INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2008; **00**:1–10 *Prepared using* acsauth.cls *[Version: 2002/11/11 v1.00]*

Prior knowledge processing for initial state of Kalman filter

E. Suzdaleva*

Department of Adaptive Systems, ÚTIA AV ČR, Pod vodárenskou věží 4, 18208 Prague, Czech Republic

SUMMARY

The paper deals with a specification of the prior distribution of the initial state for Kalman filter. The subjective prior knowledge, used in state estimation, can be highly uncertain. In practice, incorporation of prior knowledge contributes to a good start of the filter. The present paper proposes a methodology for selection of the initial state distribution, which enables eliciting of prior knowledge from the available expert information. The proposed methodology is based on the use of the conjugate prior distribution for models, belonging to the exponential family. The normal state-space model is used for demonstrating of the methodology. The paper covers processing of the prior knowledge for state estimation, available in the form of simulated data. Practical experiments demonstrate processing of prior knowledge from the urban traffic control area, which is the main application of the research. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: *Kalman filtering; prior knowledge; state-space model; initial state distribution*

*Correspondence to: ÚTIA AV ČR, Pod vodárenskou věží 4, 18208 Prague, Czech Republic

Contract/grant sponsor: GA ČR 201/06/P434

Received

 Revised

1. INTRODUCTION

The paper is devoted to the problem of selection of the initial state distribution for Kalman filtering. Kalman filter [1] can be well implemented due to good specification of the initial state and the assumed knowledge of variances of involved noises. The application area of the research is traffic control, where a lot of expert knowledge is usually available, for instance, precision of sensors, limited range of involved signals, past measurements of sensors, simulations etc. Translation of this knowledge into the respective probability density functions (pdf) and choice of the adequate initial conditions for Kalman filtering is the task, addressed in the paper.

Research in the field of prior knowledge processing is primarily concerned with the input-output models and directed at Bayesian estimation. The methodology, proposed in [2], used a complicated procedure of weighting of the prior knowledge pieces. Subsequently, the paper [3] presented a methodology of incorporation of external knowledge for parameter estimation. One of its potential applications is quantification of prior knowledge. The proposed approach [3] indicated a chance, that the underlying concept of fictitious data, used in [2], can be replaced by the more general, but less elaborated problem formulation as optimization under informational restriction. The work [4] extended a general idea of [3] and evolved a methodology of translation of the specific expert knowledge into the prior pdf, used by Bayesian state estimation. However, the paper [4] considered a special case of the state-space model, restricted by a single output.

The present paper continues a line, oriented at Bayesian state estimation and improves the methodology, described in [4], getting rid of most inconsistencies and inaccuracies. The methodology, proposed in the present paper, does not require restricting a dimension of the state-space model. Moreover, the present paper proposes a technique of processing of one of the prior knowledge types – simulated data, which brings additional specific problems.

The outline of the paper is as follows. Basic facts about the models used and Bayesian state estimation are provided in Section 2. The main emphasis of the paper is on Section 3. It describes a methodology of the prior knowledge elicitation, based on the usage of the conjugate prior distribution for models, belonging to the exponential family. The proposed methodology enables to obtain the *transformed prior pdf*, expressing the initial state, conditional on the available prior knowledge, provided by experts. Its application to Gaussian state-space model, described in Subsection 3.1, results in the mean and covariance matrix of the initial state with the incorporated processed prior knowledge. Section 4 is devoted to the processing of the prior knowledge, available in the form of past simulated data. It describes the problems, which this type of prior knowledge can cause, and specifies the methodology of the processing. Section 5 provides the illustrative experiments with the traffic control area simulated data. Remarks in Section 6 close the paper.

The version of the methodology, proposed in the present paper, assumes the model parameters and noise covariances to be known. The prior knowledge processing for the joint parameter-and-state Bayesian estimation, meanwhile, remains the open problem.

2. PRELIMINARIES

The probabilistic description of the system, which state is to be estimated, is provided by a state-space model in the form of the following pdfs.

2.1. State-space model

The *model of observation*

$$f\left(y_t|u_t, x_t\right), \tag{1}$$

relates the system output y_t to the system input u_t and system state x_t at discrete time moments $t \in t^* \equiv \{0, \ldots, \mathring{t}\}$, where $\mathring{t}$ is the cardinality of the set t^* and $\equiv$ means equivalence. The *model of state evolution*

$$f\left(x_t | u_t, x_{t-1}\right), \tag{2}$$

describes the evolution of the state x_t.

The *model of control strategy*

$$f\left(u_t \middle| d^{\,t-1}\right), \tag{3}$$

describes, generally randomized, generating of inputs u_t, based on $d^{\,t-1}$, where $d^{\,t} = (d_0, \ldots, d_{\mathring{t}})$ and $d_t \equiv (y_t, u_t)$.

Proposition 2.1 (Assumptions) *It is assumed that neither the output y_t nor the state x_t depend on the past data $d^{\,t-1}$, and control strategies ignore the unobservable system states. It means, that it holds*

$$f\left(y_t \middle| u_t, d^{\,t-1}, x_t\right) = f\left(y_t | u_t, x_t\right), \tag{4}$$

$$f\left(x_t \middle| u_t, d^{\,t-1}, x_{t-1}\right) = f\left(x_t | u_t, x_{t-1}\right), \tag{5}$$

$$f\left(u_t \middle| x_t, d^{\,t-1}\right) = f\left(u_t \middle| d^{\,t-1}\right). \tag{6}$$

The finite-dimensional system state has to be estimated and the system output has to be predicted. These operations call for application of Bayesian prediction and filtering.

2.2. Prediction and filtering

Bayesian predictor of the output is given by the formula

$$f\left(y_t \middle| u_t, d^{\,t-1}\right) = \int f\left(y_t | u_t, x_t\right) f\left(x_t \middle| u_t, d^{\,t-1}\right) dx_t. \tag{7}$$

Bayesian filtering, estimating the state x_t, includes the following coupled formulas.

Data updating

$$f\left(x_t\middle|d^{\,t}\right) = \frac{f\left(y_t|u_t,x_t\right)f\left(x_t\middle|u_t,d^{\,t-1}\right)}{f\left(y_t\middle|u_t,d^{\,t-1}\right)}, \quad (8)$$
$$\propto f\left(y_t|u_t,x_t\right)f\left(x_t\middle|u_t,d^{\,t-1}\right),$$

($\propto$ means proportionality) that incorporates the experience contained in the data $d^{\,t}$.

Time updating

$$f\left(x_{t+1}\middle|u_{t+1},d^{\,t}\right) = \int f\left(x_{t+1}|u_{t+1},x_t\right)f\left(x_t\middle|d^{\,t}\right)\,dx_t, \quad (9)$$

which fulfills the state prediction.

The filtering does not depend on the control strategy $\{f(u_t|d^{\,t-1})\}_{t\in t^*}$ but on the generated inputs only.

The application to Gaussian state-space model with Gaussian prior on x_0 and Gaussian observations provides Kalman filter. The prior pdf $f(x_0)$, that expresses the subjective prior knowledge on the initial state x_0, starts the recursions. The choice of the mentioned pdf is the main question of the paper.

3. PRIOR KNOWLEDGE ELICITATION

In traffic control, which is a target application domain of the work, the prior pdf $f(x_0)$ reflects uncertain knowledge about the initial length of a car queue (*system state*) on the intersection (*traffic system*). The length of the queue expresses a state of the transport network most adequately, but it is not directly observable and has to be estimated. Selection of the pdf, expressing the initial state, can be based on the prior knowledge about the intersection, provided by experts. The expert information can include specific traffic characteristics (a saturated flow of an intersection lane, a turn rate, time of the green light, measurements of the input and output sensors etc). To avoid a narrow specialization of the subject matter, the paper assumes, that the provided prior knowledge comprise the inputs and outputs of the system.

Let the prior knowledge, provided by experts, be described by the pdfs $f_{\tau^*} \equiv \{f(d^{\,\tau})\}_{\tau\in\tau^*}$, where τ^* is a finite set of time moments. The index τ emphasizes, that the quantities, denoted by it, are related to the prior (not current) knowledge. The observed values are denoted by index t.

The distribution of the initial state x_0 should be chosen, taking into account the provided prior knowledge f_{τ^*}. The methodology, proposed in [3], solves the analogous problem of incorporating of external knowledge for the case of parameter estimation. Modifying it for Bayesian *state* estimation, one can transform the prior (flat) pdf $f(x_0)$ into the following one

$$f(x_0|f_{\tau^*}) = \frac{f(x_0)\exp[\mathring{\tau}\Omega_{\tau^*}(x_0)]}{\int f(x_0)\exp[\mathring{\tau}\Omega_{\tau^*}(x_0)]\,dx_0}, \quad \text{with} \tag{10}$$

$$\Omega_{\tau^*}(x_0) = \frac{1}{\mathring{\tau}}\sum_{\tau\in\tau^*}\int f(d^{\,\tau})\ln[\mathcal{Z}(d^{\,\tau}|x_0)]\,dd^{\,\tau}, \tag{11}$$

where $\mathcal{Z}(d^{\,\tau}|x_0)$ relates to the local state-space model, and $\Omega_{\tau^*}(x_0)$ can be interpreted as the expectation of the logarithm of the state-space model with respect to the average pdf $\hat{f}(d^{\,\tau})$, representing the pdf f_{τ^*}

$$\hat{f}(d^{\,\tau}) = \frac{1}{\mathring{\tau}}\sum_{\tau\in\tau^*} f(d^{\,\tau}). \tag{12}$$

The relation (10) reflects the incorporation of the knowledge, contained in f_{τ^*}, into the distribution of the initial state. Mathematically it presents the extension of the ordinary Bayes rule [5] for processing the crisp values of data $\tilde{d}^{\,\tau}$, $\tau\in\tau^*$. Knowledge of these values is equivalent to $f_{\tau^*} \equiv \{f(d^{\,\tau}) = \delta(d^{\,\tau} - \tilde{d}^{\,\tau}) \equiv \text{Dirac delta on } \tilde{d}^{\,\tau}\}_{\tau\in\tau^*}$. Originally such a definition of the knowledge incorporation has been proposed in [3], where the detailed explanations can be found.

The forms of the function $\Omega_{\tau^*}(x_0)$ and the model $\mathcal{Z}(d^{\ \tau}|x_0)$ in (11) depend on the cardinality of the set τ^*, denoted by $\mathring{\tau}$. Recursively, the model $\mathcal{Z}(d^{\ \tau}|x_0)$ can be expressed as

$$\begin{aligned} \mathcal{Z}(d^{\ \tau}|x_0) &= \int f\left(y_\tau|\, x_\tau, u_\tau\right) f\left(x_\tau|\, u_\tau, x_{\tau-1}\right) f\left(x_{\tau-1}|\, u_{\tau-1}, x_{\tau-2}\right)\ldots \\ &\times\ \ f\left(x_{\tau-i}|\, u_{\tau-i}, x_0\right)\, dx_\tau \ldots\, dx_{\tau-i}\mathcal{Z}(d^{\ \tau-1}|x_0), \end{aligned} \tag{13}$$

where $i = \tau - 1$, $\tau = 1, \ldots, \mathring{\tau}$. For example, let f_{τ^*} include only the system input and output (y_τ, u_τ) with index $\tau = 0$, i.e. $f_{\tau^*} = f(y_0, u_0)$. In such a case, according to (11), the pdf (10) has a relatively simple form.

$$f(x_0|f(y_0,u_0)) \ \propto \ f(x_0)\exp\left[\underbrace{\int f(y_0,u_0)\ln[\underbrace{f\left(y_0|\, u_0, x_0\right)}_{\mathcal{Z}(d^{\ \tau}|x_0)}]\, dy_0 du_0}_{\Omega_{\tau^*}(x_0)}\right]. \tag{14}$$

The function $\mathcal{Z}(d^{\ \tau}|x_0) \equiv \mathcal{Z}(d^{\ 0}|x_0)$ in (14) relates to the local model, corresponding to $\tau = 0$. Since the only prior knowledge $f(y_0, u_0)$ is available, only the model of observation (1) should be used in (14). The situation becomes more complicated, when $f_{\tau^*} = f(y_0, y_1, u_0, u_1)$, i. e. $\tau = 1$. In this case the model is denoted by $\mathcal{Z}(d^{\ \tau}|x_0) \equiv \mathcal{Z}(d^{\ 1}|x_0)$ and requires the following modified form.

$$\begin{aligned} f\left(y_0, y_1, u_0, u_1|\, x_0\right) &= f\left(y_1, u_1|\, y_0, u_0, x_0\right) f\left(y_0, u_0|\, x_0\right), \\ &= \int f\left(y_1, x_1, u_1|\, y_0, u_0, x_0\right)\, dx_1 f\left(y_0, u_0|\, x_0\right), \\ &= \int f\left(y_1|\, x_1, u_1\right) f\left(x_1|\, u_1, x_0\right)\, dx_1 f\left(y_0, u_0|\, x_0\right), \\ &= \int f\left(y_1|\, x_1, u_1\right) f\left(x_1|\, u_1, x_0\right)\, dx_1 \underbrace{f\left(y_0|\, u_0, x_0\right)}_{\mathcal{Z}(d^{\ 0}|x_0)}, \end{aligned} \tag{15}$$

which is obtained with the help of operation of marginalization, chain rule [5], Proposition (2.1) and Dirac delta. All the pdfs in (15) are known from models (1-2). After integrating and some algebraic rearrangements the proposed form of $\mathcal{Z}(d^{\,1}|x_0)$ provides the initial state, conditional only on the prior knowledge $f(y_0, u_0, y_1, u_1)$, while the state x_1 is being integrated out. The pdf (10) for the case, when when the prior knowledge includes $f(y_0, y_1, u_0, u_1)$, takes the form

$$f(x_0|f(y_0, y_1, u_0, u_1)) \quad \propto \quad f(x_0) \exp\left[\int f(y_0, y_1, u_0, u_1) \ln[\mathcal{Z}(d^{\,1}|x_0)]\, dy_0 dy_1 du_0 du_1\right]. \tag{16}$$

The function $\Omega_{\tau^*}(x_0)$ has a simple form in the case, when the model (13) belongs to the exponential family. Gaussian model, which Kalman filter deals with, does belong to the exponential family. In this case the model (13) can be expressed in the form, proposed in [8]

$$\mathcal{Z}(d^{\,\tau}|x_0) = \mathcal{A}(x_0) \exp\left\langle \mathcal{B}(d^{\,\tau}), \mathcal{C}(x_0)\right\rangle, \tag{17}$$

where $\mathcal{A}(x_0)$ is a non-negative scalar function; $\mathcal{B}(d^{\,\tau})$ and $\mathcal{C}(x_0)$ are multivariate functions of compatible and finite dimensions; the functional $\langle \cdot, \cdot \rangle$ is linear in the first argument. For the model in the exponential family, the pdf (10) will get the following form [8]

$$f(x_0|f_{\tau^*}) \quad \propto \quad f(x_0)\mathcal{A}(x_0) \exp\left\langle \mathring{\tau} V, \mathcal{C}(x_0)\right\rangle, \tag{18}$$

where the array V is

$$V \quad \equiv \quad \frac{1}{\mathring{\tau}} \sum_{\tau \in \tau^*} \int f(d^{\,\tau}) \mathcal{B}(d^{\,\tau})\, dd^{\,\tau}. \tag{19}$$

As it has been said in Section 2, Kalman filter requires *Gaussian* prior $f(x_0)$, i.e. also belonging to the exponential family. Therefore, choosing the conjugate Gaussian prior $f(x_0)$ in (18), one can preserve the form of the exponential family for the transformed prior pdf $f(x_0|f_{\tau^*})$. The conjugate prior pdf can be chosen as

$$f(x_0) \quad = \quad \frac{\bar{\mathcal{A}}(x_0) \exp\left\langle \bar{V}, \mathcal{C}(x_0)\right\rangle}{\int \bar{\mathcal{A}}(x_0) \exp\left\langle \bar{V}, \mathcal{C}(x_0)\right\rangle dx_0}. \tag{20}$$

With such a conjugate prior the transformed pdf $f(x_0|\tau^*)$ in (18) keeps the exponential family form with recursively calculated array

$$V_\tau = V_{\tau-1} + \mathcal{B}(d^{\,\tau}),\ V_0 \equiv \bar{V} + \mathring{\tau}V, \tag{21}$$

naturally obtained via multiplication of (17) and (20).

The array V in (21) defines the pdf $f(x_0|\tau^*)$ in a "general" case of Kalman filter, when the variances are supposed to be known. The specification of the initial state distribution for the case with the unknown variances is not considered at the present paper and meanwhile related to the open problems. It could be expected, that the additional statistics – a degree of freedom – would be involved in the processing, and the other conjugate prior pdf would be used.

3.1. Prior knowledge elicitation with normal state-space model

The Gaussian models (1-2), used for demonstrating of the proposed methodology, are given by

$$\textit{observation model} \qquad y_t = Cx_t + Du_t + v_t, \tag{22}$$

$$\textit{state evolution model} \qquad x_{t+1} = Ax_t + Bu_{t+1} + \omega_{t+1}, \tag{23}$$

where x_t, y_t and u_t are the column vectors with dimensions $\mathring{x}$, $\mathring{y}$ and $\mathring{u}$ respectively; v_t is a measurement (Gaussian) noise with zero mean and covariance R_v; ω_t is a process (Gaussian) noise with zero mean and covariance R_w; A, B, C and D are known matrices of appropriate dimensions. The task is to incorporate the prior knowledge into the distribution of the initial state x_0 in (22-23).

Examples of application of the proposed methodology to Gaussian models correspond to the usage of the model $\mathcal{Z}(d^{\,\tau}|x_0)$ in relations (14), (15) and (13).

3.1.1. Example 1: available prior knowledge $f(y_0, u_0)$ In the case, when the only available prior knowledge is $f_{\tau^*} = f(y_0, u_0)$, the following sequence of calculations is necessary. As it was noted in Section 3, the model $\mathcal{Z}(d^{\ \tau}|x_0)$ should be treated only as the observation model (1), and respectively (22). It means, that being Gaussian one and according to (14) and (22), the model is

$$\begin{aligned}
\mathcal{Z}(d^{\ \tau}|x_0) &= f\left(y_0|\, u_0, x_0\right), \\
&= \underbrace{(2\pi)^{-\frac{\mathring{y}}{2}}|R_v|^{-0.5}}_{\mathcal{A}_{z0}(x_0)} \exp\left\{-\frac{1}{2}[y_0 - Du_0 - Cx_0]' R_v^{-1} [y_0 - Du_0 - Cx_0]\right\}, \\
&= \mathcal{A}_{z0}(x_0) \exp\left\{ \overbrace{-\frac{1}{2} tr\left(\underbrace{\xi_0' R_v'^{-1} \xi_0}_{\mathcal{B}(d^{\ 0})} \underbrace{[1 \ \ x_0']'[1 \ \ x_0']}_{\mathcal{C}(x_0)} \right)}^{<\mathcal{B}(d^{\ 0}), \mathcal{C}(x_0)>} \right\},
\end{aligned} \tag{24}$$

where $\mathcal{A}_{z0}(x_0)$ corresponds to $\mathcal{A}(x_0)$ in (17), $\xi_0 = [d^0 \quad - C]$, $d^0 = y_0 - Du_0$, tr is a trace of the matrix. The calculation is proved by the straightforward multiplication of components inside the exponent in (24) and properties of the inner product.

The conjugate Gaussian prior pdf (20) with the mean (column) vector μ and covariance R_x is chosen as follows.

$$f(x_0) = \underbrace{(2\pi)^{-\mathring{x}}|R_x|}_{\bar{\mathcal{A}}(x_0)} \exp\left\{-\frac{1}{2}\left[x_0 - \mu\right]' R_x^{-1} \left[x_0 - \mu\right]\right\}, \tag{25}$$

$$= \bar{\mathcal{A}}(x_0) \exp\left\{-\frac{1}{2} tr\left(\underbrace{M' R_x'^{-1} M}_{\bar{V}} [1 \ \ x_0']'[1 \ \ x_0'] \right)\right\}, \tag{26}$$

where M is a matrix of the form $[-\mu \quad I]$ with I as the unit matrix of dimension $(\mathring{x} \times \mathring{x})$. The values of the mean μ and covariance R_x are provided by experts from the application domain. In the considered domain they mostly present the physical values, that can be used directly.

With such a conjugate Gaussian prior, the transformed prior pdf $f\left(x_0|\, \tau^*\right)$ also remains a

Gaussian one and preserves its form with the following array V_0, corresponding to (21),

$$V_0 = \bar{V} + \underbrace{\xi_0' {R'}_v^{-1} \xi_0}_{\mathcal{B}(d\ ^0)}, \text{ and} \tag{27}$$

$$\mathcal{A}_0(x_0) = \mathcal{A}_{z0}(x_0)\bar{\mathcal{A}}(x_0), \tag{28}$$

and being obtained with the help of multiplication of (26) and (24). The proposed array (27) defines the update of the initial state x_0, incorporating the prior knowledge, contained in $f(y_0, u_0)$. In order to specify, in fact, the mean and covariance matrix of the initial state, the array (matrix) V_0 has to be partitioned as $[V_{1(0)} \;\; V'_{2(0)}; \;\; V_{2(0)} \;\; V_{3(0)}]$, where $V_{1(0)}$ and $V_{3(0)}$ are square matrices. The partition is naturally obtained by the straightforward calculation of (27). Such a partition of the matrix V_0 enables to apply Lemma 3 about the completion of squares [5] to the quadratic form inside the exponent in the relation, obtained according to (18).

$$f(x_0|f_{\tau^*}) \propto \mathcal{A}_0(x_0) \exp\left\{-\frac{1}{2} tr\left(V_0[1 \;\; x_0']'[1 \;\; x_0']\right)\right\}. \tag{29}$$

According to Lemma 3 [5], the completion of squares for x_0 in (29) gives the following result.

$$\begin{aligned} & tr\left([1 \;\; x_0'] \begin{bmatrix} V_{1(0)} & V'_{2(0)} \\ V_{2(0)} & V_{3(0)} \end{bmatrix} [1 \;\; x_0']'\right), \\ = \;& tr\left([x_0 + V_{3(0)}^{-1} V_{2(0)}]' V_{3(0)} [x_0 + V_{3(0)}^{-1} V_{2(0)}]\right) + tr\left(V_{1(0)} - V'_{2(0)} V_{3(0)}^{-1} V_{2(0)}\right), \end{aligned} \tag{30}$$

which gives the following Gaussian distribution for $f\left(x_0 | \tau^*\right)$

$$\textit{covariance matrix} \qquad \hat{P}_{0(0)} = V_{3(0)}^{-1}, \tag{31}$$

$$\textit{mean} \qquad \hat{x}_{0(0)} = -V_{3(0)}^{-1} V_{2(0)}. \tag{32}$$

3.1.2. Example 2: available prior knowledge $f(y_0, y_1, u_0, u_1)$ In the case, when the available prior knowledge contains $f_{\tau^*} = f(y_0, y_1, u_0, u_1)$, the model $\mathcal{Z}(d\ ^{\tau}|x_0)$ is given by the pdf (15).

Being a Gaussian one, it takes the following form, according to (22), (23) and (24).

$$
\begin{aligned}
\mathcal{Z}(d^{\,\tau}|x_0) &= f\left(y_0, y_1, u_0, u_1 | x_0\right), \\
&= \int (2\pi)^{-\frac{\mathring{y}}{2}} |R_v|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[y_1 - Du_1 - Cx_1]' R_v^{-1} [y_1 - Du_1 - Cx_1]\right\} \\
&\times \; (2\pi)^{-\frac{\mathring{x}}{2}} |R_\omega|^{-\frac{1}{2}} \\
&\times \; \exp\left\{-\frac{1}{2}[x_1 - Ax_0 - Bu_1]' R_\omega^{-1} [x_1 - Ax_0 - Bu_1]\right\} dx_1 \\
&\times \; \mathcal{A}_{z0}(x_0) \exp\left\{-\frac{1}{2} tr\left(\xi_0' R'^{-1}_v \xi_0 [1 \;\; x_0']' [1 \;\; x_0']\right)\right\}. \qquad (33)
\end{aligned}
$$

In order to integrate (33) over x_1, it is necessary to rearrange the quadratic forms inside the exponents in the integral and fulfill the completion of squares for x_1. After that and applying the matrix inversion lemma [5], two quadratic forms inside the integral are modified as

$$
\begin{aligned}
& [y_1 - Du_1 - Cx_1]' R_v^{-1} [y_1 - Du_1 - Cx_1] + [x_1 - Ax_0 - Bu_1]' R_\omega^{-1} [x_1 - Ax_0 - Bu_1] \\
= \; & [x_1 - \hat{x}_1]' \hat{R}_{x(1)}^{-1} [x_1 - \hat{x}_1] + [y_1 - \hat{y}_1]' \hat{R}_y^{-1} [y_1 - \hat{y}_1], \qquad (34)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{R}_{x(1)}^{-1} &= R_\omega^{-1} + C' R_v^{-1} C, &(35)\\
\hat{x}_1 &= \hat{R}_{x(1)} [R_\omega^{-1}(Ax_0 + Bu_1) + C' R_v^{-1}(y_1 - Du_1)], &(36)\\
\hat{R}_y &= R_v + C R_\omega C', &(37)\\
\hat{y}_1 &= Du_1 + C(Ax_0 + Bu_1). &(38)
\end{aligned}
$$

Substituting quadratic forms (34) into the model (33) and integrating (33), one obtains the following form of the model $\mathcal{Z}(d^{\,\tau}|x_0)$.

$$\begin{aligned}\mathcal{Z}(d^{\,\tau}|x_0) &= f\left(y_0, y_1, u_0, u_1|\, x_0\right), \\ &= \mathcal{A}_{z1}(x_0)\exp\left\{-\frac{1}{2}[y_1-\hat{y}_1]'\hat{R}_y^{-1}[y_1-\hat{y}_1]\right\} \\ &\times \ \exp\left\{-\frac{1}{2}tr\left(\xi_0' {R'}_v^{-1}\xi_0[1 \;\; x_0']'[1 \;\; x_0']\right)\right\}, \\ &= \mathcal{A}_{z1}(x_0)\exp\left\{-\frac{1}{2}tr\left(\underbrace{(\xi_1'(\hat{R}_y^{-1})'\xi_1}_{\mathcal{B}(d^{\,1})}+\mathcal{B}(d^{\,0}))\mathcal{C}(x_0)\right)\right\},\end{aligned} \tag{39}$$

where

$$\mathcal{A}_{z1}(x_0) = (2\pi)^{-\frac{\mathring{y}}{2}}|R_v|^{-\frac{1}{2}}|\hat{R}_{x(1)}^{-1}|^{-\frac{1}{2}}|R_\omega|^{-\frac{1}{2}}\mathcal{A}_{z0}(x_0), \tag{40}$$

$$\xi_1 = [d^1 \quad -CA], \tag{41}$$

$$d^1 = y_1 - Du_1 - CBu_1. \tag{42}$$

The conjugate Gaussian prior pdf is already chosen in (25). With such a conjugate prior and model (39), the pdf $f\left(x_0|\,\tau^*\right)$ preserves the form of the exponential family (18) with the following array V_1

$$V_1 = \underbrace{\bar{V}+\mathcal{B}(d^{\,0})}_{V_0}+\mathcal{B}(d^{\,1}), \tag{43}$$

which is obtained through multiplication of (39) and (26). The scalar function $\mathcal{A}_1(x_0)$ is defined by the product of $\mathcal{A}_{z1}(x_0)$ and $\bar{\mathcal{A}}(x_0)$.

The normal distribution of the initial state x_0 is obtained similarly to the previous example in Subsubsection 3.1.1, i.e. with the help of partition of array $V_1 = [V_{1(1)} \;\; V_{2(1)}'; \;\; V_{2(1)} \;\; V_{3(1)}]$, where $V_{1(1)}$ and $V_{3(1)}$ are square matrices. The covariance matrix and mean of x_0 are calculated as

$$\text{covariance matrix} \qquad \hat{P}_{0(1)} = V_{3(1)}^{-1}, \tag{44}$$

$$\text{mean} \qquad \hat{x}_{0(1)} = -V_{3(1)}^{-1} V_{2(1)}. \tag{45}$$

3.1.3. Recursive prior knowledge processing for normal state-space model When the available prior knowledge includes the finite set $f_{\tau^*} \equiv \{f(d^{\,\tau})\}_{\tau \in \tau^*}$, unrestricted unlikely two previous examples, the calculating of the initial state distribution should be expressed recursively. The model $\mathcal{Z}(d^{\,\tau}|x_0)$ is defined according to (13) and takes the form, depending on cardinality $\mathring{\tau}$ of the set f_{τ^*}.

$$\mathcal{Z}(d^{\,\tau}|x_0) = \mathcal{A}_{z\tau}(x_0) \exp\left\{-\frac{1}{2} tr\left\{(\mathcal{B}(d^{\,\tau}) + \mathcal{B}(d^{\,\tau-1}) + \ldots + \mathcal{B}(d^{\,0}))\mathcal{C}(x_0)\right\}\right\}, \tag{46}$$

where

$$\mathcal{A}_{z\tau}(x_0) = (2\pi)^{-\frac{\mathring{y}}{2}} |R_v|^{-\frac{1}{2}} \ldots \mathcal{A}_{z0}(x_0), \tag{47}$$

$$\mathcal{B}(d^{\,\tau}) = \xi'_\tau (\tilde{R}_\tau^{-1})' \xi_\tau, \tag{48}$$

$$\xi_\tau = [d^{\,\tau} \quad - \tilde{C}_\tau], \tag{49}$$

$$d^{\,\tau} = y_\tau - \tilde{D}_\tau u_\tau, \tag{50}$$

and with $\tilde{R}_\tau$, $\tilde{C}_\tau$ and $\tilde{D}_\tau$, resulted during calculating, subject to $\mathring{\tau}$. The formula (10) or, more precisely, (18) is applied with the conjugate Gaussian prior, chosen in (25). In this way, the form of the transformed prior Gaussian pdf $f(x_0|\tau^*)$ is preserved and obtained by multiplication of (46) and (26). The calculation of $f(x_0|\tau^*)$ includes

$$V_\tau = \underbrace{\bar{V} + \mathcal{B}(d^{\,0}) + \ldots + \mathcal{B}(d^{\,\tau-1})}_{V_{\tau-1}} + \mathcal{B}(d^{\,\tau}), \quad (51)$$

$$V_\tau = \begin{bmatrix} V_{1(\tau)} & V'_{2(\tau)} \\ V_{2(\tau)} & V_{3(\tau)} \end{bmatrix}, \quad (52)$$

$$\mathcal{A}_\tau(x_0) = \mathcal{A}_{z\tau}(x_0)\bar{\mathcal{A}}(x_0). \quad (53)$$

where $V_{1(\tau)}$ and $V_{3(\tau)}$ are square matrices. Finally, the initial state distribution is obtained as

$$\textit{covariance matrix} \qquad \hat{P}_{0(\tau)} = V^{-1}_{3(\tau)}, \quad (54)$$

$$\textit{mean} \qquad \hat{x}_{0(\tau)} = -V^{-1}_{3(\tau)} V_{2(\tau)}. \quad (55)$$

4. PRIOR KNOWLEDGE IN THE FORM OF SIMULATED DATA

In the traffic control area the prior knowledge can be often available in the form of past (historical) data, mostly obtained by means of simulation or observed on a similar traffic system. This section is focused on processing of simulated data. In practice, this type of knowledge can cause an additional problem of predominance of the prior knowledge over observed values. It means, that the observed data would not be able much to influence the final result of the filtering. The problem is not critical, when the observed data are informative, and the number of simulated ones $\mathring{\tau}$ is not too large. In that case the formula (10) and, respectively, (18), can be applied directly. The predominance of prior knowledge occurs, when the number of simulated data is large. To avoid it, the prior knowledge must include only the most representative simulated data $\{d^{\,\tau}\}_{\tau\in\tau^*}$. The representative data are understood as the "closest" to the observed data $d^{\,t}$. Selection of the representative data from the simulated set is based on the *just-in-time modelling* technique [6], which measures the "closeness" of pdfs of the past simulated and real-time observed values. Adapted to the normal state-space model (22-23), the technique of the representative data selection requires the following calculations.

The real-time system data $d^{\,t} = d^{\,0}$ start to be observed at the time moment $t = 0$, $t \in t^* \equiv \{0, \ldots, \mathring{t}\}$. The experience, contained in the data $d^{\,0}$, is incorporated into the prior distribution $f(x_0)$ of the initial state, defined in (25). The resulted pdf $f\left(x_0 \middle| d^{\,0}\right)$ is obtained with the help of the data updating (9), applied to Gaussian prior (25) and Gaussian models (22-23), i.e. by the data updating of Kalman filter. The updated distribution $f\left(x_0 \middle| d^{\,0}\right)$ includes the following mean and covariance matrix.

$$\textit{mean} \qquad \hat{x}_0 = \mu + K_0(y_0 - C\mu - Du_0), \tag{56}$$

$$\textit{covariance matrix} \qquad \hat{P}_0 = R_x - K_0 C R_x, \tag{57}$$

where

$$K_0 = R_x C'(C R_x C' + R_v)^{-1}, \tag{58}$$

is the Kalman gain.

The simulated data $\{d^{\,\tau}\}_{\tau \in \tau^*}$ are considered to be "close" to the observed data $d^{\,0}$, if the distribution of the initial state x_0, updated by the data $\{d^{\,\tau}\}_{\tau \in \tau^*}$ is close to the distribution $f\left(x_0 \middle| d^{\,0}\right)$. Incorporation of the experience, contained in $\{d^{\,\tau}\}_{\tau \in \tau^*}$, into prior distribution (25) results in the Kalman filter data updating, evolved similarly to (56), (57) and (58). The resulted mean and covariance matrix of the state x_0, updated by simulated data $\{d^{\,\tau}\}_{\tau \in \tau^*}$ are as follows.

$$\textit{mean} \qquad \hat{x}_{0(\tau)} = \mu + K_{0(\tau)}(y_\tau - C\mu - Du_\tau), \tag{59}$$

$$\textit{covariance matrix} \qquad \hat{P}_{0(\tau)} = R_x - K_{0(\tau)} C R_x, \tag{60}$$

with the Kalman gain

$$K_{0(\tau)} = R_x C'(C R_x C' + R_v)^{-1}. \tag{61}$$

The "closeness" of Gaussian distributions, defined by (56-57) and (59-60) can be measured by the Kullback-Leibler divergence [7], which takes the following form [8]

$$\begin{aligned} \mathcal{D}_{t\tau} &= \frac{1}{2}\left[\ln|\hat{P}_{0(\tau)}\hat{P}_0^{-1}| - \mathring{x} + tr[\hat{P}_0\hat{P}_{0(\tau)}^{-1}] + (\hat{x}_0 - \hat{x}_{0(\tau)})'\hat{P}_{0(\tau)}^{-1}(\hat{x}_0 - \hat{x}_{0(\tau)})\right], & (62)\\ &= \frac{1}{2}\left[(\hat{x}_0 - \hat{x}_{0(\tau)})'\hat{P}_{0(\tau)}^{-1}(\hat{x}_0 - \hat{x}_{0(\tau)})\right]. & (63) \end{aligned}$$

The result (63) is obtained according to the elementary properties of the logarithm and the matrix trace. The condition $\mathcal{D}_{t\tau} \leq \mathring{x}$ is taken as the criterion of "closeness" of distributions (56-57) and (59-60). It means, that if $\mathcal{D}_{t\tau} \leq \mathring{x}$, the simulated data $\{d^{\,\tau}\}_{\tau\in\tau^*}$, involved in (59-60), are "close" to the observed data $d^{\,0}$ and are selected as the most representative data for the prior knowledge processing. The model $\mathcal{Z}(d^{\,\tau}|x_0)$ is built just-in-time, fitted to the selected data $\{d^{\,\tau}\}_{\tau\in\tau^*}$, and the formula (10) or, respectively, (18), is applied to the reduced set f_{τ^*}.

The reducing of the number of the simulated data can be done in the alternative way by weighting of the prior knowledge quantities. However, the present paper is focused on the proposed methodology of the representative data selection.

5. EXPERIMENTS

The prior knowledge, used for illustrative experiments, is provided by the traffic microsimulator AIMSUN [9]. The simulated traffic system represents the intersection with four arms, each with one input and one output lane. Each lane is equipped by a measuring detector. The input detectors are placed about 100 meters before the stop line at the input lane of the intersection arm, and the output detectors – behind the stop line at the output lane. The detectors measure the following quantities: intensity, expressing a number of the cars, passing through an intersection lane per hour $[c/h]$, and occupancy, reflecting a proportion of a time period of activating the detector by cars [%]. According to [10], the state-space model (22-23) is specialized to the traffic control area in the following way.

$$\textit{observation model} \qquad y_t = C_t x_t + D_t u_t + v_t, \tag{64}$$

$$\textit{state evolution model} \qquad x_{t+1} = A_t x_t + B_t u_{t+1} + F_t + \omega_t. \tag{65}$$

The system output y_t in (64) relates to the vector $[Y_t, O_t]'$, where Y_t is the (column) vector of output intensities, provided by the output detectors of the intersection lanes, i.e. $Y_t = [y_{1;t}, \ldots, y_{n;t}]'$, $n = 4$ is a number of lanes (identical to the number of arms for the given system), and $O_t = [o_{1;t}, \ldots, o_{4;t}]'$ is the vector of occupancies of the output detectors. The state x_t expresses the length of the car queue at the intersection lanes in cars $[c]$. One car is supposed to have about 6 meters. The queue length is not directly observed and has to be estimated. The state vector x_t relates to $\left[\xi_t, O_t^I\right]'$, where $\xi_t = [\xi_{1;t}, \ldots, \xi_{4;t}]'$ is a queue length to be estimated. The input occupancy $O_t^I = [o_{1;t}^I, \ldots, o_{4;t}^I]'$ is added to the state vector in order to ensure the observability of the model [12]. It means, that the utilization of the occupancy in the model expresses its proportionality to the length of the queue. In general, the occupancy in the state vector x_{t+1} can be also estimated, but the present paper is focused on estimation of the length of the queue.

The scope of the paper does not allow to describe the specific features of the traffic control in details, but the physical interpretation of the state (a car queue length) evolution in (65) can be explained. According to [10, 11, 12], the general idea of the car queue length evolution lies in the statement, that the queue length at the i-th intersection lane is equal to the previous queue plus arrived cars minus departed cars. It can be expressed in the following way.

$$\xi_{i;t+1} = \delta_{i;t}\xi_{i;t} - \underbrace{[\delta_{i;t}S_i + (1-\delta_{i;t})I_{i;t}]u_t}_{departed\ cars} + \underbrace{I_{i;t},}_{arrived} \ , i = \{1,\ldots,n\}, \tag{66}$$

where S_i is the known saturated flow of the i-th lane (the maximal number of cars, which can pass through the lane per hour in the case of the green light); $I_{i;t}$ is the input intensity; u_t is the time of the green light in seconds $[s]$ with an appropriate dimension; $\delta_{i;t}$ is a queue indicator so that $\delta_{i;t} = 1$, if the queue exists, and $\delta_{i;t} = 0$ otherwise. Its value is naturally defined from the relations

$$\delta_{i;t} = 0, \ if \ (\xi_{i;t} + I_{i;t}u_t) < S_{i;t}u_t \ \text{(no queue)}, \tag{67}$$

$$\delta_{i;t} = 1, \ if \ (\xi_{i;t} + I_{i;t}u_t) \geq S_{i;t}u_t \ \text{(a queue exists)}. \tag{68}$$

Such a formulation of the queue length evolution is applied in (64-65). With its help, the time-varying matrices C_t, D_t, A_t, B_t and F_t for the considered simulated system are composed from the specific (supposed to be known) traffic parameters and input intensities as follows [10, 11, 12].

$$C_t = \begin{bmatrix} 0 & \alpha_{21}(1-\delta_{2,t}) & \alpha_{31}(1-\delta_{3,t}) & \alpha_{41}(1-\delta_{4,t}) & 0 & 0 & 0 & 0 \\ \alpha_{12}(1-\delta_{1,t}) & 0 & \alpha_{32}(1-\delta_{3,t}) & \alpha_{42}(1-\delta_{4,t}) & 0 & 0 & 0 & 0 \\ \alpha_{13}(1-\delta_{1,t}) & \alpha_{23}(1-\delta_{2,t}) & 0 & \alpha_{43}(1-\delta_{4,t}) & 0 & 0 & 0 & 0 \\ \alpha_{14}(1-\delta_{1,t}) & \alpha_{24}(1-\delta_{2,t}) & \alpha_{34}(1-\delta_{3,t}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{69}$$

where α_{ij} is the known (constant) parameter of the turn rate, reflecting the ratio of cars going from the i-th arm to the j-th arm, $j \neq i$, in percent [%]. The provided values of this parameter are $\alpha_{12} = 0.3,\ \alpha_{13} = 0.5,\ \alpha_{14} = 0.2,\ \alpha_{21} = 0.3,\ \alpha_{23} = 0.2,\ \alpha_{24} = 0.5,\ \alpha_{31} = 0.5,\ \alpha_{32} = 0.2,\ \alpha_{34} = 0.3,\ \alpha_{41} = 0.2,\ \alpha_{42} = 0.5,\ \alpha_{43} = 0.3$. The queue indicator $\delta_{i;t}, i = \{1, \ldots, n = 4\}$ is defined according (67-68).

$$D_t = \begin{bmatrix} D1_t & 0 \\ 0 & D2_t \\ D3_t & 0 \\ 0 & D4_t \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{70}$$

where

$$D1_t = \sum_{k=2}^{m=4} \alpha_{k1}((1-\delta_{k,t})I_{k,t} + \delta_{k,t}S_k), \tag{71}$$

$$D2_t = \alpha_{12}((1-\delta_{1,t})I_{1,t} + \delta_{1,t}S_1) + \sum_{k=3}^{m=4} \alpha_{k2}((1-\delta_{k,t})I_{k,t} + \delta_{k,t}S_k), \tag{72}$$

$$D3_t = \sum_{k=1}^{m=2} \alpha_{k3}((1-\delta_{k,t})I_{k,t} + \delta_{k,t}S_k) + \alpha_{43}((1-\delta_{4,t})I_{4,t} + \delta_{4,t}S_4), \tag{73}$$

$$D4_t = \sum_{k=1}^{m=3} \alpha_{k4}((1-\delta_{k,t})I_{k,t} + \delta_{k,t}S_k), \tag{74}$$

with the following values of the saturated flows in cars: $S_1 = 37.5,\ S_2 = 34.625,\ S_3 = 37.5,\ S_4 = 34.625$. The provided values of the control system input u_t, expressing the time of the green light in seconds $[s]$, are $u_t = [0.5\ \ 0.4]'$.

For the state evolution model (65) the matrices A_t, B_t and F_t are as follows.

$$A_t = \begin{bmatrix} \delta_{1,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_{2,t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_{3,t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_{4,t} & 0 & 0 & 0 & 0 \\ \kappa_{1,t} & 0 & 0 & 0 & \beta_{1,t} & 0 & 0 & 0 \\ 0 & \kappa_{2,t} & 0 & 0 & 0 & \beta_{2,t} & 0 & 0 \\ 0 & 0 & \kappa_{3,t} & 0 & 0 & 0 & \beta_{3,t} & 0 \\ 0 & 0 & 0 & \kappa_{4,t} & 0 & 0 & 0 & \beta_{4,t} \end{bmatrix}, \; F_t = \begin{bmatrix} I_{1,t} \\ I_{2,t} \\ I_{3,t} \\ I_{4,t} \\ \lambda_{1,t} \\ \lambda_{2,t} \\ \lambda_{3,t} \\ \lambda_{4,t} \end{bmatrix} \quad (75)$$

where $\kappa_{i,t}, \beta_{i,t}, \lambda_{i,t}$ are the known parameters, time varying according to the traffic during a week and a workday.

$$B_t = \begin{bmatrix} -(\delta_{1,t}S_1 + (1-\delta_{1,t})I_{1,t}) & 0 \\ 0 & -(\delta_{2,t}S_2 + (1-\delta_{2,t})I_{2,t}) \\ -(\delta_{3,t}S_3 + (1-\delta_{3,t})I_{3,t}) & 0 \\ 0 & -(\delta_{4,t}S_4 + (1-\delta_{4,t})I_{4,t}) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \quad (76)$$

The noises v_t and ω_t are defined according to (22-23) with the covariance matrices R_v and R_w respectively. The covariances are computed as a mean of squares of differences between the state (or output) value and its conditional mean. The mean is substituted by the samples of the daily (or for the corresponding time of a day) course of the state (or output), which is constructed as a spline approximation of several last periodic courses (e.g. courses during the workdays of a week). The resulted covariance matrices, used for the experimental part of the work, are respectively

$$R_v = \begin{bmatrix} 5.7173 & -0.0480 & -0.5697 & -0.7774 & -1.5699 & 0.1818 & 7.9483 & 0.6499 \\ -0.0480 & 6.9066 & 0.9071 & -0.4308 & 2.9482 & 0.0505 & 2.5980 & 1.3818 \\ -0.5697 & 0.9071 & 6.4584 & -0.1795 & 2.7968 & 0.2606 & -1.8104 & -0.0864 \\ -0.7774 & -0.4308 & -0.1795 & 6.2847 & 0.7385 & 0.7319 & 0.5578 & -0.4431 \\ -1.5699 & 2.9482 & 2.7968 & 0.7385 & 17.4974 & -0.5426 & -0.7768 & -0.4907 \\ 0.1818 & 0.0505 & 0.2606 & 0.7319 & -0.5426 & 1.2827 & 0.0387 & 0.3265 \\ 7.9483 & 2.5980 & -1.8104 & 0.5578 & -0.7768 & 0.0387 & 57.3942 & 1.7242 \\ 0.6499 & 1.3818 & -0.0864 & -0.4431 & -0.4907 & 0.3265 & 1.7242 & 7.4311 \end{bmatrix}, \tag{77}$$

and

$$R_w = \begin{bmatrix} 2.8967 & -0.3371 & -0.1246 & -0.0522 & -0.0075 & 0.0125 & -0.0082 & 0.0092 \\ -0.3371 & 1.9210 & 0.1181 & -0.0133 & 0.0104 & 0.0065 & -0.0369 & 0.0019 \\ -0.1246 & 0.1181 & 1.8617 & -0.1010 & 0.0029 & 0.0025 & -0.0614 & -0.0005 \\ -0.0522 & -0.0133 & -0.1010 & 2.2410 & -0.0252 & -0.0040 & 0.0091 & -0.0126 \\ -0.0075 & 0.0104 & 0.0029 & -0.0252 & 0.0208 & 0.0038 & 0.0042 & 0.0015 \\ 0.0125 & 0.0065 & 0.0025 & -0.0040 & 0.0038 & 0.0025 & 0.0042 & 0.0009 \\ -0.0082 & -0.0369 & -0.0614 & 0.0091 & 0.0042 & 0.0042 & 0.1182 & 0.0008 \\ 0.0092 & 0.0019 & -0.0005 & -0.0126 & 0.0015 & 0.0009 & 0.0008 & 0.0010 \end{bmatrix}. \tag{78}$$

The simulated traffic system, constructed in the described way, has been used for experiments. The prior knowledge to be incorporated into the distribution of the initial state x_0 in (64-65) for such the system is available in the form of the past simulated data about the traffic, dynamically changed according to the time of a day, with the time period, equal to ninety seconds. The exploitation of the data (not distributions) supposes Dirac delta to be used during implementation. The character of the traffic differs during the day, therefore, a series of experiments with the different prior knowledge (the morning, afternoon and the night traffic) has been conducted.

The experiments included the following steps. The representative data have been selected among available simulated ones according to the technique, proposed in Section 4. The selected prior knowledge has been processed and incorporated into the initial state distribution for Kalman filter according to the methodology, proposed in Section 3. To compare the results of the filtering, the Kalman filter (on current simulations) was run two times: with the initial state, obtained after the proposed prior knowledge processing, and starting with the state, corresponding to (25). In the case of the estimation, made for the *daily* course of the traffic, the usual practice in the traffic control area is to start the Kalman filter about 4 a.m. with the zero-mean initial state. It is naturally caused by the low night intensities. However, when the filtering should be done for a certain time of a day, the initial state would be chosen by experts as the average state values, taken from the past simulations. The paper presents the experiment with the morning peak-hours traffic, which obtained the most significant results. For this experiment the initial state distribution without application of the proposed methodology (i.e. corresponding to (25)) has the mean

$$\mu = [\underbrace{4.6134 \quad 3.4377 \quad 3.4058 \quad 3.2268}_{[\xi_{1;0},\ldots,\xi_{4;0}]'} \quad \underbrace{13.8555 \quad 3.4226 \quad 32.3499 \quad 1.8240}_{[o^I_{1;0},\ldots,o^I_{4;0}]'}]', \tag{79}$$

and covariance matrix, chosen by experts

$$R_x = \begin{bmatrix} 2.8967 & -0.3371 & -0.1246 & -0.0522 & -0.0075 & 0.0125 & -0.0082 & 0.0092 \\ -0.3371 & 1.9210 & 0.1181 & -0.0133 & 0.0104 & 0.0065 & -0.0369 & 0.0019 \\ -0.1246 & 0.1181 & 1.8617 & -0.1010 & 0.0029 & 0.0025 & -0.0614 & -0.0005 \\ -0.0522 & -0.0133 & -0.1010 & 2.2410 & -0.0252 & -0.0040 & 0.0091 & -0.0126 \\ -0.0075 & 0.0104 & 0.0029 & -0.0252 & 0.0442 & 0.0067 & 0.0018 & 0.0024 \\ 0.0125 & 0.0065 & 0.0025 & -0.0040 & 0.0067 & 0.0071 & 0.0050 & 0.0022 \\ -0.0082 & -0.0369 & -0.0614 & 0.0091 & 0.0018 & 0.0050 & 0.2220 & -0.0000 \\ 0.0092 & 0.0019 & -0.0005 & -0.0126 & 0.0024 & 0.0022 & -0.0000 & 0.0030 \end{bmatrix}. \tag{80}$$

The current simulated data for the experiments were identified with the real measurements. In order to judge about quality of the filtering, the estimation error was defined as the difference between the estimated and real state values. The errors of estimation with and without processed prior knowledge have been compared for the first twenty time periods of filtering.

5.1. Experiment with the morning peak-hours traffic

The past simulated data about the morning peak-hours traffic from 7 till 11 a.m. have been used as the prior knowledge for this experiment. The available data set contained 300 data, and 62 representative data have been selected and processed. After application of the proposed methodology, the initial state obtains the following mean values of initial queue lengths at the intersection arms (i.e. vector $\hat{\xi}_{0(\tau)} = [\hat{\xi}_{1;0(\tau)}, \ldots, \hat{\xi}_{4;0(\tau)}]'$).

$$\hat{x}_{0(\tau)} = [\underbrace{6.7328 \;\; 6.1853 \;\; 5.9099 \;\; 5.1026}_{[\hat{\xi}_{1;0(\tau)}, \ldots, \hat{\xi}_{4;0(\tau)}]'} \;\; [o^I_{1;0}, \ldots, o^I_{4;0}]']', \tag{81}$$

where the vector of occupancies $[o^I_{1;0}, \ldots, o^I_{4;0}]'$ is not a task of estimation and remains the same as in (79). The covariance matrix of the obtained initial state is as follows.

$$\hat{P}_{0(\tau)} = \begin{bmatrix} 0.6334 & -0.0787 & -0.0244 & -0.1122 & 0.0475 & 0.0214 & -0.0070 & 0.0058 \\ -0.0787 & 0.6100 & -0.1478 & -0.0365 & 0.0024 & 0.0339 & 0.0191 & -0.0088 \\ -0.0244 & -0.1478 & 0.5885 & -0.0881 & -0.0102 & 0.0191 & 0.0448 & 0.0136 \\ -0.1122 & -0.0365 & -0.0881 & 0.6362 & 0.0349 & 0.0066 & 0.0186 & 0.0282 \\ 0.0475 & 0.0024 & -0.0102 & 0.0349 & 0.5871 & -0.0191 & 0.0035 & -0.0108 \\ 0.0214 & 0.0339 & 0.0191 & 0.0066 & -0.0191 & 0.1398 & -0.0007 & 0.0178 \\ -0.0070 & 0.0191 & 0.0448 & 0.0186 & 0.0035 & -0.0007 & 0.7169 & 0.0048 \\ 0.0058 & -0.0088 & 0.0136 & 0.0282 & -0.0108 & 0.0178 & 0.0048 & 0.4468 \end{bmatrix}. \quad (82)$$

The Kalman filter has been started with the values (81-82), obtained due to the proposed methodology, and then with the initial state (79-80). Fig. 1 shows the difference of the results at the beginning of the filtering at the intersection arm 4 of the considered traffic system (the estimation results for the rest of the arms are very similar). It can be seen, that the incorporation of prior knowledge (81-82) influenced the beginning of the estimation by the initial value, more close to the real one, than in the case with incorporation (79-80).

Table I confirms this influence, demonstrating the error of the estimation, defined as the difference between the estimated and real values of the state for the first twenty time periods at all intersection arms. The error of estimation with the prior knowledge, processed according to the proposed methodology, is 0.8181 cars less than in the case without it for arm 1, about 1.0696 cars less for arm 2, and 0.9898 cars for arm 3. The improving of the estimate for arm 4 is 0.1162 cars. With respect to the initial queue lengths (79) and (81) and a character of

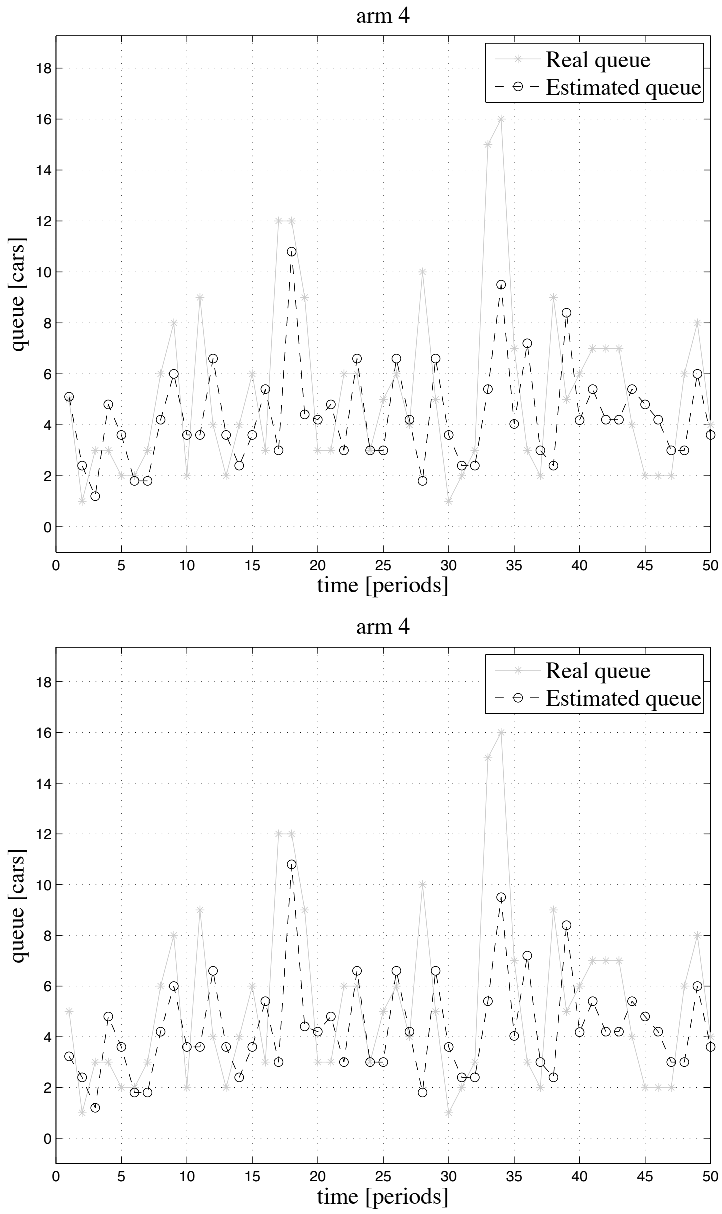

Figure 1. Start of the filter at arm 4: with the applied methodology (top) and without (bottom)

Table I. Comparison of the morning estimation results for time periods $t \in \{0, \ldots, 20\}$

Arm	Estimation error with applied methodology	Without it
1	12.8563	13.6744
2	15.3506	16.4202
3	10.4904	11.4802
4	13.4096	13.5258

the morning traffic (the car queue length starts at $5 - 10$ cars at 7 a.m. and is going down by 11 a.m.), such the improvements are rather promising, and the proposed methodology gives a good start for estimation. The rest of the estimates for time periods $t \in \{20, \ldots, \mathring{t}\}$ remains practically unchanged in comparison to Kalman filter without proposed processing of the prior knowledge. It is explained by a character of Kalman filter, i.e. updating by current measurements and their natural predominance over prior knowledge.

The covariance matrix (82) demonstrates a less degree of uncertainty of the initial state estimates with comparison to (80).

To summarize the experimental part of the work, it can be said, that the processed prior knowledge has the more influence on the filtering results during really peak hours. It means, the more length of the car queue, the less error of estimation with applying of the prior knowledge processing. The experiments with the afternoon peak-hours traffic also give the optimistic results, but less improved in comparison with the morning ones. Naturally, the noted dependence is explained by high informativeness of the peak-hours data. The night-traffic experiment, which has demonstrated unchanged, very slightly improved or even worse estimation results, confirms this statement. Such the conclusions after testing of the proposed methodology generally correspond to the global aim of the work to improve the state estimates on the peak hours with long queues of cars, awaiting at the intersection.

6. CONCLUSION

The paper proposes a methodology of the prior knowledge processing for the initial state of the Kalman filter. The proposed methodology enables to obtain the specified Gaussian distribution of the initial state with the respective mean and covariance matrix. The obtained initial state distribution with the incorporated processed prior knowledge gives the better start to the Kalman filter and improves most results of filtering. The examples of calculation of the initial state distribution are shown.

Various types of prior knowledge bring additional specific problems. The paper is focused on processing of historical data, available by means of simulation or observed on a similar system. The technique of the simulated data processing is proposed. Application of the technique is demonstrated on the traffic system simulated data.

The proposed methodology assumes the system model parameters and noise covariances to be known. The prior knowledge processing for the joint parameter-and-state Bayesian estimation can be meanwhile related to the open problems.

ACKNOWLEDGEMENTS

The research was supported by GA ČR grant No. 201/06/P434.

REFERENCES

1. Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. Tech. Rep. 95-041, UNC-CH Computer Science, 1995.
2. M. Kárný, P. Nedoma, N. Khailova, and L. Pavelková. Prior information in structure estimation. In *IEEE Proceedings — Control Theory and Applications*, vol. 150, no. 6, pp. 643–653, 2003.
3. Kárný, M. and Andrýsek, J. and Bodini, A. and Guy, T. V. and Kracík, J. and Ruggeri, F. How to exploit external model of data for parameter estimation? *International Journal of Adaptive Control and Signal Processing*, vol. 20 no. 1, pp.41–50, 2006
4. E. Suzdaleva. Initial conditions for Kalman filtering: Prior knowledge specification. In *Proceedings of the 7th WSEAS International Conference on Systems Theory and Scientific Computation (ISTASC'07)*, M.H. Le, M. Demiralp, V. Mladenov, and Z. Bojkovic, Eds., Athens, Greece, August 24-26 2007, Electrical and Computer Engineering, pp. 45–49, WSEAS Press.
5. V. Peterka. Bayesian system identification. In *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.
6. J.Y. Li, G.Z. Dong, K. Ramamohanarao, and L.S. Wong. Deeps: A new instance-based lazy discovery and classication system. *Machine Learning*, vol. 54, no. 2, pp. 99–124, 2004.
7. S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, vol. 22, pp. 79–87, 1951.
8. M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.
9. TSS – Transport Simulation Systems. AIMSUN microscopic traffic simulator: a tool for the analysis and assessment of ITS systems. *www.tss-bcn.com*
10. J. Homolová and I. Nagy. Traffic model of a microregion. In *Preprints of the 16th World Congress of the International Federation of Automatic Control*, P. Horáček, M. Šimandl and P. Zítek, Eds., Prague, July 2005, pp.1–6, IFAC.
11. J. Homolová. Linear programming approach to traffic flow control: optimization on horizon. In *Proceedings of the 11th International Conference of Hong Kong Society for Transportation Studies*, S.C. Wong, T.D. Hau, J.J. Wang, K.W. Chau and H.W. Ho, Eds., Hong Kong, December 2006, pp.697–705, Hong Kong Society for Transportation Studies Ltd.
12. J. Duník, P. Pecherková and M. Flídr. State Space Model of Traffic System and its Estimation Using Derivative-Free Methods. *ÚTIA AV ČR*, Prague, 2006.

