Quantification and Compensation of Systematic Errors in Pressure Measurements Applied to Oil Pipelines

Gregor Thiele
Department for process automation and robotics
Fraunhofer IPK
Berlin, Germany
e-mail: gregor.thiele@ipk.fraunhofer.de

Martin Liu
PSI AG
Berlin, Germany
e-mail: mliu@psi.de

Moritz Chemnitz
Department for process automation and robotics
Fraunhofer IPK
Berlin, Germany
e-mail: moritz.chemnitz@ipk.fraunhofer.de

Jörg Krüger
Department for process automation and robotics
Fraunhofer IPK
Berlin, Germany
e-mail: joerg.krueger@ipk.fraunhofer.de

***Abstract*—The monitoring of pipeline operation is an important research topic, especially for the detection and localization of leaks as well as for an efficient control. For these purposes, physical quantities in pipelines are calculated from measurement data on the basis of a mathematical model. In contrast to static models, adaptive models vary their parameters or even their structure to reach the most probable solution. But in most cases, even the best fit will hold residuals caused by discrepancies between the real system and its model. These residuals allow an estimation of travel-time delays of pressure waves and offsets in pressure values. The basic idea of our approach is to interpret these systematic, time-invariant errors of pressure measurements in pipelines either as sensor displacements or as technical defects. The proposed procedure leads to a hypothesis for a model update, regarding the sensor positions. This displacement compensation as well as a variance analysis was successfully applied to real data from a crude oil pipeline in Europe. A cross validation proves the general capability of the developed method to reduce the uncertainties.**

***Keywords* - —system identification, systematic errors, measurement technology, oil pipeline, distributed parameter system, state estimation**

I. Introduction

For the transport of a large amount of fluid, like in the case of crude oil and oil products, pipelines are the most efficient and secure method. With a total length of 1.9 million km, they are referred to as the "hidden veins of the economy" [1]. The observation of the hydraulic state in pipelines is an important topic as the base for observation and control. Especially leak detection and localization are of major interest [2], [3], [4]. Typical techniques are algorithms, which combine sensor information and a physical model to achieve the most probable system state. In most approaches, the uncertainties in the data are assumed to be noise with a normal distribution, as criticized by [5]. This allowes a precise and current calibration of all sensors. Furthermore the physical model information is not completely reliable, especially in older pipelines. A reflection about this issue leads to the phenomenon of systematic error. But often the resulting solution is not published: "In most cases, the problems of [systematic errors] are solved individually, depending on the skills of the experimenter." [6] However, an explicit exploration of this subject seems worthwhile. To resolve this issue we propose a separate approach for a special kind of systematic errors, which is suitable for all types of pressure sensors in pipelines. After a short a short overview of related work in section II, our procedure is introduced in III. IV comprises our experimental work and its results. Conclusively, V reflects on the essential aspects of this study.

II. Current questions about state observation in pipelines

The field of pipeline observation mainly aims at risk prevention. The central questions are: Is there a leak? Where is it? How large is it? To obtain this information, sensor data must be interpreted from several perspectives.
The mass flow balance can point to a leak and give a hint as to its size. High-resolution pressure measurements can indicate the position of the leak if the propagation speed is known.

A. Observation of physical states in pipelines

In general, the measurement of physical phenomena underlies different errors. To get rid of these uncertainties, observers or estimators are in use to predict the most likely value, using measurement data and model information. An overview of the model types is given by [7]. Because of the noise in all measurements in pipelines, [5] discourages from the well known Luenberger observer and recommends a stochastic approach. The most popular estimation algorithm is the Kalman Filter (KF), which is based on Bayesian inference.

The simplest kind is the steady-state KF, in [8] exemplarily combined with a Finite-Element-Model to determine the fluid flow from pressure values. Naturally, this neglects potential non-linearities of the dynamics and leads to huge system matrices if an acceptable resolution for long pipelines is required. The absent consideration of non-linearities can be remedied with an extended KF which performs a linearization in every time step [5].

B. Measuring wave travel-time

Given is a set of time series, each per sensor, which cover a pressure drop. In here, the delay of a wave passing two sensors must be found. In further explanations, we will call this value the travel-time of the wave. This measurement task is the key issue in leak localization per negative pressure wave method. Some approaches utilize a common identifiable mark of the wave - for example the inflection point. For appropriate signal processing, it is proposed to use the extreme values of a wavelet transformed pressure signal [9]. Thereby peaks in higher order wavelet levels are caused by a steep change in pressure, which can indicate a wave. This delay can also be found using a cross-correlation [10], which utilizes the entire wave shape and gives the delay value for the best similarity between two signals. A precise measurement of the delay is an important question because it limits the accuracy of upcoming processing. Three different techniques will be examined experimentally in III-D.

III. Procedure for analysis of systematic errors

A. Modeling of the system and included errors

Important previous research is done by James P. Modisette. In a study about state estimation he mentioned the challenge of isolating the individual errors: "One good reason not to [tune offsets] is that it would introduce ambiguity into the results: there would be no way [...] to distinguish between the situation with all the flow meters reading a bit high and the situation with the pressure meters underestimating the pressure drops." [11] An example of the pressure profile during a constant fluid flow is illustrated in fig. 1. The sensor is shifted either in height h or along the length coordinate z. Obviously, both of the physical discrepancies can have the same effect: a lower pressure is measured.

B. Data preparation

The starting point of the procedure is a collection of pressure values, declared $p_i(t)$ for the i-th sensor. A trivial offset component results from the barometric pressure. With given assumptions for the fluid density ρ and the gravity constant g, the height information h of each sensor is used to calculate the pressure difference to the first sensor

$$\Delta p = -\rho g \cdot \Delta h \tag{1}$$

in order to compensate this effect. Thereby all sensors lay on a virtual common height level.

In a next step, the data is split into two groups: suitable intervals containing either an insignificant flow or a pressure wave. The first ones are used for offset analysis and the second ones to determine the travel-time.

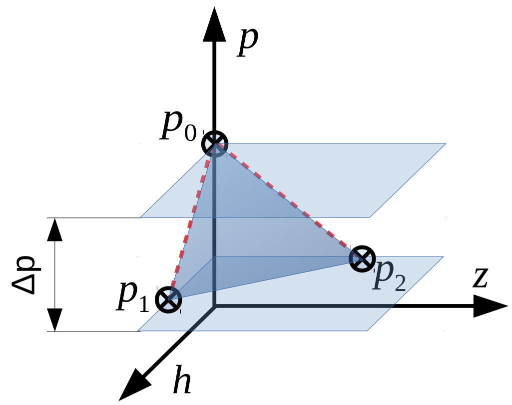

Fig. 1. Vertical (p_1) and horizontal (p_2) sensor displacements can lead to an equal pressure-decrease in cases of a certain flow rate.

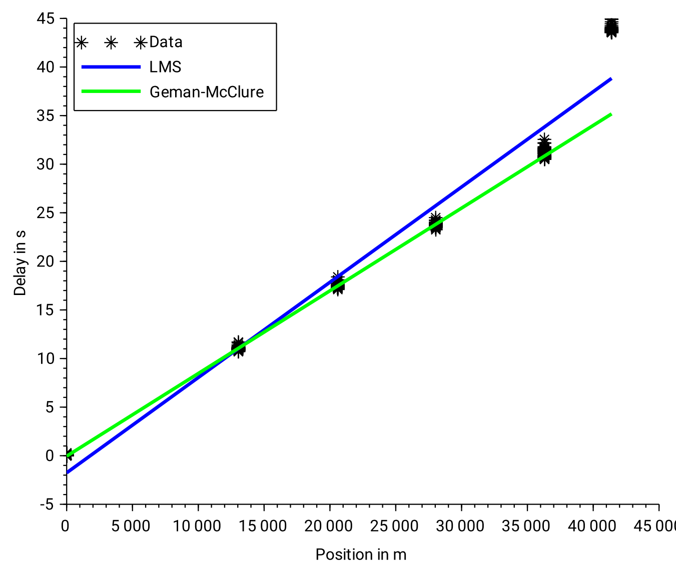

Fig. 2. In contrast to the quadratic cost function, the GM-cost function goes in saturation whereby a robust behaviour is reached.

C. Data regression

Instantiating the mathematical model requires a data regression algorithm. Depending on the measurement technique, outliers can occur. The most popular regression is the least-mean-squares technique (LMS). An example in fig. 2 illustrates the performance of LMS: the deviation of one group of points strongly influences the dashed regression line.

An alternative regression technique is the Geman-McClure estimator, which was developed for the purpose of robustness [12]. A comparison of the results in fig. 2 shows the benefit of robust regression, which can be understood with the help of the cost functions depicted in fig. 3. The quadratic function (cost function of the LMS regression) grows with an increasing slope and thus, high deviations yield to enormous influence.

In contrast, the Geman-McClure function has the form of the

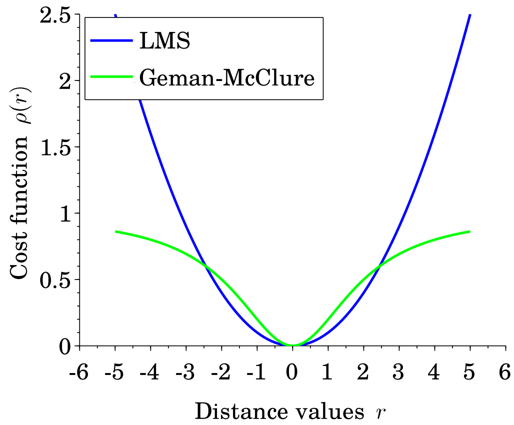

Fig. 3. In contrast to the quadratic cost function, the GM-cost function goes in saturation whereby a robust behavior is reached.

equation

$$\rho(r) = \frac{r^2}{\sigma^2 + r^2}. \tag{2}$$

at which ρ is the functional, r is the residual and σ is tuned to set the expected scope. Due to the inflection point, the influence of outliers goes to saturation. Caused by this saturation in the cost function, the GM-estimator handles outliers with less relevance. The obvious delays in fig. 2, at a position of 40 km, were caused artificially to demonstrate their effect.

The same regression principle covers also the problem of time independent errors. Whenever a sensor breaks the model with significant deviations, the regression "ignores" them better than the LMS-approach does.

This study concerns time delay in travel-time as well as pressure offset. Both introduced kinds of systematic model deviations can be quantified for compensation of prospective measurement data. A side benefit is the variance analysis during the regression: exorbitant variance of certain sensors is an indicator for technical defects. This helps to reject misleading sensor data.

D. Methods for measuring wave travel-time

In II-B we motivated the need for precise travel-time measurements. We choose three different algorithms which calculate the delay between the sensors from their pressure dynamics. In every case, the vector `travelTime` has $n-1$ entries and gives the arrival-time in the n-th sensor, whereby the wave starts in the first sensor at $t = 0$.

1) Inflection Point: The most intuitive approach starts from the assumption, that each falling pressure edge can be represented by one single point: that of the steepest slope. Therefore all pressure time series are differentiated with respect to time. The maximum of the derivative gives the inflection point. The time difference between the inflection points in each sensor gives the travel-time, as drafted in algorithm 1.

Algorithm 1 Inflection Point

```
tInflect = []
for i = 1 to numSensors do
    tInflect(i) = find(max(abs(diff(filteredSignal(i)))));
end for
travelTime = tInflect(2:end)-tInflect(1);
```

2) Cross-Correlation to the first sensor: In contrast to the first approach, the cross-correlation uses the entire signal shape. For all possible deferrals between the first and the n-th sensor, the product of the signals is summed up. The deferral with the highest cross-correlation value gives the best similarity between the two wave-shapes.

Algorithm 2 XCOR to the 1st sensor

```
tInflect = []
for i = 2 to numSensors do
    c = xcorr(diff(filteredSignal(1) , diff(filteredSignal(i));
    travelTime(i) = find(max(diff(c));
end for
```

3) Cross-Correlation to the predecessor sensor: The second algorithm compares every sensor to the first one. But in case of a significant influence of dispersion, the shape of the wave can alter during the travel-time. For this reason, the third approach takes the cross-correlation of every sensor with the particular predecessor. This assumes, that the change of the wave shape between to sensors is negligible but involves the distortion over the whole length. To achieve the absolute time values for each sensor, the cumulated sum is calculated gradually.

Algorithm 3 XCOR to sensor $n-1$

```
deltaT = []
for i = 2 to numSensors do
    c = xcorr(diff(filteredSignal(i-1) , diff(filteredSignal(i));
    deltaT(i) = find(max(diff(c));
end for
travelTime = cumsum(deltaT);
```

E. Concept for analysis of horizontal sensor displacement

The final target is a reliable improvement of the model information. Hereby, the focus lays on systematic errors which can be modeled similar to sensor displacements. The proposed procedure consists of a latency analysis to identify horizontal displacements and an offset analysis, corresponding to a vertical sensor displacement (caused by incorrect information about the pipeline altitude rather than the sensor position within the pipe). The first one requires an information about the duration a traveling wave requires between two sensors. The development of a travel-time measurement algorithm was discussed in III-D. This signal processing method takes the intervals containing pressure drops and provides the latency between the incoming of the wave at the individual sensors.

Assuming uniform conditions along the pipeline, the resulting velocity c should be constant over all segments. So we expect a linear function t for the elapsed time depending from the length coordinate z:

$$t(z) = \frac{1}{c} \cdot z + t_0. \tag{3}$$

Accordingly, a linear regression is applied to the extracted travel-time data. For every single sensor, the Geman-McClure-Estimator gives the mean deviation from the assumed pressure profile along the pipeline. As a consequence, the remaining residuals can be interpreted as horizontal sensor displacements, independent from the factual cause. Once quantified, the model information can be corrected to remedy the systematic error.
This procedure uses just the time-derivative of the pressure values, not the pressure level itself. So it is independent from possible static offset errors; thus this error type is *isolated*.

F. Concept for analysis of vertical sensor displacement

The procedure for pressure offsets is built based on the fitting of the theoretical solution, similar to the wave analysis. For times of resting fluid, the measured pressure is determined by the geodetic altitude of the pipeline. The section III-B deals with that aspect.
After removing this influence as explained above, a constant pressure level at all sensors is expected. Every given data series of resting fluid gets averaged to eliminate noise. This leads to a single pressure value for each interval. By using just the mean value, we abolish the time domain, represented by the chronological order of data points. Therefore, this procedure is fundamentally independent from the travel-time analysis (and its potential corrections); this error type is isolated from other errors as well.

G. Precision and amount of samples

Both presented concepts are designed for batch processing, whereby the precision increases with a growing amount of used data. A common statistical tool is the confidence interval which is created from the statistical properties of the random variable. For n samples of the travel-time-analysis (to find horizontal displacements) with an empirical standard deviation s and a critical value of t^* the confidence interval is

$$\left(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right). \tag{4}$$

The critical value for an unknown distribution is calculated from the student-t-distribution using the confidence level $C = 0.95$, which gives the proportion of true values inside this interval. For the sample number $n = 30$ and $\alpha = \frac{1-C}{2}$ the critical value is $t_\alpha(29) = 2.042$.
In the upcoming results in IV-A we will see that the empirical standard deviation of a single measurement is $s = 0.23s$. So the confidence interval has a radius of 0.085 s around the mean. An uncertainty in travel-time leads to an uncertainty in localization. With the determined velocity c, the confidence interval radius corresponds with a localization precision of about 100 m.
The 30 samples were taken within one month. If the model was based on more samples from a longer observation period, the accuracy will further increase. The same can be done for the pressure offset analysis.

H. Validation of the model updates

Following the law of large numbers, the influence of random errors can be reduced with a higher amount of involved samples. To check, if enough data is in use, we applied cross-validation: the overall data is separated into several subsets. Then one set is hold back while the others get used to determine the displacement. The proposed model changes are formulated as an alternative hypothesis, competing the null hypothesis, which is the a priori model. The not involved set serves as verification. In this manner, the null hypothesis as well as the alternative hypothesis is applied to the verification set. This procedure is repeated for every subset. So n subsets yield to n iterations of the cross-validation. If the alternative hypothesis improves the plausibility of the model, the suggested update can be done.
The procedure we presented in this section is summed up in a block diagram (fig. 4). Our experiments with the developed modules are documented in the next section.

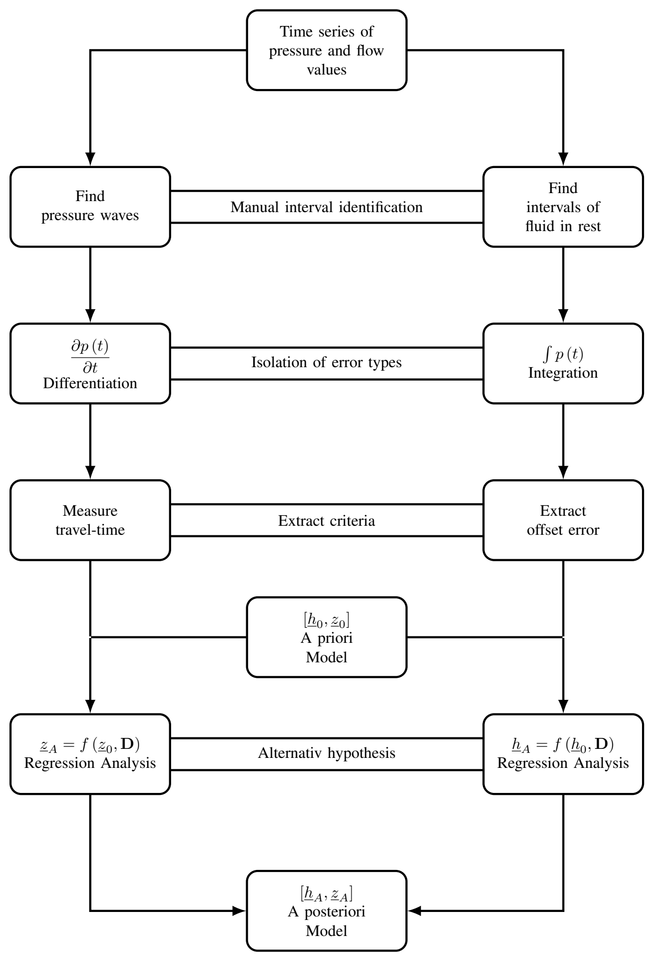

Fig. 4. Block diagram of the entire procedure

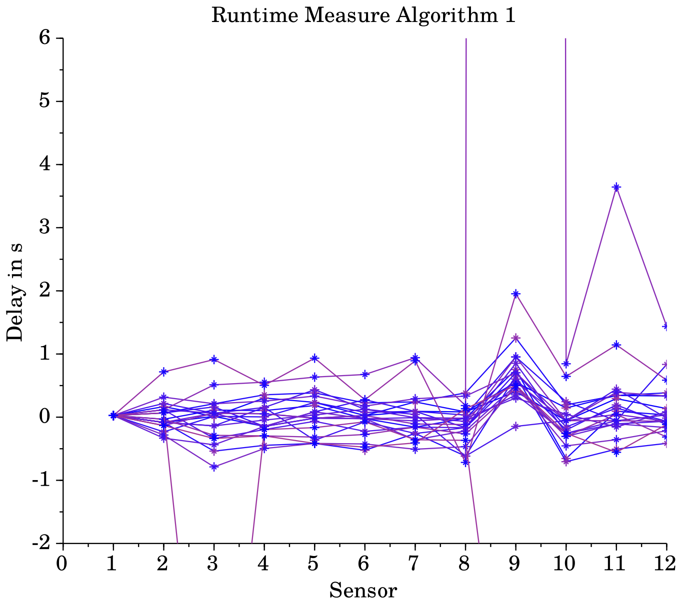

Fig. 5. The inflection point algorithm yields a high noise level and several outliers.

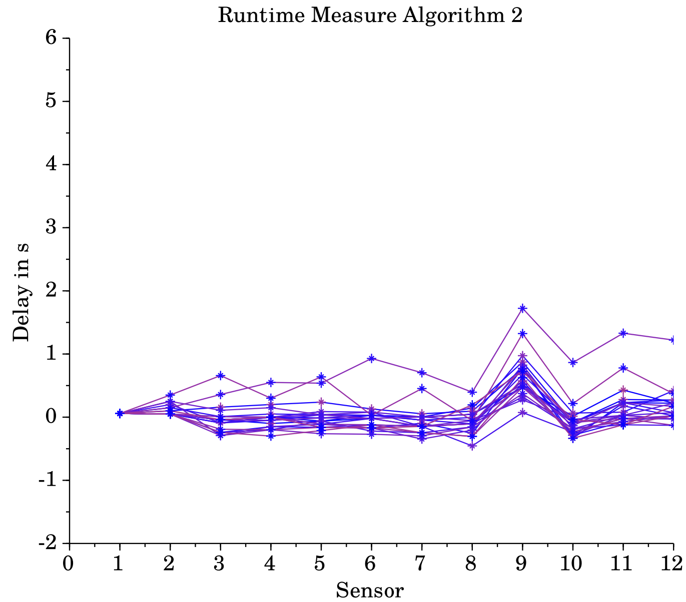

Fig. 6. The cross-correlation to the first sensor scores a robust model fit with less noise in the delay.

IV. Experimental methods and results

The explained technique was tested with data from a crude oil pipeline. From a one month time series, several intervals of fluid at rest or pressure drops were chosen. During tests with the delay measurements, we reached better results with waves from pressure drops - commonly caused by pump switch-offs - than from intervals with a rapid pressure increase. In this way, we got 33 waves for travel-time analysis and 18 intervals with resting fluid, about one hour each.

A. *Measuring wave travel-time*

The three different techniques for calculating the travel-time were presented in III-D. In this step we state a suitable comparison to judge the candidates.
For 33 different pressure drops, we calculated the travel-time between the sensors. As the most probable solution we fitted a linear regression in the position-time-diagram. The standard deviation of the remaining residuals gives a comparison. In section III-G was shown, that this precision is improved substantially by taking more data into account.

The first technique, basing on the reflection point identification, leads to the values in fig. 5. An ideal run would hold the zero-line without any disturbances. But in the results, we see noise as well as systematic errors. The tested algorithm is just responsible for the random errors because the common trends, like in sensor 9, may be caused by physical model deviations. The outliers in the sensors 3, 6 and 8 reach almost -100 s, the positive outlier in sensor 9 goes up to 450 s, which is a total fail of the algorithm (both not in the image scope). The standard deviation amounts to a mammoth value of 15 s and indicates the harsh inaccuracy of this approach. These results document the weak information of a single point: it is not trivial to find the correct inflection point from a noisy signal, in spite of a low-pass filtering.

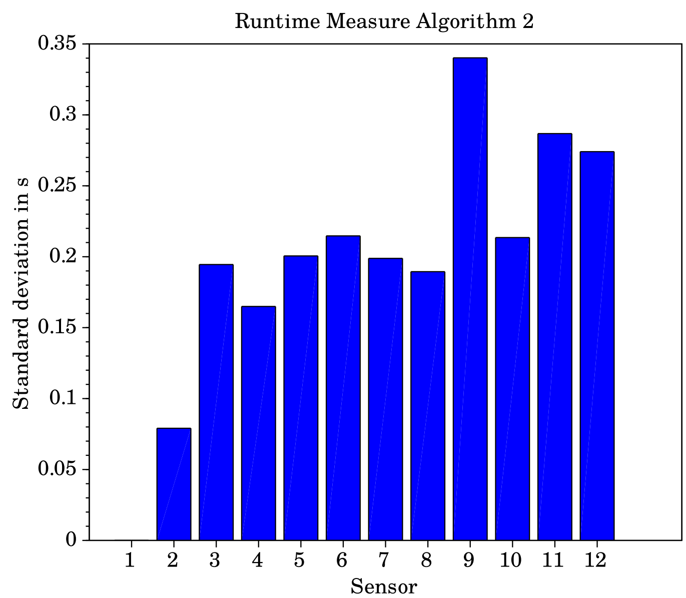

Fig. 7. The standard deviation of the results from the second procedure show a high precision. The mean value 0.23 s corresponds to an accuracy of 250 m for a single wave.

The second proposed approach utilizes the full signal shape. The fig. 6 shows a more pleasant outcome. The graphs follow a narrow corridor and the systematic deviation in sensor 9, as seen in fig. 5, is still visible, even more clearly. The standard deviation has an absolute maximum of 0.33 s and a mean value of about 0.23 s. This is an obvious advantage in comparison to the first approach.

The theoretical consideration of dispersion leads us to the

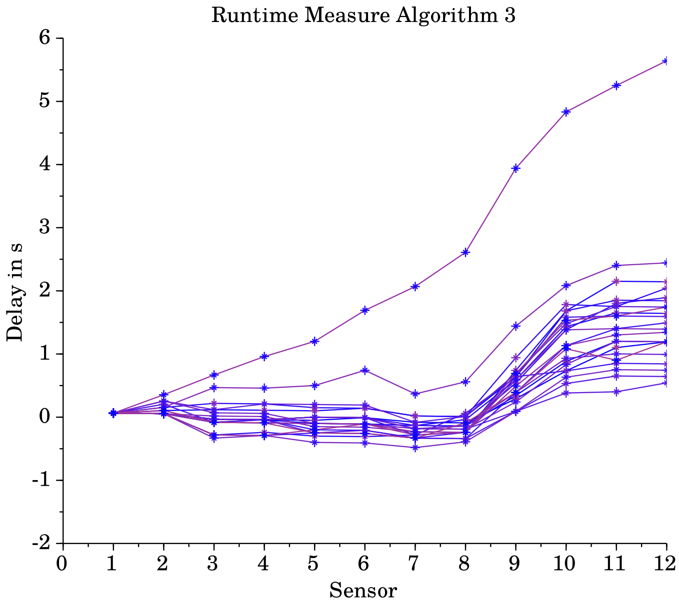

Fig. 8. The cumulated sum of the single sections provokes a divergence of the observed waves, which deteriorates the result.

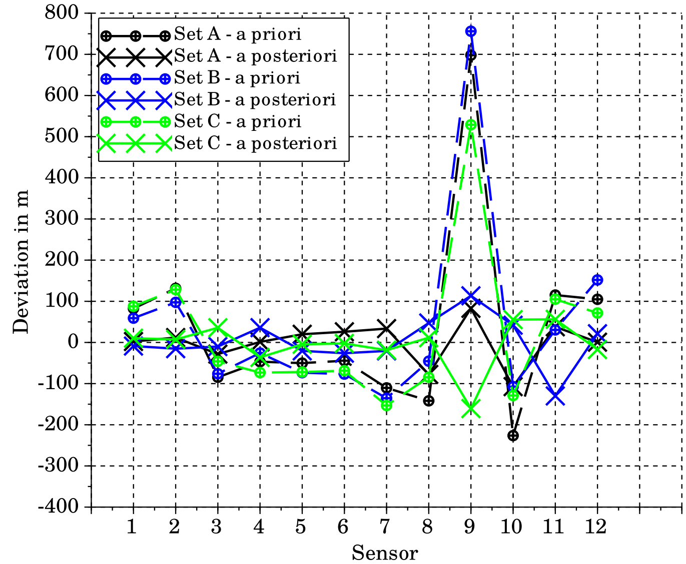

Fig. 9. For three independent subsets, the given model (dots) and the suggested model update (crosses) is compared, whereby the aim is a small distance to the zero-line.

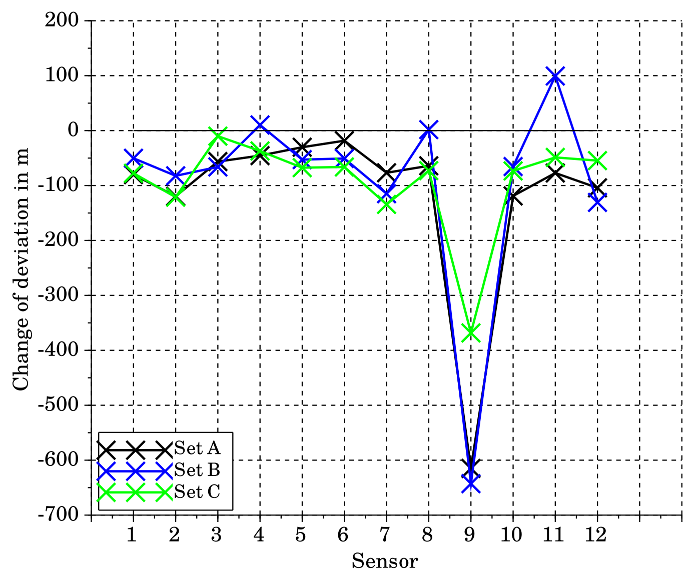

Fig. 10. The difference between the residuals in the null hypothesis and the alternative hypothesis is plotted for each sensor. A negative value indicates an improvement - like in most of the points.

third case, like explained in III-D3. But instead of improving the results, this technique accumulates the particular errors due to the cumulated sum. The evoked divergence ruins the outcome.

The comparison of the three suggested approaches shows that the second one suits best to our system and aim. But for other dimensions of pipelines or other fluids, another approach may be better.

B. Horizontal displacement of the sensor

The previous part focused the robustness of the delay measurement. Now, a further interpretation of the remaining residuals is possible. Firstly, we consider the travel-time analysis using 33 waves which are separated into three subsets for the cross-validation. Fig. 9 depicts the results for these subsets. Each dataset is interpreted using the null hypothesis (marked with dots) and the alternative hypothesis (marked with crosses). A first view gives an idea about the effect: for all sets, similar systematic residuals can be seen. The most impressive one occurs at sensor 9 with a huge deviation of about 0.7 s, which equates about 800 m. These exorbitant error seems to be caused rather by technical defect than by deviating model information, like explained in IV-C. In this case, even the updated model does not yield robust results. But for a clearer judgment of the other sensors, fig. 10 shows the distance between the a-priori-results and the a-posteriori-results. A cross below the zero-line means an improvement of the model-consistency.

The procedure improves the model for 33 of 36 points whereupon all three degradations happen in set B. This fact suggests that set B is not representative by accident. But due to the cross validation, we know, which sensor coordinates become more consistent. Only these should be updated. A study, how the data amount effects on the model quality, is planned for the future.

C. Vertical displacement of the sensor

The concept for offset analysis III-F was tested with 18 intervals of fluid at rest. For each interval, the mean was calculated using the Geman-McClure-Estimator. The result is a set of 18 parallel pressure curves, shown in fig. 11. After the preparation using the given height information, the mean values of all sensors should be on a common level. But the graphs show remaining residuals, similar in all intervals. This similarity confirms the fundamental assumption of the

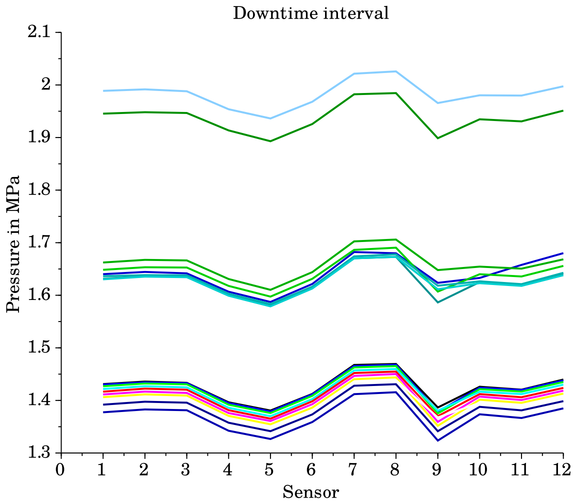

Fig. 11. Every Graph shows an interval pressure mean over the sensor number for times of rest. The similar form on different absolute levels is obvious.

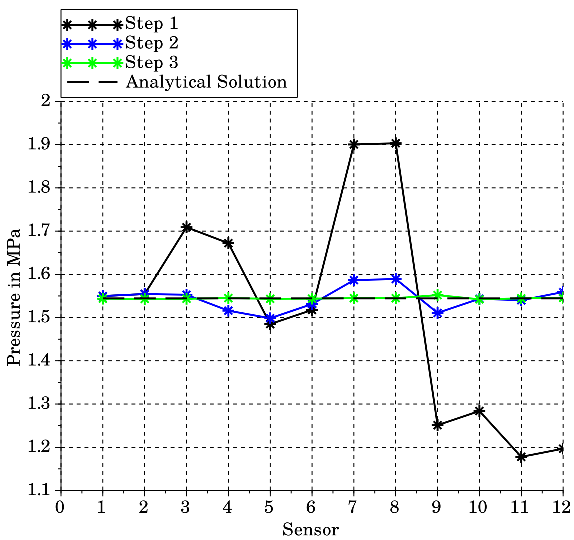

Fig. 12. Pressure residuals without any correction (step 1), corrected with the a priori knowledge (step 2) and after the proposed compensation procedure (step 3)

concept in III-F. The results of the proposed procedure prove the concept as well: a betterment is reached in almost every point. This is illustrated in fig. 12, where the three stages of correction are shown:

- step 1 gives the original data,
- step 2 represents the use of data sheet knowledge,
- and step 3 uses the new compensation approach.

The major part of the compensation is done using the given a priori height information to transform the raw data with residuals of about 3700 hPa (black graph) to the blue colored line, with remaining residual of less than 460 hPa. The proposed data driven procedure – in fig. 12 labeled as step 3 – yields to marginal residuals of maximal 75 hPa (green line).

V. Conclusion

In this study we showed that in pipeline observation, systematic errors are a significant issue. Our data-driven approach is developed to a proof-of-concept for an analysis and compensation procedure. As a necessary component, a method for the measurement of travel-time was developed, whereby one algorithm clearly excels the two competitors. To get rid of potential outliers, the procedure was advanced with the Geman-McClure estimator for robust means. The main procedures, based on the independent modeling of two types of systematic errors, yield to meaningful results. The detected model errors are a harmful weakness of every simulation or observation and its further applications. This study proved the capability to compensate these errors. From specific intervals from historical data, the modeled error-terms can be estimated. For the presented batch processing procedure, the statistical properties were discussed to estimate the precision depending from an amount of used data. Plenty of data will guarantee an accurate correction of the considered systematic errors.

ACKNOWLEDGMENT

This research was funded in part by the German Ministry for Economy and Energy in context of the project *EnEffReg*.

References

[1] A. Feizlmayr and S. Goestl, “Pipelines Adern der Wirtschaft,” *ERDL ERDGAS KOHLE*, vol. 127, no. 2, 2011.

[2] O. M. Aamo, “Leak detection, size estimation and localization in pipe flows,” *Automatic Control, IEEE Transactions on*, vol. 61, no. 1, pp. 246–251, January 2016.

[3] J. Wang, L. Zhao, T. Liu, Z. Li, T. Sun, and K. T. V. Grattan, “Novel negative pressure wave-based pipeline leak detection system using fiber bragg grating-based pressure sensors,” *Journal of Lightwave Technology*, vol. 35, no. 16, pp. 3366–3373, August 2017.

[4] S. Datta and S. Sarkar, “A review on different pipeline fault detection methods,” *Journal of Loss Prevention in the Process Industries*, vol. 41, pp. 97 – 106, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950423016300651

[5] J. Gong, J. Cai, X. Li, and S. Song, “Research on state estimation of oil pipeline considering adaptive extended kalman filtering,” in *International Conference on Mechatronics and Automation (ICMA)*, 2007, pp. 1294–1298.

[6] V. A. Granovskiy and T. N. Siraya, “Systematic errors: methodology of detection, elemination and evaluation,” in *XX IMEKO World Congress*. International Measurement Confederation, September 2012.

[7] G. Geiger, D. Matko, and T. Werner, “Extending classical pipeline models with neural nets,” in *Proceedings of the 11th Mediterranean Conference on Control and Automation*, 2003.

[8] A. Ozawa and K. Sandana, “An indirect measurement method of transient pressure and flow rate in a pipe using steady state kalman filter,” in *Proceedings of the 8th JFPS International Symposium on Fluid Power*, Okinawa, October 2011.

[9] Z. Yu, L. Jian, Z. Zhoumo, and S. Jin, “A combined kalman filter - discrete wavelet transform method for leakage detection of crude oil pipelines,” in *9th International Conference on Electronic Measurement Instruments ICEMI ’09.*, August 2009, pp. 3–1086–3–1090.

[10] S. B. M. Beck, N. J. Williamson, N. D. Sims, and R. Stanway, “Pipeline system identification through cross-correlation analysis,” *Journal of Process Mechanical Engineers*, pp. 133–142, August 2002.

[11] J. P. Modisette, “State estimation of pipeline models using the ensemble kalman filter,” in *PSIG Annual Meeting*, April 2013.

[12] S. Geman and D. McClure, “Statistical methods for tomographic image reconstruction,” *Proceedings of the 46th Session of the International Statistical Institute*, vol. LII(4), pp. 5–21, 1987.