2013 International Conference on Process Control (PC)
June 18–21, 2013, Štrbské Pleso, Slovakia

Bayesian Model Mixing for Cold Rolling Mills: Test Results

Pavel Ettler and Ivan Puchr
COMPUREG Plzeň, s.r.o.,
30634 Plzeň, Czech Republic
Email: ettler@compureg.cz, puchr@compureg.cz

Kamil Dedecius
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
18208 Praha, Czech Republic
Email: dedecius@utia.cas.cz

***Abstract*—The contribution presents the results of a collaborative R&D effort of two private companies and two national research institutions, joined at the European level. It was aimed to develop an enhanced on-line predictor of the strip thickness in the rolling gap. The issue dealt with is the absence of a reliable delay-free measurement of the outgoing strip thickness or the gap size, making the thickness control a challenging task. Although several satisfactory solutions have been used for decades, and modern control theory has been exploited as well, the pervasive competition in the field of metal strip processing emphasizes the need of a novel, more precious measuring method. The solution developed within the completed project is based on a parallel run of several adaptive Bayesian predictors whose outputs are continuously mixed to provide the best available rolling gap size prediction. The system was already tested in open loop in a real industrial environment for two reversing cold rolling mills processing steel and copper alloys strips, respectively.**

I. Introduction

Thickness control (Automatic Gauge Control – AGC) counts traditionally among key and challenging tasks in the field of cold rolling of metal strips. Its well elaborated solutions have been routinely used for decades but permanent competition and economic pressures motivate research for even slight improvements of control quality, especially for initial phases of rolling and for dealing with non-standard situations.

A company with two decades long expertise in control of rolling mills joined forces with two renowned research institutes and another automation-oriented company in order to design an innovative method, combining a group of existing solutions within a novel framework based on Bayesian treatment of uncertainty. The main achievement of the completed project accomplished by the Czech-Slovenian consortium is a functional sample of the so-called Probabilistic Bayesian Soft Sensor (ProBaSensor). This sample was extensively tested in two industrial plants. The present paper describes main principles and achievements.

II. Nature of the problem

A. Reversing cold rolling mill

A reversing cold rolling mill is a machine serving for reduction of the metal strip thickness during one or several passes through the rolling gap while alternating direction of rolling – see Fig. 1.

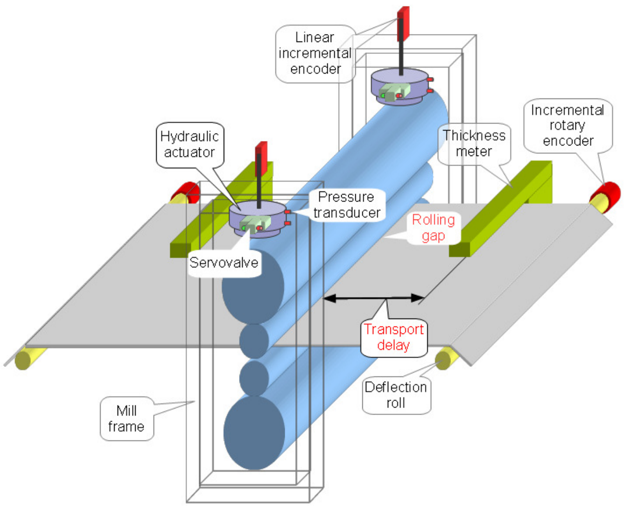

Fig. 1. Scheme of a four-high rolling mill.

The problem is that exact and reliable measurement of the rolling gap size which directly influences the outgoing strip thickness is not available. The roll positioning system provides just measurement of the position z of its actuator against the mill frame and the output strip thickness H_2 is measured with a significant transport delay.

Values of z and corresponding H_2 differ principally because of significant stretch of the rolling mill due to applied rolling force. Elongation of the rolling mill frame can reach more than 2 millimetres while the required strip thickness tolerance counts in tens of microns or even less, depending on parameters of the strip. Among usually measured signals belong also the input and output strip speeds v_1 and v_2, respectively, the rolling force F (or at least hydraulic pressures for its indirect evaluation), electric currents of the main and coiler drives and input and output strip tensions in some cases. Mostly, the input strip thickness H_1 can be measured as an auxiliary variable.

B. Existing solutions

Basic feedback-type AGC based entirely on measurement of the output strip thickness H_2 suffers from obvious limitations imposed by the transport delay. Another classical method for evaluation of the gap size, known as the *gaugemeter*, utilizes an empirically determined approximative stretch function $\mathcal{F}(F, z)$ to estimate the actual size of the rolling gap. Yet another method based on the so called mass-flow principle relies on theoretical equality of ratios of input–output thicknesses and output–input speeds,

$$\frac{H_1}{H_2} \doteq \frac{v_2}{v_1}.$$

These principles can be considered as members of a wide family of model-based methods, which utilize relevant measured signals in various combinations. There exist also other more or less successful approaches, whose description would exceed the scope of this paper.

The model-based methods [1] often employ a regression-type models of the form

$$h_2(k) = P'(k)\ D(k) + e(k), \tag{1}$$

where $k = 1, 2, \ldots$ stands for discrete time index, $h_2(k)$ is deviation of the output thickness from its nominal value,

$$D(k) = [d_1(k),\ d_2(k),\ \ldots,\ d_{m-1}(k), 1]'$$

is the regression vector of length m (the last term represents the offset) and

$$P(k) = [p_1(k),\ p_2(k),\ \ldots,\ p_m(k)]'$$

is the vector of m unknown parameters. The stochastic variable $e(k) \sim \mathcal{N}(0, r(k))$ stands for Gaussian white noise.

Under appropriate choice of a particular model type, for instance the classical ARX, the main advantage of this approach is the possibility of recursive estimation of model parameters $\Theta(k) = \{P(k), r(k)\}$, in the ARX case using the least squares method.

ProBaSensor exploits four proven models $M_i, i = 1, \ldots 4$, from which M_1 and M_2 represent the gaugemeter and mass-flow principles respectively. Models are uniquely represented by regressors (time indices k are omitted):

$$M_1:\ D_1 = [\mathcal{F}(F, z), z, 1] \tag{2}$$

$$M_2:\ D_2 = \left[\frac{v_1}{v_2} h_1, \frac{v_1}{v_2}, 1\right] \tag{3}$$

$$M_3:\ D_3 = [h_1, z, 1] \tag{4}$$

$$M_4:\ D_4 = \left[h_1, z, \frac{v_1}{v_2}, 1\right] \tag{5}$$

III. ProBaSensor's main idea

The classical approach to AGC and to evaluation of the rolling gap size is *model switching*, that is choosing a single most suitable model according to actual working conditions (initial phase, stable rolling, rolling of welds, final part of the strip, non-standard situations, etc.).

The idea of the novel approach consists in *model averaging* [1]. All models are run in parallel and at each time instant, their estimates of the gap size (equivalently of the actual non-delayed output thickness) are mixed in a form of a convex combination, whose weights are proportional to models' likelihoods.

Generally speaking, the idea prefers mixing of all available information to selection of the best piece of it.

A. Employed theory – Bayesian approach

The Bayesian approach expresses the model (1) of $h_2(k)$ given $\Theta(k)$ and $D(k)$ in the form of a conditional probability density function (pdf) [2]

$$f\left(h_2(k)|\Theta(k), D(k)\right). \tag{6}$$

Its parameters $\Theta(k) = \{P(k), r(k)\}$ are modelled with another distribution with pdf $f(\Theta(k)|\mathcal{D}^{1:k-1}, \mathcal{H}_2^{1:k-1})$, where

$$\mathcal{D}^{1:k-1} = \{D(1), \ldots, D(k-1)\}$$
$$\mathcal{H}_2^{1:k-1} = \{h_2(1), \ldots, h_2(k-1)\}$$

are used as an accumulated knowledge of the past development. Verbally it expresses the distribution of our knowledge about model parameters $\Theta(k)$, based on our observations of pairs 'previous regressor–previous measurement'. The Bayes' rule then recursively corrects this knowledge and gradually updates the distribution by incorporation of new data,

$$f(\Theta(k)|\mathcal{D}^{1:k}, \mathcal{H}_2^{1:k}) = \frac{f\left(h_2(k)|\Theta(k), D(k)\right) f\left(\Theta(k)|\mathcal{D}^{1:k-1}, \mathcal{H}_2^{1:k-1}\right)}{f\left(h_2(k)|\mathcal{D}^{1:k}, \mathcal{H}_2^{1:k-1}\right)}. \tag{7}$$

The denominator in (7) serves as a normalizing constant, independent of $\Theta(k)$. If the prior parameter pdf is chosen to be conjugate to the model, the posterior attains the same form as the prior and can be used as the prior in the next time instant [3]. Furthermore, the existence of sufficient statistics implies that the knowledge is aggregated without increasing the dimension.

The predictive pdf $f\left(h_2(k+1)|D(k+1), \mathcal{D}^{1:k}, \mathcal{H}_2^{1:k}\right)$ provides the Bayesian prediction. It follows from the marginalization equation

$$f\left(h_2(k+1)|D(k+1), \mathcal{D}^{1:k}, \mathcal{H}^{1:k}\right) = \int f(h_2(k+1)|\Theta(k), D(k+1))\, f\left(\Theta(k)|\mathcal{D}^{1:k}, \mathcal{H}_2^{1:k}\right) \mathrm{d}\Theta(k), \tag{8}$$

cf. with the denominator in (7). The point estimate is then equivalent to the mean of the distribution.

B. Parameter estimation in regressive models

The Bayesian regressive models (6) equivalent to the deterministic model (1) has the form of a normal distribution

$$h_2(k)\Big| P(k), r(k), D(k) \sim \mathcal{N}\big(P'(k)D(k), r(k)\big). \tag{9}$$

Under the lack of knowledge of $r(k)$, the conjugate prior to this model is of the normal inverse-gamma distribution $\mathcal{N}i\mathcal{G}(V(k), \nu(k))$ with the extended (symmetric square) information matrix $V(k)$ of dimension $N = m + 1$ and scalar degrees of freedom $\nu(k)$ as hyperparameters, i.e. a compound distribution where the normal part serves for estimation of regression coefficients P given r and h_2, and the inverse-gamma part is used for the unknown variance r.

The Bayes' rule (7) updates the hyperparameters by new data,

$$V(k) = V(k-1) + \begin{bmatrix} h_2(k) \\ D(k)' \end{bmatrix} \begin{bmatrix} h_2(k) \\ D(k)' \end{bmatrix}' \qquad (10)$$

and

$$\nu(k) = \nu(k) + 1. \qquad (11)$$

The point estimator of $P(k)$ is [2]

$$\hat{P}(k) = \begin{bmatrix} V_{21} \\ \vdots \\ V_{N1} \end{bmatrix}' \begin{bmatrix} V_{22} & \dots & V_{2N} \\ \vdots & \ddots & \vdots \\ V_{N2} & \dots & V_{NN} \end{bmatrix} \qquad (12)$$

where N is dimension of the positive definite square matrix $V(k)$. The point prediction of $h_2(k+1)$ given the regressor $D(k+1)$, equivalent to the mean of the predictive pdf (8) is then

$$\hat{h}_2(k+1) = \hat{P}(k)D(k+1). \qquad (13)$$

The relation (12) is equivalent to the recursive least squares (RLS), as can be found together with other details in [2]. Since the information matrices are often ill-conditioned and their inversions can lead to significant numerical issues, they are usually evaluated in factorized forms, e.g. Cholesky's LU or LDL' (the latter is used in ProBaSensor).

The described approach to modelling neglects the potential variability of estimated parameters $P(k), r(k)$. If the parameters vary slowly, they can be estimated using various techniques, e.g. the exponential forgetting [2], directional forgetting [4] or partial forgetting [5]. Another possibility is finite data window approach, however, at the cost of higher computational burden.

It is often advisable to restrict estimates of the parameters based on knowledge of the modelled process. The idea of *bounded estimation* consists in running two models, one ordinary and the other reduced. When the ordinary model's parameter of interest exits the set of allowed values, the reduced model takes the responsibility for further estimation, leaving the problematic parameter fixed. Sensitive parts of the algorithm [6] lie in situations when the estimates cross their boundaries.

ProBaSensor enables to switch manually among three types of estimators differing in concrete implementation of the above-mentioned principles.

C. Model mixers

The model mixer follows the idea of *dynamic model averaging* [1]. Assume, that the models $M_1, \dots, M_4$, i.e. (2) – (5), are run in parallel and that their suitability for online prediction of the variable of interest is uncertain. This uncertainty can be reflected by *weights* $w_i(k), i = 1, \dots, 4$, expressed as the probability that the true model $\tilde{M}(k)$ at time k is the ith one,

$$w_i(k) = \Pr\left(\tilde{M}(k) = M_i(k) | \mathcal{D}_i^{1:k}, \mathcal{H}_i^{1:k}\right).$$

Clearly,

$$w_i(k) \in [0,1] \quad \text{and} \quad \sum_{i=1}^{4} w_i(k) = 1. \qquad (14)$$

The Bayes' rule then incorporates the predictive ability of each model (measured by likelihood, cf. (7) and (8)) into its weight,

$$w_i(k) \propto w_i(k-1) f_i\left(h_2(k)|D(k)\right),$$

where $\propto$ denotes proportionality, i.e. equality up to a normalizing constant, ensuring (14). Similarly to the estimation of models' parameters, it is necessary to reflect a potential drift of weights w_i in time. Again, exponential or other forgetting methods can be used.

Finally, the ProBaSensor's output is the point estimate of h_2, yielded by the convex combination of point predictions $\hat{h}_{2,i}$ of models $M_i, i = 1, \dots, 4$,

$$h_2(k) = \sum_{i=1}^{4} w_i(k)\hat{h}_{2,i}(k).$$

It was shown in [7] that averaging over several available models leads to better average predictive ability than any single model.

Model mixing is potentially sensitive to transition states, when a measured variable, used in a dominating model, accidentally changes its value. During the stabilization phase, the dominant model and its weight continually adapt to the new state. To smoothen the prediction of h_2, it can be advantageous to add yet one more modelling level. The autoregressive model was tested for this purpose in [8].

Again, ProBaSensor offers alternatives in the form of two available mixer types differing in the way how to guarantee the conditions (14).

IV. Implementation

The ProBaSensor's framework described above was implemented into industrial hardware in the form of a software package, deployed together with a set of supportive applications as a part of the rolling mill control system. The structure and its integration is shown in Fig. 2.

The actual ProBaSensor system runs on a dedicated computer under the real-time Linux OS. ProBaSensor MMI (Man-Machine Interface) node runs a monitoring application which provides on-line information about the system and enables adjustments of various options and parameters.

Analysis of the network traffic showed that the network load is still well bellow its limits even for the shortest intended sampling period of 2 ms and that provisional extension of the existing control system has no negative impact on its operation.

Block diagram of the ProBaSensor core can be seen in Fig. 3. Data buffers are realized by the memory resident

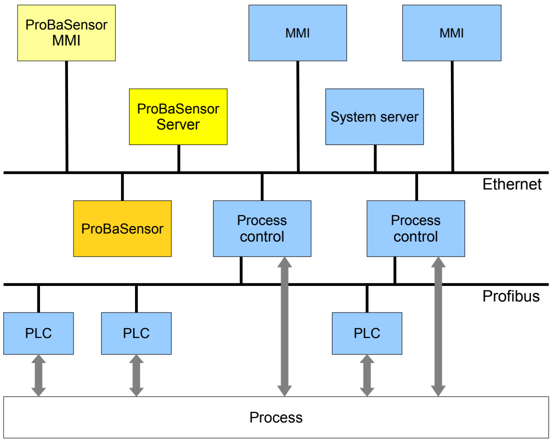

Fig. 2. Scheme of the experimental integration of new nodes into a real control system of a rolling mill.

database which includes interfaces for both local and remote (networked) types of access, comprising the mutex (mutual exclusion) and FIFO queuing mechanisms. Main modules of the system are executed at the kernel level of the operating system to ensure hard real-time timing.

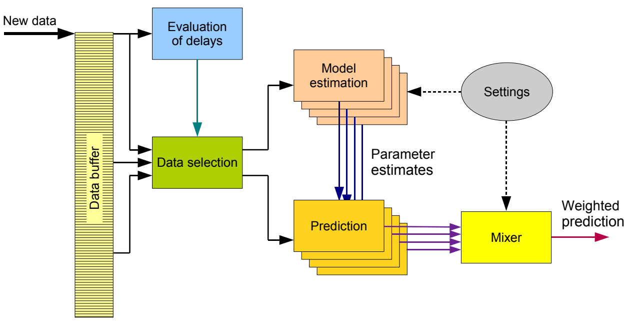

Fig. 3. Block diagram of the ProBaSensor core.

The module responsible for data selection and pre-processing provides – together with the delay elimination module – data for particular models, parameters of which are estimated recursively and used for evaluation of predictions. The mixer includes estimation of weights of each of the models and constructs the overall weighted prediction of the variable in question. The modules enable various settings concerning types of estimators, inclusion or exclusion of particular models, introduction of the externally computed offset, forgetting factors, limits, etc. These options and parameters can be selected or changed in the supporting monitoring application whose main window is shown in Fig. 4. Here the situation corresponds to experiments on the rolling mill producing relatively thick copper strips (see below). The prediction of the output thickness is displayed in microns by the largest numbers while the other labels display various groups of input data, estimated model parameters, etc.

Fig. 4. Main window of the monitoring application.

V. Tests

The ProBaSensor system was in the first instance tested within a simulated environment on real data originating from various rolling mills. Afterwards, the system was successively integrated into the control system of two real rolling mills. The first set of experiments was conducted on the four-high reversing rolling mill $\mathcal{S}$ processing steel strips. The second on the rolling mill $\mathcal{C}$ with similar arrangement of rolls processing copper, brass and other copper alloys strips. Each testing period lasted several weeks.

A. *Combinations of settings*

As mentioned above, the experimental ProBaSensor system is able to switch among 3 types of estimation algorithms for each process model and 2 types of the model mixer. Together with other options, there exist altogether 256 theoretical combinations of system settings. A reasonably smaller subset of combinations was selected for experiments.

B. *Real-data tests within simulated environment*

Initial period of testing exploiting real data within the simulated environment alternated successively 30 most promising combinations of settings. Two examples comparing predicted output thickness with its measured values can be seen in Fig. 5 and 6.

The first plot in Fig. 5 shows the predicted deviation of thickness in the rolling gap by the red line while the actual measured value h_2 is drawn in blue. Both signals were shifted in time to allow direct comparison. Here the prediction is almost excellent when we disregard the obviously increased high frequency noise in the predicted value which would be naturally filtered by dynamic properties of the roll positioning

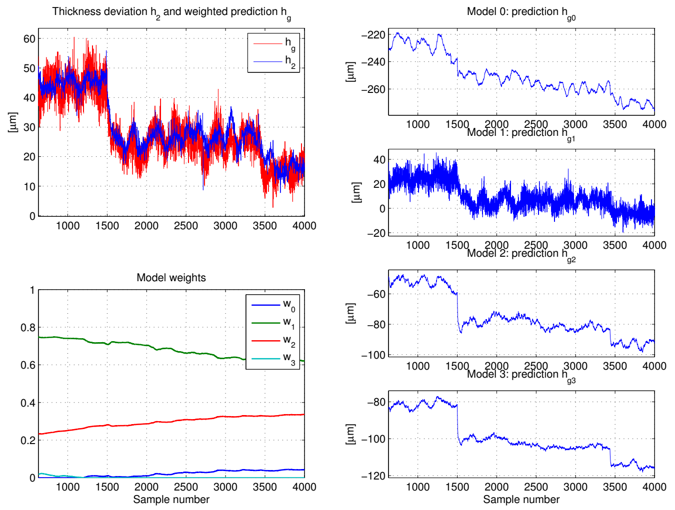

Fig. 5. Example of predictions and model weights for one of favorite settings of the system.

system in case of a feedback control. Achievement of the proper synchronization of the prediction and h_2 is the most important. The plot bellow shows progress of weights w_i of particular predictions $h_{g\,i}$ which are plotted separately in the right column.

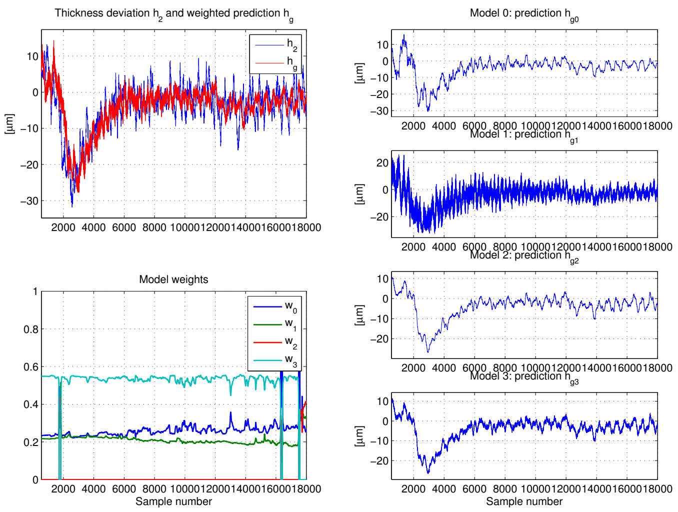

Fig. 6. Another example of predictions and model weights for one of favorite settings of the system.

Fig. 6 shows another situation with a fair prediction of a "wave" of h_2. Single predictions are not so different when we neglect increased noise of the second one.

Fig. 7 corresponds to the best combination of settings among the 30 tested adjustments and it depicts the proportion of successfulness of particular process models. The biggest plot represents histogram of identified model weights, the bar plot shows averages of model weights and the pie plot enables

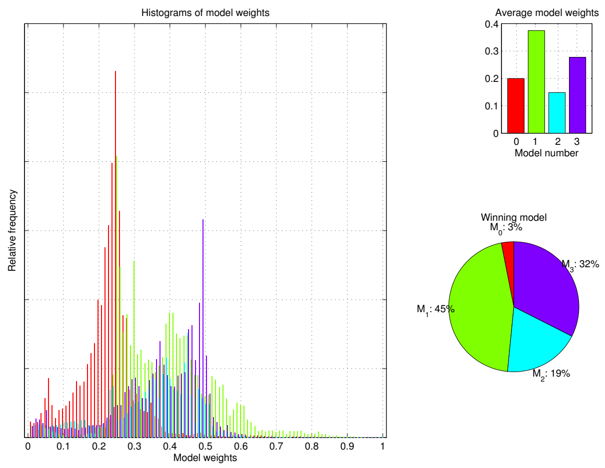

Fig. 7. Winning model for the best group of settings.

to evaluate how successful were the models in average.

C. Tests on the rolling mill $\mathcal{S}$

After statistical analysis of preceding results and their visual inspection, the number of combinations for the first period of industrial tests on rolling mill $\mathcal{S}$ was decreased to 14. Histograms of prediction errors shown in Fig. 8 for various configurations of settings were used to allow simple comparison of results. Again, this statistical method was just a basic one, additional expert evaluation was unavoidable.

The left plot shows histograms of prediction errors for all 14 tested combinations while the right plot represents the winning adjustment.

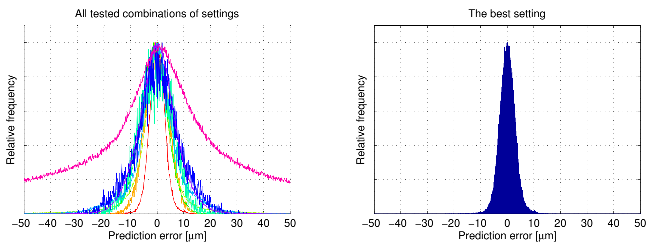

Fig. 8. Histograms of prediction error for experiments on the rolling mill $\mathcal{S}$.

D. Tests on the rolling mill $\mathcal{C}$

Fig. 9 shows histograms for experiments on the rolling mill $\mathcal{C}$ for which the number of tested settings was decreased again to 5. The results for particular combinations do not differ dramatically here but they are noticeably worse in comparison with the preceding set of experiments because of different processed material and thicker target thicknesses prescribed for the rolling mill $\mathcal{C}$.

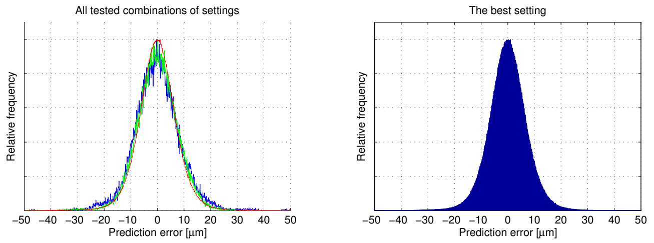

Fig. 9. Histograms of prediction error for experiments on the rolling mill $\mathcal{C}$.

VI. Conclusion

Extensive tests were accomplished to evaluate functionality of the ProBaSensor system. The system enables number of various adjustments which were compared in the evaluation. Main purpose of accomplished industrial experiments was to compare prediction of the strip thickness in the rolling gap with its measured value taking the inherent transport delay into account. The system must be further elaborated to allow unsupervised operation in closed loop, that is, to be used for automatic gauge control.

The experiments proved functionality of the developed system. At the same time, they confirmed the expectation that coping with uncertainty close to the level of measurement accuracy leads rarely to clear determination of the best method to be utilized. Continuous decision making from a set of few favorite methods seems to provide the best possible results.

Acknowledgment

This research was supported by project MŠMT 7D09008 (ProBaSensor). Some of its generic results are being elaborated in project MŠMT 7D12004 (ProDisMon) support of which is gratefully acknowledged.

References

[1] A. Raftery, M. Kárný, and P. Ettler, “Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill,” *Technometrics*, vol. 52, no. 1, pp. 52–66, 2010.

[2] V. Peterka, “Bayesian Approach to System Identification In P. Eykhoff (Ed.) Trends and Progress in System Identification,” 1981.

[3] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory (Harvard Business School Publications)*. Harvard University Press, Jan. 1961.

[4] R. Kulhavý and M. B. Zarrop, “On a general concept of forgetting,” *International Journal of Control*, vol. 58, no. 4, pp. 905–924, 1993.

[5] K. Dedecius, I. Nagy, and M. Kárný, “Parameter tracking with partial forgetting method,” *International Journal of Adaptive Control and Signal Processing*, vol. 26, no. 1, pp. 1–12, 2012. [Online]. Available: http://dx.doi.org/10.1002/acs.1270

[6] P. Ettler and M. Kárný, “Parallel estimation respecting constraints of parametric models of cold rolling,” in *Proceedings of the 13th IFAC Symposium on Automation in Mineral, Mining and Metal Processing (IFAC MMM 2010)*, Cape Town, South Africa, 2010, pp. 63–68.

[7] D. Madigan and A. E. Raftery, “Model selection and accounting for model uncertainty in graphical models using occam’s window,” *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1535–1546, 1994. [Online]. Available: http://www.jstor.org/stable/2291017

[8] K. Dedecius and P. Ettler, “Adaptive continuous hierarchical model-based decision making for process modelling with realistic requirements,” in *Proceedings of the 8th International Conference on Informatics in Control, Automation and Robotics - ICINCO 2011*, Noordwijkerhout, The Netherlands, 2011.