Flow Measurement and Instrumentation 48 (2016) 29–35

Contents lists available at ScienceDirect

Flow Measurement and Instrumentation

journal homepage: www.elsevier.com/locate/flowmeasinst

Comparison of different approaches for detection and treatment of outliers in meter proving factors determination

Elcio Cruz de Oliveira [a,b,*], Alcir de Faro Orlando [b], Anderson Luiz dos Santos Ferreira [b], Carlos Eduardo de Oliveira Chaves [c]

[a] Petrobras Transporte S.A., Project Management, 20091 060 Rio de Janeiro, RJ, Brazil
[b] Postgraduate Programme in Metrology: Metrology for Quality and Innovation, Pontifical Catholic University of Rio de Janeiro, 22453 900 RJ, Brazil
[c] Petróleo Brasileiro S.A., Supply, Measurement Control of Custody Transfer and Audit Losses, 20231 030 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form
21 December 2015
Accepted 4 February 2016
Available online 6 February 2016

Keywords:
Meter proving factor determination
Detection and treatment of outliers
Nonparametric data
Turbine meter

ABSTRACT

A meter proving factor can be considered as a calibration parameter, by expressing the ratio the reference volume and the gross volume of liquid passed through a meter. The international guideline recommends Dixon's test for outliers to a meter proving factor set. However, the literature says that this statistic test is restricted only to data with Gaussian behavior, besides of not to be able to detect and treat two outliers at the same time. Here, Gaussian behavior of the meter proving factor set is evaluated, then different parametric and nonparametric approaches for detection and treating outliers applied to turbine meter proving factors for custody transfer of liquefied petroleum gas are compared. Afterwards, this effect is evaluated in relation to the number of outliers and how this handling affects the variable range criteria for expanded uncertainty in average meter proving factor. The results show that different average meter factors can be reached for each nonparametric and parametric test; anyway, no statistically significant effect between them is noticed.

1. Introduction

Although not a renewable fuel, liquefied petroleum gas (LPG) is an alternative fuel, safe, with few atmospheric emissions, low-cost and considerable social benefits [1]. Recent studies in many global markets indicate that LPG can be an excellent alternative fuel for the road transport sector [2].

Despite of the growing use of ultrasonic meters, turbine flow meters still are widely used to measure this valuable fuel [3].

Pipeline transportation companies consider the meter calibration as one of the most important parameter in order to guarantee reliability in commercial transactions. Meter calibrations can be more relevant when the device meter is mechanical as a turbine meter that is more sensitive to friction and wear [4]. In this scenario, the term "proof" represents tests in meters. A volume meter is considered as proved, when a materialized measure of volume, called as prover, is compared to the totalized volume indication of the meter. The totalized volume by the meter and the prover are submitted to several calculations, using correction factors to convert volumes to reference conditions, establishing a meter factor (MF). The meter factor is a dimensionless number obtained by ratio of the volume of liquid passing through the meter and the volume of the prover, both at reference conditions for a particular flow rate [5].

Meter proving factors or simply meter factors (MF) are usually monitored to detect and track down trends or sudden shifts as indications when carrying out maintenance and calibration of the meter or of the auxiliary measuring equipment.

The turbine flow meter response depends mainly on changes in flow rate, mechanical condition of the meter, physicochemical properties of the fluid, contaminants and flow impurities. These parameters can change the pulse numbers for each volumetric unit of liquid passing through the turbine flow meter, i.e., the meter factor [6].

Well defined acceptance criteria are usually used to evaluate the meter factor. Custody transfer players in petroleum industry reach a consensus in relation to a minimum number of proving runs that agree within a maximum range between high and low meter factors, to a meter proving interval and deviation limit between consecutive meter factors [7]. This latter parameter is very useful to shed lights in the reliability of the complete metering system, meter and proving systems. Fixed limits are based on the operator experience, however statistics methods can be used to decide if the variability of the meter proving is suitable or not.

In order to improve this essential control, a statistical analysis

* Corresponding author at: Petrobras Transporte S.A., Project Management, 20091 060 Rio de Janeiro, RJ, Brazil.
E-mail address: elciooliveira@petrobras.com.br (E.C. de Oliveira).

http://dx.doi.org/10.1016/j.flowmeasinst.2016.02.002
0955-5986/

of a single set of meter proving factor is required. A statistically based uncertainty criterion is recommended to determine the acceptability of a set of meter proving factor. The most representative guidelines to describe these tests, API MPMS 13.1 [8] and API MPMS 13.2 [7] calculate an average or arithmetic mean and indicating, exclusively, Dixon's test for outliers. However, it is worth mentioning that this approach is specific when data are parametric, i.e. they have Gaussian distribution [9].

The guideline API MPMS 13.2, Appendix B, recommends Dixon's test for outlier and makes the following consideration: "Alternate outlier tests are not expected to duplicate the exact results provided by these procedures; however, alternate computational methods should achieve the same purpose intended by the outlier tests in this appendix [7]".

Contradicting this guideline, this paper suggests, first, that the Gaussian data behavior be evaluated based on Shapiro–Wilk test [10]. Then, beyond this proposed test, other parametric and nonparametric approaches are discussed and compared.

The aim of this paper is to evaluate the whether or not the meter proving factor set follows a Gaussian or normal distribution, and then compares different parametric and nonparametric approaches for detection and treating outliers applied to turbine meter proving factors for custody transfer of liquefied petroleum gas.

2. Methodology

The methodology is divided into four parties. The two first parts are revisions of the Shapiro–Wilk procedure and parametric tests for detection and treating of outliers; the third part is an introduction to the nonparametric tests for detection and treating of outliers and the last one is the variable range criteria for expanded uncertainty in average meter factor.

Revision of the nonparametric statistical analysis

2.1. Shapiro–Wilk test [10]

To verify if one data set can be treated as Gaussian or normal distribution, this paper uses Shapiro–Wilk test.

In this test, the data number (n) is a limitation, $3 < n \leq 50$.

The test procedure is:

- Primarily, set of n data X_i ($i=1, 2, \ldots, n$) is arranged in ascending order;
- The subtraction are calculated: $(X_{(n+1)-i} - X_i)$;
- The index i varies from 1 to $n/2$ or from 1 to $(n+1)/2$, according to n being even or odd, respectively;
- The multiplications are calculated: $a_i(X_{(n+1)-i} - X_i)$;
- The coefficients a_i are tabulated;
- The sum is calculated: $SW = \sum a_i(X_{(n+1)-i} - X_i)$;
- The sum of squared is calculated: $SQT = \sum (X_i - \bar{X})^2$, or $(n-1)S^2$, considering $\bar{X}$ as the arithmetic mean of the set data;
- The ratio is calculated: $W_{calculated} = \frac{SW^2}{SQT}$;
- Compare the calculated value $W_{calculated}$ to the value $W_{critical}$. If $W_{calculated} > W_{critical}$, the data set can be treated as Gaussian or normal distribution.

2.2. Parametric tests for outliers

Scientific studies derived from petroleum industry flow measurements have inserted outlier tests in their approaches, always considering the data behavior as Gaussian, without testing them [11–15].

2.2.1. Dixon's test [16]

Dixon's Q test, or simply the Q test, is one way to assess if suspected data belong to a population. Dixon Q value is defined as the ratio of the difference between the suspect value and the closest to this value and the difference between the largest and the smallest value of the set.

Q calculated value by Eq. (1), Eq. (2) or Eq. (3), depends on the sample size, is compared to the Q critical value for a desired level of confidence. If it is not greater than the critical value, the suspect value is kept, otherwise it is rejected.

Considering a set of n data x_i ($i=1, 2, \ldots, n$) arranged in ascending order. The statistical test, for $3 \leq n \leq 7$, Eq. (1) (depending if x_1 or x_n is the suspect value):

$$Q_{3\leftrightarrow 7} = \frac{x_2 - x_1}{x_n - x_1} \quad \text{or} \quad \frac{x_n - x_{n-1}}{x_n - x_1} \tag{1}$$

For $8 \leq n \leq 12$, Eq. (2):

$$Q_{8\leftrightarrow 12} = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad \text{or} \quad \frac{x_n - x_{n-1}}{x_n - x_2} \tag{2}$$

For $13 < n \leq 40$, Eq. (3):

$$Q_{13\leftrightarrow 40} = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad \text{or} \quad \frac{x_n - x_{n-2}}{x_n - x_3} \tag{3}$$

One limitation of this approach is when it has two suspect results in high or low part of the sample data and when there are two suspect results one at each end of the data set [17].

Another drawback of Dixon's test is the fact that increasing the number n of measures, it also increases the probability of occurrence of large gaps in the set of measures. For example, to two thousand measures, the probability of detecting a deviation greater than 3.29 is large and there is no sense to discard the measure once the probability of a value from standard normal distribution greater than 3.29 is about 0.0005, i.e., $P(X > 3.29) \approx 0.0005$, where $X \sim N(0, 1)$. Chauvenet's criterion eliminates this problem.

2.2.2. Chauvenet's criterion [18]

Based on this criterion, one measure must be rejected if $|d_j| = |(y_j - \bar{y})| > d_{ch}$, where d_{ch} is the Chauvenet's limit for rejection, defined by: $p_o = \int_{-\infty}^{-d_{ch}} G(\eta)d\eta + \int_{+d_{ch}}^{+\infty} G(\eta)d\eta = \int_{-d_{ch}}^{+d_{ch}} G(\eta)d\eta = \frac{1}{2n}$, where G($\eta$) is the Gaussian function. In other words, a measure may be excluded if the probability of obtaining the specific deviation from the mean is less than $1/(2n)$.

This criterion establishes that a measure x_i must be discarded if the r calculated value by Eq. (4) is larger than the critical value for those degrees of freedom, considering $\bar{X}$ as the arithmetic mean and $s(X)$ as the standard deviation:

$$r = \frac{|x_i - \bar{X}|}{s(X)} \tag{4}$$

2.2.3. Grubbs' test [19]

Grubbs' test is firstly performed to verify the existence of a dispersed value in each extremity of the data set. If this first analysis, one of the two values is considered to be scattered, it is refused, withdrawn from data set and new test, checking for the existence of a dispersed value in each extremity of the data set is carried out and so on. Otherwise, if this first analysis, both values are accepted as not dispersed, the test is finished and the remaining data set is used for analysis. If in the second analysis, the two results of extremity are considered as dispersed, they must be discarded, removed from the data set and new test is carried out, verifying the existence of two outliers in each extremity of the data set, and so on, until both values are accepted as not dispersed.