Development of data reconciliation system for billing process at PT Perusahaan Gas Negara

R Hadianti¹, A Ihsan², A Mujiyanto³, S. Uttunggadewa⁴

¹Industrial and Financial Mathematics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha no. 10 Bandung 40132, Indonesia
²PT Scada Prima Cipta, Bandung, Indonesia
³Energy Management Division, PT. Perusahaan Gas Negara, Jakarta, Indonesia
⁴Combinatorial Mathematics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha no. 10 Bandung 40132, Indonesia

Corresponding author: hadianti@math.itb.ac.id

Abstract. We study a problem faced by PT. Perusahaan Gas Negara (abbreviated as PGN), the Indonesian company in the natural gas transportation and distribution segment, in developing data reconciliation system for billing process. PGN serves its residential, commercial and industrial customers' natural gas need through transmission pipelines. We focus our study on the billing process for industrial customers, where gas consumptions are measured by combining measurement results from turbine meter, temperature measurement device and pressure measurement device. Normally, customers’ bills can be obtained directly from volume metered by volume measurement device. But in every month there are a number of anomaly behaviour in customers’ gas consumption data or reports on meter failure, although statistically there are less than 1% of meter failure or anomaly behaviour from all customers’ data. Due to the large number of on-site meters which located in very wide area in Sumatera and Java islands, sometimes defective meter problems are already resolved after several days since the problem was reported/indicated, and this can generate loss for PGN. In this paper, we perform analysis to answer: 1. how to determine the beginning of the period where a volume meter is fail, 2. how to estimate gas volume delivered to the customers during the meter failure. The first problem is solved by analysing historical volume, temperature, and pressure data of each customer. The analysis results, in line with Spitzglass (Energy) Equation, that gas flow is linear to the pressure drop between the pressure at distribution pipe and the pressure at the customer’s pipe. We derive a procedure for detecting anomaly in gas volume data based on this phenomenon, where we reconcile the volume data and the pressure data and categorize the reconciled data into normal type or anomaly type data. The period during the meter failure is then determined as the period where the data in this period are of anomaly type. As soon as the period during the meter failure is determined, we then need to estimate the gas volume delivered during this period. We apply some forecasting methods for time series data such as AR(p) and Holt Winters models, and we choose the estimation from the forecasting model with the smallest error. The solution of those problems will be used for calculating bills for those customers which experience meter defective in the billing month. We develop a graphical user interface based on the anomaly detection method and the forecasting method. This system has
been implemented, tested, and used by Energy Management Division of PGN, and will be developed into an energy management system.

1. Introduction

PT. Perusahaan Gas Negara (abbreviated as PGN) is the largest national company in the natural gas transportation and distribution segment. This company has pipeline networks in North Sumatera, Southern part of Sumatera, Riau Islands, and Java Sea to connect the natural gas sources location with the natural gas end user location through the transmission pipelines mode. PGN plays a significant role in meeting domestic natural gas needs, serves natural gas to 196,221 residential, commercial and industrial customers in Indonesia’s main population centers (in the end of 2017). For the past 50 years, PGN has strengthened its energy business by investing in upstream oil and gas, mid-stream transmission and downstream distribution. Today, PGN has transformed from a natural gas transmission and distribution company into a provider of integrated energy solutions. By integrating the assets and expertise bound within its subsidiaries, PGN is unlocking new synergies across its enterprise, enabling it to develop effective solutions for customers’ specific energy needs.

In the competitive market situation, PGN needs to focus on value-added services and service options that can differentiate PGN from competitors. The customer measurement system is now fully digitized, ensuring greater efficiency and transparency. Customers can also access their accounts, support services, and other products through PGN mobile applications [1]. All innovations are strived in order to enhance current PGN level of service.

An effort in ensuring greater efficiency, accuracy and transparency is performed by developing data reconciliation system for billing process. Industrial customers’ gas consumptions are measured by combining measurement results from turbine meter, temperature measurement device and pressure measurement device. All measurement results are converted to base condition with Electronic Volume Converter. Normally, customers’ bills can be obtained directly from volume metered by volume measurement device. But this device, also the other two devices as well, can be defect although happen rarely. Statistically, in every month there are only less than 1% anomaly behaviour in customers’ gas consumption data or reports on meter failure from all customers’ data. But if it happens to the volume meters of industrial customers which consume large volume of gas, the unmetered consumption during the defective meter problem unresolved may lead to a massive revenue loss for PGN. Due to the large number of on-site meters which located in very wide area in Sumatera and Java islands, sometimes defective meter problems are already resolved after several days since the problem was reported/indicated. No one knew when the meter start to failure so that it is difficult to estimate gas volume delivered to the customers during the meter failure. Briefly, there are two data validation problems to be solved due to meter failure: 1. how to determine the beginning of the period of meter failure, 2. how to estimate gas volume delivered to the customers during the meter failure. These problems are common problems in pipeline network, so that we may find many approaches to solve the problems [2][3][4].

The solutions of the two problems will be used for calculating bills for those customers which experience meter defective in the billing month. Without any scientific methods for solving the problems, PGN is always difficult to convince customers the gas volume receipt by customers. A collaborative research[1] among PGN, PT. Scada Prima Cipta and Research Consortium Optimization on Pipeline Network (OPPINET) P2MS ITB is then performed to solve the problems, and to develop an application based on these data reconciliation.

The first problem is solved by analysing historical volume, temperature, and pressure data of each customer. The analysis results, in line with the Spitzglass (Energy) Equation, that gas flow is linear to the pressure drop between the pressure at distribution pipe and the pressure at the customer’s pipe. We derive a procedure for detecting anomaly in gas volume data based on this phenomenon, where we reconcile the volume data and the pressure data and categorize the reconciled data into normal type or

1 Among PGN, PT. Scada Prima Cipta and Research Consortium Optimization on Pipeline Network (OPPINET) P2MS ITB
anomaly type data. The period during the meter failure is then determined as the period where the data in this period are of anomaly type. We perform some validations to historical data, and we can say that our procedure work effectively in determining the period of anomaly data which indicate the period during the meter failure. As soon as the period during the meter failure is determined, we then need to estimate the gas volume delivered during this period. The estimation is performed by first analysing the trend, seasonal and cyclical behaviour of the customers’ historical consumption data. Most of the data show strong seasonal behaviour so that we try to estimate the gas volume by applying Exponential Smoothing Forecasting Methods such as Holt Winters model [5]. Estimation on the data without non-seasonal behaviour is performed by applying Auto Regressive Forecasting Model \((g_18 \in 8)\) for various values of \(p\). We estimate the gas volume based on the forecasting model with the smallest forecast error.

For practical purpose, we develop a graphical user interface based on the anomaly detection method and the forecasting method described above. This system has been implemented, tested, and used by Energy Management Division of PGN. Up to now, this system has been helping PGN in monthly billing process. In the future, this system will be developed into an energy management system. PGN needs to know consumption behaviour of each PGN customers’ sector such as manufacture sector, energy sector, chemical sector, etc. By knowing the consumption behaviour and the economic development prediction for each sector, PGN can decide the optimal plan for future natural gas supply for each sector. Moreover, by the system PGN can offer dynamic pricing to the customers. Dynamic pricing is reported can reduce unutilized supply and can give an enormous potential cost saving both for PGN and the customers [6][7][8].

This paper is written as follows. After this introduction, in section 2 we discuss the historical data and its analysis. The result of the data analysis lead us to a derivation of anomaly detection procedure and the forecast method we have to choose for forecasting gas consumption. In section 3 we then discuss how it results when we implement the anomaly detection procedure and the forecast method in real case billing process. This paper is ended by a summary and discussions that we present in section 4.

2. Data analysis, anomaly detection, and forecasting

The PGN data reconciliation system is constructed by first performing three mathematical steps, i.e. data analysis, anomaly detection and estimation/forecasting. This section gives brief explanations for those three steps.

2.1. Data analysis

We analysed the trend, seasonal and cyclical behaviour of the industrial customers’ historical consumption data. Most of the data have strong cyclical or intermittent behaviour as shown in the following example.

Figure 1 and 2, in the following, show an example of hourly consumption volume and pressure of one customer of PGN. In longer period of time the data show that this customer consumption has a cyclical or intermittent behaviour with weekly period, as can be seen on figure 1 from the 968th data to 1135th data. We can say that every cycle is started by a period of time with zero volume and followed by a period of time with non-zero volume. For brevity, we call the period of time with zero volume by Zero Volume Period, and the period of time with non-zero volume by Non-Zero Volume Period.

Due to a customer’s production system shut down, the gas flow was blocked as can be seen on figure 1 from the 285th data to 529th data. If we refer to the pressure data, we can see that in the same period the pressure are as high as the pressure in other periods of time with zero volume. This bring us to a conclusion that there is a strong correlation between the behaviour of volume data and the pressure data. This conclusion scientifically can be explained by the Spitzglass energy equation, which described in the following.
For gas flowing in a pipeline between two points (A and B), there are a number of energy equations which relate the properties of the gas to the flow rate, pipe diameter and length, the temperature and the pressure in the pipe. One of the energy equation is the Spitzglass equation, i.e.
\[ Q = c_1 c_2 E \left( \frac{T_f}{P_B} \right) \left( \frac{P_A - P_B}{G T_j / L_s Z (1 + \frac{T_f}{\theta} c_s D)} \right)^{0.5} D^{2.5}, \]

where

- \( Q \): is the gas flow rate
- \( T_f \): is the average gas flowing temperature
- \( E \): is the pipeline efficiency,
- \( L_s \): is the pipe segment length between point A and point B
- \( T_b \): is the base temperature,
- \( Z \): is the gas compressibility factor at the flowing temperature,
- \( P_b \): is the base pressure,
- \( D \): is the pipeline inside diameter,
- \( P_A \): is the gas pressure at point A,
- \( c_1, c_2, c_3, c_4 \): are some constants (see (Error! Reference source not found.).)
- \( G \): is the gas gravity,

In the condition where \( Q \) goes to zero, from the Spitzglass equation we will have \( P_A - P_B \) goes to zero since the remaining parameter in the right-hand side of the equation are constants. If A and B are the starting point and the end point of the segment from the distribution pipeline to the customer’s location, respectively, then when there is no flow to the customer the pressure at a customer’s meter is almost the same as the pressure at the distribution pipe. So, normally during Zero Volume Period the volume is around zero and the pressure is high (around the pressure at the distribution pipeline). During Non-Zero Volume Period, if the volume is far from zero then the pressure is also far from the high-pressure band.

The problem we study in this paper is motivated by our finding on the other customers’ data where we found certain long periods of time, even longer than one cycle, where the hourly volumes are zero but the pressure is far from high-pressure band. This phenomenon breaks the Spitzglass equation, and we have to categorize it as an anomaly. This anomaly behaviour indicates volume meter failure in anomaly period. Our discussion on the anomaly behaviour can be summarized in the following table.

### Table 1. Normal and anomaly conditions

| Data behavior | Volume and pressure conditions |
|---------------|--------------------------------|
| Normal        | 1. Volumes tends to zero & pressures are in the high-pressure band, or |
|               | 2. Volumes are far from zero & pressures are far from the high-pressure band. |
| Anomaly       | Volumes tends to zero & pressures are far from the high-pressure band |

#### 2.2. Anomaly detection

The anomaly detection procedure we derive in this sub section is based on Table 1, and here we articulate it in mathematical model.

Let \( T \) be a set of time where historical data are observed, \( P \) be a set of gas pressure data in \( T \), and \( V \) be a set of gas volume data in \( T \). For each \( t \in T \), let \( v_t \in V \) be the volume at time \( t \) and \( p_t \in P \) be the pressure at time \( t \). We model the statement ‘volume tends to zero’ as \( v_t \leq 1 \) and the statement
‘pressures are in the high-pressure band’ as \( p_t \geq \rho \), where \( \rho \) is the 95th percentile of \( P \). Then, we have mathematical expression of Table 1 in the following.

**Table 2. Normal and anomaly conditions in mathematical expressions**

| Data behavior | Volume and pressure conditions |
|---------------|--------------------------------|
| Normal        | 1. \( v_t \leq 1 \) and \( p_t \geq \rho \), or |
|               | 2. \( v_t > 1 \) and \( p_t < \rho \) |
| Anomaly       | \( v_t \leq 1 \) & \( p_t < \rho \) |

The range of data in \( P \) depends on the diameter of customer’s pipe, so that we set the value of \( \rho \) as 95th percentile of \( P \) in order to accommodate the variant of \( P \) among customers.

### 2.3. Forecasting and estimation

If the anomaly period has been determined by the anomaly detection procedure, we then need to determine a number of forecasting methods that can capture the behaviour of customers’ gas consumption. We mean by the behaviour of customers’ gas consumption is the behaviour in a certain period of time close to the anomaly period, which we call the training period. If the training period is close to the anomaly period, we can assume that the behaviour of customers’ gas consumption in both period are the same. Mathematically, the behaviour are characterized by the trend, seasonal and cyclical behavior of the time-series data in the training period. For the data that has strong seasonal behavior we can choose Exponential Smoothing Forecasting Methods such as Holt Winters model [5]. For the data without non-seasonal behavior is performed by applying Auto Regressive Forecasting Model \((\rho)\) for various values of \( \rho \). Since we do not know the behavior of customers’ gas consumption data beforehand, we perform a number of forecasting with various forecast methods to accommodate any behaviour.

In order to measure forecast accuracy, we tried some forecast accuracy measures such as Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Relative Geometric Root Mean Square Error (RGRMSE), Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE). But after analyze the errors between the training data and the fitting data resulted by any forecasting method we use, we think that most suitable forecast accuracy measure for our need to estimate the total consumption value in the anomaly period is Percentage Cumulated Forecast Error (PCFE), which we create based on the Cumulated Forecast Error (CFE).

If we note the training data as \( X_1, X_2, \ldots, X_T \) and its forecast data as \( \hat{X}_1, \hat{X}_2, \ldots, \hat{X}_T \), the CFT is given by

\[
CFE = \sum_{i=1}^{T} (X_i - \hat{X}_i),
\]

and our PCFT is given by

\[
PCFE = \frac{\sum_{i=1}^{T} (X_i - \hat{X}_i)}{\sum_{i=1}^{T} X_i}.
\]

We come up to consider the CFE after we realize that MSE and MAD values are quite difficult to be understood by the PGN customers. It is though that the customers will easy to understand a measure in term of percentage, so that we then consider MAPE as an alternative for the measure. But the MAPE value depends on the training data scale [10][11], and furthermore the existence of zero
volume data make the really high value of MAPE even for the case where the forecast results is very close to the training data. Since all we need is to estimate the total consumption value in the anomaly period, we then consider the CFE. The ‘percentage’ version of CFE is given by the PCFE. We have test that there is a strong linear correlation between MSE and PCFE values, and between MAPE and PCFE as well.

We estimate the volume during the anomaly period from the forecast result with the smallest PCFE.

3. The implementation of the anomaly detection procedure and the forecast method in real case billing process

We tested the anomaly detection procedure and the forecast method in real case billing process to a large number of historical PGN customers’ data. The test showed that:

1. The anomaly detection procedure work effectively in determining the period of anomaly data which indicate the period during the meter failure. We test the procedure to some customers’ historical data in which these customers reported meter failure in certain periods of time. By implementing our procedure we obtained periods of meter failure that cover such a period reported by the customer.

![Figure 3. Anomaly data, indicated by red-cross signs.](image)

2. For cases in which the customers’ historical data have strong seasonal behavior, we obtain the best forecast model that give quite small error (under 20%, for some cases even less than 5%). Our experience show that (p) model oftenly occurs as the best model compare to the other models (Holt Winters model and the general Exponential Smoothing Models).
Figure 4. Example of forecast result, given by the green curve.

For practical purpose, we develop a web-based graphical user interface based on the anomaly detection method and the forecasting method described above. The following figure is the screen shos of the interface which show the results on the anomaly detection and forecasting. In this example, we need to forecast gas consumption during period 4 am on 1 October 2017 to 3 am on 4 October 2017 (92 hours long). We use three month historical data as training data. In this example, the best fitting model is the Holt-Winter Multiplicative with Damped model that give us 18% error in training data period. The interface shows the historical data and its fitting data, and the forecasting result shown in green colour while the hourly quantities are given on the right bottom box.
We also found other case where it is impossible to pick training data from historical data before the anomaly period. This case may occur to a new customer of PGN where there is no enough historical data and the anomaly period happen in the beginning of service. In this case, we implement backward estimation after fitting data on the period after the anomaly period. Our procedure still work effectively as shown in the following figure.

Figure 5. Graphical User Interface showing farecast result and its error.
4. Summary and discussion
A collaborative research has been performed for solving the estimation for total natural gas consumption due to volume meter failure, which is needed for billing process. By analysing historical data on gas consumption, temperature and pressure, the problem is then formulated as anomaly detection and estimation in the anomaly period. We derive a procedure for anomaly detection, and choose some appropriate forecasting methods for estimation of gas volume in the anomaly period. After some test or validation for those procedures, we developed a graphical user interface for practical purposes.

The system we derive has been implementing in monthly billing process. Although it still needs minor refinements, the end users of this new system have seen some possibilities to develop it to new energy management system. The energy management system not only content the anomaly detection procedure and the forecast procedure for billing process, but is expected to be able to analyse overall customers’ consumption behaviour. Since PGN supply natural gas to the pipeline network in bulk volume, we guess there are some period of time where the gas supply is excessive. For developing the energy management system, we first need to perform analysis for determining the periods where the gas supply are excessive, and making an inference whether such periods are periodic or not. If the analysis results show that the excessive supply periods are periodic, we will try to reduce unutilized gas supply by offering dynamic pricing to the customers that their production can be shifted to the excessive supply periods or to the customers that can increase their productivity. The energy management system will monitor the impact of the dynamic pricing implementation to the unutilized supply reduction, so that can give an enormous cost saving both for PGN and the customers as reported in some scientific articles.

Acknowledgement
We are grateful to our research assistant Muhammad Ridwan Reza Nugraha and Nita Ratnawati who have implemented the anomaly detection procedure and the forecast method in R. We also thank to Luthfi Pratama for developing the graphical user interface.

References
[1] PT PERUSAHAAN GAS NEGARA (Persero) Tbk 2017 Annual Report 2017
[2] Akpınar S and Yumusak N (F. Martinez-Alvarez, Ed.,) 2016 ) Energies 727
[3] Bagajewicz J and Cabrera E 2008 Data Reconciliation in Gas Pipeline Systems Industrial & Engineering Chemistry Research 42 pp. 5596-5606
[4] Gschneidner Jr K A and Erwin L 2015 Gas Reconciliation Requirements and Procedures Gas Industry Company Limited
[5] De Gooijer G and Hyndman R J 2006 25 Years of Time Series Forecasting International Journal of Forecasting p. 443-473
[6] Coşgun Ö and Kaya G O 2015 Stochastic Dynamic Programming for natural gas pricing in the Turkish Energy Market J. of Mult.-Valued Logic & Soft Computing 26 pp. 353–361
[7] Flaim T 2018 Fundamentals of Dynamic Pricing (Presentation file https://emp.lbl.gov/sites/default/files/oh120809.pdf ) dowloadeed on 16 July 2018
[8] Harvard’s Electricity Policy Group 2018 Dynamic Pricing (presentation file: https://sites.hks.harvard.edu/hepg/Papers/2016/June%202016/P2.%20Kelly..pdf, ) dowloadeed on 16 July 2018
[9] Carroll L and Hudkins R W 2018 Advance Pipeline Design The University Oklahoma Report, www.ou.edu/class/che-design/a.../Pipeline%20Design.pdf dowloadeed on 16 July 2018
[10] Makridakis S, Wheelwright S C and Hyndman R J 1998 Forecasting,Methods and application 3rd edition
[11] Wallström P 2009 *Evaluation of Forecasting Techniques and Forecast Errors with Focus on Intermittent Demand* Licentiate Thesis, Department of Business Administration and Social Sciences Division of Industrial Marketing, e-commerce and Logistics Industrial