these two statistics are [12]

$$V_{t|t} = V_{t|t-1} + \begin{pmatrix} y_t \\ \boldsymbol{\psi}_t \end{pmatrix} \begin{pmatrix} y_t \\ \boldsymbol{\psi}_t \end{pmatrix}' \tag{8}$$

$$\nu_{t|t} = \nu_{t|t-1} + 1 \tag{9}$$

It may be proven [12], that the estimator of regression coefficients $\boldsymbol{\theta}_t = (a_1, \ldots, a_n, k_t)'$ is

$$\hat{\boldsymbol{\theta}}_t = \begin{pmatrix} \hat{a}_{t;1} \\ \vdots \\ \hat{a}_n \\ \hat{k}_t \end{pmatrix} = \begin{pmatrix} V_{21} \\ \vdots \\ V_{m1} \end{pmatrix}' \begin{pmatrix} V_{22} & \ldots & V_{2m} \\ \vdots & \ddots & \vdots \\ V_{m2} & \ldots & V_{mm} \end{pmatrix}^{-1} \Bigg|_{t|t}. \tag{10}$$

This relation is equivalent to the recursive least-squares (RLS).

The direct use of statistics V can lead to numerical difficulties due to the inversion operation. Therefore, we prefer to use its factorized representation, characterizing alternative definition of

[1] We abstract from the expert information.

the $\mathcal{N}i\mathcal{G}$ pdf. The $\mathcal{N}i\mathcal{G}$ pdf with the decomposition $\boldsymbol{V} = \boldsymbol{L'DL}$, where $\boldsymbol{L}$ is a unit lower triangular matrix and $\boldsymbol{D}$ is a diagonal matrix, has the form [8]

$$\mathcal{G}i\mathcal{W}(\boldsymbol{L},\boldsymbol{D},\nu) \equiv \frac{\sigma^{-(\nu+n+2)}}{\mathcal{I}(\boldsymbol{L},\boldsymbol{D},\nu)} \times \exp\left\{\frac{-1}{2\sigma^2}\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\boldsymbol{C}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + D_y\right]\right\}.$$

With the generalization of $\boldsymbol{L}$ and $\boldsymbol{D}$ to block matrices of corresponding dimensions (D_y scalar)

$$\boldsymbol{L} = \begin{bmatrix} 1 & \\ \boldsymbol{L}_{y\psi} & \boldsymbol{L}_{\psi} \end{bmatrix},\ \boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_y & \\ & \boldsymbol{D}_{\psi} \end{bmatrix}$$

$\hat{\boldsymbol{\theta}} \equiv \boldsymbol{L}_{\psi}^{-1}\boldsymbol{L}_{y\psi}$ is the least-squares (LS) estimate of $\boldsymbol{\theta}$,

$\boldsymbol{C} \equiv \boldsymbol{L}_{\psi}^{-1}\boldsymbol{D}_{\psi}^{-1}(\boldsymbol{L}_{\psi}^{-1})' \in \mathbb{R}^{n\times n}$ is the LS covariance of $\hat{\boldsymbol{\theta}}$,

$D_y \in \mathbb{R}^+$ is the least squares remainder,

$\mathcal{I}$ stands for the normalization integral

$$\mathcal{I}(\boldsymbol{L},\boldsymbol{D},\nu) \equiv \Gamma(0.5\nu)\sqrt{\frac{2^{\nu}(2\pi)^n}{D_y^{\nu}|\boldsymbol{D}_{\psi}|}}. \tag{11}$$

More on properties of the distribution can be found in related literature, e.g., [8].

2.2 Prediction

Bayesian prediction with a parametric model follows from the rule

$$f_{t+1|t}(y) = \int_{\boldsymbol{\Theta}^*} f_{t+1|t}(y|\boldsymbol{\Theta}) f_{t|t}(\boldsymbol{\Theta}) d\boldsymbol{\Theta} = \frac{\mathcal{I}_{t+1}}{\mathcal{I}_t}. \tag{12}$$

Under the assumption of model normality and under the quadratic criterion

$$\sum_{t\in t^*}(y_t - \boldsymbol{\psi}_t'\hat{\boldsymbol{\theta}}_t)^2 \to \min$$

we may use the point estimates of $a_{1;t},\ldots,a_{n;t}$ and k_t to obtain the prediction of y_{t+1}. Generally, the knowledge of regression vector for any t allows us to evaluate the predictions relevant to this index. This is equivalent to multiple steps-ahead prediction, or smoothing if we regress some intermittent value.

3 Estimation with forgetting

It has already been mentioned in Section 2.1 that the potential parameters time variability often has to be taken into account. If we deal with traffic intensities, the time update becomes very important. The intensities vary during day and week, which can be expressed as the time variation of the mean value. Recall the model (1) and remind, that the mean value is modelled with the absolute term k. Hence the goal is to release the absolute term and let it vary with the true intensities.

If we further analyse the situation, the lack of knowledge of parameters' time evolution becomes evident. The only known data are the past intensity measurements and there is no other clue. In this case, we employ forgetting in place of the time update (6). Instead of explicit modelling of parameters' evolution in time, or finite data window modelling, we release the parameters by gradual discarding the old and potentially outdated information. There exist several forgetting methods, e.g., directional forgetting [10] or linear forgetting [11], however, the

most popular yet the most basic one is the exponential forgetting [5, 12]. For the Bayesian models, it is defined as follows

$$f_{t+1|t}(\boldsymbol{\Theta}) = \left[f_{t|t}(\boldsymbol{\Theta})\right]^{\alpha}; \qquad \alpha \in (0,1).$$

The term α stands for the forgetting factor; it is usually greater than 0.95. In the normal model (3), whose prior pdf is of normal inverse-gamma type, the forgetting demonstrates itself in the form

$$\boldsymbol{V}_{t+1|t} = \alpha \boldsymbol{V}_{t|t} = \alpha \boldsymbol{L}'_{t|t} \boldsymbol{D}_{t|t} \boldsymbol{L}_{t|t} \tag{13}$$

$$\nu_{t+1|t} = \alpha \nu_{t|t}. \tag{14}$$

3.1 Hypotheses of partial forgetting

The exponential forgetting is doomed to fail if used for modelling of dynamic systems with different variability of parameters, which becomes evident if we summarize the properties of traffic intensities:

1. In certain time intervals, e.g., during nights, probably no parameter varies.
2. In other time intervals all parameters vary slowly.
3. Generally, during the daytime, the mean value varies significantly.

Let these three cases label as hypotheses H_0, H_1 and H_2 and suppose, that at each time instant, the regression coefficients obey some true distribution $g_{t+1|t}(\boldsymbol{\theta})$. Now, we formalize the hypotheses as follows:

$$\begin{aligned} H_0 &: \mathbb{E}\left[g_{t+1|t}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{d}(t), H_0\right] = f_{t|t}(\boldsymbol{\theta}) \\ H_1 &: \mathbb{E}\left[g_{t+1|t}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{d}(t), H_1\right] = \left[f_{t|t}(\boldsymbol{\theta})\right]^{\alpha} \\ H_2 &: \mathbb{E}\left[g_{t+1|t}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{d}(t), H_2\right] = f_{t|t}(a_1, \ldots, a_n|, k)\left[f_{t|t}(k)\right]^{\alpha} \end{aligned} \tag{15}$$

where again $\alpha \in (0,1)$ and where $\mathbb{E}\left[g_{t+1|t}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{d}(t), H_i\right]$ has the meaning of a point estimate of the true but unknown pdf. It expresses our presumption of the true pdf under the knowledge of data $\boldsymbol{d}(t)$, parameters θ_{t+1} and the true hypothesis H_i at time t. The meaning of H_2 is simple – we decompose the pdf $f_{t|t}(\boldsymbol{\theta})$ using the chain rule and forget only the marginal pdf related to the absolute term. Each of the three hypotheses characterizes one specific case, but any of them can appear during the modelling. The conceptually correct solution is to use the mixture in which each pdf is weighted by its non-negative probability $p_{i,t|t} \leq 1$

$$\mathbb{E}\left[g_{t+1|t}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{d}(t)\right] = \sum_{i=0}^{2} p_{i,t+1|t} \mathbb{E}\left[g_{t+1|t}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{d}(t), H_i\right], \qquad \sum_{i=0}^{2} = 1. \tag{16}$$

3.2 Determination of probabilities

The probabilities $p_{i;t|t}$ quantify the uncertainty of each hypothesis at the particular time instant. If we want to recursively tune them, the data update and time update are necessary.

Data update tunes the probabilities with respect to the predictive abilities of the hypotheses,

$$p_{i;t|t} \propto p_{i;t|t-1} \int_{\boldsymbol{\Theta}^*} f_{t|t-1}(y|\boldsymbol{\Theta}) \mathbb{E}\left[f_{t|t-1}(\boldsymbol{\Theta}) \middle| \boldsymbol{\Theta}, H_i, \boldsymbol{d}(t)\right] \mathrm{d}\boldsymbol{\Theta}. \tag{17}$$

Time update suppresses the risk of degradation of the weights to a singular case, when right one hypothesis dominates with probability close to 1, while the others are close to zero. We can use the exponential forgetting with factor $\alpha \in (0,1)$, i.e.

$$p_{i;t+1|t} \propto p_{i;t|t}^{\alpha}, \tag{18}$$

3.3 Approximation

The mixture-based modelling requires a complex treatment, which discards its use for our purpose. To avoid it, we prefer to approximate the mixture (16) by a single pdf, using the Kullback-Leibler divergence [9] as a minimization criterion.

The Kullback-Leibler divergence of two pdfs f, g of a random variable X, acting on a common set X^*, holds the following form:

$$\mathrm{KL}(f||g) = \int_{X^*} f(x) \ln \frac{f(x)}{g(x)} \mathrm{d}x. \tag{19}$$

It can be shown, that the Kullback-Leibler divergence is a non-negative functional with equality for $f = g$ almost everywhere [1].

We search the argument minimizing the Kullback-Leibler divergence,

$$\tilde{g}_{t+1|t}(\boldsymbol{\Theta}) = \arg\min_{g \in g^\star_{t+1|t}} \mathbb{E}\left[\mathrm{KL}\left(g_{t+1|t}||\tilde{g}_{t+1|t}\right) \Big| \boldsymbol{\Theta}, \boldsymbol{d}(t)\right].$$

The pdf $\tilde{g}_{t+1|t}$ represents the best approximation of the mixture (16) and may be used for further modelling.

The Kullback-Leibler divergence of two $\mathcal{N}i\mathcal{G}$ pdfs has the following form [8]:

$$\begin{aligned}\mathrm{KL}(g||\tilde{g}) = \ln \frac{\Gamma(0.5\tilde{\nu})}{\Gamma(0.5\nu)} - 0.5 \ln |\boldsymbol{C}\tilde{\boldsymbol{C}}^{-1}| + 0.5\tilde{\nu} \ln \frac{D_y}{\tilde{D}_y} \\ + 0.5(\nu - \tilde{\nu})\Upsilon(0.5\nu) - 0.5n - 0.5\nu + 0.5\mathrm{Tr}\left(\boldsymbol{C}\tilde{\boldsymbol{C}}^{-1}\right) \\ + 0.5\frac{\nu}{D_y}\left[\left(\hat{\boldsymbol{\theta}} - \hat{\tilde{\boldsymbol{\theta}}}\right)' \tilde{\boldsymbol{C}}^{-1} \left(\hat{\boldsymbol{\theta}} - \hat{\tilde{\boldsymbol{\theta}}}\right) + \tilde{D}_y\right],\end{aligned} \tag{20}$$

where $\Upsilon(\cdot)$ denotes the digamma function, i.e., the first logarithmic derivative of the gamma function $\Gamma(\cdot)$.

In our application, we substitute the mixture obtained in (16) for $g_{t+1|t}$ and search for its best approximation $\tilde{g}_{t+1|t}$ by minimization of (20) with respect to the parameters of the $\mathcal{N}i\mathcal{G}$ distribution. These parameters are:

$$\hat{\tilde{\boldsymbol{\theta}}}_{t+1|t} = \left(\sum_{i=0}^{2} \lambda_{i;t+1|t} \frac{\nu_{i;t|t}}{D_{yi;t|t}}\right)^{-1} \left(\sum_{i=0}^{2} \lambda_{i;t+1|t} \frac{\nu_{i;t|t}}{D_{yi;t|t}} \hat{\boldsymbol{\theta}}_{i;t|t}\right)$$

$$\tilde{D}_{y;t+1|t} = \tilde{\nu}_{i;t|t} \left(\sum_{i=0}^{2} \lambda_{i;t+1|t} \frac{\nu_{i;t|t}}{D_{yi;t|t}}\right)^{-1}$$

$$\tilde{\boldsymbol{C}}_{t+t|t} = \sum_{i=0}^{2} \lambda_{i;t+1|t} \frac{\nu_{i;t|t}}{D_{yi;t|t}} \times \left[\left(\hat{\boldsymbol{\theta}}_{i;t|t} - \hat{\tilde{\boldsymbol{\theta}}}_{i;t|t}\right)\left(\hat{\boldsymbol{\theta}}_{i;t|t} - \hat{\tilde{\boldsymbol{\theta}}}_{i;t|t}\right)'\right] + \sum_{i=0}^{2} \lambda_{i;t+1|t} \boldsymbol{C}_{i;t|t}$$

$$\tilde{\nu}_{t+1|t} = \frac{1 + \sqrt{1 + \frac{4}{3}(A - \ln 2)}}{2(A - \ln 2)}$$

$$A = \ln\left(\sum_{i=0}^{2} \lambda_{i;t+1|t} \frac{\nu_{i;t|t}}{D_{yi;t|t}}\right) + \sum_{i=0}^{2} \lambda_{i;t+1|t} \ln D_{yi;t|t} - \sum_{i=0}^{2} \lambda_{i;t+1|t}\ \Upsilon(0.5\nu_{i;t|t}).$$

The proof can be found in [2]. A normal inverse-gamma distribution with these parameters may be used as the best approximation of the true parameters pdf in (3).

4 Practical implementation in Matlab

We use the Mixtools library developed at the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. This library allows to use regressive models and

their mixtures in Matlab. Here, we demonstrate the modelling of traffic intensities depicted in the Fig. 1. We use autoregressive model of first order with an absolute term, with non-informative prior with parameters $\mathrm{diag}V_0 = (0.1, 0.01, 0.01)$ and $\nu_0 = 10$. The forgetting factor α for H_1 is 0.95, for H_2 it is 0.9. The probabilities of hypotheses are flattened with $\alpha = 0.99$. The course of parameter estimates is depicted in the Figure 2. Evidently, the absolute term follows quite well the variations of the traffic intensity mean value. The Fig. 3 shows the course when the estimation was evaluated without forgetting. The one-step ahead prediction errors (partial forgetting) have mean -0.017, median 0.002 and standard deviation 3.673.

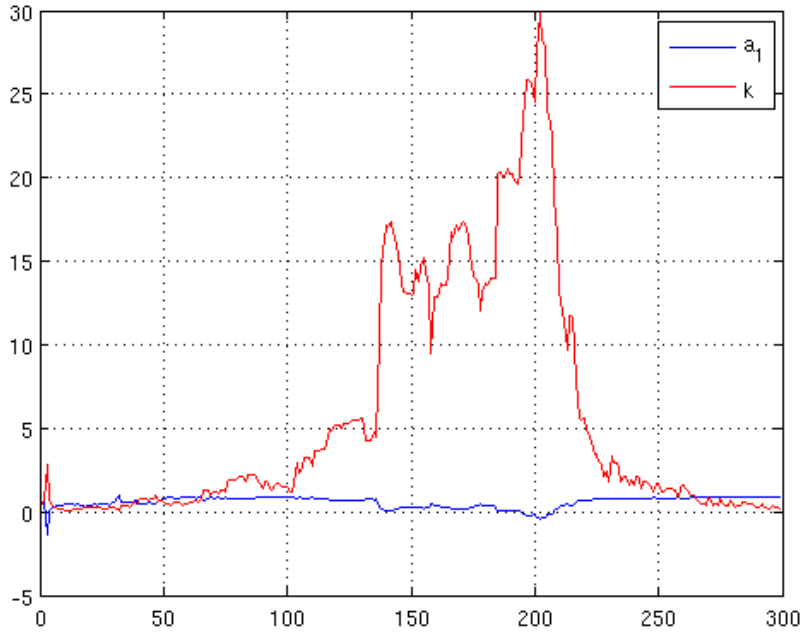

Figure 2: Evolution of parameter estimates (partial forgetting).

5 Conclusions

The paper described the Bayesian modelling of traffic intensities with low-order normal autoregressive models. As the parameters (regression coefficients) are supposed to vary with different rates, the use of partial forgetting method was proposed. The method was briefly described and the results were demonstrated in an example.

Acknowledgement

This project was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS 10/099/OHK3/1T/16 and the treaty between ÚTIA AV ČR, v.v.i. and Škoda Auto, a.s. ENS/2010/UTIA.

References

[1] Bernardo, J.M. (1979). *Expected Information as Expected Utility.* The Annals of Statistics, Vol. 7, No. 3, pp. 686–690.

[2] Dedecius, K. (2010). *Partial Forgetting in Bayesian Estimation.* PhD thesis. Czech Technical University in Prague.

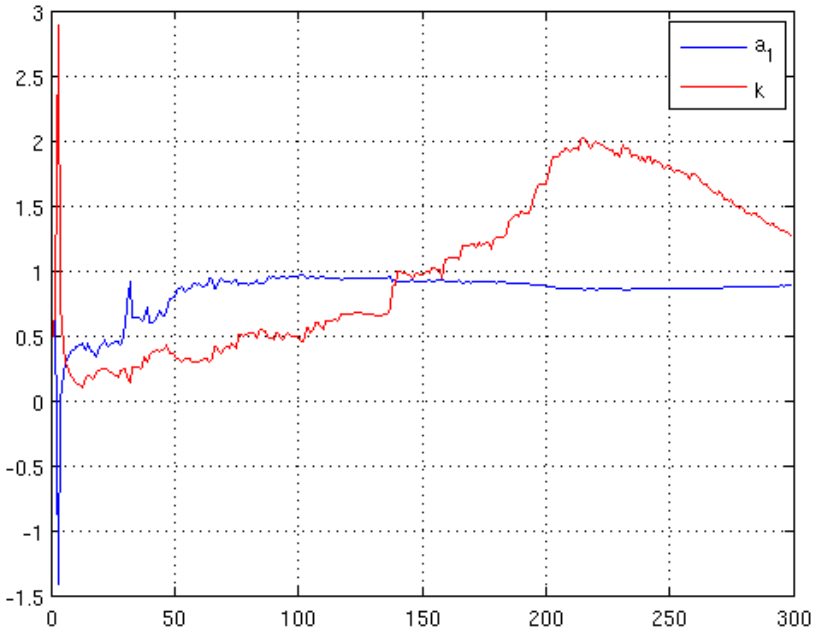

Figure 3: Evolution of parameter estimates (no forgetting).

[3] Fink, D. (1995). *A Compendium of Conjugate Priors*, Cornell University, Tech. Rep.

[4] Cao, L. & Schwartz, H. (2000). *Directional Forgetting Algorithm Based on the Decomposition of the Information Matrix*, Automatica, vol. 36, no. 11, pp. 1725–1731.

[5] Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

[6] Jing, L. and Wei, G. (2004). *A Summary of Traffic Flow Forecasting Methods*. Journal of Highway and Transportation Research and Development, vol. 3.

[7] Künzli, N. et al. (2000). *Public-health impact of outdoor and traffic-related air pollution: a European assessment*. The Lancet, vol. 356, no. 9232, pp. 795–801.

[8] Kárný, M. et al. (2005). *Optimized Bayesian Dynamic Advising*, Springer.

[9] Kullback, S. and Leibler, R.A. (1951). *On information and sufficiency*. Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79–86.

[10] Kulhavý, R. & Kárný, M. (1984). *Tracking of Slowly Varying Parameters by Directional Forgetting*, In Preprints of the 9th IFAC World Congress, Budapest, Vol. X, pp. 78–83.

[11] Kulhavý R. & Kraus, F.J. (1996). *On Duality of Regularized Exponential and Linear Forgetting*, Automatica, vol. 32/10, pp. 1403–1415.

[12] Peterka, V. (1981). *Bayesian Approach to System Identification*, in *Trends and Progress in System Identification*, P. Ekhoff, Ed., pp. 239–304. Pergamon Press, Oxford.

[13] Roorda-Knape, M.C. et al. (1998). *Air pollution from traffic in city districts near major motorways*. Atmospheric Environment, vol. 32, no. 11, pp. 1921–1930.

[14] Přibyl, P. and Svítek, M. (2001). *Inteligentní dopravní systémy*. BEN, Prague.

[15] Williams, I.D. and McCrae, I.S. (1995). *Road traffic nuisance in residential and commercial areas*. Science of the Total Environment, vol. 169, no. 1/3, pp. 75–82.

Kamil Dedecius
Faculty of Transportation Sciences, Czech Technical University in Prague. Na Florenci 25, 110 00, Prague 1.
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. Pod Vodárenskou věží 4, 182 08, Prague 8.

Ivan Nagy
Faculty of Transportation Sciences, Czech Technical University in Prague. Na Florenci 25, 110 00, Prague 1.
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. Pod Vodárenskou věží 4, 182 08, Prague 8.

Radek Hofman
Faculty of Transportation Sciences, Czech Technical University in Prague. Na Florenci 25, 110 00, Prague 1.
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. Pod Vodárenskou věží 4, 182 08, Prague 8.