

# Machine Learning Applied to Ultrasonic Flow Meters for measuring Dilute, Turbulent Water-Bentonite Suspension Flow

Thiam Wan<sup>1</sup>, Hon Chung Lau<sup>1</sup>, and Wai Lam Loh<sup>1</sup>

<sup>1</sup>National University of Singapore

September 11, 2020

## Abstract

An ultrasonic flow meter that is calibrated in single phase flow has inherent errors when applied to measure dilute water-bentonite mixture flow. This paper endeavors to use artificial intelligence for recalibration of an ultrasonic flow meter. A commercial ultrasonic transit time flow meter was tested for measuring dilute water-bentonite mixture flow of 0.1-1.0 vol% concentration at room temperature. Results show the test data had a systematic error of -8.3% and a random error of 20.3%. The machine learning LLS regression, 2D interpolation and Gaussian Naïve Bayes methods were considered in this exercise. Finally, a combined 2D interpolation method and Gaussian Naïve Bayes classifier approach was preferred. It reduced the systematic error to -0.6% and random errors to  $\pm 13.7\%$ . Our study shows a high accuracy ultrasonic flow meter with systematic errors smaller than 1% for oil and gas multiphase application is possible with the aid of artificial intelligence technology.

## 1.0 Introduction

Artificial Intelligence (AI) will reshape the future the flow metering industry. The connectivity and flow of information between flow measuring devices and sensors provide an abundance of available data. The main goal behind the artificial intelligence research is the use of technology and data to improve the flow meter accuracy and efficiency. In recent years, artificial intelligence has been applied to recalibrate utility ultrasonic flow meters (Yazdanshenashad et al. 2018). This paper investigated the use of machine learning to address the accuracy problem of ultrasonic flow meter in multiphase measurement. This includes the use of available data and extracting only useful information for the purpose of reducing costs and optimizing capacity.

According to a report by Berrebi et al. (2004), the maximum error of a typical ultrasonic flow meter is 2% or 3% in turbulent flow rate ( $Re > 4000$ ), and 5% at laminar ( $Re < 2000$ ) or transient flow ( $20000 < Re < 4000$ ). However, this kind of accuracy is only valid in single phase flow assuming the flow is homogeneous. In the case of water-bentonite suspension flow, the mixture contains particles distributed randomly throughout the fluid. Due to the nature of the processes, the particle distribution in the fluid may randomly vary in time and space. As the ultrasound interacts with such particles, some microscopic phenomena may take place between the waves, particles and the flow (Eren 1998). For example, a water-bentonite suspension up to 1 vol% consists of two or more substances of very different acoustic impedance that alter the acoustic signal and the way ultrasonic beam is transmitted through and reflected from multiphase solid-solid, solid-fluid and fluid-fluid interfaces (Sirmurda et al. 2016). Therefore, the accuracy of ultrasonic flow meter in measuring particulate flow such as water-bentonite mixture flow is still a question that needs to be addressed.

In practice, transit time ultrasonic flow meters are sensitive to many factors. They are sensitive the variation in velocity profile and the installation effects. For example, all ultrasonic flow meters assume an ideal flow profile before the disturbing geometry. However, the distance between the flow meter and elbows in reality is typically not sufficiently long to redevelop such an ideal profile. In a recent study (Weissenbrunner et al. 2016), CFD simulation is used to quantify the uncertainties of ultrasonic flow meter caused by variations of

the inflow profiles. The study revealed a bias of 1.5-4.5% as flow meter was installed at a distance smaller than 40 pipe diameters to the double elbow. Another experiment (Ma et al. 2012) showed that moving the ultrasonic transducer 0.2 to 0.6 mm axially from the correct position led to a velocity error as much as 4 - 10% in a transit time ultrasonic flow meter. Changes in fluid density, flow viscosity and flow patterns affected the dynamical characteristics in the ultrasonic flow meter measurement (Catak and Ergan 2019). Therefore, re-calibration is often necessary to reduce errors due to the possible change of the fluid density, viscosity and flow patterns in multiphase flow.

In this study, we have examined the accuracy of transit time ultrasonic flowmeters in measuring dilute water-bentonite suspension flow up to 1 vol% bentonite concentration. Error analysis has been conducted to evaluate the meter performance in terms of systematic errors (indicated by the mean relative error) and random errors (indicated by the standard deviation value of relative errors). Next, the study attempted to close the gap for both type of errors using multiple error reduction algorithms including LLS regression method, 2D interpolation method, and various Gaussian Naïve Bayes classifier algorithms. The goal is to reduce the systematic error to less than 1% and improve random errors as much as possible without compromising the resolution of data.

Typical results of the measurement errors are presented in section 4.2, and the results of various machine learning error reduction exercise will be presented in section 4.3. Finally, the best combination of machine learning models is trained to reduce both systematic errors and random errors as presented in section 5.0. This helps to re-calibrate the transit time ultrasonic flow meters and improve its accuracy for future use of ultrasonic flow meter measuring drilling fluid flow. Further evaluation will be continued in higher bentonite concentrations in the future. Another important issue in the accurate measurement of multiphase flow is that of temperature compensation. Using machine learning models for temperature compensation in an ultrasonic flow meter will be investigated in future.

## 2.0 Literature Survey

This review introduces the definitions used in error analysis, summarizes the past activities in flow meters re-calibration and error reduction exercise for ultrasonic flow meters.

### 2.1 Operating Principle of Transit Time Ultrasonic Flowmeter

The single path transit-time ultrasonic flowmeter utilizes two piezoelectric transducers, which are clamped on the outside of the closed pipes at a specified distance from each other, as depicted in Figure 1.

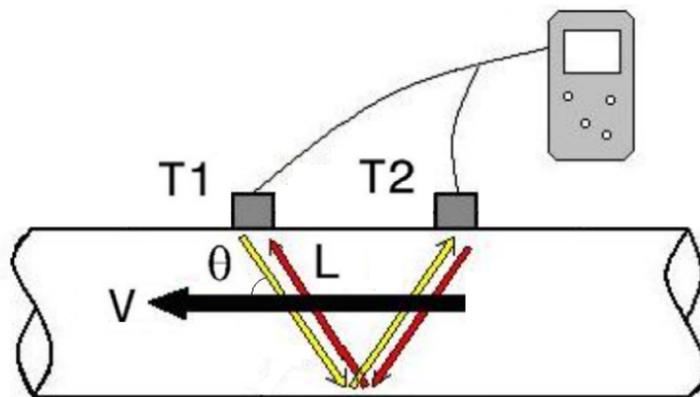


Figure 1: Transit-time ultrasonic flowmeter.

The flowmeter measures fluid velocity by transmitting acoustic signals between the two transducers alternatively, first in the opposing direction of fluid flow, and then in the direction of flow. The transit time of the ultrasonic signal from the downstream transducer to the upstream transducer ( $t_{21}$ ) and that in the opposite direction ( $t_{12}$ ) are computable using Eq. (1) and (2) as follows.

$$t_{21} = \frac{L}{(c+v\cos\theta)} \dots\dots (1) \quad t_{12} = \frac{L}{(c-v\cos\theta)} \dots\dots (2)$$

where  $c$  is the ultrasound speed in water,  $L$  is the path travelled by ultrasound, and  $\vartheta$  is the incident angle.

Ultrasonic meters are velocity meters by nature. The fluid flow velocity  $v$  is calculated from the differences between the transit times of the signals which are directly proportional to fluid velocity. The equation for the flow velocity is:

$$v = \frac{L}{2\cos\theta} \left( \frac{1}{t_{12}} - \frac{1}{t_{21}} \right) \dots\dots (3)$$

In a single-path flow meter, the ultrasonic beam path is diagonal, and the fluid flow velocity  $v$  is computed along this path. However, the velocity needed for computing volume flow rate is the average fluid velocity  $v$  across the pipe cross sectional area. Therefore, to convert velocity  $v$  to average fluid velocity  $v$ , an average velocity correction factor is used that is shown as  $k_c$  according to Eq. (4). This quantity is generally a function of the Reynolds' number.

$$v = k_c v \dots\dots (4)$$

Meter manufacturers have differing methodologies for computing velocity correction factor  $k_c$ . Some derive it by using propriety algorithms. This requires knowledge about the velocity profile patterns at different Reynolds' numbers. Nonetheless, the flow meter is still subjected to a certain degree of uncertainty. For example, fluid velocity profiles in the pipeline are not always uniform. Often there is swirl and asymmetrical flow profile within the meter (Lansing et al. 2003).

## 2.2 Error Analysis

All measurements have some degree of uncertainty that may come from a variety of sources. The process of evaluating the uncertainty associated with a measurement result is often called error analysis. In flow measurement terminology, the error is the difference between the true value of a measurement and the recorded value of a measurement, which can be classified into two broad categories.

A **systematic error** (or bias) refers to deviations that are not due to the changes in flow. It occurs with a measuring device that is faulty or improperly calibrated so that it consistently overestimates (or underestimates) the measurements by  $X$  units. Systematic errors cannot be reduced by taking more measurements. To reduce the systematic error of a data set, researchers must identify the source of the error and remove it. For example, the errors can be reduced by recalibrating the flow meter. In other circumstances, the error could be due to the inherent nature of the measuring technique. Hamouda et al (2016) reported a scenario where an ultrasound flow meter measured fluid flow rate based on transit time principle. This difference can be as low as a few picoseconds, which give rise to technical difficulties in measuring such a small time-difference with a given accuracy. This type of systematic errors can be reduced by using more sensitive sensors or avoiding conducting measurement in low flow range accuracy.

A **random error** is a deviation that randomly fluctuates over a mean value. It has no preferred direction. It occurs because there are a very large number of parameters beyond the control of the instrument that may interfere with the results of the measurement. For example, a random error occurs due to the instrument resolution (CDL 2020) and the way it is affected by changes in the surroundings (Kalla 2009). Nevertheless, the readings may be imprecise, but not inaccurate, as the averaging over large number of observations will yield a net effect of zero deviation.

According to definitions found in literature (CPL 2020), **accuracy** is defined as the closeness of agreement

between a measured value and a true or accepted value. Accuracy is often reported quantitatively by using relative error:

$$\text{Relative Error} = \frac{\text{Measured value} - \text{Expected value}}{\text{expected value}} \dots (5)$$

Meanwhile, **precision** is the degree of consistency among independent measurements of the same quantity; also described as the reliability or reproducibility of the result (CPL 2020). Precision is often reported quantitatively by using relative or fractional uncertainty:

$$\text{Relative Uncertainty} = \left| \frac{\text{Uncertainty}}{\text{measured quantity}} \right| \dots (6)$$

**Uncertainty** analyses are essential to determine whether measurement systems are capable of meeting performance targets. According to ISO-5168, the uncertainty of a flow measurement should be specified at a confidence level of 95%, which corresponds to two standard deviations.

$$v = v \pm 2\sigma_v \dots (7)$$

where  $v$  is mean velocity and  $\sigma_v$  is the standard deviation of flow data.

In summary, random error corresponds to imprecision (or repeatability), and bias to inaccuracy.

### 2.3 Error Reduction Algorithm for Self-Calibration of Ultrasonic Flow Meters

In recent years, the use of artificial intelligence in flow metering have attracted researchers' attention. For example, neural networks and support vector regression algorithms have been applied to the data from temporal and spatial ultrasonic level measurements of the drilling fluid in the open channel to estimate the flow rate (Chhantyal et al. 2017). The Least Square Error Reduction technique and neural networks method have been used for self-calibration of ultrasonic water flow meter (Yazdanshenashad et al. 2018; Catak and Ergan 2019). However, none of these self-calibration exercises involves the use of transit-time ultrasonic flow meter in multiphase flow such as water-bentonite mixture flow.

Catak and Ergan (2019) reported using the least square error method for the self-calibration of ultrasonic water flow meter. Three common least square errors calibration methods have been employed to the data obtained from DN-20 type ultrasonic flow meter, namely, Linear Least Squares (LLS), Weighted Least Squares (WLS), and Piecewise Linear Least Squares (PLR). The results presented found PLR gave the best results in all cases, while WLS was the best for higher flow rates. Both WLS and LLS were especially not adequate for low level of flowrate. For example, the flowmeter accuracy at 10 L/h (0.167 L/min) was around 5-8%, but after calibration, the improvement was only about 0.8-1.3%. (Catak and Ergan 2019).

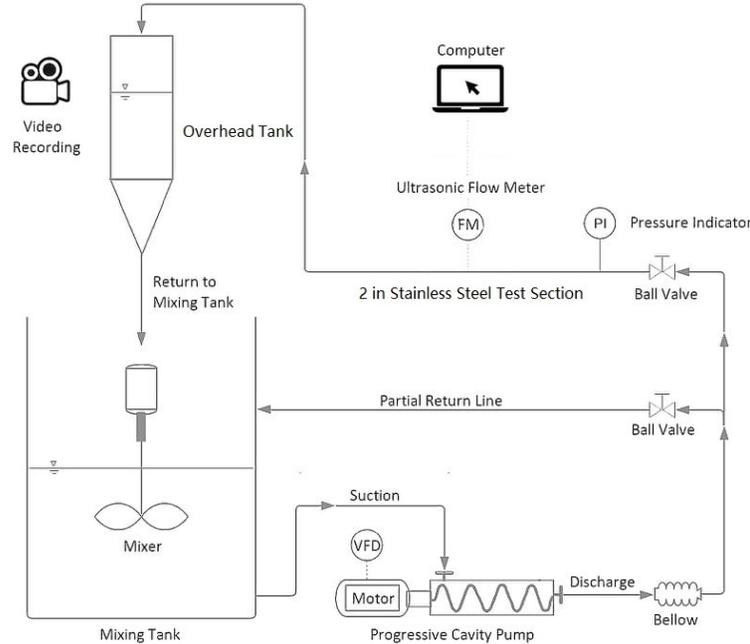
In a recent report, Yazdanshenashad et al. (2018) use the Multi-Layer Perceptron Neural Network (MLPNN) model to calibrate an ultrasonic flow meter to achieve an error smaller than 1.5%. The measured flow range was from 0.2 to 4 m<sup>3</sup> per hour. However, it was only aimed to reduce systematic errors. The authors did not report on the improvement in random errors which could be revealed by the change in standard deviation value of the errors. The authors have instead suggested reducing random errors by averaging large number of data. But the main drawback of this approach is the loss in data resolution. If it has to average 1000 data to significantly reduce the random error, it also means the resolution of data would be compromised by 1000 times.

## 3.0 The Present Work

### 3.1 Experimental Test Rig

A set of ultrasonic tests of a concentrated liquid-particle system in shear flow has been carried out using a transit-time ultrasonic flow meter. Ca-Bentonite particles (estimated nominal diameter of 60-70 μm, dried density  $\rho_f = 2500 \text{ kg/m}^3$ ) in water at shear flow of 0-1.5 m/s were studied, for volume fraction up to 1 vol% (or approximately 2.5 wt%). Tests were performed for particle volume fractions ranging from 0 %, in steps of 0.05 vol% until 1.0 vol%.

Figure 2 shows a schematic of the experimental unit used. The test loop is equipped with a mixing tank, a high-level tank, a mechanical mixer, a progressive cavity pump, ultrasonic flow meters, a bellow, valves and a pressure indicator. It consists of a 1 in (24.5 mm) diameter PVC test section and a 2 in (54.78 mm) diameter stainless steel (SS304) test section.



*Figure 2: Schematic of experimental setup.*

A mixing tank is used to generate a mixture of bentonite powder with known concentration in water. Dried calcium bentonite powder was dispersed in water with the assistance of a mechanical stirrer. The bentonite powder was added at a ratio according to Eq (15) in section 3.2. Before the experiments, the suspension was stirred for 15 minutes. The homogeneous fluid mixture was then re-circulated by a progressive cavity pump in a closed loop for ultrasonic flow measurement. The pump motor speed was regulated by a variable frequency drive (VFD) via a speed potentiometer. A bellow was installed at the pump discharge side to minimize the piping vibration which may affect the meter accuracy.

The single path transit-time ultrasonic flowmeter utilized two piezoelectric transducers, which were clamped on the outside of the closed pipes at a specified distance from each other. The flowmeter measured fluid velocity by transmitting acoustic signals between the two transducers alternatively, first in the opposing direction of fluid flow, and then in the direction of flow. The differences between the transit times of the signals were directly proportional to fluid velocity. The flow output meter signal was then fed to the PC after being converted to an appropriate digital signal. The mud fluid was slowly pumped into 2-in (54.78 mm ID) stainless steel test section, which was controlled by the manual operated ball valve.

Experimentally, the output of transit-time ultrasonic flowmeter alone was insufficient, and a joint measurement with another reference method was normally required. Therefore, the desired flow was drained to an overhead tank for volumetric measurement. Video was recorded and the video images were processed by an algorithm to acquire the time series flow rate for comparison with the transit-time ultrasonic flowmeter readings.

After completion of all measurements, The experimental data were recorded in room temperature (27~33°C) were analyzed to obtain relationship between signal patterns in the ultrasonic sensors and bentonite concentration, flow speed and Reynolds number. Each experiment was conducted at unsteady flowing condition with flow accelerating from rest to 1.2 ~1.5 m/s, and next, decelerated to rest again, within a duration of 25~30 seconds.

### 3.2 Preparation of Water Based Bentonite Fluid Mixture

Bentonite is naturally occurring clay. It is inorganic, non-toxic, and non-irritating. It is not considered hazardous on skin contact and is employed in cosmetics and skin products as a suspender. It is therefore very safe to use the bentonite clay in the experiment.

In this work, a water-based drilling mud fluid was prepared by using pure water as the base fluid and bentonite clay as a viscosifier. The bentonite clay has a density of 2500 kg/m<sup>3</sup> and the water density is 1000 kg/m<sup>3</sup>.

For mud velocity calculations, the following equations are used.

$$V_s + V_{m1} = V_{m2} \dots\dots (8) \rho_s V_s + \rho_{m1} V_{m1} = \rho_{m2} V_{m2} \dots\dots (9)$$

From equation (8):

$$V_{m1} = V_{m2} - V_s \dots\dots (10)$$

Substituting Eq (10) into Eq (9) yields:

$$V_s = V_{m2} \left[ \frac{\rho_{m2} - \rho_{m1}}{\rho_s - \rho_{m1}} \right] \dots\dots (11)$$

Multiplying Eq (11) by the density of solid yields the corresponding weight (of solids) required.

$$\rho_s V_s = \frac{\rho_s V_{m2} (\rho_{m2} - \rho_{m1})}{\rho_s - \rho_{m1}} \dots\dots (12)$$

By re-arranging Eq (11), the density of final mixture is given by:

$$\rho_{m2} = \rho_{m1} + (\rho_s - \rho_{m1}) \left( \frac{V_s}{V_{m2}} \right) \dots\dots (13)$$

Re-arranging Eq (9) and dividing by  $V_{m2}$ , the volume percentage of final mixture is given by:

$$\frac{V_s}{V_{m2}} = \left[ \frac{\rho_{m2} V_{m2} - \rho_{m1} V_{m1}}{\rho_s V_{m2}} \right] \dots\dots (14)$$

Bentonite swells when it is wet. This expansion is due to the adsorption of water. Hence the clay volume increases and the total volume (clay plus water) is, for practically purposes, unchanged (Othman 2007).

The particle concentration in weight percentage of final mixture is given by:

$$\frac{w_s}{w_{m2}} = \left( \frac{\rho_s}{\rho_{m2}} \right) \left( \frac{V_s}{V_{m2}} \right) \dots\dots (15)$$

According to Birgersson et al. (2009), Ca-bentonite forms a stable sediment with water at a solid mass ratio well below 10, which has sand-like properties. Due to low swelling capacity, Ca-bentonite is unable to form colloids.

## 4.0 Results and Discussions

### 4.1 Flow Rate Measurement in 2-in diameter Stainless Steel pipe

During the experiments, a homogeneous bentonite-water mixture was pumped through a 2-in pipe. Variable frequency drive (VFD) was installed to regulate the pump speed. This was to minimize the disturbance resulting from throttling the valve to improve the flow meter accuracy. Figure 3 shows typical experimental readings acquired from transit-time ultrasonic flow meter and its comparison to the rate of the flow as collected in the volume tank. Each experiment was conducted at unsteady flowing condition with flow

accelerating from rest to 1.2 ~1.5 m/s, and next, decelerated to rest again, within a duration of 20~25 seconds.

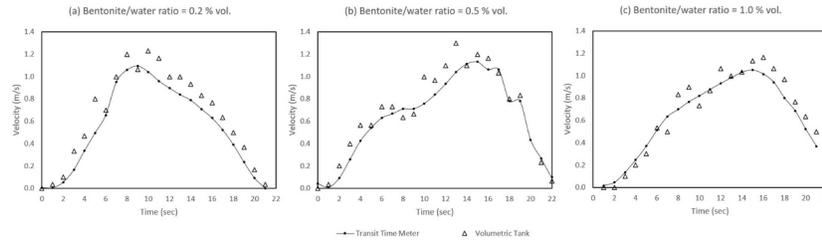


Figure 3: Typical time series results of transit-time flow measurement for various bentonite concentrations.

The validation was carried out by a comparison between the measurement made by the device under test and that by a reference method. In this study, the reference readings were acquired by the ‘static method’ based on collecting fluid in a volumetric tank and determining its quantity by a static measurement (measuring the change in fluid level in tank).

Figure 4 shows a scattered plot for the transit-time ultrasonic flow meter reading versus the actual flow measured by the volume tank. As expected, the scattered points show a positive linear pattern as they move from left to right. Transit-time ultrasonic flow meter exhibits good linearity, giving a predictable response for a known change in the flow rate. The Pearson coefficient R is 0.9529 indicating a strong positive correlation.

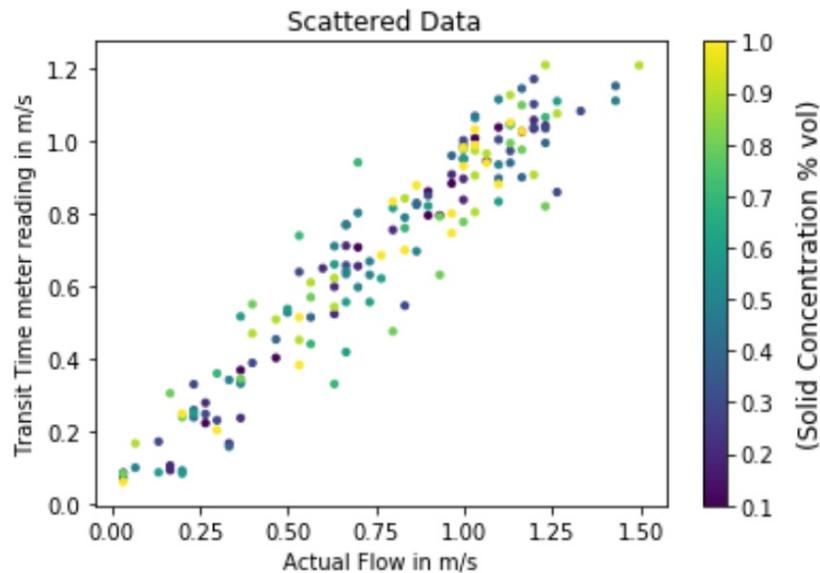


Figure 4: Scattered plot of transit-time ultrasonic flowmeter performance on 2-in horizontal bentonite-water mixture flow for bentonite concentration of 0.1 – 1.0 vol%.

In Figure 5, the scattered points of flow coefficient  $C$  for transit-time ultrasonic flowmeter have been found to stabilize as the flow speed exceeds 0.6 m/s. The flow coefficient  $C$  is defined as the ratio of measured flow rate to the actual fluid rate. The equation is  $C = Q_{\text{meter}}/Q_{\text{actual}}$  or  $V_{\text{meter}}/V_{\text{actual}}$ . The value of  $C$  should approach unity ( $C = 1$ ) if the meter is accurate. But the mean flow coefficient  $C_{\text{All}}$  is obtained as 0.9651.

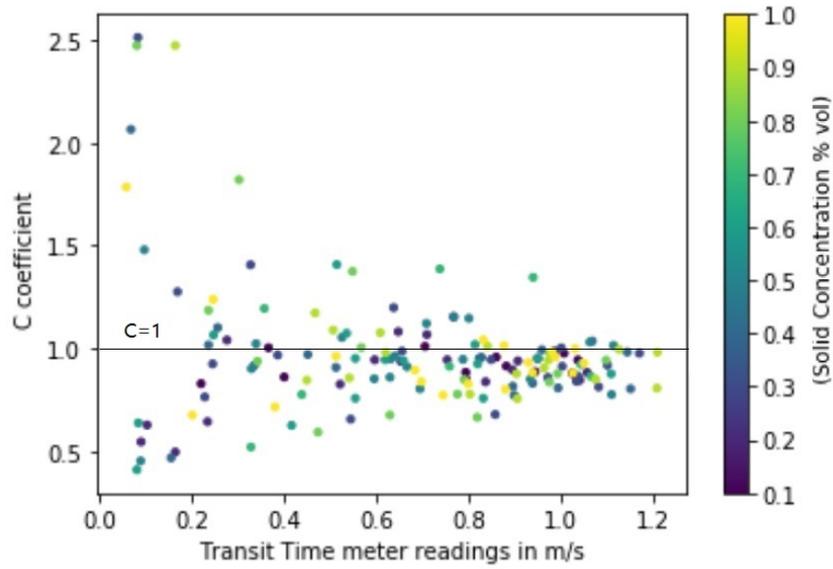


Figure 5: Flow coefficient versus Flow velocity for bentonite concentration of 0.1 – 1.0 vol%

#### 4.2 Error Analysis

An error is defined as a deviation from its truth value (Abedjan et al. 2016). The relative error is determined by subtracting ultrasonic flow meter readings from the reference velocity divided by reference velocity. Figure 6 show the relative error for all experimental data acquired.

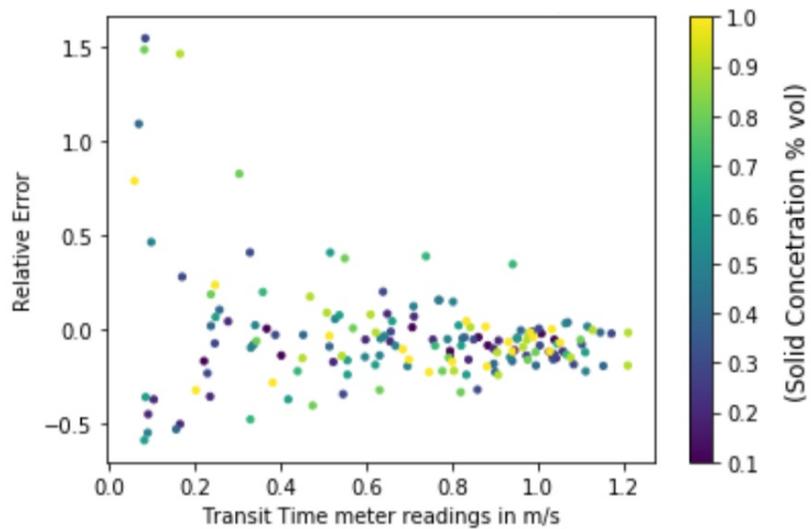


Figure 6: Relative error versus flow velocity for bentonite concentration of 0.1 – 1.0 vol%.

In this study, a total of 160 sample data were taken from experiments in a range of flow velocity (0 -1.2~1.5 m/s) and bentonite concentration (0.1-1.0 % volume). The machine learning models went under training.

About 62% of the data were used for training, and the remaining 38% were used for testing.

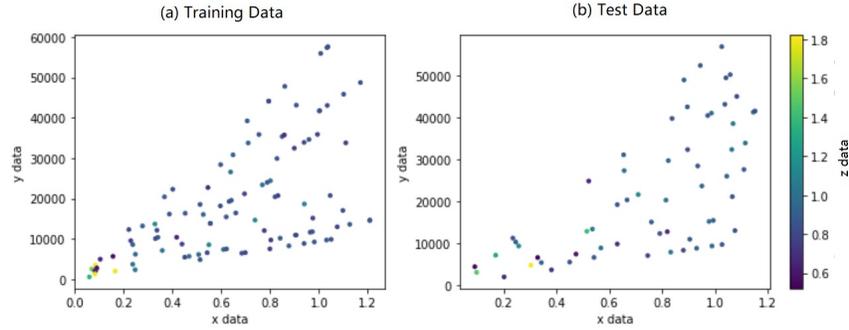


Figure 7: The training data set and test data set used to conduct recalibration exercise.

From the error analysis in Table 1, the mean relative error  $\epsilon_{All}$  is -0.033 (-3.03%), with a standard deviation  $\sigma_{All}$  of 0.3047 (30.47%). Before applying the error reduction algorithm, the test dataset has a mean flow coefficient  $C_{Test}$  of 0.917, a mean relative error  $\epsilon_{r_{Test}}$  of -0.0831 (-8.31%) and a standard deviation  $\sigma_{r_{Test}}$  of  $\pm 0.2030$  (20.30%).

Table 1: Summary of error information of sampled data.

Item	Subject	Mean Relative Error, $\epsilon$	Std. Deviation, $\sigma$	Uncertainty, $\epsilon \pm 2\sigma$
1	All Data	-0.033 (-3.3 %)	$\pm 0.305$ ( $\pm 30.5\%$ )	$\pm 0.610$ (61.0%)
2	Train Data Set	-0.007 (0.7%)	$\pm 0.344$ ( $\pm 34.3\%$ )	$\pm 0.688$ (68.8%)
3	Test Data Set	-0.0830 (-8.3%)	$\pm 0.203$ ( $\pm 20.3\%$ )	$\pm 0.406$ ( $\pm 40.6\%$ )

### 4.3 Machine Learning Error Reduction Exercises

#### 4.3.1 Linear Least Square Regression Technique (LLS)

The least square approach is employed to determine the parameters for obtaining the "best fitting line" for a series of observations. It aims to train parameter  $\beta_o$  and  $\beta_1$  in a linear equation as given below:

$$y = \beta_o + \beta_1 x \dots\dots (16)$$

Let us consider n observations of x and y to build a mathematical model using the parameters  $\beta_o$  and  $\beta_1$ , such that:

$$y_1 = \beta_o + \beta_1 x_1 + r_1 \quad y_2 = \beta_o + \beta_1 x_2 + r_2 \dots\dots\dots (17) \quad y_n = \beta_o + \beta_1 x_n + r_n$$

where  $r_i$  is the residual error terms,  $\beta_o$  is the intercept of the line with the y axis, and  $\beta_1$  is the slope of the line.

The corresponding cost function is defined as:

$$h(\beta_o, \beta_1) = \frac{1}{2} \sum_{i=1}^n r_i^2 = \frac{1}{2} \sum_{i=1}^n (y_i - \beta_o - \beta_1 x_i)^2 \dots\dots\dots (18)$$

As the cost function  $h(\beta_o, \beta_1)$  is at its minimum, the partial derivatives of  $h(\beta_o, \beta_1)$  with respect to  $\beta_o$  and  $\beta_1$  is equal to zero. This ultimately leads to the solutions of  $\hat{\beta}_o$  and  $\hat{\beta}_1$  which satisfies the least square error to give the best fitting line for a number of given data.

In this machine learning study, the least square approach is used to train "best fitting line" for the scattered data as depicted in Figure 8 below. The purpose is to obtain the value of  $\beta_o$  (the intercept of the line with

the y axis) and  $\beta_1$  (the slope of the line), as previously introduced in Eq. (16).

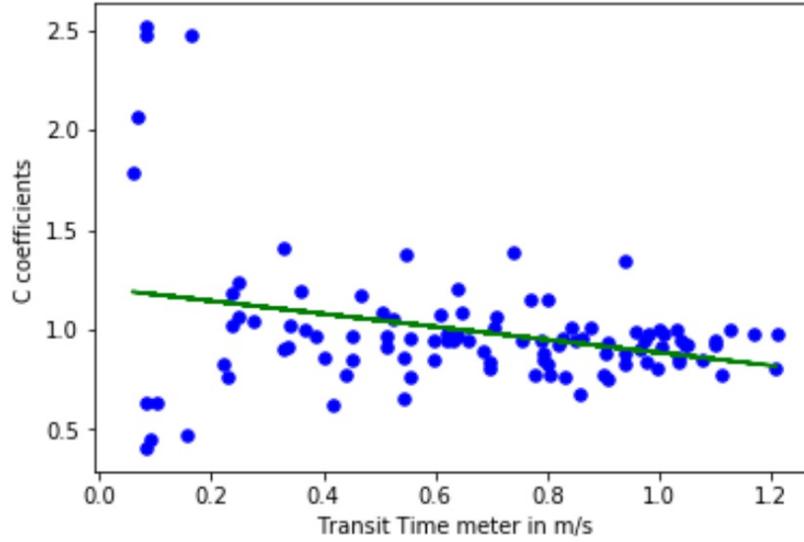


Figure 8: Linear least square error reduction for training data.

Results give  $\beta_o = 1.2082$ , and  $\beta_1 = -0.3241$ . Eq (16) is then modified to yield  $C_{LLS}$ :

$$C_{LLS} = \beta_o + \beta_1 v_{Test} \dots (19)$$

where  $v_{Test}$  is the test data given by transit time ultrasonic flow meter.

Next, the revised flow coefficient  $C_{LLS}$  was applied to the correction equation, which corrects the value of ultrasonic flow measurement.

$$v_{Rev} = \frac{v_{Test}}{C_{LLS}} \dots (20)$$

By applying the correction equation, the mean flow coefficient  $C_{Rev}$  has been revised from 0.917 to 0.957. As shown in Table 2, the mean relative error has reduced from -0.0830 (-8.3%) to -0.0422 (-4.2%). However, the revised standard deviation of relative error  $\sigma_{Rev}$  is about the same at  $\pm 0.2004$  (20.04%). This indicates the LLS regression method has reduced both the systematic error and the relative error only slightly.

Table 2: Summary of error information before and after the LLS Regression model was implemented.

Item	Subject	Mean Relative Error, $\epsilon$	Std. Deviation, $\sigma$	Uncertainty, $\epsilon \pm 2\sigma$
1	Test Data Set	-0.0830 (-8.3%)	$\pm 0.203$ ( $\pm 20.3\%$ )	$\pm 0.406$ ( $\pm 40.6\%$ )
2	LLS regression method	-0.0422 (-4.2%)	$\pm 0.200$ ( $\pm 20.0\%$ )	$\pm 0.401$ ( $\pm 40.1\%$ )

### 4.3.2 Linear 2D Interpolation Error Reduction

In numerical analysis, a multivariate interpolation is interpolation on a function of more than one variable. The function to be interpolated is known at given points and the interpolation problem consists of yielding values at arbitrary points. Multivariate interpolation could be useful for flow meter self-calibration, where it is used to create a digital model from a set of error information (such as flow coefficient  $C$ ) based on flowmeter measurements. It can be used to improve meter output at points where the error information is

unknown.

Python coded 2D grid data algorithm has been used to fix a surface of the form  $z = f(x,y)$  from the training data  $(x, y, z)$ . For example,  $(x, y, z)$  can be information extracted from flow meter outputs such as  $x$  as flow meter readings,  $y$  as Reynolds number and  $z$  as flow coefficient  $C$ .  $(x_q, y_q)$ . The grid-data function is then used to interpolate the surface at the query points or test points  $(x_q, y_q)$  and returns the interpolated values  $(z_i)$ . As illustrated in Fig. 9, the returned interpolated values can be visualized by using 2D Delaunay Triangulation method.

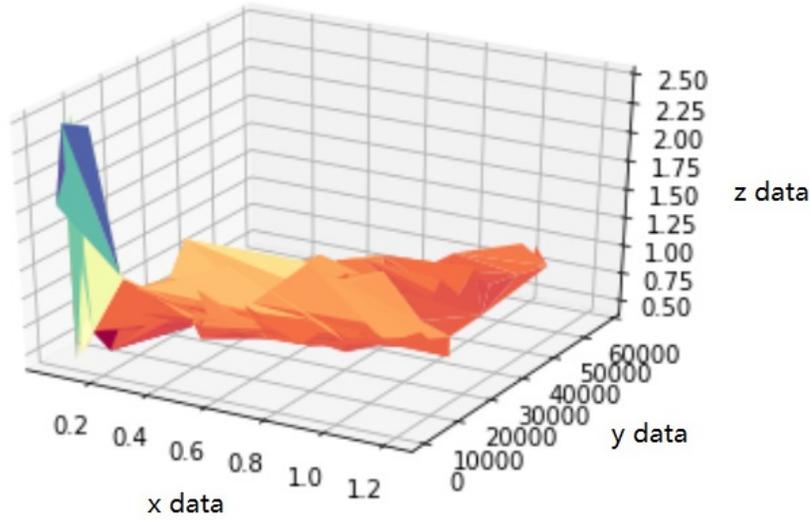


Figure 9: The interpolated plot as visualized by 2D Delaunay triangulation method.

The returned interpolated values  $z_i$  based on test point  $(x_q, y_q)$  are plotted in Figure 10. However, 2D interpolation method does not extrapolate but fill the values which are not within the input area with 'nan' values by default, illustrated in Fig. 10 (a) as the area in white color. As can be seen in Fig. 10(b), these 'nan' values can be replaced by numeric '1', illustrated as the area in blue color.

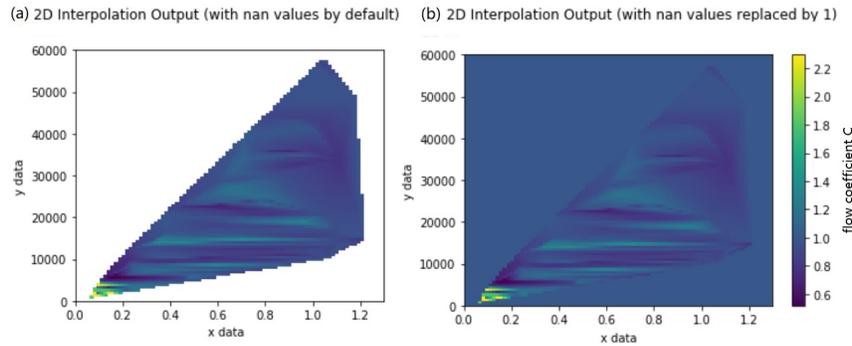


Figure 10: The interpolated error information based on training points.

The interpolated error information will be used to revise the flowmeter output as follow:

$$v_{Rev} = \frac{v_{TTUF}}{Z_i} \dots (21)$$

By applying the correction equation (21), the mean flow coefficient  $C_{Rev}$  has improved from 0.917 to 1.012. The machine learning method such as 2D interpolation is inherently non-linear, so it can detect almost any kind of non-linear behaviors and produce a better estimate for the output and response. It gives a better accuracy than linear regression approach such as LLS method.

As summarized in Table 3, the mean relative error  $\epsilon_{Rev}$  was revised from -0.0830 (-8.30%) to +0.0124 (+1.24%). However, the error was reduced to less than 1.5% but at the cost of increased standard deviation,  $\sigma_{Rev}$  from  $\pm 0.2030$  (20.30%) to  $\pm 0.2702$  (27.02%).

Table 3: Summary of error information before and after implementing the 2D interpolation model.

Item	Subject	Mean Relative Error $\epsilon$	Std. Deviation $\sigma$	Uncertainty $\epsilon \pm 2\sigma$
1	Test Data Set	-0.0830 (-8.3%)	$\pm 0.203 (\pm 20.3\%)$	$\pm 0.406 (\pm 40.6\%)$
2	2D Interpolation method	+0.0124 (+1.2%)	$\pm 0.270 (\pm 27.0\%)$	$\pm 0.540 (\pm 54.0\%)$

### 5.0 Filtering off Data with Large Relative Error

Data cleaning has played a critical role in ensuring data quality for industrial applications. With the increasing prevalence of data-centric approaches to business and scientific problems with data as a crucial asset, data cleaning has become even more important (Abedjan et al. 2016).

As can be seen in preceding Table 1 and Table 2, the results from both LLS regression method and 2D interpolative model has failed to reach the accuracy below 1% error, moreover, it does not improve on the standard deviation of the relative errors. As reviewed in preceding section 2.2, the improvement of mean relative error can be associated with the improvement of the **systematic error** (or bias). However, the failure in reducing the standard deviation of relative error is believed to be the inherent nature of random errors.

As depicted in Fig. 11, it was clearly shown that standard deviation can increased drastically due to the presence of random errors. Yazdanshenashad et al. (2018) believed the random errors cannot be reduced by flow meter recalibration. They suggested to reduce the errors by averaging large number of data.

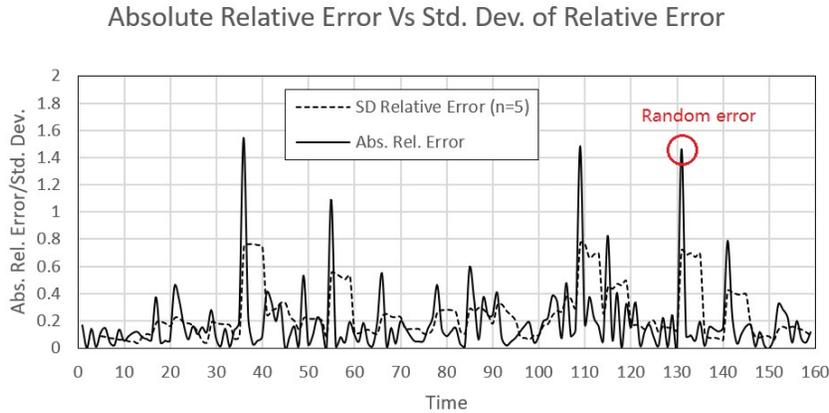


Figure 11: Summary of error information of 160 sampled data.

In this study, we decided to address the issue by the filtering off sampled data which had large relative errors. However, these incoming with random errors were not known in real-time operation. Therefore, it requires

a way to train the artificial intelligence models to predict data errors accurately before the data are cleaned. A successful exercise should lead to an outcome of lower standard deviation of relative errors.

### 5.1 Error Data Prediction using Bayes Theorem Classifier Algorithm

Error data prediction can be categorized as a type of anomaly detection (or outlier detection), which is the identification of data or observations which differs significantly from majority of the data. The most common way to perform anomaly detection is using the classification algorithm.

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems (Brownlee, 2016). This method is a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable (Zhang, 2004). The Theorem was named after English mathematician Thomas Bayes (1701-1761). Bayes' Theorem is stated as:

$$P(A/B) = \frac{P(B/A).P(A)}{P(B)} \dots \quad (22)$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \dots \quad (23)$$

where  $P(A/B)$  is the probability of class (A) given the provided data (B).

Bayes' theorem allows users to figure out  $P(A|B)$  from  $P(B|A)$ . Rather than attempting to calculate the probabilities of each attribute value, the data are assumed to be conditionally independent given the class value. As a result, Naïve Bayes classifier approach requires only a small amount of training data to predict and classify the outcome. In spite of the assumption that the attributes do not interact (which is most unlikely in real data), the approach has worked quite well in many real-world situations.

Naive Bayes can be extended to real-valued attributes by assuming a Gaussian distribution. This extension of naive Bayes is called Gaussian Naive Bayes (Brownlee, 2016). A collection of data points typically has a certain distribution (e.g. a Gaussian distribution). To detect error data (anomalies), the probability distribution  $p(x)$  from the data points is first calculated. As a new datum,  $x$ , comes in, we compare  $p(x)$  with a threshold  $r$ . If  $p(x) < r$ , it is considered as an error or anomaly. This is because normal examples tend to have a large  $p(x)$  while anomalous examples tend to have a small  $p(x)$  (Flovik 2019). This is the easiest way to proceed because users only need to estimate the mean and the standard deviation from the training data.

In this exercise, 62% of our experimental data were used as the training data, while the remaining 38% were used as the test data. The filtration task is defined as removing data which contains large errors (i.e.  $C [?] 0.85$  or  $C [?] 1.15$ ). Three types of Gaussian Naive Bayes classifier algorithms were used to predict the probabilities of having data with large errors. The probabilities will be used to classify whether the data is the error data which should be removed or retained. Three types of Gaussian Naive Bayes algorithm have been implemented, which are (a) Gaussian NB without calibration, (b) Gaussian NB based on a non-parametric isotonic regression calibration, and (c) Gaussian NB based on Platt's sigmoid model calibration. This probability calibration is reported to improve the confidence on the prediction (Metzen, 2015).

Figure 12 (a) shows the background information of the test. The red-colored points represent data with an acceptable error (where  $0.85 < C < 1.15$ ), while dark colored points represent the targeted points to be filtered because of excessive error (i.e.  $C [?] 0.85$  or  $C [?] 1.15$ ). After the Gaussian Naive Bayes scheme is implemented, the results are plotted in Figure 12 (b). The dark colored points have been filtered.

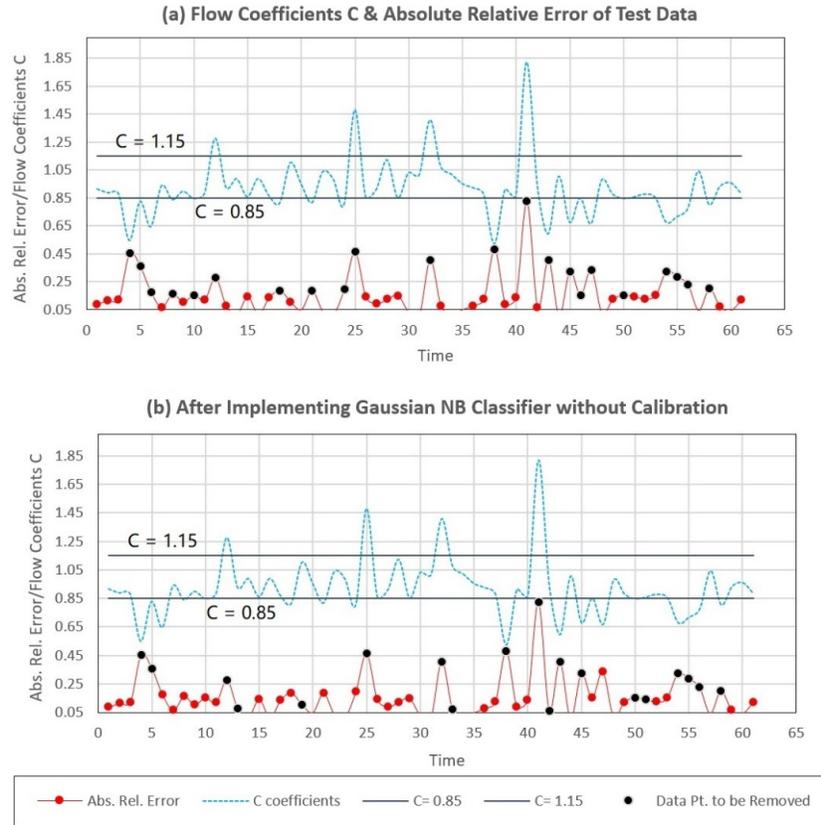


Figure 12: Graph (a) shows the flow Coefficient  $C$  and Absolute Relative Error of the test data. Dark colored points indicate the targeted points to be removed because of excessive errors (i.e.  $C \leq 0.85$  or  $C \geq 1.15$ ); Graph (b) shows the results after the Gaussian Naïve Bayes Classifier Scheme was implemented.

Comparing the two graphs, most of the data points with significant errors have been predicted and removed. The exercise was repeated with Gaussian NB methods based on isotonic calibration and sigmoid calibration. The final outcomes were summarized in Table 4.

Table 4: Summary of error before and after implementing the Gaussian Naïve Bayer Scheme.

Item	Subject	Mean Relati
1	Train Data Set	-0.007 (0.7%)
2	<b>Test Data Set</b>	<b>-0.0830 (-8.3%)</b>
3	Manual removing of test data with flow coefficient $C \leq 0.85$ or $C \geq 1.15$ (as depicted in Fig. 12 (a)).	-0.0588 (-5.9%)
4	Filtering Test Data using <b>Gaussian Naïve Bayes algorithm with no calibration</b>	<b>-0.0945 (-9.4%)</b>
5	Filtering Test Data using Gaussian Naïve Bayes algorithm with isotonic calibration	-0.0928 (-9.3%)
6	Filtering Test Data using Gaussian Naïve Bayes algorithm with sigmoid calibration	-0.0839 (-8.4%)

The results in Table 2 indicate the Gaussian Naïve Bayes algorithm (with no calibration) is good enough to filter raw data with large errors. A significant reduction of the standard deviation value of relative error  $\sigma_{Rev}$  from  $\pm 20.3\%$  to  $\pm 8.0\%$  shows the Naïve Bayes method is effective to reduce the random errors. However, it may not be useful to correct the systematic error, as the mean relative error  $\epsilon_{Rev}$  after the exercise has

increased from -8.3% to -9.5%.

### 5.3 Combined Strategies using 2D Interpolation and Bayes Theorem Classifier Algorithm.

There are multiple types of errors associated with machine learning and predictive analytics and different types of errors may coexist in the same data set (such as systematic errors and random errors). It might be necessary to run more than one kind of tool. In the final resort to improve both type of errors, it was decided to implement the 2D interpolative error correction algorithm on the test data which have been filtered by the Gaussian Naïve Bayes algorithm. The results are summarized in Table 5.

Table 5: Summary of error information before and after implementing the combined strategies.

Item	Subject
1	Test Data Set
2	Test Data after being filtered using Gaussian Naïve Bayes algorithm with no calibration
3	Test Data after being filtered using Gaussian Naïve Bayes algorithm with no calibration+ followed by 2D Interpolation

Results show the combined 2D interpolative method and Gaussian Naïve Bayes algorithm has reduced both types of errors. From Table 5, it was obvious that the mean relative error  $\epsilon_{Rev}$  was revised from -8.3% to -0.6%. The standard deviation  $\sigma_{Rev}$  has decreased from  $\pm 20.3\%$  to  $\pm 13.7\%$ . It has improved the mean flow coefficient  $C_{Rev}$  from 0.900 to 0.994.

### 6.0 Conclusions

The following conclusions may be drawn from our study.

- The linear least square error method (LLS) and 2D interpolation method exhibit lower mean relative error (-4.2% and +1.2%) compared to that of the test dataset before calibration (-8.3%). Either modeling approaches was able to improve accuracy (indicated by the improvement in systematic error) but not the precision (indicated by the standard deviation of relative errors, which is associate with the distribution patterns of random errors).
- Three types of Gaussian Naïve Bayes modeling approaches were tested. The best result has succeeded to reduce the standard deviation value of relative error from  $\pm 20.3\%$  (of test data) to  $\pm 8.0\%$ . However, the reduction is achieved at the cost of increased mean relative error from -8.3% (of test data) to -9.5% (after data cleaning). Results also show the Gaussian Naïve Bayes method is useful to improve random errors, but not the systematic errors.
- Since different types of errors may coexist in the same data set (such as systematic errors and random errors), we show that use of multiple modeling approaches can successfully reduce both types of errors. When the 2D interpolation method was applied to the test data using the Gaussian Naïve Bayes algorithm, the mean relative errors were reduced from -8.3% (of test dataset) to -0.6% (after data cleaning), and standard deviation of the relative errors was reduced from  $\pm 20.3\%$  to  $\pm 13.7\%$ .
- These results proved that multiple machine learning models can be trained to evaluate the sampled data from flow meter, thus re-calibrating both the systematic errors and random errors.
- Our study shows a high accuracy ultrasonic flow meter with systematic errors less than 1% for oil and gas multiphase application is possible with the aid of artificial intelligence technology.

### Nomenclature

Latin Characters

$C$  = flow coefficient.

$C_{LLS}$  = predicted flow coefficient based on LLS method.

$C_{Rev}$  = revised mean flow coefficient after machine learning algorithm was implemented.

$c$  = ultrasound speed in water (m/s).

$D$  = pipe internal diameter (m).

$k_c$  = velocity correction factor.

$L$  = the path travelled by ultrasound (m).

$m_p$  = total mass of suspended particles (kg)

$u$  = flow velocity (m/s).

$u_f$  = fluid velocity (m/s).

$V_s$  = volume of solids (m<sup>3</sup>).

$V_{m1}$  = volume of initial mud or fresh water (m<sup>3</sup>).

$V_{m2}$  = volume of final mixture (m<sup>3</sup>).

$v$  = mean flow velocity (m/s)

$v_{Test}$  = flow velocities from test data set.

Re = Reynolds number.

$t_{21}, t_{12}$  = The transit time of the ultrasonic signal from downstream transducer to the upstream transducer and that is in opposite direction (s).

$w_b$  = weight of bentonite particles (kg).

$w_w$  = weight of water (kg).

*Greek Characters*

$\beta_0, \beta_1$  = coefficient in the least square error method.

$\epsilon_{Rev}$  = revised mean relative error after machine learning algorithm was implemented.

$\vartheta$  = the incident angle of ultrasound wave.

$\mu_f$  = fluid viscosity (Pa.s)

$\nu$  = kinematic viscosity (m<sup>2</sup>/s).

$\rho_f$  = density of fluid (kg/m<sup>3</sup>).

$\rho_s$  = density of solids (kg/m<sup>3</sup>).

$\rho_p$  = density of particle (kg/m<sup>3</sup>).

$\rho_{m1}$  = density of initial mud or fresh water (kg/m<sup>3</sup>).

$\rho_{m2}$  = density of final mixture (kg/m<sup>3</sup>).

$\sigma_v$  = standard deviation of flow velocities (m/s).

$\sigma_{Test}$  = standard deviation of relative error of test data set.

$\sigma_{Rev}$  = revised standard deviation of relative error after machine learning algorithm was implemented.

## References

Abedjan, Z. et al. (2016). Detecting Data Errors: Where are we and what needs to be done. [Web]. Cited 01.07.2020. Available from: <http://www.vldb.org/pvldb/vol9/p993-abedjan.pdf>

- Berrebi, J. et al (2004). Ultrasonic Flow Metering Errors Due to Pulsating Flow. EISLAB, Dept. of Computer Science and Electrical Engineering, Lulea University of Technology, Sweden. [Web]. Cited 06.05.2020. Available from: <https://www.diva-portal.org/smash/get/diva2:989127/FULLTEXT01.pdf>
- Birgersson, M. et al. (2009), Bentonite erosion, Final Report, SKB Technical report TR-09-34 [Web]. Cited 12.04.2020. Available from: <https://www.skb.se/publication/1975568/TR-09-34.pdf>
- Brownlee, J. (2016). Naive Bayes for Machine Learning. [Web]. Cited 25.06.2020. Available from: <https://machinelearningmastery.com/naive-bayes-for-machine-learning/#>
- Catak, M. and Ergan, C. (2019). Self-Calibration of Ultrasonic Water Flow Meter. International Journal of Recent Technology and Engineering (IJRTE). Vol. 9, Issue 4., November 2019. [Web]. Cited 05.05.2020. Available from: [https://www.researchgate.net/publication/338886029\\_Self-Calibration\\_of\\_Ultrasonic\\_Water\\_Flow\\_Meter](https://www.researchgate.net/publication/338886029_Self-Calibration_of_Ultrasonic_Water_Flow_Meter)
- College Physics Labs (2020). Measurement and Error Analysis. The University of North Carolina at Chapel Hill. Retrieved from [webassign.net](http://webassign.net). [Web]. Cited 18.05.2020. Available from: [https://www.webassign.net/question\\_assets/unccolphysmechl1/measurements/manual.html](https://www.webassign.net/question_assets/unccolphysmechl1/measurements/manual.html)
- Eren, H. (1998). Accuracy in Real Time Ultrasonic Applications and Transit-time Flow Meters. IEEE Instrumentation and Measurement Technology Conference, Ottawa, Canada, May 18-21.
- Flovik, V. (2019). How to use machine learning for anomaly detection and condition monitoring. [Web]. Cited 29.06.2020. Available from: <https://towardsdatascience.com/how-to-use-machine-learning-for-anomaly-detection-and-condition-monitoring-6742f82900d7>
- Hamouda, A. et al (2016). An Enhanced Technique for Ultrasonic Flow Metering Featuring Very Low Jitter and Offset. *Sensors* 16. 1008.
- International Organization for Standardization - ISO 5168. Measurement of fluid flow: Procedures for the evaluation of uncertainties. Geneva, Switzerland, 65 p., 2005.
- Kalla, S. (2009). Random Error. Retrieved from [Explorable.com](http://Explorable.com). [Web]. Cited 18.05.2020. Available from: <https://explorable.com/random-error>
- Lansing, J. (2003). Principles of Operation for Ultrasonic Gas Flow Meters. Daniel Measurement and Control, Inc. [Web]. Cited 08.05.2020. Available from: <https://asgmt.com/wp-content/uploads/pdf-docs/2003/1/11.pdf>
- Metzen, J.H. (2015). Probability Calibration. [Web]. Cited 25.06.2020. Available from: <https://jmetzen.github.io/2015-04-14/calibration.html>
- Othman, A. (2007). Drilling Engineering - SKM3413. [Web]. Cited 20.02.2020. Available from: <https://www.scribd.com/doc/89388268/Drilling-Mud-Calculations>
- Simurda, M. et al. (2016). Modelling of Transit-Time Ultrasonic Flow Meters Under Multiphase Flow Conditions. 2016 IEEE International Ultrasonics Symposium (IUS). DOI: 10.1109/ULTSYM.2016.7728844
- Weissenbrunner, A. et al (2016). Simulation-Based Determination of Systematic Errors of Flow Meters Due to Uncertain Inflow Conditions. *Flow Measurement and Instrumentation*. Vol. 52, 25-39. [Web]. Cited 06.05.2020. Available from: <https://www.sciencedirect.com/science/article/pii/S0955598616300875>
- Yazdanshenashad, B. et al. (2018). Neural-network-based Error Reduction in Calibrating Utility Ultrasonic Flow Meters. *Flow Measurement and Instrumentation* 64, 54-63.
- Zhang, H. (2004). The optimality of Naive Bayes. *Proc. FLAIRS*. Cited 18.05.2020. Available from: <https://www.cs.unb.ca/~hzhang/publications/FLAIRS04ZhangH.pdf>