Soft Sensor for Faulty Measurements Detection and Reconstruction in Urban Traffic

Juš Kocijan [#1], Jan Přikryl [*2]

 # Jozef Stefan Institute, Ljubljana, Slovenia and
University of Nova Gorica, Nova Gorica, Slovenia
[1] `jus.kocijan@ijs.si`

* Czech Technical University in Prague and Department of Adaptive Systems, UTIA AVČR
Prague, Czech republic
[2] `prikryl@fd.cvut.cz`

Abstract—**Soft sensors are a valuable alternative to the traditional hardware sensors, which are indispensable in configuration of modern systems. They are often used in process industry, but other applications are possible. This paper describes a possible application of soft sensor for faulty measurement detection and reconstruction in urban traffic. One of the key indicators of traffic control quality in urban traffic control (UTC) systems is the queue length. With the exception of some expensive surveillance equipment, the queue length itself cannot be measured directly. Many methods that estimate the queue length from detector measurements are used in engineering practice, ranging from simple to elaborate ones. The proposed method is a soft sensor based on Gaussian process (GP) model of the traffic queue length in a traffic cross-section. The resultant soft sensor detects single measurement outliers as well as longer lasting faults and replaces the faulty measurements with model prediction.**

I. INTRODUCTION

Soft sensors are models that based on measurements of various variables provide estimation of another variable. The expression soft sensor is mainly used in the field of process industry, but the utility of mathematical modelling is widespread in engineering fields.

Surveys of use of soft sensors that mainly focus to process industry can be found in [1],[2] and [3]. The approach is however not limited to process industries and similar ideas are utilised in other engineering fields, for example in traffic engineering, e.g. [4] and [5]. Soft sensors are used where their hardware counterparts are not available, are very costly or their installation is very costly.

Soft sensors can in general be divided to two different classes [1] which are model-driven and data-driven soft sensors. The model-driven family of soft sensors is most commonly based on theoretical or so called first principle models or on extended Kalman filter. The data-driven family of soft sensors is based on data measured within the systems. Most frequent methods [1] for data-driven soft sensors are principle component analysis (PCA), partial least squares, artificial neural networks, neuro-fuzzy systems and support vector machines.

Most frequent applications of soft sensors are: on-line prediction, monitoring and process fault detection, sensor fault detection and reconstruction. Despite of relatively frequent soft sensor applications there exist some issues that are still not enough addressed [1]. These are mainly related to measured data which need a lot of pre-processing due to: missing data, data outliers, drifting data, data co-linearity, different sampling rates and measurement delays. Even though authors of [1] list these issues in context of process control, they are present to the certain extent everywhere where data are measured and urban traffic is no exception.

The problem of detecting of faulty measurements and its reconstruction, which is the focus of this paper has been partially addressed in the literature. For example, papers [6],[7] and [8] give surveys of methods dealing with outlier detection and removal, while [9] describes faulty detection and signal reconstruction for dynamic systems with lagged PCA.

In urban traffic control systems, different sensors are usually employed in order to collect relevant information about traffic situation in the city. These sensors range from inductive loops over video-based systems to elaborate ITS installations providing toll collection and enforcement. Typical data collected by traffic control systems include count of vehicles travelling over a detector, occupancy of the detector, or velocity of vehicles.

Despite their higher installation costs, inductive loops are still the sensor of choice for providing vehicle counts and occupancy when installing signalized intersections. When properly designed, they last long and are virtually maintenance-free (compared to other alternatives like video-based sensors). Nevertheless, the inductive loop itself, the connecting cables, or the interface board which is a part of an intersection controller may eventually fail, and sometimes the inductive loop interface gradually de-tunes from the proper working point. When the detector stops working the hardware usually recognises the failure and reports it to the maintenance staff — however, it takes some time (our practical experience ranges from several days to several weeks) before the system gets repaired.

In order to control an intersection in traffic-actuated mode, measurements provided by detectors are crucial. Typical solution in the case of recognised failure is to program the interface board of 'important' detectors to generate permanent

requests (artificially generate signal 'vehicle is present') in the case of loop rupture. This provisional solution may have severe impact on the throughput of the controlled system of intersections in the case that the detector is located on lane experiencing low traffic (and short green lengths). In such a case, employing a soft sensor that is able to identify and reconstruct measurements of the failing detector may significantly improve the traffic situation in the transient period before the detector gets repaired.

The paper is organised as follows. In the next section problem description is given. Gaussian process model and its utility for dynamic systems are described in Section 3. Section 4 describes modelling and validation of queue-length predictor and conclusions are gathered in the last section.

II. PROBLEM DESCRIPTION

Typical control scenario of an urban network similar to the one depicted in Figure 1 would collect counts of incoming vehicles on the border of the controlled area, use them to identify the current traffic demands, and finally set the signals at particular intersection to set-points proving an optimum of some control criterion. In the case that some of the input sensors start to fail and become unreliable, the control system has no means to react and the traffic model used for control of the urban network drifts off the current situation, resulting in incorrect control actions to be applied.
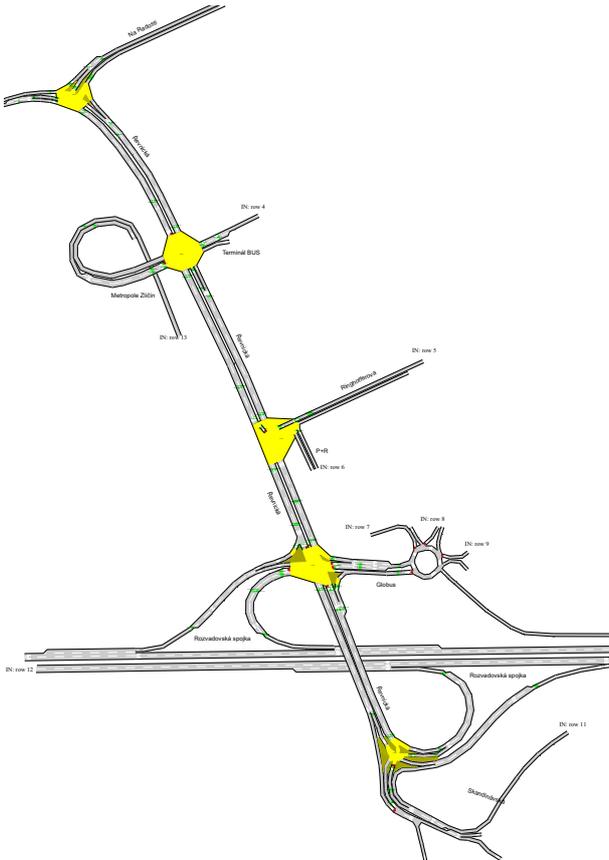

Fig. 1.   Example of a controlled network (Zličín shopping centre, Prague).

This is the reason why we seek a simple solution to make the sensor more reliable without the need to store historic data. Our aim is to devise a soft sensor for vehicle count measurements, with the ability of evaluation of the incoming data, on-line outlier and inconsistent data detection, and, if necessary, reconstruction of the input signal to its maximum likely value.

The measurements of queue length are from microscopic view measurements of discrete events. Only from macroscopic view they look like a signal that is very noisy. The candidates for modelling such time series would be methods that are applicable and give acceptable results in the presence of substantial noise, give prediction without major delay or are quick enough that can be used for on-line calculations and that obtained prediction confidence band is calculated with very modest amount of calculation. Modelling of Gaussian process models that proved itself as successful in various regression applications was selected in our case.

III. GAUSSIAN PROCESS MODELS

The Gaussian process model is an example of the use of a flexible, probabilistic, nonparametric model with uncertainty predictions. It fits naturally in the Bayesian modelling framework in which instead of parameterising a mapping function $f(\mathbf{x})$, a prior is placed directly on the space of possible functions $f(\mathbf{x})$ which could represent the nonlinear mapping from input vector $\mathbf{x}$ to output y. Its use and properties for modelling are reviewed in [10].

A Gaussian process is a generalisation of the Gaussian probability distribution. It can be viewed as a collection of random variables $f(\mathbf{x}_i)$ with joint multivariate Gaussian distribution: $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n) \approx \mathcal{N}(0, \mathbf{\Sigma})$, where $\mathcal{N}(\cdot, \cdot)$ stands for Gaussian distribution determined with mean value and variance, where covariance matrix element $\Sigma_{pq} = \mathrm{Cov}(y_p, y_q) = C(\mathbf{x}_p, \mathbf{x}_q)$ gives the covariance between values of the functions $y_p = f(\mathbf{x}_p)$ and $y_q = f(\mathbf{x}_p)$. Thus, the mean $\mu(\mathbf{x})$ (usually assumed to be zero) and the covariance function $C(\mathbf{x}_p, \mathbf{x}_q)$ fully specify the Gaussian process. Note that the covariance function $C(\cdot, \cdot)$ can be any function having the property of generating a positive definite covariance matrix.

A common choice is

$$C(\mathbf{x}_p, \mathbf{x}_q) = v_1 \exp\left[-\frac{1}{2} \sum_{d=1}^{D} w_d (x_{dp} - x_{dq})^2 \right] + \delta_{pq} v_0, \quad (1)$$

where $\mathbf{\Theta} = [w_1, \ldots, w_D, v_1, v_0]^T$ are the 'hyperparameters' of the covariance functions, v_0 is estimated noise variance, v_1 is the estimate of the vertical scale of variation, D is the input dimension, x_{dp} and x_{dq} are d^{th} components of input vectors $\mathbf{x}_p, \mathbf{x}_q$ and $\delta_{pq} = 1$ if $p = q$ and 0 otherwise. The covariance function (1) is composed of two parts: the Gaussian covariance function for the modelling of system function and the covariance function for the modelling of noise. The noise, in our case, is presumed to be white. Other forms of covariance functions suitable for different applications can be found in [10].

For a given problem, the parameters in $\mathbf{\Theta}$ are learned (identified) using the data at hand. After the learning, one can use the w parameters as indicators of 'how important' the corresponding input components (dimensions) are: if w_d is zero or near zero it means that the inputs in dimension d contain little information and could possibly be removed.

Consider a set of N D-dimensional input vectors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ and a vector of output data $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$. Based on the data $(\mathbf{X}, \mathbf{y})$, and given a new input vector $\mathbf{x}^*$, we wish to find the predictive distribution of the corresponding output y^*. Unlike other models, there is no model parameter determination as such, within a fixed model structure. With this model, most of the effort consists in *tuning* the parameters of the covariance function. This is done by maximisation of the log-likelihood

$$
\begin{aligned}
\mathcal{L}(\mathbf{\Theta}) &= \log(p(\mathbf{y}|\mathbf{X})) \\
&= -\frac{1}{2}\log(|\ \mathbf{K}\ |) - \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{N}{2}\log(2\pi) \quad (2)
\end{aligned}
$$

where $\mathbf{\Theta}$ is the vector of hyperparameters and $\mathbf{K}$ is the $N \times N$ training covariance matrix. The calculation of the log-likelihood and its derivatives due to the optimisation algorithm involves the computation of the inverse of the $N \times N$ covariance matrix $\mathbf{K}$ at every iteration, which can become computationally demanding for large N. Nevertheless, the number of parameters to be optimised is small ($D + 2$, see Equation (1)), which means that optimisation convergence might be faster and that the 'curse of dimensionality' so common to black-box identification methods is circumvented or at least decreased.

The described approach can be easily utilised for regression calculation. Based on training set $\mathbf{X}$ a covariance matrix $\mathbf{K}$ of size $N \times N$ is determined. As already mentioned, the aim is to find the distribution of the corresponding output y^* at some new input vector $\mathbf{x}^* = [x_1(N+1), x_2(N+1), \ldots, x_D(N+1)]^T$.

For a new test input $\mathbf{x}^*$, the predictive distribution of the corresponding output y^*, over cases in the training set $(\mathbf{X}, \mathbf{y})$ is $y^*|(\mathbf{X}, \mathbf{y}), \mathbf{x}^*$ and is Gaussian, with mean and variance

$$
\begin{aligned}
\mu(\mathbf{x}^*) &= \mathbf{k}(\mathbf{x}^*)^T\ \mathbf{K}^{-1}\ \mathbf{y}, \quad (3)\\
\sigma^2(\mathbf{x}^*) &= \kappa(\mathbf{x}^*)\ -\ \mathbf{k}(\mathbf{x}^*)^T\ \mathbf{K}^{-1}\ \mathbf{k}(\mathbf{x}^*),
\end{aligned}
$$

where $\mathbf{k}(\mathbf{x}^*) = [C(\mathbf{x}^1, \mathbf{x}^*), \ldots, C(\mathbf{x}^N, \mathbf{x}^*)]^T$ is the $N \times 1$ vector of covariances between the test and training cases, and $\kappa(\mathbf{x}^*) = C(\mathbf{x}^*, \mathbf{x}^*)$ is the covariance between the test input and itself. Vector $\mathbf{k}(\mathbf{x}^*)^T\ \mathbf{K}^{-1}$ in (3) can be interpreted as a vector of *smoothing* terms which weights the training outputs from $\mathbf{y}$ to make a prediction at the test point $\mathbf{x}^*$.

The reasons to select modelling with Gaussian process models are small amounts of data relative to the number of selected regressors, data corrupted with noise and measurement errors and the need for the measure of model prediction confidence. If there is not enough data or it is heavily corrupted with noise, even the Gaussian process model cannot perform well, but in that case the inadequacy of the model and the identification data is indicated through higher variance of the predictions.

The utility to provide the information about the model confidence made Gaussian process models attractive for modelling case studies in various domains like: chemical engineering [11] and process control [12], biomedical engineering [13], biological systems [14], environmental systems [15], power systems [16] and engineering [17], motion recognition [18], etc., to list just a few. We believe that this utility makes it interesting also for use in the domain of traffic modelling and its applications.

We should note that there is no assumption that signals measured in a traffic process will have Gaussian distribution. The Gaussian process prior is put over the space of functions meaning that every prediction that is made by GP model has Gaussian distribution. This implies that every prediction that is made has some most likely value and the less likely values are equally possible on both sides of the most likely value. This is, by our convenience, not unrealistic.

Gaussian processes can, like neural networks, be used to model static nonlinearities and can therefore be used for modelling of dynamic systems [14], [19], [11] as well as time series if lagged samples of output signals are fed back and used as regressors. In such cases an autoregressive model is considered, such that the current output depends on previous outputs.

$$
\begin{aligned}
\mathbf{x}(k) &= [y(k-1), y(k-2), \ldots, y(k-L)]^T, \\
\hat{y}(k) &= f(\mathbf{x}(k)) + \epsilon, \quad (4)
\end{aligned}
$$

where k denotes the consecutive number of data sample. Let $\mathbf{x}$ denote the state vector composed of the previous outputs y up to a given lag L, and ϵ is white noise.

IV. MODELLING

A. Data Acquisition for modelling

Data samples studied in this paper are extracted from a collection of measurements taken at the test site of the Czech DAR project at Zličin area in Prague. This area contains mostly shopping centres and a bus terminal serving both long-haul busses and regular public transport.

The whole collection of vehicle counts covers period from 2007/12/10 to 2008/09/02 for all detectors installed at the site. From this collection, two separate data sets of working detector data for a typical weekday (Tuesday-Thursday) were extracted for GP model training and validation – see example data in Fig. 2. Another data set containing failing detector data from two consecutive weeks was used to test the failure identification and reconstruction capabilities of the model.

B. One-step-ahead predictor

One working day of data was selected as data for estimation. Covariance function (1) was selected because we treated the system as stationary and smooth. With these presumptions we wanted to assure certain level of smoothness. A systematic selection of regressors, so called *validation based regressor selection* was pursued starting with larger number of regressors, which was then decreased. At the regression selection a
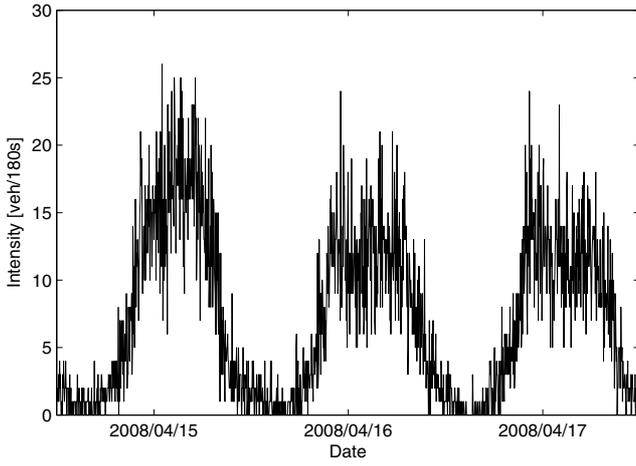
Fig. 2. The profile of vehicle arrival data for three consecutive working days, sampled with 180s interval.

care was taken that the model was not too fitted to estimation data, which in our case exposed as total averaging of validation signal resulting in constant predictions. On the other hand prediction confidence band should not be too narrow, which was also the case when the model was too fitted to estimation data.

Excessive amount of noise contributed to the relatively low quality of finally selected model predictions. The structure of AR model with four delayed output values was selected. One-step-ahead predictions on estimation data are given in Fig. 3. It can be seen from Fig. 3 that residuals are contained in 3σ band. The residuals' distribution is close to the one of white noise.
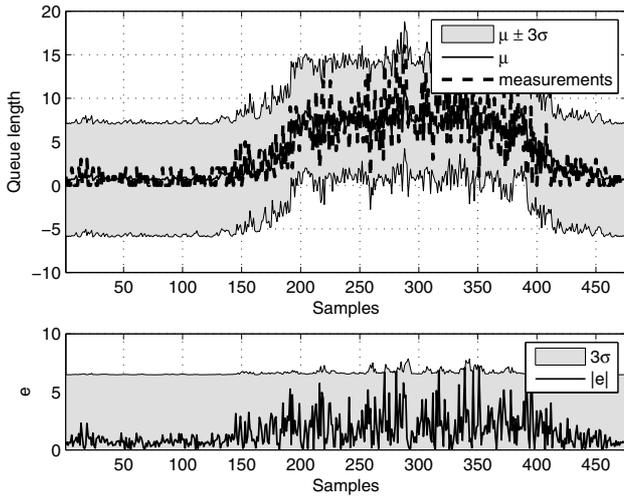

Fig. 3. One-step-ahead predictions on data used for estimation (upper figure) and residuals of predictions with 3σ band (lower figure)

C. Model validation

One-step-ahead predictions on validation data that were taken from measurements on a different day in a different week are depicted in Fig. 4. Again it can be said that the
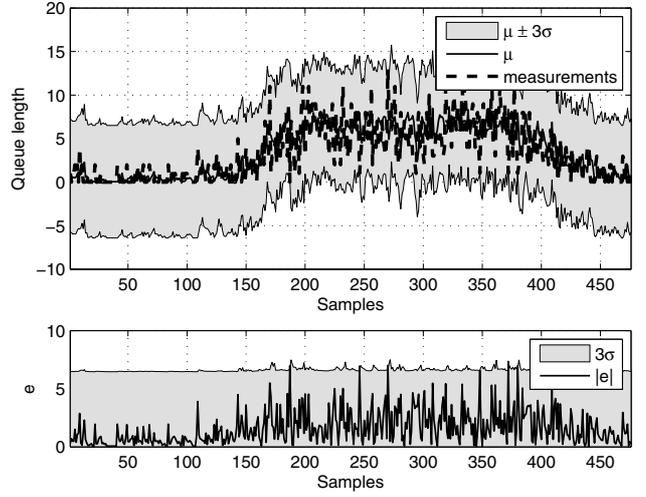

Fig. 4. One-step-ahead predictions on data used for validation (upper figure) and residuals of predictions with 3σ band (lower figure)

residuals' distribution is close to the one of white noise.

As we are interested predominantly in mean value prediction mean relative square error (MRSE) was used for validation of prediction fit. It compares only the mean prediction of the model to the output of the process:

$$\text{MRSE} = \sqrt{\frac{\sum_{i=1}^{N} e(i)^2}{\sum_{i=1}^{N} y(i)^2}} \qquad (5)$$

where $y(i)$ and $e(i) = \hat{y}(i) - y(i)$ are measurement and prediction error in i-th step. In general log predictive density [14] performance measure is more suitable, when validating predictions of Gaussian process model, but since we are focused on mean values prediction MRSE will do.

Model purposiveness or usefulness tells whether or not the obtained model satisfies its purpose, which means the model is validated when the problem that motivated the modelling exercise can be solved using the obtained model. In our case the model has to give a prediction output where standard deviation of the prediction will be used as a measure of measurement validity and in the case the measurement is recognised as 'not valid' replaced with a prediction that might not be mirroring the actual value, but will be close enough so that the control system will continue its regular operation without exhibiting any extreme modes. Our model, even though not expressing high prediction accuracy, satisfies all described requirements and is able to calculate predictions fast.

D. Algorithm for data removal and reconstruction

The threshold at which a measurement is characterised as not realistic is set at 3σ, i.e. at three times standard deviation of output given by Gaussian process model prediction. This is the threshold frequently used in off-line outlier's detection algorithms, e.g. [7].

The proposed algorithm that runs in soft sensor and on-line in every time sample detects irregularities and tries to reconstruct the data is as follows:

loop
    take the measurement M
    calculate the prediction $P = \mathcal{N}(\mu, \sigma)$
    if $|M - P| > 3\sigma$ **then**
        $M \leftarrow P$
        indicate possible fault
    end if
end loop

Fig. 5 shows detection of faulty data and their reconstruction by model, i.e. soft sensor.

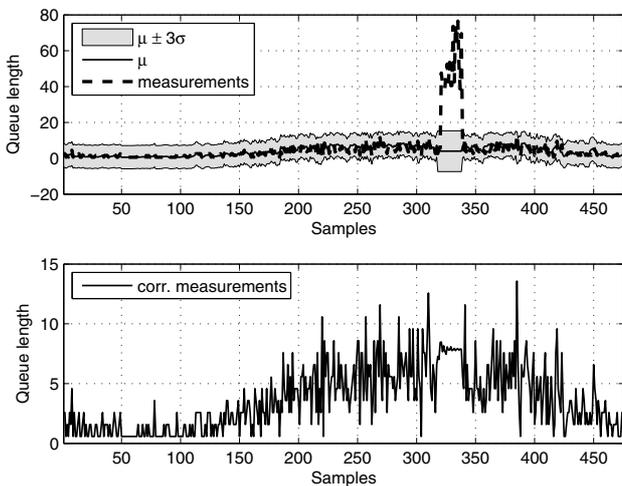Fig. 6. The comparison of MRSE for k-step-ahead predictions

Fig. 5. Soft sensor applied on faulty data. Upper figure depicts measurements that are corrupted from 318st sample until 338th sample and predictions of soft sensor. Bottom figure depicts measurements with reconstructed samples of one to 21-steps-ahead predictions.

As it can be seen from the proposed algorithm, the reconstruction is done as a replacement of faulty measurement with model predictions. In the case the fault is not just an outlier, but a longer period fault then model starts to predict the queue length until operator detects and removes the fault. Fig. 6 shows how mean values of predictions deteriorate with progressing number of predictions.

V. CONCLUSION

The paper describes a possible application of a time series autoregressive model as the soft sensor for faulty measurement detection and reconstruction in urban traffic. The proposed method for developing a computational model that serves as a soft sensor is based on G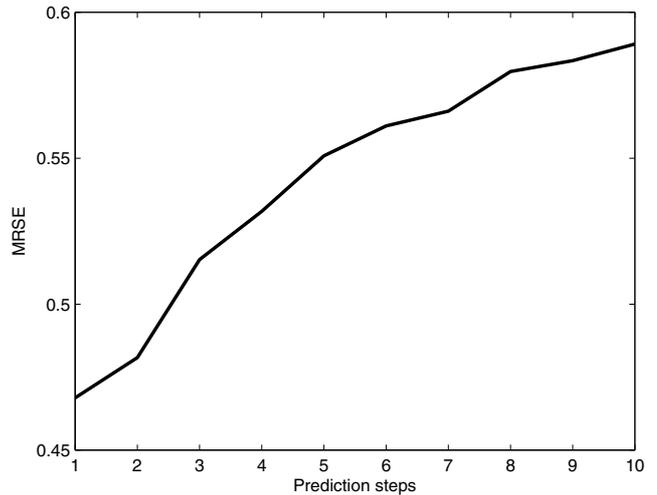aussian process (GP) model. The resultant soft sensor detects single measurement outliers as well as longer lasting faults and will be used to replace the faulty measurements with model prediction. This way it will increase reliability of the hardware sensors that are used in urban traffic control system.

Main conclusions can be listed as follows:

- GP models can be used as a method for modelling data when excessive amount of noise is present.
- Obtained model that will serve as soft sensor was validated as appropriate for fulfilling its task, i.e. outlier detection and reconstruction of measurements.
- The excessive noise limits the possibility to develop better predictor, which only confirms known fact.
- Soft sensors are useful technology for fault detection and signal reconstruction also in traffic control engineering.

The presented study is the initial feasibility study which purpose is to show potentials of the soft sensor approach. Prediction models for data measurements over weekends and holidays, when traffic density is much different from working days have to be developed and combined with the existing ones and a field test has to be pursued to confirm successful operation in practice.

Making").

REFERENCES

[1] P. Kadlec, B. Gabrys, and S. Strand, "Data-driven soft sensors in the process industry," *Computers and Chemical Engineering*, vol. 33, pp. 795–814, 2009.

[2] G. D. Gonzalez, "Soft sensors for processing plants," in *Proceedings of the second international conference on intelligent processing and manufacturing of materials, IPMM99*, 1999.

[3] L. Fortuna, *Soft sensors for monitoring and control of industrial processes*. Berlin: Springer, 2007.

[4] A. Adamski and S. Habdank-Wojewodzki, "Traffic congestion and incident detector realized by fuzzy discrete dynamic system," *Archives of Transport*, vol. 17, no. 2, pp. 5–13, 2004.

[5] W. B. Zhu, D. S. Li, and Y. Lu, "Real time speed measure while automobile braking on soft sensing technique," *Journal of Physics: Conference Series*, vol. 48, no. 1, pp. 730–733, 2006.

[6] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[7] B. Lin, B. Recke, J. Knudsen, and S. B. Joergensen, "A systematic approach for soft sensor development," *Computers and Chemical Engineering*, vol. 31, no. 5, pp. 419–425, 2007.

[8] P. H. Menold, R. K. Pearson, and F. Allgower, "Online outlier detection and removal," in *Proceedings of the 7th Mediterranean on control and automation (MED99)*, Haifa, Israel, 1999, pp. 1110–1133.

[9] C. Lee, S. Choi, and I.-B. Lee, "Sensor fault identification based on time-lagged pca in dynamic processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 70, no. 2, pp. 165–178, 2004.

[10] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[11] J. Kocijan and B. Likar, "Gas-liquid separator modelling and simulation with gaussian process models," *Simulation Modelling Practice and Theory*, vol. 16, no. 8, pp. 910–922, 2008.

[12] B. Likar and J. Kocijan, "Predictive control of a gas-liquid separation plant based on a gaussian process model," *Computers and Chemical Engineering*, vol. 31, no. 3, pp. 142–152, 2007.

[13] S. Faul, G. Gregorčič, G. Boylan, W. Marnane, G. Lightbody, and S. Connolly, "Gaussian process modelling of eeg for the detection of neonatal seizure," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2151–2162, 2007.

[14] K. Ažman and J. Kocijan, "Application of gaussian processes for black-box modelling of biosystems," *ISA Transactions*, vol. 46, no. 4, pp. 443–457, 2007.

[15] B. Grašič, P. Mlakar, and M. Z. Božnar, "Ozone prediction based on neural networks and gaussian processes," *Nuovo Cimento della Societa Italiana di Fisica, Sect. C*, vol. 29, no. 6, pp. 651–662, 2006.

[16] D. J. Leith, M. Heidl, and J. Ringwood, "Gaussian process prior models for electrical load forecasting," in *Proceedings of 2004 International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2004, pp. 112–117.

[17] W. E. Leithead, Y. Zhang, and K. S. Neo, "Wind turbine rotor acceleration: Identification using gaussian regression," in *Proceedings of International conference on informatics in control automation and robotics (ICINCO)*, Barcelona, 2005, pp. 84–91.

[18] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283–298, 2008.

[19] J. Kocijan, A. Girard, B. Banko, and R. Murray-Smith, "Dynamic systems identification with gaussian processes," *Math. comput. model. dyn. syst.*, vol. 11, no. 4, pp. 411–424, 2005.