

Research Article

Systematic Prioritization of Sensor Improvements in an Industrial Gas Supply Network

Onur Babat,¹ Ali Esmaili,² Joshua D. Isom,² Camilo Mancilla,² and Luis F. Zuluaga¹

¹Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015, USA

²Industrial Optimization/Control, Air Products and Chemicals Inc., Allentown, PA 18195, USA

Correspondence should be addressed to Onur Babat; onur.babat@lehigh.edu

Received 18 September 2016; Accepted 13 December 2016; Published 31 January 2017

Academic Editor: Bhaskar Kulkarni

Copyright © 2017 Onur Babat et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper analyzes the impact of the sensor reading errors on parameters that affect the production costs of a leading US industrial gas supply company. For this purpose, a systematic methodology is applied first to determine the relationship between the system output and input parameters and second to identify the assigned input sensors whose readings need to be improved in a prioritized manner based on the strength of those input-output relationships. The two main criteria used to prioritize these sensors are the decrease in production costs and the decrease in production costs' volatility obtained when the selected sensor's precision is improved. To illustrate the effectiveness of the proposed approach, we first apply it to a simplified version of the real supply network model where the results can be readily validated with the simulated data. Next, we apply and test the proposed approach in the real supply network model with historical data.

1. Introduction

The US industrial gas companies provide indispensable products like oxygen, nitrogen, and hydrogen to manufacturing, health care, transportation, and other essential industries worldwide. In a recent study by the American Chemistry Council, it was shown that industrial gas companies produced approximately \$17 billion worth of products in 2010 and employed approximately 60,000 American workers. Furthermore, the study showed that industrial gas companies supply products to industries in the US that account for 42% of America's Gross Domestic Product [1].

One of the key decision support techniques used by this type of companies is the solution of *industrial gas supply network optimization* models (see, e.g., [2]). These models are extremely helpful to identify optimal operating settings, by expressing real life constraints and conditions mathematically. Also, they are extensively used to describe and integrate all components of the industrial gas supply network within a single framework. However, industrial gas supply network optimization models are often difficult to

solve due to the presence of physical and quality constraints, which result in discontinuities and other nonconvexities (cf. [3, 4]) on the mathematical formulation of the problem (cf. [2]).

There is a vast literature on industrial gas supply network optimization models. For example, consider the work of Almansoori and Shah [5], Fonseca et al. [6], Kumar et al. [7], Ding et al. [8], Jiao et al. [9], and Jiao et al. [10]. Moreover, due to their nature, these models are subject to the presence of uncertainties at various levels. Integrating these uncertainties in industrial gas supply network models has also taken wide attention in the literature (see, e.g., [2, 11–14]). However, often times in practice, the uncertainties in the system are disregarded because their integration into the models increases model's complexity which is already very difficult to solve under deterministic assumptions. Alternatively, a common approach to avoid increasing the complexity of solving the industrial gas supply network model is to fix the value of uncertain parameters using a point estimate, in order to obtain a deterministic model that approximates the uncertain model.

While many of the uncertainties that appear in industrial gas supply network optimization models are due to the intrinsically random nature of the input parameters in the network such as flow rate, temperature, and pressure levels, industrial gas supply network optimization models are also affected by uncertainties resulting from incorrect sensor readings in the system. Some of these sensors provide feedback signals which are crucial for the control and efficiency of the network. Thus, erroneous sensor readings could degrade the performance of the network significantly.

Incorrect sensor readings could be due to the following:

- (i) Outliers in sensor readings: sometimes sensors may read a value that happens to be outside of the normal range of operation. This could be caused by an inherent error in the corresponding sensor or a sudden change in the supply network conditions (e.g., a sudden pressure drop). We identify the ones caused by inherent sensor errors, which we name as sensor outliers. Outliers are detected by determining whether an out of range reading is due to sudden changes in the network. This can be verified by other sensors readings (e.g., pressure sensors) in the network. Thus, by looking at the correlation between the out of range readings of associated sensor readings, a system change can be distinguished from an outlier reading due to inherent sensor error in the supply network.
- (ii) Bias in sensor readings: if the measured signal is shifted by a constant from the actual signal throughout the sensor reading time period, then the sensor has a constant offset or bias. The bias could be negative or positive.
- (iii) Noise in sensor readings: it refers to the high-frequency error component in the sensor measurements. This type of uncertainty is unavoidable and inherent to every sensor, but it can be improved through maintenance or upgrade.

These incorrect sensor readings can be critical because they are used in measuring the key input parameters of the system. In turn, the aggregate effect of inaccuracies in the model parameters leads to inaccuracies in the optimization model's output. This brings a negative effect on customer satisfaction and puts an unnecessary strain in the industrial gas supply network operation. However, these effects can be mitigated by upgrading and maintaining sensors to improve their reading's accuracy.

In this paper, we focus on the impact of improving sensor reading errors on the model's main output, namely, production costs. From now on, we will regularly refer to production costs as the output of the model. In particular, we use two key performance indicators to prioritize the improvement of sensors in the network.

- (i) Key performance indicator-1 (KPI-1): it measures the average change in the production costs' value over a time horizon when the sensor reading accuracy is improved. The change in the production costs' value

over the time is a result of the presence of outliers and bias in the sensor readings, and it has a direct financial meaning for the company because the elimination of these errors can bring savings on the production cost value.

- (ii) Key performance indicator-2 (KPI-2): this KPI measures the change in production costs' volatility when the sensor reading accuracy is improved. The volatility change in production costs' is due to the amelioration of outliers and random noise in sensor readings.

Production facilities prefer to have low uncertainty in their production systems. Often times these uncertainties cause the total production cost of the system to exceed the defined companies' limited operational budget. Moreover, planning under uncertainty is a difficult task for the companies because plant production levels are very sensitive to input parameters' volatility, as any change in the inputs may cause new production settings in these facilities. Shifting from a defined setting to another could bring extra hidden costs and even infeasibilities in the system. For all these reasons, while the average change in the production cost's value (KPI-1) is crucially important for the company's financial benefits, the change in production costs' variability (KPI-2) is equally crucial.

The faulty sensors affecting production cost accuracy can be addressed by upgrades and maintenance to have more precise readings. However, this is not possible for every single sensor in the system due to limitations on the budget allocated for sensors' upgrade and maintenance. Thus, the purpose of this study is detecting the sensors in a leading US industrial gas supply network whose inaccuracies have the biggest impact in the supply network.

To detect, identify, and determine the sensor faults, a systematic approach is needed. Traditionally, two ways to deal with sensor faults have been used: preventive maintenance and condition-based maintenance. Preventive maintenance is accomplished by regular checking and calibration of sensors, while condition-based maintenance is based on monitoring a process's real-time condition and automatically detecting sensor faults [15]. Sensor fault detection and identification methodologies have focused on the condition-based maintenance aiming at the development of automated sensor fault detection systems, which offer cost advantages over the preventive maintenance systems.

For example, consider the work of [16–20]. These methodologies construct complex predictive models to replicate actual sensors' behavior. Such predictive models can be constructed based on these methodologies using techniques such as principal component analysis (PCA), neural networks, and Bayesian belief networks, among others (see, e.g., [15]). However, the resulting models based on these methodologies make it difficult for the user to interpret the underlying relation between the input and output elements.

Here, we follow a similar idea and develop a methodology to approximate the relationship between outputs and inputs in the model by using appropriate sensitivity analysis tools (cf. [21, 22]). Sensitivity analysis methods have been frequently applied to chemical process optimization models

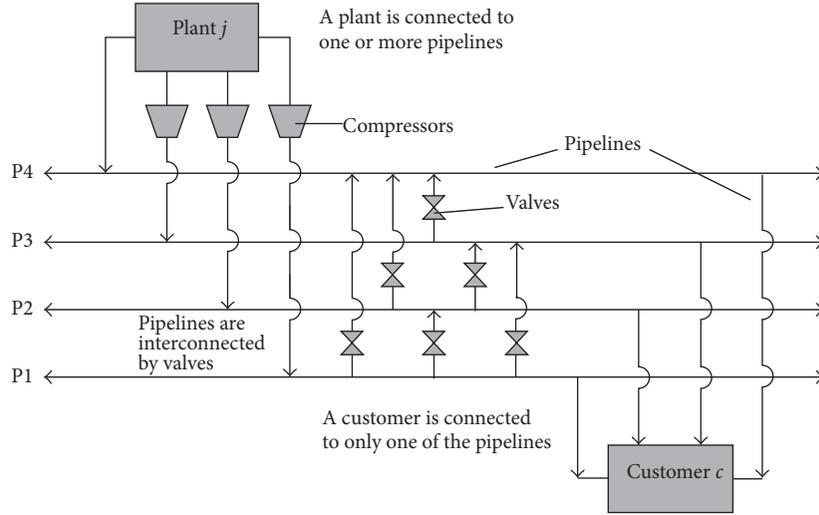


FIGURE 1: Network of plants, pipelines, and customers from [2].

in the literature (cf. [23, 24]). However, to the best of our knowledge, these techniques have not been applied in the context considered here.

After modeling the relationship between model inputs and outputs, we design a heuristic approach to eliminate sensor malfunctions and approximate the true signal. By integrating the true signal to the constructed predictive model, we calculate the relative change in the system outputs when a sensor is improved by computing the proposed KPIs. We will discuss the methodology in Sections 3 and 4.

The rest of the paper is structured as follows. In Section 2, we briefly describe the industrial gas supply network and review some relevant sensitivity analysis methods. The methodology used in this paper is described in Section 3. In Section 4, the chosen methodology is applied to a simplified gas network model, where the corresponding results can be readily validated. In Section 5, we discuss the results obtained after applying this methodology to a real industrial gas network in the US. We conclude the paper in Section 6 with some final remarks.

2. Industrial Gas Supply Network

The optimization model of an industrial gas supply network can be mathematically formulated as follows [2]:

$$\begin{aligned}
 \min_{f,p} \quad & y(f) \\
 \text{st.} \quad & Af = d, \\
 & g_i(f, p) = 0, \quad \forall i \in I \\
 & h_j(f, p) \leq b, \quad \forall j \in J \\
 & L \leq f \leq U, \\
 & p^- \leq p \leq p^+,
 \end{aligned} \tag{1}$$

where A is a matrix representing the structure of the network, f represents the vector of flows in the network, and p is the vector of pressures. The objective function minimizes the cost of gas production y as a function of the network flows f (more precisely, flows at production facilities). The first set of constraints $Af = d$ ensures the customer demands and flow balances are satisfied. The second set of constraint $g_i(f, p) = 0, \forall i \in I$ represents physical constraints relating flows and pressures throughout the network. The constraints $h_j(f, p) \leq b, \forall j \in J$ ensure that the model solution satisfies operational quality standards. Finally, in model (1), both pressures and flows should remain between allowed bounds.

Model (1) is a nonconvex, nonlinear deterministic optimization problem, which for real-sized networks is very difficult to solve to optimality and typically can be only approximately solved. Although we cannot provide here the actual customer and pipeline layout due to confidentiality, Figure 1 shows a representative layout of the industrial gas network. The network in the company involves tens of customers and plants, which causes the pipeline flow model to be computationally expensive and takes in the order of minutes to be approximately solved. Referring to Figure 1, sensors in the network are typically located at each customer node, each plant node, and at the intermediate nodes where the different pipeline branches intersect. Those sensors read, among others, the gas flow, as well as pressure and CO concentration levels. The industrial gas supply network model is set to work in real-time, and it is an important advisory tool for setting the physical plant production levels.

In order to obtain the desired analysis of sensors in the industrial gas network, the first step is to obtain a predictive model that approximates the relationship between the sensor readings used as inputs in model (1) (e.g., demands and pressures) and the main output of model (1), namely, production costs.

To construct predictive models, several sensitivity analysis methods have been proposed in the literature, which can be divided into two main groups: local sensitivity analysis

methods and global sensitivity analysis methods (cf. [25]). The local methods provide a measure of the local effect on the model output under small changes of the model inputs (cf. [26]). When the relationship between the inputs and output can be described using a simple differentiable function, we can look at the partial derivative of the output function with respect to the input parameters to find out the local impact of the parameters on the model output.

However, the industrial gas supply network model we analyze is nonconvex, nonlinear and computationally expensive to solve (e.g., [2]). In this case, partial derivatives are not easily available. Thus, in order to obtain approximate partial derivatives in a simple effective way, we use sample historical runs of the optimization model for different parameter settings.

This type of approach based on historical or simulated data to gain information about partial derivatives is typically referred to as global sensitivity analysis (cf. [27, 28]).

The total cost of the industrial gas network can be defined as

$$\mathbf{y}(\mathbf{x}) = [y^1(x^1), y^2(x^2), \dots, y^m(x^m)] \quad (2)$$

as the functions of uncertain inputs $\mathbf{x} = [x^1, x^2, \dots, x^m]$, where $x^i = [x_1^i, x_2^i, \dots, x_k^i]$ represents the values of k parameters in the system. From now on, we denote these uncertain inputs with \mathbf{x} . In turn, uncertainty in the \mathbf{x} parameters results in a corresponding uncertainty in the output $\mathbf{y}(\mathbf{x})$.

Different sensitivity analysis techniques will do well on different types of problems. The important aspect here is choosing the most suitable methodology to determine a predictive model between input parameters and output costs. For linear models, linear relationship measures like partial correlation coefficients (PCC) (cf. [29]) will be adequate. For nonlinear but monotonic models, measures based on rank transforms like partial rank regression coefficient (PRCC) (cf. [30]) will perform well. For nonlinear and nonmonotonic models, methods based on decomposing the variance of the output are the best choice. One of the examples of these methods is Sobol's method (cf. [31]).

Here, we chose regression coefficients (RC) (cf. [32]) as the tool to estimate the desired predictive model. The magnitude of the coefficients in this predictive model will be the indicators for identifying the key input parameters in the system. Favorable results of this least squares approach can be verified in Section 4.

After identifying the key parameters in the industrial gasses optimization model using an appropriate sensitivity analysis, we need to do further analysis to measure the impact of the errors in sensor readings. For this purpose, we develop a series of heuristic approximation methodologies to eliminate these sensor reading errors step by step that helps us to conduct the sensor improvement analysis in the network.

3. Methodology

A simple approach to address the problem considered here is to upgrade a sensor based on its current reading accuracy which aims to reduce only process variation. Specifically,

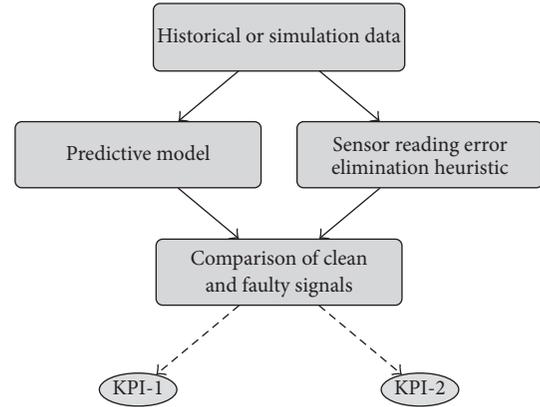


FIGURE 2: Sensor fault detection process flowchart.

this approach allocates a budget so that the sensors with the highest inaccuracies are upgraded to decrease model output's variation. However, this would not necessarily be the best way to eliminate errors from the system because of two main reasons. First, high input uncertainty does not always lead to high output uncertainty (in our case the production cost). Second, this simple approach also fails to consider the importance of the evaluation criterion KPI-1, by ignoring the complex impact of the sensor reading errors in the total production costs.

Here we propose a methodology which is suitable to work in real-time and with scarce data that provides an effective advisory tool to make decisions regarding the improvement of sensors in the network within the budget limitations. The historical data of the network in the company's database is enormous. However, the older the data gets, the more meaningless it becomes. Thus, the methodology has to be run with recent data.

Here we provide a brief explanation of each step of the methodology represented in Figure 2.

- (i) The first step is to generate samples of the input parameters $\mathbf{x} = [x^1, x^2, \dots, x^k]$ using simulation or by obtaining historical data from previous runs from the industrial gas network model. In our case study problem in Section 4, we generate the samples using Monte Carlo simulation while we use historical data from the gas supply network in the real case implementation given in Section 5.

After collecting the data, we run the following two steps in parallel.

- (ii) A predictive model is constructed to approximate the relationship between production costs and input parameters. The least squares approach provides a good approximation to get the desired sensor improvement prioritization in the case study problem. However, due to the possible multicollinearity between the input parameters in the real system, we apply the ridge regression (cf. [33]) to consider this factor. The ridge regression technique basically penalizes the size of

regression coefficients due to multicollinearity in the model.

- (iii) A heuristic methodology for sensor reading error elimination is applied. The heuristic methodology is applied as a univariate analysis: that is, it is applied to a single sensor at a time. For the selected sensor reading, the heuristic starts by looking at the outliers to eliminate. Note that outliers definitely produce an increase in the production cost's volatility and may cause an increase in the production cost's value depending on the outlier's position and input-output relationship. After detecting and eliminating the outliers, we eliminate the constant bias in the selected sensor reading. Constant bias in sensors can bring a substantial amount of extra cost in the system. Since it has a financial impact on the output, this needs to be certainly handled in the analysis. Finally, the heuristic approach eliminates the noise in sensor reading. Every sensor noise is assumed to follow a Gaussian distribution $\mathcal{N}(0, \sigma)$ with mean 0 and standard deviation σ . Noise in sensors can cause extra volatility in the system outputs (e.g., production cost, optimal production settings of the plants), and thus it brings up some indirect financial impact to the company.
- (iv) After running the predictive model and the three-step sensor reading error elimination heuristic to approximate the real sensor signal, we reproduce the production costs with the improved signal.
- (v) In the final step of the model, we calculate two key performance measures previously defined as KPI-1 and KPI-2 that are used to support the decision-making process. These KPIs help us to capture the marginal contribution of a single sensor reading error and assign the priorities to the sensors for upgrading purposes. KPIs are calculated based on the comparison of the original production cost values and the regenerated production cost values for every improved sensor one at a time.

To validate the methodology, we first apply it to a case study problem which is introduced in Section 4. After its validation with the simplified network, in Section 5, we show how this methodology together with some adaptations can be applied to a real industrial gas network.

4. Case Study Problem

In this section, we illustrate the details and the results of applying the aforementioned sensitivity methodology by implementing it to a simplified version of an industrial gas network model. This allows providing a clear description of the proposed methodology and validates the approach by providing the desired results regarding the prioritization of sensor improvements in an industrial gas supply network.

4.1. Problem Setting. Consider a simplified version of the industrial gas network depicted in Figure 3, where there are

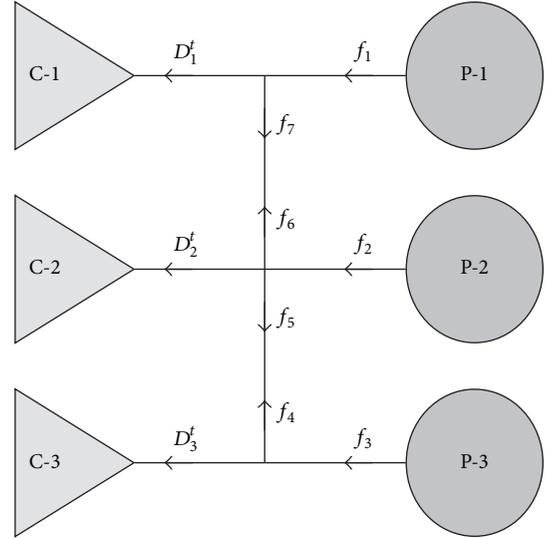


FIGURE 3: Simplified industrial gas network model with three (3) plants (P-1, P-2, and P-3) and three (3) customers (C-1, C-2, and C-3). Arrows indicate the directions of gas flows.

three plants (P-1, P-2, and P-3) and three customers (C-1, C-2, and C-3) with the decision variables and parameters described in Abbreviations.

In the simplified network, the customers submit their demands to the system which are then fulfilled at the lowest possible cost while satisfying the flow balance and flow bound constraints. The customer demands define the flow rates sent to the customers in the system. In this case, consider the following optimization model:

$$\begin{aligned}
 \min \quad & \sum_{i=1}^7 \alpha_i f_i^2 + \beta_i f_i, \\
 \text{st.} \quad & f_1 + f_6 - f_7 = D_1^t, \\
 & f_4 + f_2 + f_7 - f_6 - f_5 = D_2^t, \\
 & f_3 + f_5 - f_4 = D_3^t, \\
 & f_i \in [L_i, U_i], \quad \{i = 1, \dots, 7\}.
 \end{aligned} \tag{3}$$

The first three constraints in formulation (3) are flow balance constraints (i.e., $Af = d$ formulation (1)). The last constraint set represents the bound constraints on the flows. Formulation (3) described above is a simplified version of model (1) where constraints defining physical and operational quality standards are disregarded, and there are no pressure constraints. The sensors subject to the analysis are chosen to be the flow rate sensors at customer nodes. Sensors measuring pressure levels at demand nodes are used only as a system check. The objective function is assumed to be convex and quadratic (i.e., $\alpha_i \geq 0$, $i = 1, \dots, 7$). This setting allows us to obtain sample data for the global sensitivity analysis easily by solving the model with state of the art optimization solvers.

4.2. *Problem Data.* After setting up the problem, we simulate the data for the input parameters of model (3), which is the first step of the flowchart in Figure 2.

The values of the input parameters, except for the demand parameters (cf. Section 4.2.1) are defined as follows: $\alpha = [10, 5, 2, 0, 0, 0, 0]$, $\beta = [10, 5, 2, 0, 0, 0, 0]$, $L = [0, 0, 0, 0, 0, 0, 0]$, and $U = [150, 70, 50, 10, 10, 10, 10]$, where α is the quadratic cost coefficient vector in \$ per square unit of flow and β is the linear cost coefficient vector in \$ per unit of flow. These values have the following characteristics that will be the key to illustrate the effectiveness of the proposed approach.

- (i) The most expensive plant is plant P-1 while the cheapest one is plant P-3.
- (ii) Similarly, plant P-1 has the largest production capacity while plant P-3 has the least.
- (iii) Note that α and β coefficients are nonzero only for flows f_i , $i = 1, 2, 3$ corresponding to plant productions. This means that the cost only increases with the production levels in plants.
- (iv) There is a limit on the demand of customers that can be supplied from nonadjacent plants. For instance, customer C-1 can only receive a certain amount of gas from plant P-2 and plant P-3. The main supplier of customer C-1 is plant P-1. This also applies to the other customer demands. In particular, flow values f_i , $i = 4, 5, 6, 7$ in the center pipeline are bounded above by 10 units.

4.2.1. *Customer Demand Simulations.* In order to simulate the customer demand, we use Autoregressive (AR) processes which are commonly used to forecast demands in industrial gas networks. These simulated demand profiles are shown in Figure 4. The AR(1) model specified for customer C-1 is given by the following equation:

$$D_1^t = 1 + 0.99D_1^{t-1} + \epsilon_1^t, \quad (4)$$

where $\epsilon_1^t \sim \mathcal{N}(0, 1)$, $D_1^0 = 130.53$, the average value of the customer C-1 demand is given by $\mu_1 = 138.11$, and the standard deviation of the demand is provided as $\sigma_{(D_1^t)} = 3.360$. The AR(1) model specified for customer C-2 is given by the following equation:

$$D_2^t = 1 + 0.95D_2^{t-1} + \epsilon_2^t, \quad (5)$$

where $\epsilon_2^t \sim \mathcal{N}(0, 1)$, $D_2^0 = 60.67$, the average value of the customer C-2 demand is given by $\mu_2 = 62.03$, and the standard deviation of the demand is provided as $\sigma_{(D_2^t)} = 2.360$. The AR(2) model specified for customer C-3 is given by the following equation:

$$D_3^t = 0.7 + 0.7D_3^{t-1} + 0.25D_3^{t-2} + \epsilon_3^t, \quad (6)$$

where $\epsilon_3^t \sim \mathcal{N}(0, 1)$, $D_3^0 = 14.29$, the average value of the customer C-3 demand is given by $\mu_3 = 17.74$, and the standard deviation of the demand is provided as $\sigma_{(D_3^t)} = 2.835$.

The Pearson correlation coefficients of the simulated demand profiles are $\rho_{12} = 0.278$, $\rho_{13} = 0.005$, and $\rho_{23} = 0.137$.

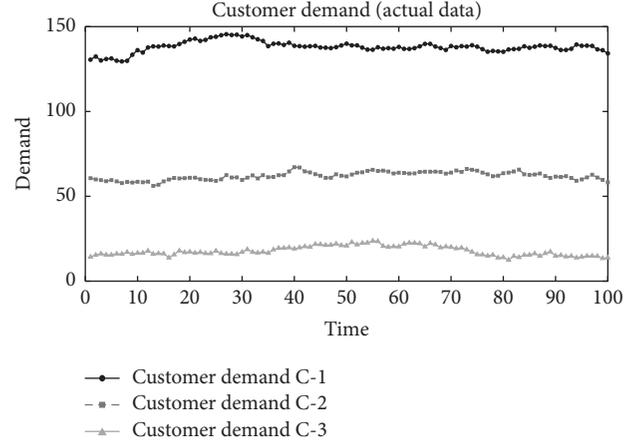


FIGURE 4: Simulated base customer demands.

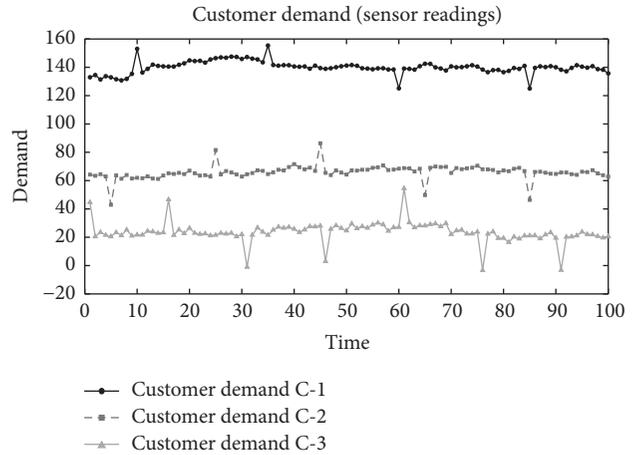


FIGURE 5: Simulated customer demand sensor readings.

4.2.2. *Customer Flow Sensor Reading Simulations.* To simulate the errors due to sensor readings of demand profiles, we add some simulated out of range readings which are the candidates to be the outliers, constant bias B , and Gaussian noise $N_i^t \sim \mathcal{N}(0, \sigma_i)$ to the demand profiles. That is, we set

$$\bar{D}_i^t = D_i^t + O_i^t + B_i + N_i^t, \quad i = 1, 2, 3, \quad (7)$$

where \bar{D}_i^t represents the sensor readings, O_i^t is the vector of out of range readings suspected to be outliers, $N_i^t \sim \mathcal{N}(0, \sigma_i)$ with $\sigma = [0.5, 1, 1.5]$ are the sensor noises, and $B = \{2, 4, 6\}$ is the constant bias vector.

The resulting sensor reading data can be seen in Figure 5, where potential outliers are highlighted by big markers over data points.

As it can be noticed from the parameter values in (7) and Figure 5, customer C-3 has the greatest sensor reading errors in terms of all three different error types. On the other hand, customer C-1 has the most accurate sensor reading among the other customers. In these conditions, one would expect to prioritize improving the sensor readings for customer C-3

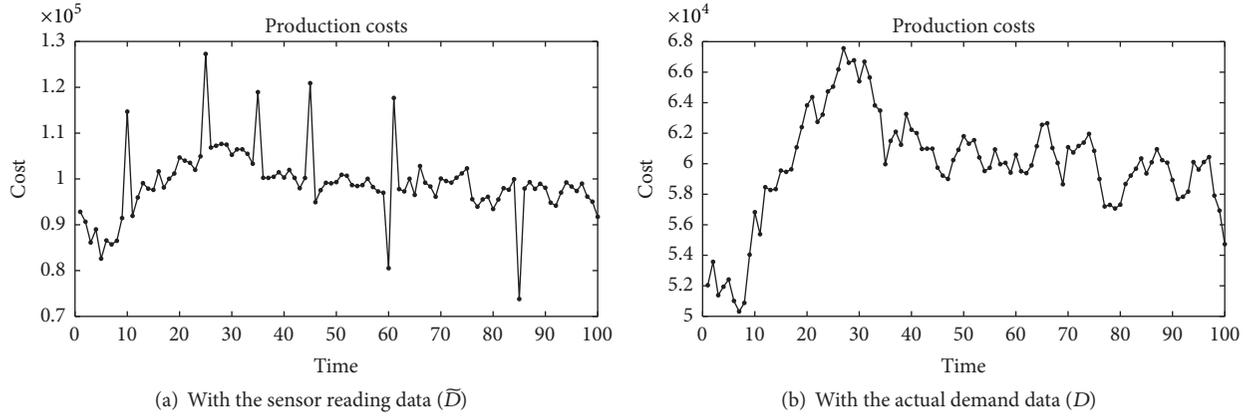


FIGURE 6: Cost function output through optimization model.

first, then customer C-2, and finally customer C-1. However, this would be deciding based on the input errors only. To consider the impact of input errors on the output, we need further analysis.

Operating the system by these faulty signals observed in Figure 5 may cause large deviations and higher values in the production costs. These deviations can be inspected by comparing the plots displaying production costs produced with the sensor reading data given in Figure 6(a), and the production costs values produced with the actual demand data given in Figure 6(b). The real life optimization model is computationally very expensive; for this reason, we need a simple way to determine the effect of input errors on the production cost. As mentioned before, this is why sensitivity analysis tools are used to construct a predictive model between model's inputs and outputs. We will compare our findings by controlling simulated sensor reading errors and show that the proposed methodology estimates the correct order of the sensors for improvement.

4.3. Predictive Model. In this section, we apply the second step of the developed methodology in Figure 2. After collecting the data in (2) for the vector of uncertain inputs denoted by same symbol as in (2), and populating the sensor readings of customer demands \bar{D} , we solve model (3) T times with uncertain input parameters \bar{D} to optimality with current optimization solvers [34]. In particular, we use the quadprog solver in a MATLAB environment [35]. The time required to solve model (3) with 3 customers and 3 flow sensors is only of about 0.01 seconds.

The resulting production costs $y(\bar{D})$ are illustrated in Figure 6(a), where we set $T = 100$ as the number of observations in the data set. In the real network, this number corresponds to the daily amount of data the company collects from the optimization model. We work with daily data because the system conditions can vary considerably for longer periods. Such high variations cannot be captured in a linear predictive model.

By using the input data created from the simulation of \bar{D} and the output of the optimization model $y(\bar{D})$, we get the following linear equation as our predictive model:

$$\hat{y} = -67044 + 1208.6 * D_1 + 445.65 * D_2 + 138.10 * D_3, \quad (8)$$

where \hat{y} represents the estimated production cost values. Since the multicollinearity is not an issue in the case study problem, ridge regression model coefficients would be expected to be identical to linear regression model coefficients for any ridge parameters.

The residuals of the regression model also satisfy statistical independence, homoscedasticity, and normality assumptions.

According to the regression equation (8), the relative importance of the input customer demands is sorted from more to less important as the demand of customer C-1, the demand of customer C-2, and the demand of customer C-3. This is expected because of the flow restrictions in the pipelines. Customer C-1 has to satisfy most of its demands by using the most expensive gas which is produced in plant P-1 while customer C-3 uses mostly the least expensive gas produced in P-3.

4.4. Sensor Reading Error Elimination Heuristic. After defining the predictive model, we run the heuristic approach to eliminate each type of sensor errors step by step. As discussed before, the predictive model and the heuristic for sensor reading error elimination are independent processes as shown in Figure 2. Moreover, the heuristic approach is a univariate analysis that is applied to each sensor individually.

4.4.1. Outlier Elimination. The first step of the heuristic approach for sensor reading error elimination is the outlier elimination. The out of range readings are often observed as unusual spikes in flow readings. As mentioned before in Section 1, these inaccuracies may happen because of one of the following reasons: due to an inherent error in the sensor

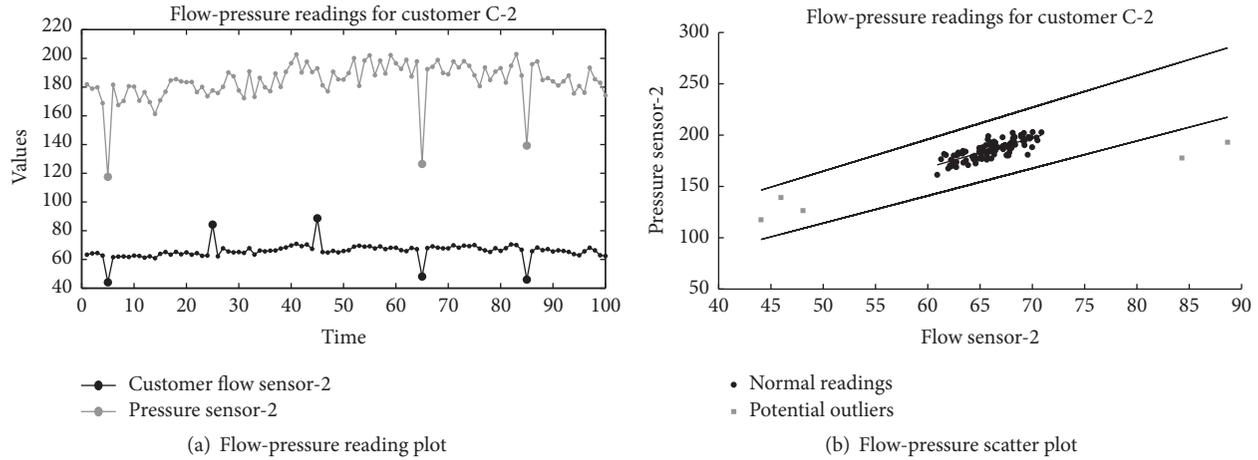


FIGURE 7: Simulated flow-pressure plots for customer C-2.

reading or due to an actual sudden change in conditions in the supply network (i.e., like sudden pressure drops). In the first case the error can be reduced by upgrading the sensor. In the latter case, such sudden changes in the network can be determined by other sensors in the network. Thus, by looking at the correlation between the out of range readings of appropriate groups of sensors, outlier readings due to reading errors can be distinguished from out of range readings due to sudden changes in the supply network. Here, we look at the pressure reading values given that pressure and flow rates are highly correlated in industrial gas supply networks.

To detect the potential outliers, we will use principal component pursuit (PCP) analysis. PCP optimally decomposes a data matrix as the sum of a low-rank matrix and a sparse matrix. Given a data matrix H , PCP is the solution of the convex optimization problem,

$$\begin{aligned} \min \quad & \|Z\|_* + \lambda \|E\|_1, \\ \text{st.} \quad & H = Z + E, \end{aligned} \quad (9)$$

where $\|Z\|_*$ is the nuclear norm; that is, the sum of the singular values of Z , and $\|E\|_1$ is equal to the sum of the absolute values of the elements of E . Under certain conditions, and with a Lagrange multiplier λ , optimization problem (9) recovers a low-rank matrix Z corresponding to a fault-free process condition and a sparse matrix E that has nonzero entities corresponding to sensor and process faults, which can be considered as sensor reading abnormalities or sharp changes in the system (cf. [36]). The matrix H is the selected flow sensor reading data vector \bar{D} in our case. The resulting nonzero values in the E vector help us to identify out of range readings.

After implementing the PCP routine above, we can identify the out of range readings in the sensor signal similar to those illustrated in Figure 5. After identifying the out of range readings, we look at the correlation of these sensor readings with pressure sensor readings associated with the same customer. If the sharp spikes in flow sensor readings are caused by a sudden change in the network, these secondary sensors for pressure readings will also have abnormal readings.

Let us illustrate this process by looking at the correlation between the flow sensor readings and the pressure readings of customer C-2. In Figure 7(a), two different signals for customer C-2 are shown: the sensor readings for customer C-2 demand and the simulated pressure sensor reading for customer C-2. Obviously, the marked points in the flow readings are strong candidates for outliers. However, notice that, in a few of these points, the pressure level also has abnormal spikes. These spikes occur at exactly the same timestamps that the spikes occur in the flow readings. These out of range readings are strong indicators of the sudden changes in the system, and they shall not be classified as sensor reading errors.

This can be better viewed by inspecting the scatter plot between the pressure and the gas flow in Figure 7(b). The PCP routine identifies out of range readings in the selected flow sensor and marks them as square dots. Dark circle dots are marked as normal readings for both pressure and flow sensors. We fit a least squares line between these flow and pressure readings excluding the out of range observations. Then, a 95% confidence interval is built around this least squares line. If the out of range readings are beyond these confidence intervals, then they are marked as sensor outliers, and we replace the sensor readings with the values from the low-rank matrix Z for these corresponding points. In Figure 7(b), the outlier points are the two square dots at the right side of the graph. These points are identified to be sensor reading errors and needed to be eliminated from the sensor signal. On the other hand, if some of these out of range readings fall between the confidence intervals provided in Figure 7(b), they are treated as sudden changes in the conditions of the supply network, not as outliers. In Figure 7(b), the out of range readings caused by system changes are denoted by the three square dots at the left side of the graph. Notice that they fall between the two confidence interval lines.

The same procedure is also applied to customer C-1's and customer C-3's sensors, and resulting outliers are eliminated from the sensor readings. Figure 8 shows the demand profiles

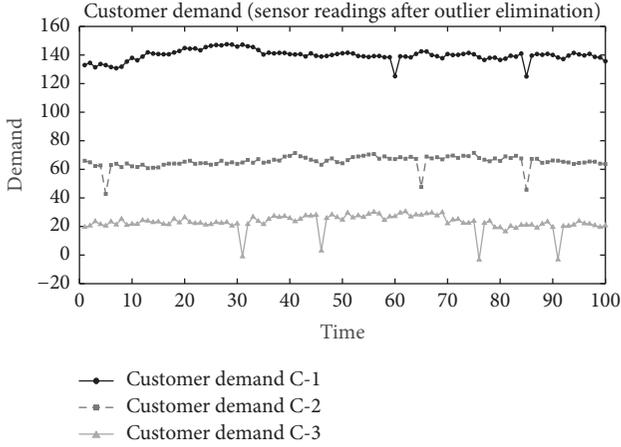


FIGURE 8: Customer demand data (without outliers).

after elimination of the outliers caused by inherent sensor errors.

4.4.2. Constant Bias Elimination. Constant bias elimination is the second step of the heuristic methodology to eliminate the sensor reading errors. The flow readings in industrial gas networks can have a constant bias in their measurement besides random noise and outliers. This inherent bias is difficult to detect with data analysis because of the constant shift in the parameters' sensor reading data.

In practice, bias is detected by putting a different but precise sensor next to the biased one. In this way, we can approximate the accurate reading of the precise sensor and detect the bias amount that the malfunctioning sensors have. For the case study problem, we need to measure the bias in a similar way. Specifically, we subtract the faulty sensor readings of the sensor from the nonerroneous sensor readings of the precise sensor. Then, we take the average of the difference over the selected period.

$$\bar{B}_i = \text{mean}(\bar{D}_i^t - D_i^t), \quad (10)$$

where $\bar{D}_i^t = D_i^t + N_i^t + B_i$ after the outliers are eliminated. To do this, we simulate the actual demand data for the new sensor, compute $D_i^t + N_i^t$, subtract it from the biased and noisy sensor readings \bar{D}_i^t , and take the average of the values for selected time horizon (n simulation points). The approximated bias values are the following: $\bar{B}_i = [1.98, 3.97, 5.81]$ under the assumption that the AR model coefficients are known in demand models (4), (5), and (6).

Recall from (7) that the bias $B = \{2, 4, 6\}$ was added to the demand profiles. According to these results, the bias approximation \bar{B} for the three customer flow sensors accurately predicts the actual bias values B with no more than 3.1% error.

4.4.3. Noise Filtering Approach. After eliminating the other sensor errors, that is, outliers and bias, we need to remove the noise from sensor readings which is the third and last steps in the heuristic approach for sensor reading error

```

Input:  $D$ 
(1) procedure NOISEFILTERING
(2)  $size \leftarrow \text{length of } D$ 
(3)  $D_{new}(1) \leftarrow D(1)$ 
(4)  $k \leftarrow 2$ 
(5) for  $j=2: size$  do
(6)    $diff \leftarrow |D(j) - D_{new}(j-1)|$ 
(7)   if  $diff \leq 3 * \sigma$  then
(8)      $avg = (\sum_1^k D(j-k+1))/k$ 
(9)      $[D_{new}(j-k+1), \dots, D_{new}(j)] = avg$ 
(10)     $k \leftarrow k + 1.$ 
(11)   else
(12)      $D_{new}(j) \leftarrow D(j)$ 
(13)      $k \leftarrow 2$ 
Output:  $D_{new}$ 

```

ALGORITHM 1: Noise filtering algorithm.

elimination. To do so, we present a naive filtering algorithm (cf. Algorithm 1). This filtering approach is similar to both the well-known Savitzky-Golay and moving average filters (cf. [37, 38]). While the proposed filter tries to separate the sensor noise from the process noise, which is not a simple task, it behaves conservatively by considering a portion of the actual demand volatility as sensor noise. However, this does not degrade the filter's performance and approximates the true demand signal as it can be seen from Figures 9(a), 9(b), and 9(c).

The filtering process starts by selecting the first observation as the starting observation. Then, it looks at the adjacent observation and decides whether the next observation differs significantly from the selected point based on the variance information of the noise. If the difference is significant, then we leave the next observation as it is. Otherwise, we take the moving average of the observations and replace the nominal values with the averaged values. The details of the noise filtering algorithm are given in Algorithm 1.

In Figures 9(a), 9(b), and 9(c), we display the simulated data for customer demands for our problem setting before and after the noise filtering.

4.5. Numerical Results and Verification of the Methodology. Having perfect knowledge of both the output of the optimization model (3) and the error-free demand profiles in Figure 4 gives us a chance to confirm our methodology's validity. We already have the original production costs values, which we got from the optimization model by running it with faulty sensor demand readings (\bar{D}_i). To compute the proposed KPIs, we calculate the average production costs values and the production costs volatility, which is the variance of the values over the selected time horizon. After that, we select a sensor, and we use the perfect error knowledge that we created in Section 4.2.2 to eliminate its inherent sensor errors. While doing that, we keep the other sensors faulty to see the effect of elimination of a single sensor reading errors on the output. Then, we integrate the correct error-free signal values for selected input parameter to the optimization model and

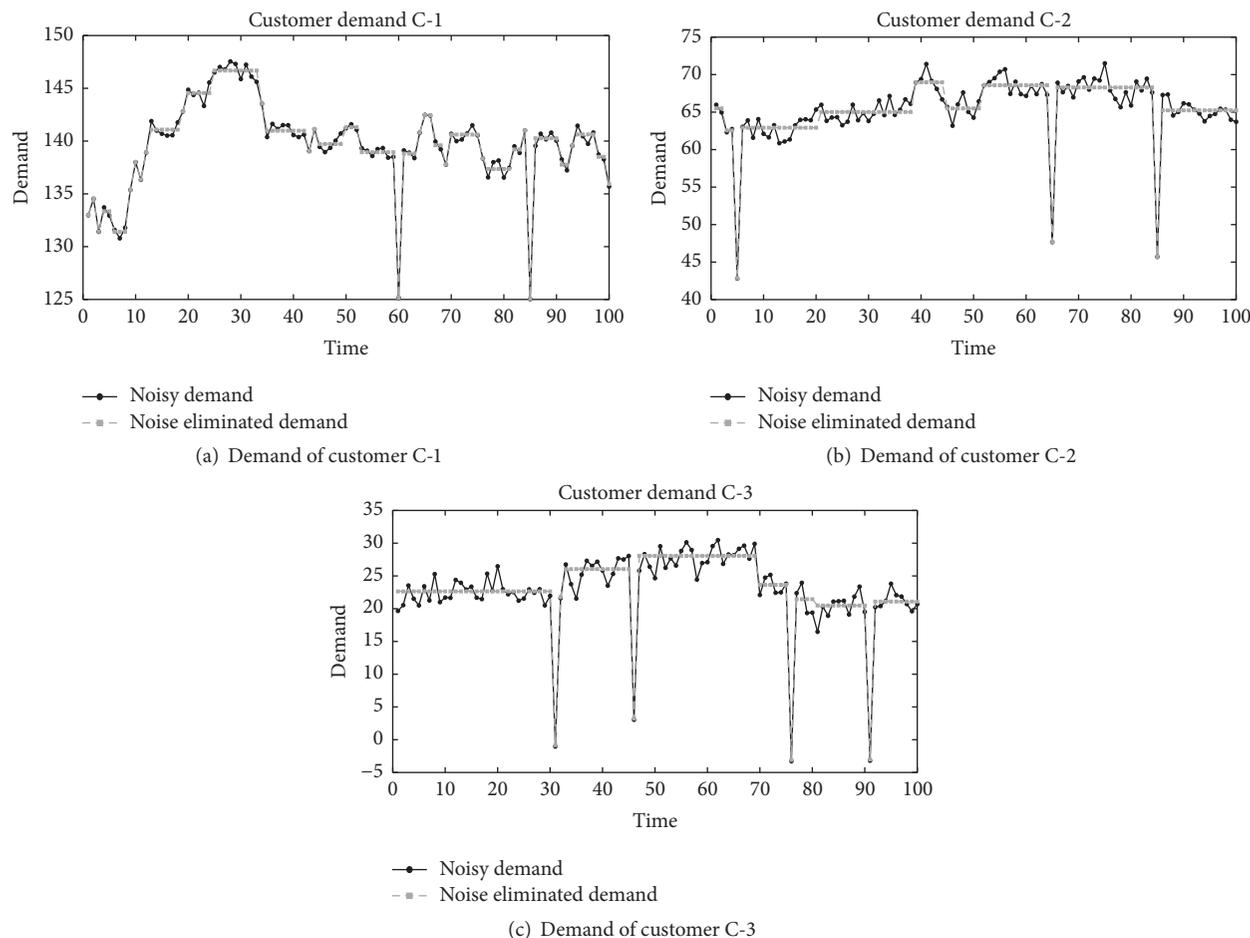


FIGURE 9: Noisy data and filtered data for customer demands.

reproduce the production costs values. Finally, we calculate the average production costs values and the production costs volatility by using the regenerated production values.

When we introduced sensor readings for customer demands (\bar{D}_i) back in Section 4.2.2, we designed the errors to be the greatest for customer C-3 and the least for customer C-1. Now, we are going to inspect the analysis results to see what sensors need to be prioritized for the improvement.

Remember from Section 3 that we have two different KPIs. First, we check the results for KPI-1, which is the average change in the production costs' value over a time horizon. Table 1 displays the impact of the sensor reading error elimination on the production costs in a controlled way by using the optimization model and the simulated error information.

After that, we apply our methodology to sensor readings as a univariate analysis to validate our methodology and approximate the true signal of the selected input parameter. While doing that, we keep the other sensors faulty to see the effect of elimination of a single sensor reading errors on the output. Then, we integrate the approximated error-free signal values for the selected input parameter to the predictive model and reproduce the production costs values. Finally, we calculate the average production costs values

and the production costs volatility by using the regenerated production values.

Table 2 displays the impact of the sensor reading error elimination on the production costs value (KPI-1) with the sensor error elimination methodology given in Figure 2.

According to the comparison of results between Tables 1 and 2, it follows that customer C-1 has the least amount of sensor reading errors, but it has the top priority to be fixed. This is because of two reasons. First, customer C-1 is using the most expensive gas in the network produced due to restrictions in the pipeline. Second, customer C-1's average demand value is much higher than the other customer demand profiles. These reasonings can be verified by the provided regression model's coefficients (8). The D_1 's coefficient is reflecting the significance of the cost related to customer C-1. Although the results in Table 2 are not numerically precise to estimate the real changes given in Table 1, the differences between the numerical results are small and produce a priority ranking in which the sensors are ordered as customer C-1, customer C-2, and customer C-3 in agreement with Table 1.

Now, we check the results for KPI-2, which measures the change in the production cost's volatility. Table 3 displays the impact of the sensor reading error elimination on the

TABLE 1: Error elimination of sensors with perfect knowledge (KPI-1).

Fixed sensors	Avg. cost function value	Avg. saving	Avg. Improvement
All errors present	\$9.90E + 04	—	—
Fix sensor-1 error only	\$9.56E + 04	\$3400	3.45%
Fix sensor-2 error only	\$9.63E + 04	\$2700	2.69%
Fix sensor-3 error only	\$9.84E + 04	\$600	0.58%

TABLE 2: Error elimination of sensors with the approximation methodology (KPI-1).

Fixed sensors	Avg. cost function value	Avg. saving	Avg. Improvement
All errors present	\$9.90E + 04	—	—
Fix sensor-1 error only	\$9.56E + 04	\$3400	3.45%
Fix sensor-2 error only	\$9.60E + 04	\$3000	3.07%
Fix sensor-3 error only	\$9.75E + 04	\$1500	1.54%

TABLE 3: Error elimination of sensors with perfect knowledge (KPI-2).

Fixed sensors	Cost function's variance	Avg. Improvement
All noise present	5.52E + 07	—
Fix sensor-1 error only	5.14E + 07	6.92%
Fix sensor-2 error only	5.25E + 07	5.05%
Fix sensor-3 error only	5.42E + 07	1.89%

TABLE 4: Error elimination of sensors with the approximation methodology (KPI-2).

Fixed sensors	Cost function's variance	Avg. Improvement
All noise present	5.52E + 07	—
Fix sensor-1 error only	5.29E + 07	4.26%
Fix sensor-2 error only	5.38E + 07	2.56%
Fix sensor-3 error only	5.43E + 07	1.71%

production costs volatility in a controlled way by using the optimization model and the error information.

Table 4 displays the impact of the sensor reading error elimination on the production costs volatility (KPI-2) with the sensor error elimination methodology given in Figure 2.

Once again, customer C-1's sensor ranks first in the priority list for the sensor improvement, while customer C-3's sensor ranks last according to Tables 3 and 4. This is again because of the customer C-1's and plant P-1's influences on the system. Similarly as for the KPI-1 results, the results in Table 4 are not numerically precise to estimate the real changes given in Table 1. However, we have the same priority ranking of the sensor maintenance ordered as customer C-1, customer C-2, and customer C-3 according to the results in both tables.

According to these results for both KPIs, we clearly see that both measures are suggesting, in order of importance, the upgrading of sensor-1 (demand of customer C-1) first, then sensor-2 (demand of customer C-2), and then sensor-3 (demand of customer C-3) in the industrial gas system.

This is contrary to the order of the sensors' input error magnitudes discussed at the beginning of Section 3. Thus, these results show that analyzing the effects of input errors on the output (production costs) helps us to take better decisions. The computing time of the methodology for the case study problem with 3 customers and 3 flow sensors takes 1.23 seconds.

5. Implementation to the Real Pipeline System

In this section, we discuss the results obtained by applying the methodology to the company's real industrial gas network. The methodology we use for the real pipeline system is slightly modified to capture the real network's properties. First of all, we use the weighted ridge linear regression approach to obtain the predictive model instead of the ordinary least squares approach used in Section 4.3. The ridge regression approach is chosen because it addresses multicollinearities by imposing a penalty on the size of coefficients. Multicollinearities are possible between the input parameters because of the number of sensors (over 400 sensors) in the real network. This large number of predictors creates low ratio of number of observations to number of variables. It is also selected to be a weighted model because based on the expert's experience, the most recent information of the system carries more explanation of this very dynamic system. The industrial gas network of interest is a real-time optimization system, where the optimal plant production flows need to be updated frequently. Thus, we desire to have a suggestion mechanism running and reporting suspicious sensor readings on a daily basis. Thus, the predictive model and heuristic analysis are implemented by using the last 100 data points reading from the optimization model.

The predictive model is based on the weights calculated using the Euclidean distance between the latest observation and the other samples in the set. The ridge regression estimate $\hat{\beta}$ is defined as the value of β that minimizes

$$\min \sum_{i=1}^T w^i (y^i - x^i \beta)^2 + \lambda \sum_{j=1}^k \beta_j^2, \quad (11)$$

where λ is chosen based on a 5-fold cross-validation and the weights are computed by the following equations:

$$d(\mathbf{x}^i, \mathbf{x}^T) = \sqrt{(x_1^i - x_1^T)^2 + (x_2^i - x_2^T)^2 + \dots + (x_k^i - x_k^T)^2}, \quad (12)$$

where $\mathbf{x}^i = \{x_1^i, x_2^i, \dots, x_k^i\}$,

$$w^i = \frac{1}{d(\mathbf{x}^i, \mathbf{x}^T)}, \quad (13)$$

$$w^T = 1.$$

That is, we assign a weight of 1 to the most recent data point and for the other samples a weight equivalent to the inverse ratio of the Euclidean distance to the most recent sample.

Another modification in the methodology is leaving the bias elimination as an option to the end user. This is to avoid an extra investment cost at this stage of the study because the bias in sensors is estimated by placing a highly precise sensor to measure the same signal as the imprecise sensor measures. The general implementation of the methodology is aimed to be a statistical analysis only, and whenever the end user has the estimated bias information, there is an option in the interface for the user to give that information as an input to the analysis.

There are a few more specific implementations designed to make the analysis more meaningful and user-friendly. Based on the selection of the time period for the analysis, the end user can track these daily reports before taking the final decision for the upgrading or maintenance decision for the sensors. The interface summarizes the daily reports for the selected time period and suggests the most important ones for a possible upgrade or maintenance based on the analysis we describe in Section 5.1. Also, highly improving the sensor precision can be very costly. Due to this fact, the end user is also given the option to see the effects of partial improvements such as eliminating 50%, 75%, or 100% of the sensor errors. The default value in the results presented in this section is 75%. This is typically the fraction of the improvement that the company considers.

In the real industrial gas network problem, the uncertain inputs are not only customer demands as in the case study in Section 4. Thus, random input variables can have very different units. In this case, parameters' regression coefficients can be easily influenced by the units in which the variables are measured, for example, gallons, pascals, bars, or grams in the real pipeline. Therefore, they do not provide a very reliable measure of the relative importance of the input variables. So, to compare the relative importance of the input parameters, input and output variables need to be standardized before drawing conclusions from their predictive model coefficients (cf. [39]).

The described methodology is implemented in MATLAB and the historical sampling information is drawn by SQL queries from the company's database. The results are transferred to the company's online environment on a daily basis.

TABLE 5: Statistical measures of the sensor readings.

Sensor #	Mean	Standard deviation
Sensor-1	12.245	0.124
Sensor-2	19.101	6.228
Sensor-3	43.006	2.132
Sensor-4	63.736	1.489
Sensor-5	42.595	1.307
Sensor-6	100.603	0.496
Sensor-7	115.115	0.307
Sensor-8	27.214	1.536
Sensor-9	15.141	0.105
Sensor-10	6.697	0.800
Sensor-11	30.167	0.441
Sensor-12	3.040	0.281
Sensor-13	85.729	5.838
Sensor-14	101.628	4.635
Sensor-15	22.418	4.364
Sensor-16	85.477	3.144
Sensor-17	39.449	5.152
Sensor-18	63.932	1.346
Sensor-19	0.807	2.064
Sensor-20	31.591	3.714

5.1. Implementation Results. Next, we present some of the results obtained by applying the methodology discussed here in the real industrial gas supply network. To protect the company's intellectual property, the sensor details are not provided. Due to the same reason, the input data for the analysis is not provided, as well as the exact results of the analysis. Instead, we provide a relative improvement on production costs and its volatility to prioritize the sensors in the system.

The pilot implementation for the analysis is chosen to be run in the business days (22 days) of October 2015. Customer demands are measured by flow sensors, and Table 5 provides the mean value and standard deviation values of the readings of 20 of these sensors.

Additionally, pairwise correlations between the sensors' readings are shown by a heatmap in Figure 10 where the individual values of the correlation matrix are represented as colors. Just by considering the pairwise correlations between these readings, the multicollinearity between the sensor readings is quite apparent.

The objective of the implementation is to run the univariate analysis daily with the updated results data coming from real-time optimization and monitoring the parameter rankings based on their effect on the uncertainty and the value of the cost function. The resulting interface displays the sensor rankings in the order of positive savings on the production cost (KPI-1) and daily sensor rankings in the order of their sensor reading errors contribution on the objective function's volatility (KPI-2). While a single day analysis may not be meaningful due to the dynamic nature of real-time optimization system, a collection of the daily analysis can

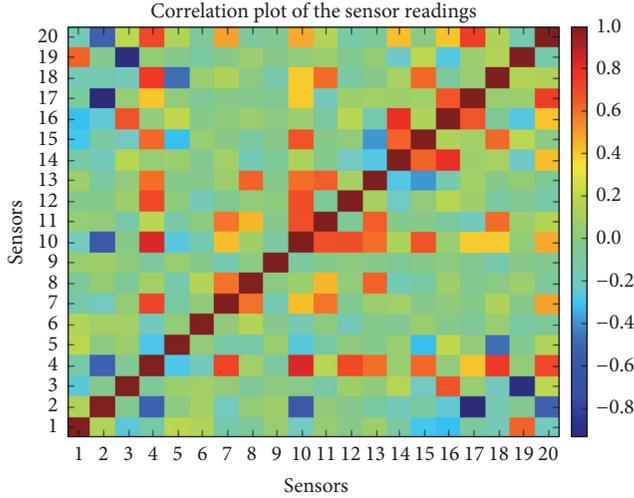


FIGURE 10: Pairwise correlations of the sensor readings.

TABLE 6: Rankings of top 5 sensors based on reduction in production cost value (KPI-1).

Parameter names	Rankings in KPI-1					Total
	1st	2nd	3rd	4th	5th	
Sensor-7	5	7	9	—	1	22
Sensor-8	3	4	1	5	1	14
Sensor-3	5	1	1	—	1	8
Sensor-11	2	1	—	1	3	7
Sensor-5	—	3	—	2	1	6

have strong suggestions to identify what sensors need to be upgraded or maintained.

Table 6 provides the top 5 sensors in this monthly collection, which ranked among the top 5 sensors based on KPI-1 in any of these daily runs. We note the number of times that any of these sensors were ranked in each of these top 5 positions. In each of the daily runs and in the summary, we only look at the top 5 sensors because based on our observations, most of volatility and cost improvements can be satisfied by upgrading or maintaining the top 5 sensors according to the daily analysis results.

According to Table 6 results, sensor-7 is leading the list by being ranked among the top-5 sensors 22 times out of 22 business days. It is ranked 5 times as the first sensor, 7 times as the 2nd sensor, and 9 times as the 3rd sensor and only once as the 5th sensor in this set of analysis. Sensor-8 also appears as one of the top sensors in almost two-thirds of the days analyzed.

The ranking of sensors displayed in Table 6 is one way of checking the significance of sensors in the system. We can also look at the average improvement in production cost's value over this month of analysis when the specified sensor is improved. These improvement values in percentage are calculated as follows:

$$\text{Avg. Improvement} = \frac{(\hat{y} - \hat{y}') / \hat{y}}{T} * 100, \quad (14)$$

TABLE 7: Ranking of sensors based on average production savings (KPI-1).

Parameter names	Avg. Improvement in KPI-1
Sensor-7	1.415%
Sensor-8	0.266%
Sensor-3	0.095%
Sensor-5	0.065%
Sensor-11	0.040%

where \hat{y} is the vector of the values of estimated objective functions before error elimination in the sensor, \hat{y}' is the vector for the values of estimated objective functions after error elimination in the sensor, and T is the number of data points in a day. "Avg. Improvement" gives us the relative difference before and after the error elimination. Positive values for "Avg. Improvement" imply positive savings based on the univariate analysis for the selected sensor.

The five highest priority sensors based on "Avg. Improvement" values in descending order are listed in Table 7. Table 7 suggests that we would reduce the daily production costs by around 1.4% on average if we go ahead and fix the errors of sensor-7. Sensor-7 is known as the flow sensor of an important customer in the system, and the sensor reading errors affect the production cost significantly. The daily average improvement is based on a month of analysis and it is an important number when it is translated to real money. For this reason, KPI-1 is definitely an important indicator that upper management would certainly consider.

The numbers in Table 7 are not additive, in other words, one cannot guarantee that the amount of improvement would be equal to the summation of percentage values if multiple sensors are decided to be upgraded. This is due to the univariate nature of the analysis; that is, we approximate the true signal for only one selected sensor at a time while keeping the other sensor reading errors present in the system.

Also, although the rankings and/or identities of the sensors in Tables 6 and 7 are almost identical to each other in this analysis, they do not necessarily have to match with each other as they rank the sensors based on different evaluations for the same key performance index. However, the similarity between the order of the sensors in Tables 6 and 7 is a strong indicator to consider prioritizing the suggested sensors for the maintenance based on the selected KPI. This also applies to the results going to be presented in Tables 8 and 9 based on KPI-2.

As we discussed in Section 1, it is also important to consider KPI-2 as a decision criterion while selecting the sensors for maintenance. Table 8 provides a summary of the analysis realized in October 2015 for the key performance indicator-2 (KPI-2). Specifically, it provides the number of times that a sensor is ranked among the top 5 sensors in terms of KPI-2 in any of these daily runs. Similar to the results of the KPI-1 analysis, sensor-7 was ranked as the first sensor for a possible precision upgrade in 12 out of 22 business days. On the other hand, it was ranked 5 times as the second sensor and 4 times as the third sensor to be improved in this monthly analysis.

TABLE 8: Rankings of top 5 sensors based on volatility reduction in production costs (KPI-2).

Parameter names	Rankings in KPI-2					Total
	1st	2nd	3rd	4th	5th	
Sensor-7	12	5	4	—	—	21
Sensor-3	5	4	3	2	2	16
Sensor-6	1	2	3	3	2	11
Sensor-11	—	5	1	1	1	8
Sensor-9	1	5	—	1	1	8

TABLE 9: Ranking of sensors based on average volatility reduction (KPI-2).

Parameter names	Avg. Improvement in KPI-2
Sensor-7	2.910%
Sensor-3	1.922%
Sensor-6	0.765%
Sensor-8	0.399%
Sensor-11	0.341%

Similar to Table 6, one can look at the average improvement in production cost's volatility over the selected time period when the specified sensor is improved. These improvement values in percentage are calculated as follows:

$$\text{Improvement} = \frac{\text{Var}(\hat{y}) - \text{Var}(\hat{y}')}{\text{Var}(\hat{y})} * 100, \quad (15)$$

where $\text{Var}(\hat{y})$ is the variance of computed objective function values before upgrading the sensor and $\text{Var}(\hat{y}')$ is the variance of estimated objective function values after upgrading the sensor. "Improvement" gives us the relative difference before and after the operation. Positive values for "Improvement" imply reduction in the objective function's volatility.

The top five sensors based on the average volatility reduction in the case of their maintenance are listed in Table 9. According to these results, volatility in production costs reduces around 2.9% in daily average when sensor-7's reading errors are tackled. Although the relationship is not obvious, due to the implied costs and inconveniences of having uncertainties in the problem, the reduction in production cost's volatility could result in great savings in the production cost's value. Similar to the results displayed in Table 7, the given percentage improvement for the sensors in Table 9 is not additive.

After reviewing the results of the analysis, the end user can decide what time period he or she is concerned with and what criteria are important for the company for the sensor improvements. In this pilot analysis summary conducted with October 2015's optimization model's historical data, sensor-7's reading errors significantly dominate the system in comparison with the other sensors in the system according to both of these KPI's and their ranking and/or average improvement criteria.

Finally, since the methodology is designed as a heuristic approach, it is crucial to report the computational run times of the methodology for different number of sensors. The

TABLE 10: Summary of 10 computational run times of heuristic approach for different number of sensors.

Number of sensors	Average run time (sec.)	Standard deviation of run times (sec.)
400	9.111	6.639
350	12.620	13.991
300	7.735	5.699
250	3.776	0.858
200	3.532	3.985
150	2.377	1.124
100	2.622	1.418

methodology is run 10 times for the instances with different number of sensors. Table 10 presents the average run time and the standard deviation of the time needed to run these experiments. Although the time required for solving an instance typically increases with the number of sensors, this solution time depends on the number of outliers, making experiments with lower number of sensors take longer time on average. For example, this is the case for the cases with 350 and 400 sensors. In general, computation times are small and in the range of 10 seconds.

On the other hand, the proprietary optimization of the real system is inflexible and cannot be run for different network sizes. This is due to the nature of the model which is highly complex and nonadjustable. We refer the reader to [2] for some recorded times of a similar optimization model on networks of different sizes.

6. Conclusion

In this paper, we presented a practical sensor fault identification and improvement methodology for an industrial gas supply network based on sensitivity and data analysis. We constructed predictive models based on global sensitivity analysis tools and then used some data analysis techniques to approach sensor's error-free signals. Then, we analyzed the benefits of having an error-free signal for each specified sensor. To validate the methodology, we presented the application of the methodology in a simple case study problem. Then, with a few modifications, we extended the same methodology to the real industrial gas network system. The verified approximation gives us the necessary tools to reduce the measurement inefficiencies in the network.

The results of the analysis are currently being used in decision-making processes to detect which sensors are providing suspect readings in a given period of time. Based on the application of the analysis, sensor repairs are going to be selected and realized based on the cost savings and/or the reduction of the volatility of the production cost for the company.

Abbreviations

Decision Variables

f_i : Flow variables in the pipeline, $i = 1, \dots, 7$

Input Parameters

- D_j^t : Random demand of customer j at time t ,
 $j = 1, \dots, 3, t = 1, \dots, T$
 U_i : Upper bound on $f_i, i = 1, \dots, 7$
 L_i : Lower bound on $f_i, i = 1, \dots, 7$
 β_i : Linear objective coefficients for each flow variable, $i = 1, \dots, 7$
 α_i : Quadratic cost coefficients for each flow variable, $i = 1, \dots, 7$.

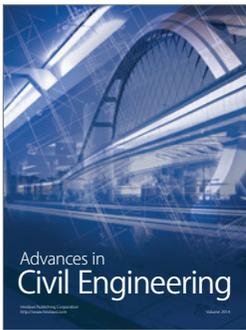
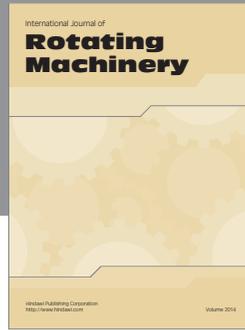
Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] A. C. Council, "Industrial gases energy and environmental benefits," 2012, <http://www.americanchemistry.com/ProductsTechnology/Industrial-Gases/Industrial-Gases-Energy-and-Environmental-Benefits.pdf>.
- [2] S. A. Van Den Heever and I. E. Grossmann, "A strategy for the integration of production planning and reactive scheduling in the optimization of a hydrogen supply network," *Computers & Chemical Engineering*, vol. 27, no. 12, pp. 1813–1839, 2003.
- [3] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, 2006.
- [4] H. W. Kuhn, "Nonlinear programming: a historical view," in *Traces and Emergence of Nonlinear Programming*, pp. 393–414, Springer, Berlin, Germany, 2014.
- [5] A. Almansoori and N. Shah, "Design and operation of a future hydrogen supply chain: snapshot model," *Chemical Engineering Research and Design*, vol. 84, no. 6, pp. 423–438, 2006.
- [6] A. Fonseca, V. Sá, H. Bento, M. L. C. Tavares, G. Pinto, and L. A. C. N. Gomes, "Hydrogen distribution network optimization: a refinery case study," *Journal of Cleaner Production*, vol. 16, no. 16, pp. 1755–1763, 2008.
- [7] A. Kumar, G. Gautami, and S. Khanam, "Hydrogen distribution in the refinery using mathematical modeling," *Energy*, vol. 35, no. 9, pp. 3763–3772, 2010.
- [8] Y. Ding, X. Feng, and K. H. Chu, "Optimization of hydrogen distribution systems with pressure constraints," *Journal of Cleaner Production*, vol. 19, no. 2-3, pp. 204–211, 2011.
- [9] Y. Jiao, H. Su, Z. Liao, and W. Hou, "Modeling and multi-objective optimization of refinery hydrogen network," *Chinese Journal of Chemical Engineering*, vol. 19, no. 6, pp. 990–998, 2011.
- [10] Y. Jiao, H. Su, W. Hou, and P. Li, "Design and optimization of flexible hydrogen systems in refineries," *Industrial & Engineering Chemistry Research*, vol. 52, no. 11, pp. 4113–4131, 2013.
- [11] J. Kim, Y. Lee, and I. Moon, "Optimization of a hydrogen supply chain under demand uncertainty," *International Journal of Hydrogen Energy*, vol. 33, no. 18, pp. 4715–4729, 2008.
- [12] A. Almansoori and N. Shah, "Design and operation of a stochastic hydrogen supply chain network under demand uncertainty," *International Journal of Hydrogen Energy*, vol. 37, no. 5, pp. 3965–3977, 2012.
- [13] Y. Jiao, H. Su, W. Hou, and Z. Liao, "Optimization of refinery hydrogen network based on chance constrained programming," *Chemical Engineering Research and Design*, vol. 90, no. 10, pp. 1553–1567, 2012.
- [14] J. Lou, Z. Liao, B. Jiang, J. Wang, and Y. Yang, "Robust optimization of hydrogen network," *International Journal of Hydrogen Energy*, vol. 39, no. 3, pp. 1210–1219, 2014.
- [15] A. Kusiak and Z. Song, "Sensor fault detection in power plants," *Journal of Energy Engineering*, vol. 135, no. 4, pp. 127–137, 2009.
- [16] R. Dunia, S. J. Qin, T. F. Edgar, and T. J. McAvoy, "Identification of faulty sensors using principal component analysis," *AIChE Journal*, vol. 42, no. 10, pp. 2797–2812, 1996.
- [17] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: part I: quantitative model-based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.
- [18] C. Lee, S. W. Choi, and I.-B. Lee, "Sensor fault identification based on time-lagged PCA in dynamic processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 70, no. 2, pp. 165–178, 2004.
- [19] N. Mehranbod, M. Soroush, and C. Panjapornpon, "A method of sensor fault detection and identification," *Journal of Process Control*, vol. 15, no. 3, pp. 321–339, 2005.
- [20] L. Guo and J. Kang, "A hybrid process monitoring and fault diagnosis approach for chemical plants," *International Journal of Chemical Engineering*, vol. 2015, Article ID 864782, 9 pages, 2015.
- [21] D. G. Cacuci, M. Ionescu-Bujor, and I. M. Navon, *Sensitivity and Uncertainty Analysis, Volume II: Applications to Large-Scale Systems*, vol. 2, CRC Press, Boca Raton, Fla, USA, 2005.
- [22] A. Saltelli, M. Ratto, T. Andres et al., *Global Sensitivity Analysis. The Primer*, John Wiley & Sons, Chichester, UK, 2008.
- [23] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Sensitivity analysis for chemical models," *Chemical Reviews*, vol. 105, no. 7, pp. 2811–2828, 2005.
- [24] P. Seferlis and A. N. Hrymak, "Sensitivity analysis for chemical process optimization," *Computers & Chemical Engineering*, vol. 20, no. 10, pp. 1177–1200, 1996.
- [25] E. Borgonovo and E. Plischke, "Sensitivity analysis: a review of recent advances," *European Journal of Operational Research*, vol. 248, no. 3, pp. 869–887, 2016.
- [26] R. A. Howard, "Uncertainty about probability: a decision analysis perspective," *Risk Analysis*, vol. 8, no. 1, pp. 91–98, 1988.
- [27] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice*, John Wiley & Sons, Chichester, UK, 2004.
- [28] T. Homma and A. Saltelli, "Importance measures in global sensitivity analysis of nonlinear models," *Reliability Engineering & System Safety*, vol. 52, no. 1, pp. 1–17, 1996.
- [29] B. L. Brown and S. B. Hendrix, *Partial correlation coefficients*, Wiley StatsRef: Statistics Reference Online, 2005.
- [30] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.
- [31] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, no. 1–3, pp. 271–280, 2001.
- [32] H. M. Wagner, "Global sensitivity analysis," *Operations Research*, vol. 43, no. 6, pp. 948–969, 1995.
- [33] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [34] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [35] T. Coleman, M. A. Branch, and A. Grace, *Optimization Toolbox for Use with MATLAB: User's Guide, Version 2*, Math Works, 1999.

- [36] J. D. Isom and R. E. LaBarre, "Process fault detection, isolation, and reconstruction by principal component pursuit," in *Proceedings of the American Control Conference (ACC '11)*, pp. 238–243, IEEE, San Francisco, Calif, USA, July 2011.
- [37] J. L. Guiñón, E. Ortega, J. García-Antón, and V. Pérez-Herranz, "Moving average and Savitzki-golay smoothing filters using mathcad," in *Proceedings of the International Conference on Engineering Education (ICEE '07)*, Coimbra, Portugal, September 2007.
- [38] R. W. Schafer, "What is a savitzky-golay filter? [lecture notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111–117, 2011.
- [39] J. Bring, "How to standardize regression coefficients," *The American Statistician*, vol. 48, no. 3, pp. 209–213, 1994.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

