Proceedings of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas Production, Norwegian University of Science and Technology, Trondheim, Norway, May 31 - June 1, 2012

On-line calibration monitoring system based on data-driven model for oil well sensors ⋆

Andre A. Boechat * **Ubirajara F. Moreno** *
Decio Haramura, Jr. *

* *Automation and Systems Department, Federal University of Santa Catarina, Brazil, (email: {boechat, moreno, deciohjr}@das.ufsc.br)*

Abstract: In the oilfield industry, data collected from well sensors plays an important role in performance and security applications. The quality of the measurements is directly related to the accuracy of the control actions and the optimisation of the production. On-line calibration monitoring systems can determine drifts in the sensors measurements and provide more reliable information to the user. In this paper, a robust on-line calibration monitoring system for drift correction/detection in well sensors is presented and evaluated for simulated and real data sets. Comparisons with a state-of-art monitoring system is also showed. The results indicate a promising applicability of the calibration monitoring system for the oilfield industry data.

Keywords: Performance monitoring, Sensor failures, Drift, Calibration, Prediction methods, Non-parametric regression, Kalman filters.

1. INTRODUCTION

In many industry sectors, the maintenance strategy is based on traditional approaches to sensor validation, which involves periodic instrument calibration. Various periodic sensor calibration techniques require the process be shut down and the instrument taken out of service, loaded and calibrated. In some cases, the costs to repair a sensor are so high that it is left faulty and its data is just ignored, and wrong decisions could be made because of the lack of information.

In the oilfield industry, permanent downhole sensors and subsea sensors, for offshore platforms, have a great importance for control actions, production optimisation and well monitoring. Due to the harsh environmental conditions in which these sensors are deployed, their reliability is degraded throughout the well life. However, as Aggrey and Davies (2007) stated, "replacement of a failed sensor rarely occurs in practice even when data is known to be incorrect (or completely missing) due to the cost implication of a workover". In addition, Eck et al. (1999) point out about permanent downhole gauges, "once in place, the devices are not routinely repaired, replaced or recovered".

For these reasons, less invasive and more efficient maintenance strategies are desirable. Condition based maintenance techniques can lead to optimal maintenance, once the steady state performance of instruments are monitored during plant operation and physical recalibrations are performed only when their performance is degraded.

On-line calibration monitoring consists essentially in estimating the correct measurements that the sensors should have read and monitoring continuously the difference between estimate values and values of the sensors. Hardware redundancy can be much expensive, and it may be not so useful to detect faulty sensors that drift in the same direction. Moreover, for downhole gauges, redundant components occupy valuable, limited space and consume precious power. Analytical redundancy by models based on physical equations can generates accurate results, but it is always hard or impossible to obtain in practice. Empirical models developed with historical data also rely on relationships between correlated measurements within a system, but these relationships are formulated in a implicitly way by training the model through analysis of fault-free training data obtained during normal operations.

Success of empirical model applications has already been reported in industry, like nuclear power industries, as reported in Gribok et al. (2000) and Ma and Jiang (2011). There are some works where artificial neural networks were applied (Aggrey and Davies (2007)), but kernel based techniques are among the most used techniques. The main techniques are based on *Auto-Associative Kernel Regression* (Hines et al. (2008) and An et al. (2011)), *Multivariate State Estimation Technique* (Gribok et al. (2000)) and *Support Vector Machine for Regression* (Gribok et al. (2000), and Takruri et al. (2008)).

In this paper is presented a robust on-line calibration monitoring system for oil and gas well sensors. The system is composed of an auto-associative empirical model, a Kalman filter and a statistical decision module, based on the main ideas of Hines et al. (2008) and Takruri et al. (2008). It is proposed that the association of these ideas could improve the results of a state-of-art monitoring system when applied to oil well sensors. The goals are correcting the readings of sensors and detecting drift occurrences. The system is evaluated for simulated and

⋆ This work was supported by Cenpes/Petrobras.

real data sets, and the results are compared to those presented by the state-of-art system.

In Section 2, details of the system architectures are described; in Section 3, the simulated and real data sets used to evaluate the monitoring systems are described; in Section 4, the evaluation results are shown; and finally, the conclusions are presented in Section 5.

2. MONITORING SYSTEM ARCHITECTURES

In the monitoring systems based on data-driven techniques correlated measurements within a system, fault free training data obtained during normal operations, are used to develop a empirical model. This data-driven model is used to predict future measurements, which are compared with the real measurements, generating residuals. Any faults in the system may cause statistically abnormal changes in these residuals and could be detected by performing statistical tests. If a sensor is working properly, the residuals normally has a zero mean and a variance related to the amount of noise in the signal of the sensor.

The state-of-art architecture is illustrated in Fig. 1. In this paper, this architecture is referenced as AAKR-SPRT, which is composed of two components: a Auto-Associative Kernel Regression model (AAKR) and a statistical decision logic module. The empirical model is developed using a fault free data set $\mathbf{X}_{n\times p}$, where n is the number of observations of p different process variables. When a new measurements vector $\mathbf{r}_{1\times p}$ is available, the prediction of its correct values, $\hat{\mathbf{x}}_{1\times p}$, is calculated. The difference between the sensors readings $\mathbf{r}_{1\times p}$ and the predictions $\hat{\mathbf{x}}_{1\times p}$ is called residual $\mathbf{d}_{1\times p}$. Finally, $\mathbf{d}_{1\times p}$ is analysed by a statistical module where the *sequential probability ratio test* (SPRT) is implemented, indicating a possible drift occurrence, $D \in [0, 1]$.

The architecture of the robust system, AAKR-KF-SPRT, include a Kalman filter (KF) for drift estimation, as in Takruri et al. (2008). The KF is used to track the amplitude of the drift over time, allowing a correction of the sensor readings used by the AAKR model. The idea is to reduce the drift effects on the AAKR predictions.

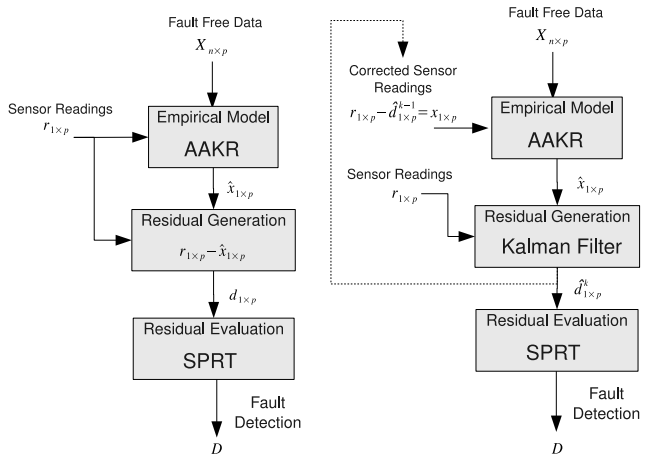

Fig. 1. Process diagrams of the monitoring system architectures: the AAKR-SPRT on left, the AAKR-KF-SPRT on right.

Like the AAKR-SPRT, a AAKR model is developed using a fault free data set $\mathbf{X}_{n\times p}$. A new measurements vector $\mathbf{r}_{1\times p}$ is corrected using a estimation of the drift at the previous stage $\hat{\mathbf{d}}^{k-1}_{1\times p}$. The corrected measurements $\mathbf{x}_{1\times p}$ are inputted into the AAKR model. The estimation of the corrected measurements, $\hat{\mathbf{x}}_{1\times p}$, and $\mathbf{r}_{1\times p}$ are used by the KF to calculate a new estimation of the drift, $\hat{\mathbf{d}}^{k}_{1\times p}$, which will be used to correct the next sensor measurements. Then, $\hat{\mathbf{d}}^{k}_{1\times p}$ is analysed by the SPRT algorithm, indicating a possible drift occurrence, $D \in [0, 1]$. The architecture is illustrated in Fig. 1.

The following subsections give more details about each component of the architectures and some performance metrics.

2.1 AAKR Model

Auto-Associative Kernel Regression (AAKR) is a type of similarity based model, a nonparametric modelling technique that uses the similarity of a query vector to memory or exemplar vectors to infer the response of the model, as described in Hines et al. (2008). The derivation of the AAKR model architecture used in this work is based on multivariate, inferential kernel regression, that uses historical, fault free observations to correct drifts in the current observations. These fault free observations are stored in a matrix $\mathbf{X}_{n\times p}$, where $X_{i,j}$ is the i^{th} observation of the j^{th} variable, n is the number of observations of the p variables. A query vector $\mathbf{x}$ is a $1 \times p$ vector of the sensors measurements, which is the input of the model.

The prediction of the corrected input is calculated as a weighted average of the memory vectors stored in $\mathbf{X}_{n\times p}$, following some basic steps. First, the distance between a query vector $\mathbf{x}$ and each of the memory vectors $X_{i,j}$ is calculated. The most common measure is the Euclidean distance. This calculation results in a vector $\mathbf{u}_{n\times 1}$:

$$\mathbf{u} = \begin{bmatrix} u_1(X_1, \mathbf{x}) \\ \vdots \\ u_n(X_n, \mathbf{x}) \end{bmatrix}, \tag{1}$$

$$u_i(X_i, \mathbf{x}) = \sqrt{(X_{i,1} - x_1)^2 + \cdots + (X_{i,p} - x_p)^2}$$

These distances are converted into similarity measures $\mathbf{w}_{n\times 1} = [w_1 \ \cdots \ w_n]^T$ by using a kernel function, like the Gaussian kernel

$$\mathbf{w} = \frac{1}{\sqrt{2\pi h^2}} \exp\left(-\frac{\mathbf{u}^2}{h^2}\right) \tag{2}$$

where h is called *kernel bandwidth*. Finally, the prediction of the corrected input is calculated by using this similarities, or weights, to form a weighted average of the memory vectors:

$$\hat{x}_j = \frac{\sum_{i=1}^{n} (w_i \cdot X_{i,j})}{\sum_{i=1}^{n} w_i}, \quad \hat{\mathbf{x}} = \frac{\mathbf{w}^T \mathbf{X}}{\sum_{i=1}^{n} w_i} \tag{3}$$

where $\hat{\mathbf{x}} = [\hat{x_1} \ \cdots \ \hat{x_p}]$ is the prediction vector.

It is import to note that the memory vectors must include all the condition operations that are expected to be included into the future query vectors. As discussed by Hines et al. (2008), "several circumstances, including equipment repair or failure, seasonal variations and system-operating changes, can cause a change in operating conditions". If these operating conditions are not included in the memory vectors, no confidence can be given to predictions of the model and the memory matrix must either be appended or replaced with new data. The distance between the query vector and the most similar memory vector could indicate if the operating condition is inside of the training region.

In this work, the memory vectors were selected from a training set by a combination of min-max and vector ordering methods, as described in Hines et al. (2008). The bandwidth were chosen using grid-search and k-fold cross-validation, as in An et al. (2011). Large bandwidths produce smoother model predictions, as many memory vectors are used to infer a parameters value. Conversely, small bandwidths produce rough and/or inconsistent predictions because a limited number, if any, of the memory vectors are used to infer a parameters value.

2.2 Kalman Filter Model

As illustrated in Fig. 1, the objective of the Kalman filter is to estimate or to track the drift embedded in the measurements. The mathematical model used to estimate the drift $\mathbf{d}^k_{p\times 1}$ is

$$\mathbf{d}^k = \mathbf{I}\,\mathbf{d}^{k-1} + \mathbf{v}^k, \quad \mathbf{v}^k \sim \mathcal{N}(\mathbf{0}, \mathbf{V}) \tag{4}$$

where $\mathbf{v}^k$ is assumed to be Gaussian noise and $\mathbf{V}$ is the state noise covariance. In *target tracking*, equation (4) is known as the mathematical model of the dynamics behaviour of the target. In this work, the sensor drift is the target, and the objective is to track the amplitude of the drift over time. Like Takruri et al. (2008), assuming the sensor drifts in a smooth, slowly increasing, linear or exponential fashion, the model of equation (4) is a reasonable approximation. Another assumption is that the drifts are not correlated with each other, despite the correlation between the process variables.

The available observations of the drifts, $\mathbf{z}$, is given by

$$\mathbf{z} = \mathbf{r} - \hat{\mathbf{x}} \tag{5}$$

The vector $\mathbf{z}$ is not the true values of the drifts, since the true values of the process variables are unavailable. So, the measurement equation is

$$\mathbf{z} = \mathbf{I}\,\mathbf{d}^k + \mathbf{q}^k, \quad \mathbf{q}^k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{6}$$

where $\mathbf{q}^k$ is Gaussian noise and $\mathbf{Q}$ is the measurement noise covariance.

When a reading vector $\mathbf{r}$ and the estimation of its correct value, $\hat{\mathbf{x}}$, are available, the following steps are executed to obtain a new estimation of the drifts, $\hat{\mathbf{d}}^k$:

(1) prediction, $\hat{\mathbf{d}}^{k|k-1} = \mathbf{I}\,\hat{\mathbf{d}}^{k-1}$
(2) minimum prediction MSE (mean squared error) matrix, $\mathbf{M}^{k|k-1} = \mathbf{M}^{k-1} + \mathbf{V}$
(3) Kalman gain, $\mathbf{K}^k = \mathbf{M}^{k|k-1}\,(\mathbf{Q} + \mathbf{M}^{k|k-1})^{-1}$
(4) correction, $\hat{\mathbf{d}}^k = \hat{\mathbf{d}}^{k|k-1} + \mathbf{K}\,(\mathbf{z} - \hat{\mathbf{d}}^{k|k-1})$
(5) minimum MSE matrix, $\mathbf{M}^k = (\mathbf{I} - \mathbf{K})\mathbf{M}^{k|k-1}$

In Section 4, the Kalman Filter parameters, $\mathbf{Q}$ and $\mathbf{V}$, were chosen using trial and error. If $\mathbf{Q}$ is set to a high value, the estimated drift takes longer to follow the real drift, whereas if $\mathbf{Q}$ is set to a small values, the estimated drift follows any little difference between the AAKR estimation and the sensor readings. The $\mathbf{V}$ values have a inverse effect: high values yield unstable drift estimates that follow any little difference between the AAKR estimation and the sensor readings; for small values, the estimated drift takes longer to follow the real drift.

2.3 Statistical Decision Logic Module

The statistical decision logic module is responsible for evaluating the residuals and then making a decision about the operating condition of the sensor. In this work, the SPRT algorithm is applied to the residual analysis, detecting statistical changes in the residuals between the measurements and the predict values.

The SPRT is a statistical technique for system anomaly detection (Hines et al. (2008)), which consists of testing two possible hypotheses: the system is more likely to be in a normal mode H_0 or in a degraded mode H_1. For each new residual d^i_p of the process variable p (i is the instant time), the following procedure is executed to accept one of the two hypotheses:

(1) calculation of the log-likelihood ratio

$$\Lambda^i_p = \ln \frac{P(d^i_p|H_1)}{P(d^i_p|H_0)} \tag{7}$$

(2) calculation of the cumulative sum of Λ_p

$$S^i_p = S^{i-1} + \Lambda^i_p \tag{8}$$

(3) application of the stopping rule, a simple thresholding scheme
- $\ln A < S^i_p < \ln B$, continue monitoring and calculating $S^{i+1}_p = S^i_p + \Lambda^{i+1}_p$
- $S^i_p \leq \ln A$, H_0 is accepted
- $S^i_p \geq \ln B$, H_1 is accepted and a alarm is emitted

where A and B are the lower and upper bound, and $P(d^i_p|H_1)$ is the probability of observing d^i_p given H_1 is true. A and B could be defined by the false alarm probability α and missed alarm probability β as

$$A = \frac{\beta}{1-\alpha}, \quad B = \frac{1-\beta}{\alpha} \tag{9}$$

If $S^i_p \leq \ln A$, it is determined to belong to the normal mode H_0 of the system and $D = 0$. Conversely, if $S^i_p \geq \ln B$, it is determined to belong to the degraded mode H_1 of the system and $D = 1$, indicating a drift occurrence.

Here the residuals are assumed to be normally distributed with zero mean and variance of σ^2, which is an estimate of the sensor noise. Therefore, the probability distribution function for the normal mode of the residuals is given by

$$P(d^i_p|H_0) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-d^{i\,2}_p}{2\sigma^2}\right) \tag{10}$$

Supposing the degraded mode represented as a mean shift up ($+M$) or a mean shift down ($-M$), where M is the

amplitude of the change, the log-likelihood ratio is given by

$$\Lambda_p^i = \begin{cases} M/\sigma^2(d_p^i - M/2), & \text{for a positive change} \\ M/\sigma^2(-d_p^i - M/2), & \text{for a negative change} \end{cases} \quad (11)$$

As suggested in Hines et al. (2008), the optimal M value can be determined numerically by applying the SPRT to unfaulted, test data and locating the M value that results in a false-alarm probability that is nearest the theoretical false alarm probability α.

2.4 Performance Metrics

As stated in Hines et al. (2008), the performance of auto-associative on-line monitoring systems has traditionally been measured in terms of three metrics: accuracy, auto sensitivity, and cross sensitivity.

The ability of a model to correctly and accurately predict sensor values is measured by the accuracy, and it is is normally presented as the MSE between sensor predictions and the measured sensor values. It is important to note that this metric compares the unfaulted, or error corrected, predictions with the target, or error free, data. The equation for a single variable is given by

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2 \quad (12)$$

where N is the number of test observations.

Auto sensitivity (S_A) measures the ability of a model to make correct sensor predictions when the respective sensor value is incorrect due to some sort of fault. The auto sensitivity for a sensor k is given by

$$S_{Ak} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{x}_{ik}^{drift} - \hat{x}_{ik}|}{|x_{ik}^{drift} - x_{ik}|} \quad (13)$$

where $\hat{x}_i^{drift}$ is the drifted prediction, $\hat{x}_i$ is the unfaulted prediction, and x_i^{drift} and x_i are the drifted and unfaulted input, respectively.

The effect a faulty sensor input has on other sensor predictions is measured by the cross sensitivity (S_C). For a unfaulted sensor j and a drifted sensor k, the cross sensitivity is calculated as

$$S_{Ck} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{x}_{ij}^{drift} - \hat{x}_{ij}|}{|x_{ik}^{drift} - x_{ik}|} \quad (14)$$

Good monitoring systems have lower MSE, S_A and S_C values.

3. DATA SETS FOR SYSTEMS EVALUATION

The two monitoring systems were evaluated using simulated and real data sets. Following, major details are given about each data set.

3.1 Simulated Data Set

Simulated data were generated by a model developed using OLGA [1]. The model represent a well operated by gas-lift with a 2500 meters deep pipe production. The pressure at the separator was keep at 1000 kPa and the reservoir was set to 18000 kPa.

The simulated data were collected during 225 hours of well production, while the injected gas flow rate was set to different values, representing different operating conditions. The data set is composed of the pressure at the bottom hole (PT_f), at the top of the production pipe (PT_t), at the top of the annular space (PT_g) and at the upstream of the injection choke (PT_m). The sample rate was 1 sample per minute. White noise of NSR (noise to signal ratio) equal to 0.3 was added to the measurements.

The data set was divided into two parts: the first 90 hours for AAKR training and the rest of data for test. All the data set is presented in Fig. 2.

3.2 Real Data Set

The real data set is composed of pressure measurements at the bottom hole (PDG), at the Christmas tree (TPT) and at the upstream of the injection choke (PT_m). The data was collected during 955 hours at the sample rate of 1 sample per minute. The first 500 hours were used for training and the rest of the data for test.

It is not known if the data set includes some faulty data, except the visible outliers equal to zero.

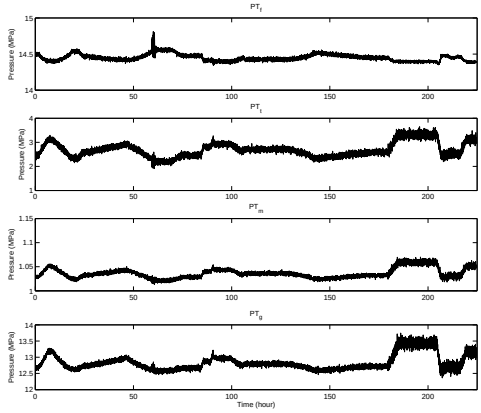

Fig. 2. Simulated data set.

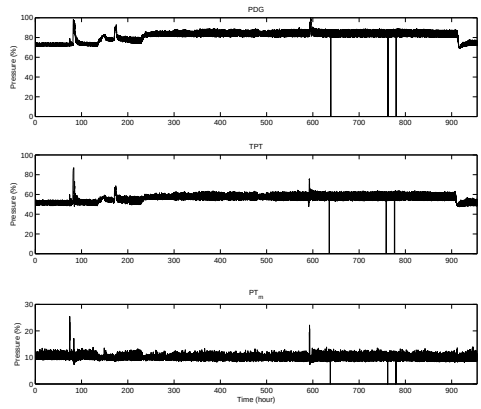

Fig. 3. Real data set.

[1] http://www.sptgroup.com/en/Products/olga/Multiphase-Flow-Simulator/

4. RESULTS

4.1 Simulated Data

In both architectures, the AAKR model and the SPRT module were the same, using the following configurations: an AAKR model with $h = 0.2$ and $25\% \times 5400 = 1350$ memory vectors (almost the most accurate configuration); and a SPRT module with false alarm probability set to 5%, missed alarm probability set to 10% and $M = 3$ (hypothetical values) for all process variables. In the AAKR-KF-SPRT architecture, the Kalman Filter covariance matrices Q and V were set to 0.3 and 0.0001, respectively.

In order to evaluate the ability to correct/detect sensors faulty data, an artificial linearly increasing sensor drift ending at a magnitude of 5σ (five times the standard deviation of the sensor signal) was introduced in the testing data set at the time approximately equal to 12 hours. The results are presented in Fig. 4, 5, 6 and Table 1.

It can be seen in Fig. 4 that the AAKR-KF-SPRT results follow the true values with just a small bias, even at the

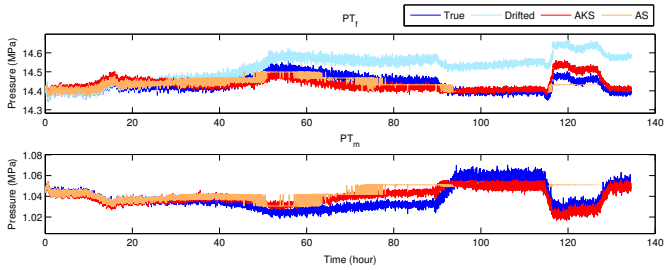

Fig. 4. Predictions of PT_f and PT_m for three drifting sensors, PT_f, PT_t and PT_g, in simulated data set. AKS means AAKR-KF-SPRT and AS means AAKR-SPRT.

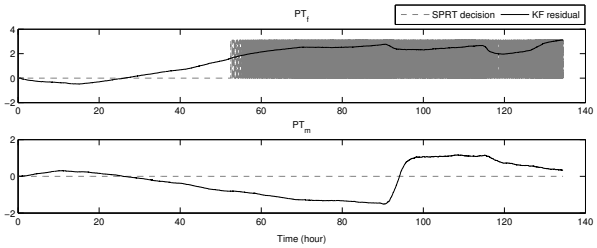

Fig. 5. Drift estimates and SPRT decision of PT_f and PT_m for the AAKR-KF-SPRT on simulated data set.

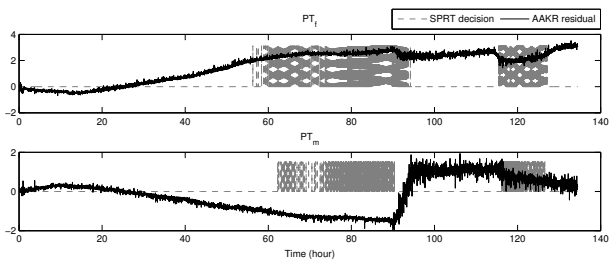

Fig. 6. Residual and SPRT decision of PT_f and PT_m for the AAKR-SPRT on simulated data set.

Table 1. Performance metrics for the system monitoring architectures.

	AAKR-KF-SPRT	AAKR-SPRT
MSE	9.1293×10^{-3}	2.8491×10^{-3}
S_A	1.9431×10^{-5}	5.0840×10^{-5}
S_C	8.1022×10^{-5}	11.8895×10^{-5}

maximum drift amplitude. The AAKR-SPRT can not generate good predictions for severe drift in its inputs, mainly for small bandwidth values. The S_A and S_C values in Table 1 confirm the higher sensitivity of the AAKR-SPRT for drifting inputs. From the MSE values, the AAKR-SPRT is more accurate for unfaulted data, but the AAKR-KF-SPRT can generate better predictions using faulty data. Larger bandwidths could improve the performance of the AAKR-SPRT, but its predictions tend to be biased, with values always near a mean value of the memory vectors.

A comparison of the drift detection performance is showed in the Fig. 5 and 6 [2]. The AAKR-SPRT correctly detected a drift occurrence in PT_f at the 56^{th} hours, but it emitted a false alarm about PT_m. In addition, the variance assumed for the sensor noise appear to be in discordance with the real variance, as it can be seen in the decision results of PT_f at the time instances around 100 hours, a large interval between two decisions $D = 1$. The AKKR-KF-SPRT detected correctly the drift occurrence without any false alarms, and its residuals are cleaner, allowing stable drift detection. Since its predictions kept near the true values, the residuals reach the limit early.

The better generalisation performance of the AAKR-KF-SPRT allows a greater extension of the useful life sensor, once its predictions are very near the true values even on a severe drift occurrence. However, it has a cost, biased predictions even on normal sensor readings. Since the KF can not discern between bias and drift, incorrect drift predictions can cause false SPRT alarms.

4.2 Real Data

The following configurations were used for the AAKR model and the SPRT module: an AAKR model with $h = 0.6$ and $5\% \times 30000 = 1500$ memory vectors (almost the most accurate configuration); and a SPRT module with false alarm probability set to 5%, missed alarm probability set to 10% and $M = 3$ for all process variables. In the AAKR-KF-SPRT architecture, the Kalman Filter covariance matrices Q and V was set to 10 and 0.0001, respectively.

Again, an artificial linearly increasing sensor drift ending at a magnitude of 5σ (five times the standard deviation of the sensor signal) was introduced in the testing data set at the time approximately equal to 42 hours. The results are presented in Fig. 7, 8, 9 and Table 2.

The predictions for two drifting sensors, TPT e PT_m, are illustrated in Fig. 7. It can be seen that a similar behaviour of the monitoring systems when compared to Fig. 4. The AAKR-KF-SPRT present better performance

[2] The shaded area means the occurrence of many close high and low values. High values indicate $D = 1$ (degraded mode), whereas low values indicate $D = 0$ (normal mode).

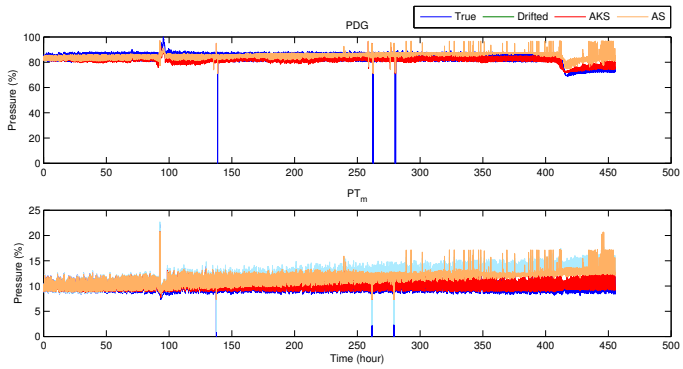

Fig. 7. Predictions of PDG and PT_m for two drifting sensors, TPT and PT_m in real data set. AKS means AAKR-KF-SPRT and AS means AAKR-SPRT.

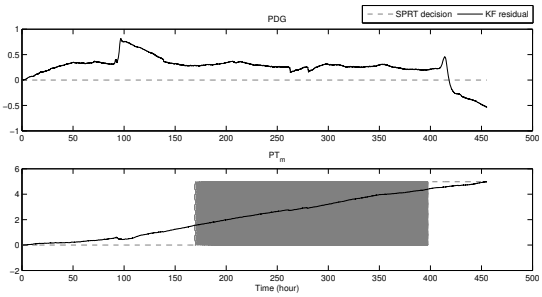

Fig. 8. Drift estimates and SPRT decision of PDG and PT_m for the AAKR-KF-SPRT on real data set.

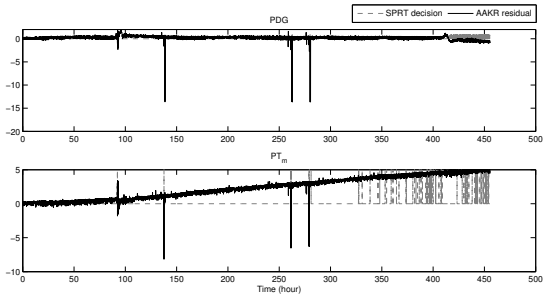

Fig. 9. Residual and SPRT decision of PDG and PT_m for the AAKR-SPRT on real data set.

Table 2. Performance metrics for the system monitoring architectures on real data set.

	AAKR-KF-SPRT	AAKR-SPRT
MSE	2.7013	1.4061
S_A	0.0747	0.6956
S_C	0.0466	0.1208

for drift correction/detection. The S_A and S_C values in Table 2 indicate the lower sensitivity for faulty inputs presented by the AAKR-KF-SPRT, whereas the AAKR-SPRT predictions presented some inconsistent peaks, even for the PDG data. However, as indicated by the MSE values, for unfaulted data, the AAKR-SPRT is more accurate, considering the original data as error free. For the visible outliers equal to zero, both systems generated reasonable predictions.

As it can be noted in the Fig. 8 and 9, the residual generated by the AAKR-SPRT is affected by outliers in the test data; its SPRT module gave false alarms for the PDG and missed some alarms for the other sensor. The drift estimates of the AAKR-KF-SPRT are much more clean, resulting in correct SPRT decisions.

Although the performance of the systems have some differences, both systems show promising results for the oilfield industry data.

5. CONCLUSION

In this work, a robust on-line calibration monitoring system for drift correction/detection in oil and gas well sensors is presented. The system is based on an empirical model for prediction, a Kalman filter for drift tracking and a statistical decision module for drift detection. The evaluation for simulated and real data sets demonstrated promising results for the oilfield industry. With a slight loss of accuracy, the predictions of the corrected readings of the sensors showed to be much less sensitive for drifts than a state-of-art monitoring system. This characteristic allows a greater extension of the useful life of the sensors, playing an important role on the production optimisation of the well and planning. For future work, the use of the empirical model uncertainty in the drift detection and improvements in Kalman filter will be analysed.

ACKNOWLEDGEMENTS

The authors would like to thank Cenpes/Petrobras for provision of data and financial support.

REFERENCES

Aggrey, G. and Davies, D. (2007). Tracking the state and diagnosing Down Hole Permanent Sensors in Intelligent Well Completions with Artificial Neural Network. *Offshore Europe*.

An, S.H., Heo, G., and Chang, S.H. (2011). Detection of process anomalies using an improved statistical learning framework. *Expert Systems with Applications*, 38(3), 1356–1363.

Eck, J., Ewherido, U., Mohammed, J., and Ogunlowo, R. (1999). Downhole monitoring: the story so far. *Oilfield Review*, 3(3), 18–29.

Gribok, A., Hines, J.W., and Uhrig, R. (2000). Use of kernel based techniques for sensor validation in nuclear power plants. In *NPIC&HMIT 2000*.

Hines, J.W., Garvey, J., Garvey, D., and Seibert, R. (2008). Technical Review of On-Line Monitoring Techniques for Performance Assessment. Volume 3. Limiting Case Studies. Technical report, U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research, Washington, DC.

Ma, J. and Jiang, J. (2011). Applications of fault detection and diagnosis methods in nuclear power plants: A review. *Progress in Nuclear Energy*, 53(3), 255–266.

Takruri, M., Rajasegarar, S., and Challa, S. (2008). Online drift correction in wireless sensor networks using spatio-temporal modeling. *Information Fusion, 2008 11th International Conference on*, 1–8.