Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Multirate fusion of data sources with different quality

Joel Sansana[a], Ricardo Rendall[a], Zhenyu Wang[b], Leo H. Chiang[b], Marco S. Reis[a,*]

[a] *CIEPQPF - Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, Pólo II, 3030-790 Coimbra, Portugal*
[b] *Continuous Improvement Center of Excellence, Dow Inc., Lake Jackson, TX, United States of America*

Abstract

The chemical process industry makes increasingly use of a diversity of data collectors, that should be properly integrated to build effective solutions for process monitoring, control and optimization. Concerning the assessment of products properties, one of the most common scenarios involve the collection of data from plant laboratories that provide more accurate measurements at lower rates, together with more frequent measurements or predictions of lower quality. Soft sensors and online analyzers are examples of viable alternatives for acquiring more frequent and updated information, although with a higher uncertainty. All of these data collectors have informative value and should be considered when it comes to estimate key product attributes. This is the goal of fusion methods, whose importance grows together with the increase in the number of sensors and data sources available. In this article, two fusion schemes that address prevailing characteristics of industrial data are proposed and compared: one version of the classic tracked Bayesian fusion scheme (TBF) and a novel modification of the track-to-track algorithm, designated as bias-corrected track-to-track fusion (BCTTF). The proposed methodologies are able to cope with the multirate nature of data and irregularly sampled measurements that present different uncertainty levels. An application to a real industrial case study shows that BCTTF presents better prediction performance, higher alarm identification sensitivity and leads to a smoother estimated signal.

Keywords: Sensor fusion, Bayesian fusion, Kalman filter, Machine learning, Industrial case study

1. Introduction

Soft sensors and online analyzers (OAs) are effective tools for timely and frequent process monitoring [1]. Soft sensors, also known as inferential sensors, virtual online analyzers and observer-based sensors, are predictive models (mechanistic, data-driven or hybrid) that provide information on a process or a relevant key performance indicator (KPI) [2]. These KPIs are often variables that cannot be directly measured (*e.g.* concentrations and quality attributes) [3]. Furthermore, these quality variables difficult to estimate online, require complex lab analysis that often introduce significant time delays [4]. By combining information from different sources, the quality and timeliness of state estimation can be improved, improving process monitoring, control and operational decision-making [5, 6, 7]. Soft sensors require regular maintenance as its prediction accuracy tend to degrade over time due to various factors, such as changes of process conditions, process drifting and equipment degradation [8].

In the literature, two types of approaches have been adopted more intensively to address the multi-sensor data fusion problem: Bayesian inference and state estimation. The former approach models the behavior of each sensor with a Gaussian probability density function (PDF) and combines the individual sensor PDF into a joint PDF. A likelihood function is then established based on the joint PDF, and its parameters are estimated using data. This state estimate corresponds to the maximum a posteriori (MAP) estimate [9]. As an example, Wang and Chiang have applied Bayesian inference based methods to improve the monitoring of industrial distillation column [10, 11].

The second approach, state estimation, is based on Kalman filtering [12], known as track-to-track fusion (TTF) [13]. The Kalman filter also has a Bayesian origin, but the specificity of TTF algorithms is high, justifying a separate treatment. More specifically, TTF makes use of a Kalman filter to track each sensor and fusing the sensors' state estimates into a new state es-

*Corresponding author
Email address: `marco@eq.uc.pt` (Marco S. Reis)
URL: `https://www.eq.uc.pt/~marco/` (Marco S. Reis)

Preprint submitted to IFAC 2020 *May 4, 2020*

timate generated from a static linear estimation equation [14]. Gao and Harris [5] reviewed some measurement fusion approaches and proposed a modified track-to-track fusion (MTF) that guaranteed the optimality of the estimate. On the oter hand, Haque et al. [15] proposed a Monte Carlo Bayesian inference approach to improve floor localization in indoor positioning systems. This method processes the signal strengths of Wi-Fi measurements and fuse them with barometric altimetry through a Kalman filter scheme.

Previous works have essentially considered the commonly used linear, time-invariant, discrete-time model to make state predictions. In this work, state predictions will be given by a stochastic partial least squares (PLS) model which presents some challenges, particularly for estimating the noise variance-covariance matrix [16]. The key contribution is a new fusion scheme for handling multirate and irregularly sampled data, that weights the quality of the different measurement sources while computing the fused estimates. This problem has been getting more attention recently, with researchers proposing different approaches to deal with asynchronous sensors including smoothing [17], combining KF with a neural network [18] and discrete-time linear modeling [19, 1].

The proposed methodologies were tested using real data from a refining unit of a two trayed distillation columns from a Dow production facility. Process measurements include pressures, temperatures and flow rates across the columns, as well as reflux ratio and online analyzer readings from upstream reactors. Quality variables are concentrations of impurity components. These responses are determined every hour by an online analyzer and a soft sensor. Furthermore, every 12 hours a new sample is collected for laboratory analysis [20].

This article is organized as follows. In the next section, the proposed fusion methodologies are described in detail. Then, in the third section we present the results obtained by the application of the fusion schemes to real plant data, where different sources of information are available for the relevant response variables: online analyzer, gas chromatography and soft sensor. The main conclusions are summarized in the final section of this paper.

2. Materials and Methods

In this section, the algorithms for the classical tracked Bayesian fusion and for the proposed bias-corrected track-to-track fusion methodologies are presented.

2.1. Tracked Bayesian fusion

Measurements can be fused to track model predictions using Kalman filters. When two or more sensor measurements are available, it is common practice either to concatenate all measurements into a vector of observations for the Kalman filter update step, or alternatively, to combine the measurements using minimum mean square error estimation [5]. The latter approach is followed in this work, but instead of minimum mean square error approach, the estimate is obtained through Bayesian fusion. Given the multirate nature of the problem, Bayesian fusion takes place at sparser time intervals while measurements are inputted directly into the Kalman filter when data collected from the more frequent sensor becomes available. (See Figure 1).

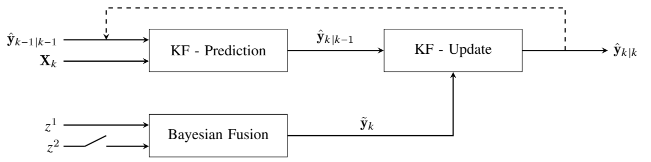

Figure 1: A schematic representation of the TBF process. As inputs, the Kalman filter prediction step takes process variables ($\mathbf{X}_k$) and the previous estimated state ($\hat{\mathbf{y}}_{k-1|k-1}$). Measurements from sensors (z^1,z^2), once fused ($\tilde{\mathbf{y}}_k$) update the Kalman filter's prediction $\hat{\mathbf{y}}_{k|k-1}$), yielding the new estimated state ($\hat{\mathbf{y}}_{k|k}$).

It is often the case in the chemical industry that a detailed description of the process is not available because the phenomena is not well understood or due to difficulties in estimating the model parameters. This leads to the impossibility to build first principle models with the necessary rigor and robustness to be applied in practice. On the other hand, machine learning and statistical methods can help to model the system behavior using data from the processes' normal operating conditions. A candidate data-driven modeling technique for this task is partial least squares (PLS), given its simplicity, robustness and effectiveness in handling highly correlated datasets [21].

2.2. Bias-corrected track-to-track fusion

As previously discussed, sensors may have different sampling rates with different associated uncertainties. Furthermore, there is usually one measurement source that assumes the role of the "reference source", "golden standard" or "ground truth". It is important to incorporate this knowledge in the fusion algorithm, as well as the different measurement uncertainties involved. This is the main goal of the new devised fusion framework, called bias-corrected track-to-track fusion (BCTTF).

This approach is able to correct for systematic biases that often occur in the more frequent sensors, while still outputting reliable predictions of the response variable. The rationale behind this method is that since it is known that one sensor is much more reliable and less biased than the others (the reference sensor), this sensor's measurements should be used to correct any bias the more frequent sensors may have. In industrial practice, the sources acting as golden standard devices are located in the laboratories where they are operated under strict conditions and according to the best metrology practices [22]. Their quality is continuously monitored and compared against metrological standards in order to detect and immediately correct any malfunction, including measurement bias. Under these circumstances, it is reasonable to assume or consider them to be bias free data sources. In order to accommodate this correction scheme, the Kalman filter update step is modified in order to include a bias correction term, θ, yielding,

$$\hat{\mathbf{y}}^m_{k|k} = \hat{\mathbf{y}}_{k|k-1} + \mathbf{K}^m_k \left(z^m_k - \mathbf{H}^m \hat{\mathbf{y}}_{k|k-1} - \theta^m \right) , \qquad (1)$$

where

$$\theta^m = \begin{cases} E\left[z_{\mathrm{F}} - z_{\mathrm{NF}}\right], & \text{if } m = \text{Frequent sensor (F)} \\ 0, & \text{if } m = \text{Non-frequent sensor (NF)}, \end{cases} \qquad (2)$$

with $E[.]$ being the expected value, z_F is the measurement from the frequent sensor and z_{NF} is the measurement from the non-frequent sensor.

The assumption that laboratory measurements are the most accurate makes them the benchmark to estimate the process noise variance matrix, computed as variance of the difference between laboratory measurements and PLS predictions.

Estimation of the laboratory noise variance matrix, $\hat{R}_{\mathrm{NF}}$, is subjected to prior knowledge from laboratory technicians. Thus, computing $\hat{R}_{\mathrm{NF}}$ follows a simple model for the heteroscedastic behavior of measurement noise. Note that $\hat{R}_{\mathrm{NF}}$ can be used both by TBF and BCTTF in the Kalman filter update step. It is also used in the TBF's Bayesian fusion step to estimate the NF noise standard deviation.

Regarding the more frequent sensor, we propose the use of the wavelet transform as non-linear filters that separate noise from the true underlying signal. Then, the variance of the noise component of the signal is used as an estimate of $\hat{R}_{\mathrm{F}}$. These variances are computed in a dyadic moving window.

One way to improve TBF, is to integrate some information connected to the systematic bias of the frequent sensor. This can be achieved by adding a squared bias term, θ^2, to the variance of the noise from this sensor, leading to an estimated variance that is different from the one computed with bias correction and that reflects the decreasing reliability of their outcomes. In this context, the bias is not corrected, but is consistently incorporated in the TBF framework. Furthermore, the computations of $\hat{Q}$, $\hat{R}_{\mathrm{NF}}$ and $\hat{R}_{\mathrm{F}}$ quantify the quality of the sensor measurements, constituting by itself an interesting element of analysis, as can be verified on the Results section.

Finally, the proposed track-to-track fusion algorithm fuses the state estimates, $\hat{\mathbf{y}}^1_{k|k}$ and $\hat{\mathbf{y}}^2_{k|k}$, each one of them associated with a sensor type, into a new optimal state estimate $\hat{\mathbf{y}}_{k|k}$ that is the input to the Kalman filter in the next stage [5, 1]. The multirate challenge is tackled within this methodology by bypassing this fusion center when only the more frequent sensor is available. In this case, no fusion takes place and the algorithm proceeds by making a prediction using information only from the soft sensor and updating it with the more frequent measurements. Furthermore, it is assumed that a frequent measurement is always available with the non-frequent one. A scheme of the present framework is presented in Figure 2.

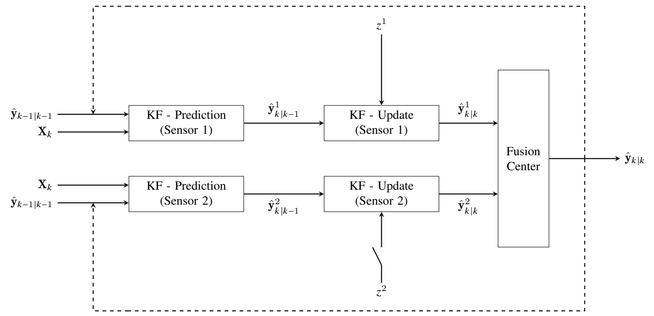

Figure 2: The BCTTF fusion process. As inputs, both Kalman filter prediction steps take process variables ($\mathbf{X}_k$) and the previous estimated state ($\hat{\mathbf{y}}_{k-1|k-1}$). Measurements from sensors (z^1,z^2), update each of the Kalman filter's predictions $\hat{\mathbf{y}}^{1,2}_{k|k-1}$) which are fused in the fusion center, yielding the new estimated state ($\hat{\mathbf{y}}_{k|k}$).

3. Results

In this section, the aforementioned fusion schemes are applied in a real industrial case study and compared. A brief description of the case study is first presented followed by the results of the fusion schemes and an analysis of their advantages and limitations.

3.1. Case study

The dataset provided for this study comes from the refining section of a production process. The process comprises two trayed distillation columns, namely a primary column and a rectifying column, as shown in Figure 3. The dataset includes measurements of pressures, temperatures and flow rates across the columns, as well as reflux ratio and online analyzer readings from upstream reactors [20]. The quality outputs (target responses) are the concentration of impurities y_1 and y_2 which are measured by two sensors: an online analyzer (OA) sampling at an hourly rate and a gas chromatography (GC) analysis performed every 12 hours. The measurements units can not be disclosed for proprietary reasons. During 8 months of operation data, plant technicians reported a true alarm in the third month. The true alarm occurs when the sum of the concentrations of impurities is above a threshold value.

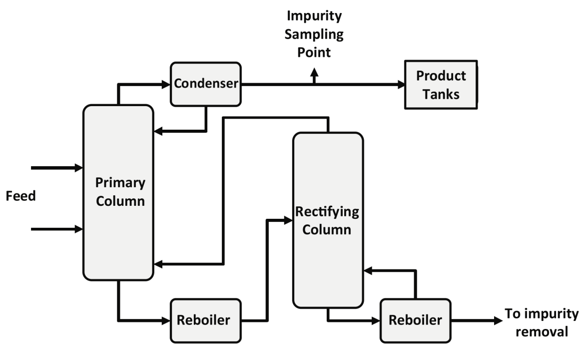

Figure 3: Simplified process flow diagram of the refining process studied.

3.2. A comparative analysis of sensor fusion using TBF and BCTTF

This section presents the results for both fusion schemes. Every plot of the response is paired with a plot of the relative sensor uncertainty (RSU) computed by the fusion scheme for each one of the three sensors. These range from 0 (small uncertainty) to 1 (high uncertainty), representing the relative magnitude of the uncertainty for the data streams, using the estimation of $\hat{R}_{\mathrm{OA}}$, $\hat{R}_{\mathrm{GC}}$ and $\hat{Q}$ (note that, in the present case all these quantities are scalar), as follows:

$$\mathrm{RSU}_{\mathrm{OA}} = \frac{\hat{R}_{\mathrm{OA}}}{\hat{R}_{\mathrm{OA}} + \hat{R}_{\mathrm{GC}} + \hat{Q}}, \tag{3}$$

$$\mathrm{RSU}_{\mathrm{GC}} = \frac{\hat{R}_{\mathrm{GC}}}{\hat{R}_{\mathrm{OA}} + \hat{R}_{\mathrm{GC}} + \hat{Q}}, \tag{4}$$

$$\mathrm{RSU}_{\mathrm{PLS}} = \frac{\hat{Q}}{\hat{R}_{\mathrm{OA}} + \hat{R}_{\mathrm{GC}} + \hat{Q}}. \tag{5}$$

Results for TBF

Figure 4 shows the results for TBF. On top, we can see that the true alarm at sample 2371 is correctly identified by the method and three GC incorrect measurements (samples 3327, 3735 and 3741) are rejected by the sensor fusion algorithm. The bottom plot shows how the RSU changes over time. While GC measurements are consistently close to zero, OA and PLS soft sensor RSUs vary, compensating each other. In this method, an increase of the OA's noise and its deviation from GC measurements, causes $\hat{R}_{\mathrm{OA}}$ to be higher, increasing the OA relative uncertainty. Upon penalization of the OA, the method relies more in the PLS soft sensor.

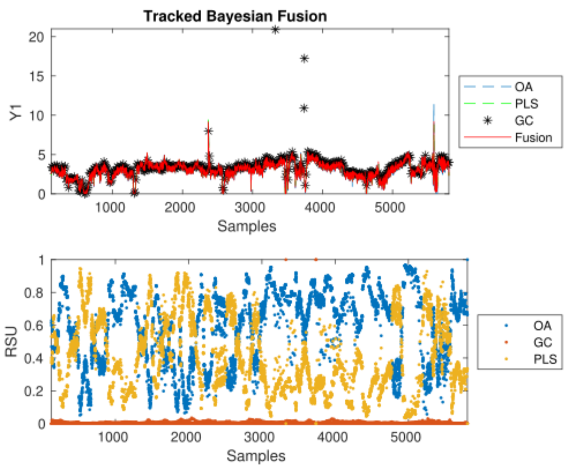

Figure 4: TBF fused signal (top) and the relative weight of each sensor for Y1 (bottom).

Results for BCTTF

Figure 5 shows how the proposed BCTTF method performed. The true alarm at 2371 is correctly detected and the GC and OA incorrect measurements are successfully rejected. In this method, the OA RSU are lower than in TBF, because θ is not added as a penalty to $\hat{R}_{\mathrm{OA}}$. Instead, θ is used to correct the OA measurements in the Kalman filter update step.

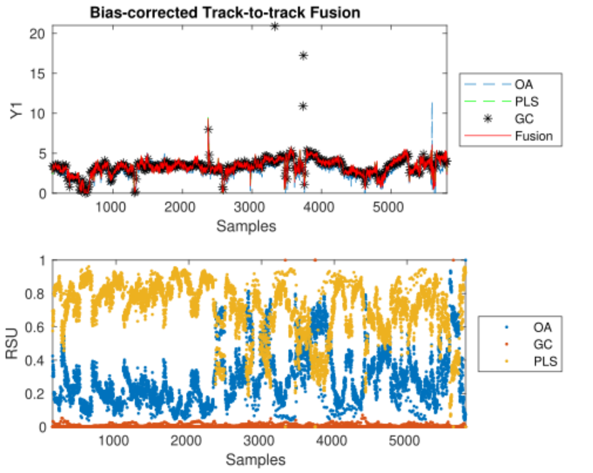

Figure 5: BCTTF fused signal (top) and the relative weight of each sensor for Y1 (bottom).

BCTTF is also capable of detecting when the OA fails but as it corrects its trajectory, this method does more than just discarding the failing sensor. BCTTF can still collect some information from the OA relative to its dynamic trend. This feature can be observed in sample 5591, where the OA signals a false alarm and BCTTF is capable of disregarding it because of the bias correction made to the OA measurements. On the other hand, TBF does not correct the OA measurement, and issues a false alarm.

Discussion

The sensor fusion methods presented in this work provide an efficient way to improve the quality of difficult to measure quantities. From three different signals (although the fusion schemes can be expanded to deal with n sensors) measuring the same variable, these methods allow for the reconstruction of a new signal which is smoother, has the fastest sensor's sampling rate and is more accurate than any of the sensors alone.

The quantification of uncertainty is an interesting byproduct of these methods. It is possible to follow the evolution of the relative sensor uncertainty in the fusion processes through the estimation of the noise variance matrices. This feature can even be used as an indicator for the need of sensor maintenance.

Finally, regarding the ability to detect alarms, both fusion schemes performed well and no miss-identifications are reported. Nevertheless, BCTTF is shown to be the best fusion scheme between the two because it did not report any false alarms while TBF was influenced by the OA values (see Table 1).

Table 1: List of alarms signaled by sensors or fusion schemes.

Sample Index	Alarm Type	Lab Analyzer	Online Analyzer	PLS	TBF	BCTTF
2371	True	✓	✓	✓	✓	✓
3327	False	✓				
3735	False	✓				
3741	False	✓				
5591	False		✓		✓	

4. Conclusions

In this work, we report the development and test of two new sensor fusion schemes that integrate information collected from different sources, namely a data-driven dynamic soft sensor, an online analyzer and a laboratory analysis (GC). The proposed schemes share the capability of managing asynchronous sampling rates and faulty measurements while reducing false alarms and miss-identifications. They also characterize the different measurement sources according to their relative uncertainty. The quality of the resulting fused signal is higher than for each of the measurement sources, yielding estimates closer to the ground truth. Furthermore, the fused signal is less noisy and more robust which are positive features for control and process supervision applications.

Regarding the multirate problem, both fusion schemes were able to handle quite well the different rates of information sources, capitalizing on the precision of the non-frequent sensor to improve the quality of estimates. Note that there is no requirement for the acquisition rates to be regular in either of the fusion schemes. The sampling rate can be changed without the need of major retuning, except perhaps for the adjustment of the moving window size. Overall, the BCTTF has proved to be superior to the TBF in our case study, where the non-frequent sensor is much more precise than the more frequent one.

While the application of the new method to an industrial dataset grants it validity from a practical point of view, the method's performance can be more finely assessed comparing the fused signal with the true state of the system, which is only possible in a simulation study. Therefore, future activities contemplate the analysis of simulated industrial systems.

References

[1] A. Fatehi, B. Huang, Kalman filtering approach to multi-rate information fusion in the presence of irregular sampling rate and variable measurement delay, Journal of Process Control 53 (2017) 15–25 (May 2017). doi:10/f96r7r.

[2] P. Kadlec, B. Gabrys, S. Strandt, Data-driven Soft Sensors in the process industry, Computers & Chemical Engineering 33 (4) (2009) 795–814 (Apr. 2009). doi:10/cpcgqq.

[3] F. A. A. Souza, R. Araújo, J. Mendes, Review of soft sensor methods for regression applications, Chemometrics and Intelligent Laboratory Systems 152 (2016) 69–79 (Mar. 2016). doi:10.1016/j.chemolab.2015.12.011.

[4] M. Pottmann, B. A. Ogunnaike, J. S. Schwaber, Development and Implementation of a High-Performance Sensor System for an Industrial Polymer Reactor, Industrial & Engineering Chemistry Research 44 (8) (2005) 2606–2620 (Apr. 2005). doi:10/cc2gkf.

[5] J. Gao, C. Harris, Some remarks on Kalman filters for the multisensor fusion, Information Fusion 3 (3) (2002) 191–201 (Sep. 2002). doi:10/dr8r8w.

[6] Z. Geng, Y. Han, X. Gu, Q. Zhu, Energy Efficiency Estimation Based on Data Fusion Strategy: Case Study of Ethylene Product Industry, Industrial & Engineering Chemistry Research 51 (25) (2012) 8526–8534 (Jun. 2012). doi:10/f33wzb.

[7] K. Sivaramakrishnan, A. Puliyanda, D. T. Tefera, A. Ganesh, S. Thirumalaivasan, V. Prasad, A Perspective on the Impact of Process Systems Engineering on Reaction Engineering, Industrial & Engineering Chemistry Research 58 (26) (2019) 11149–11163 (Jul. 2019). doi:10/gf7k5m.

[8] K. Chen, I. Castillo, L. H. Chiang, J. Yu, Soft Sensor Model Maintenance: A Case Study in Industrial Processes$_{**}$The authors would like to acknowledge the support from the DOW chemical company and the natural sciences and engineering research council of Canada (NSERC)., IFAC-PapersOnLine 48 (8) (2015) 427–432 (Jan. 2015). doi:10.1016/j.ifacol.2015.09.005.

[9] M. Kumar, D. P. Garg, R. A. Zachery, A Method for Judicious Fusion of Inconsistent Multiple Sensor Data, IEEE Sensors Journal 7 (5) (2007) 723–733 (May 2007). doi:10/bs5cf7.

[10] Z. Wang, L. Chiang, Monitoring Chemical Processes Using Judicious Fusion of Multi-Rate Sensor Data, Sensors 19 (10) (2019) 2240 (May 2019). doi:10/gf69mm.

[11] Z. Wang, L. Chiang, Hard and soft sensors fusion for process monitoring: An industrial application, in: Proceedings of the ISA 63rd Analysis Division Symposium, Galveston, TX, USA, 2018, pp. 22–26 (2018).

[12] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of basic Engineering 82 (1) (1960) 35–45 (1960). doi:10/dmftj3.

[13] Y. Bar-Shalom, L. Campo, The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance, IEEE Transactions on Aerospace and Electronic Systems 22 (6) (1986) 803–805 (Nov. 1986). doi:10/dzw46n.

[14] H. Chen, T. Kirubarajan, Y. Bar-Shalom, Performance Limits of Track-to-Track Fusion vs. Centralized Estimation: Theory and Application, IEEE Trans. Aerosp. Electron. Syst. 39 (2) (2003) 386–398 (2003). doi:10/ff7njx.

[15] F. Haque, V. Dehghanian, A. O. Fapojuwo, J. Nielsen, A Sensor Fusion-Based Framework for Floor Localization, IEEE Sensors Journal 19 (2) (2019) 623–631 (Jan. 2019). doi:10.1109/JSEN.2018.2852494.

[16] B. J. Odelson, M. R. Rajamani, J. B. Rawlings, A new autocovariance least-squares method for estimating noise covariances, Automatica 42 (2) (2006) 303–308 (Feb. 2006). doi:10/bb7hcw.

[17] A. Smyth, M. Wu, Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring, Mechanical Systems and Signal Processing 21 (2) (2007) 706–723 (Feb. 2007). doi:10/b9bt7f.

[18] S. Safari, F. Shabani, D. Simon, Multirate multisensor data fusion for linear systems using Kalman filters and a neural network, Aerospace Science and Technology 39 (2014) 465–471 (Dec. 2014). doi:10/gfzbgk.

[19] Y. Guo, Y. Zhao, B. Huang, Development of soft sensor by incorporating the delayed infrequent and irregular measurements, Journal of Process Control 24 (11) (2014) 1733–1739 (Nov. 2014). doi:10/f6qqmt.

[20] B. Lu, L. Chiang, Semi-supervised online soft sensor maintenance experiences in the chemical industry, Journal of Process Control 67 (2018) 23–34 (Jul. 2018). doi:10/gdvc2j.

[21] S. Wold, M. Sjöström, L. Eriksson, PLS-Regression: A basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2) (2001) 109–130 (Oct. 2001). doi:10/cv6jxq.

[22] Working Group 1 of the Joint Committee for Guides in Metrology, Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement, 2008 (Sep. 2008).