NEXRES 2 (2025) 100915

Contents lists available at ScienceDirect

Next Research

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

Modelling over-reading correction factors for ultrasonic flow meters in wet gas measurement using advanced regression and machine learning techniques

Ishigita Lucas Shunashu [a,*], Osmund Kaunde [b]

[a] Department of Mechanical Engineering, College of Engineering and Technology, Mbeya University of Science and Technology, Tanzania
[b] College of Engineering and Technology, Mbeya University of Science and Technology, Tanzania

ARTICLE INFO

Keywords:
Machine learning
Ultrasonic flow meter
Wet gas measurement
Over-reading correction
Advanced regression

ABSTRACT

Accurate wet gas measurement is essential for optimizing production, transmission, and reservoir management in oil and gas operations. Ultrasonic flow meters, though non-intrusive and versatile, often overestimate flow rates due to the presence of liquid phases, leading to significant operational and economic errors. To address this, a data-driven correction model was developed using computational techniques to predict and compensate for over-reading. This study evaluates the performance of several advanced regression and machine learning approaches, including polynomial regression, random forest regression, nonlinear curve fitting, neural networks, multiple linear regression, ridge regression, and lasso regression, using an extensive experimental dataset. Key input variables include liquid volume fraction, Lockhart–Martinelli parameter, Froude number, Weber number, slip ratio, and density ratio. Among the models tested, random forest regression and multiple linear regression achieved the highest accuracy, with average relative absolute errors of 3.02% and 3.20 % respectively. These findings demonstrate the potential of data-driven modeling to enhance the reliability of ultrasonic flow meters in complex wet gas environments.

1. Introduction

Wet gas, characterized by the simultaneous presence of gas and liquid phases, presents significant challenges for accurate flow measurement. Ultrasonic flow meters (USMs) are widely adopted in the oil and gas industry due to their non-intrusive design, wide turndown ratio, and suitability for high-pressure applications [1]. However, in wet gas conditions, USMs are prone to over-reading errors, where the measured flow rate exceeds the actual gas flow rate due to the influence of entrained liquid on acoustic signal propagation and velocity profile distortion [2]. These errors can lead to substantial inaccuracies in production allocation, reservoir management, and custody transfer. Recent studies have explored the integration of machine learning (ML) and advanced regression techniques to model and correct over-reading by capturing the nonlinear interactions between flow parameters and sensor responses [3,4]. Despite these advancements, the development of robust, interpretable, and field-deployable correction models remains an open research challenge.

In recent years, machine learning and advanced regression techniques such as Polynomial Regression, Random Forest Regression, Neural Networks, and Multilinear Regression have been employed to model and correct over-reading in wet gas measurement [5–7]. These techniques offer promising avenues for enhancing the precision of gas flow rate measurements by capturing the complex interactions between the gas and liquid phases. The integration of machine learning and advanced regression techniques into flow measurement systems represents a significant advancement in the field, providing more accurate and reliable solutions for the oil and gas industry [8]. Despite advancements in flow measurement technologies, the accurate quantification and correction of over-reading in wet gas conditions remain unresolved. The complexity of two-phase flow dynamics, coupled with the variability in gas-liquid ratios, necessitates the development of robust correction models that can accurately predict and mitigate over-reading.

This research aims to develop and compare various regression models to determine the most effective approach for over-reading correction in ultrasonic flow meters using machine learning and regression techniques, including Polynomial Regression (PR), Random

* Corresponding author.
E-mail addresses: ishigita.shunashu@cbe.ac.tz (I.L. Shunashu), osmund.kaunde@must.ac.tz (O. Kaunde).

https://doi.org/10.1016/j.nexres.2025.100915
Received 23 September 2025; Accepted 8 October 2025
Available online 12 October 2025

Forest Regression (RFR), Scikit-learn Neural Network (SNN), and Multilinear Regression techniques (MLR). However, this research is guided by three specific objectives: (1) To develop over-reading correction models for ultrasonic flow meters in wet gas measurement using multiple regression and machine learning techniques; (2) To perform a comparative analysis of model performance across varied flow conditions using experimental data; and (3) To evaluate the practical implications of the correction models for improving flow measurement accuracy in industrial applications.

2. Literature review

Accurate modeling of over-reading (OR) in wet gas measurement has gained significant attention due to its implications for enhancing wet gas flow metering precision, particularly in ultrasonic flow meters. Traditional methods have focused on physics-based approaches, as demonstrated by van Putten and Dsouza [1], who developed OR correction models tailored to horizontal configurations using extensive experimental datasets. Similarly, the Joint Industry Project by DNV GL [1] introduced correction algorithms addressing liquid-induced OR, while Xu et al. [2] investigated droplet film thickness in gas-liquid two-phase flow for DN80 two-path ultrasonic meters, achieving a relative uncertainty of ±15 %. However, these methods lacked adaptability to dynamic flow conditions and fell short in leveraging modern computational advancements.

Recent breakthroughs have highlighted the integration of machine learning (ML) techniques into OR modeling. Gryzlov et al. [9] explored Artificial Neural Networks (ANN) in virtual flow metering, demonstrating solid baseline performance with moderate training times. ANN was effective for dynamic scenarios, particularly with Gated Recurrent Units (GRU), yet faced challenges with noisy data and dataset generalization. Wang et al. [8] extended this exploration by comparing Deep Neural Networks (DNN), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN) for flow rate estimation, with SVM achieving the lowest Mean Squared Error (MSE) and highest accuracy. Conversely, CNN exhibited limitations due to batch averaging techniques and poor handling of data constraints. Similarly, Hosseini [4] emphasized the robustness of Long Short-Term Memory (LSTM) and GRU networks, which demonstrated low errors by RMSE: 0.0349 and adaptability to flow regime changes. However, these techniques often struggled with generalization, especially in unseen or high-dimensional datasets.

Hybrid approaches combining data-driven models with physical principles have been proposed as promising solutions. Mofunlewi et al. [10], utilized Random Forest Regression (RFR) to predict oilfield fluid flow rates, highlighting its robustness despite computational demands. Jiang et al. [11] revealed that integrating DNN and SVM achieved reliable results for gas and liquid flow rates, although Gradient Boosting Decision Trees (GBDT) failed due to poor dataset adaptability. Similarly, Nguyen and Saga [12] leveraged multilinear regression (MLR) to model pressure drops in two-phase flows, while Qiu et al. [13] employed neural networks for spatiotemporal interface modeling, enhancing accuracy. Despite their advancements, these models required significant data pre-processing and were limited by their sensitivity to noisy conditions.

Regression models have traditionally played a significant role in OR correction, but they face inherent limitations. Multilinear regression, as studied by Chen et al. [14], struggled with multicollinearity among predictor variables, leading to unstable estimates and diminished accuracy. Random Forest Regression, as demonstrated by Mofunlewi et al. [10], provided robust predictions for oilfield fluid flow rates but required substantial computational resources for training. Neural networks, such as Scikit Neural Network, excelled in modeling spatiotemporal evolution in two-phase flows, as highlighted by Qiu et al. [13]. However, the interpretability of these models remains a challenge, and they often struggle with extrapolation beyond training ranges, leading to inaccuracies in real-world applications.

The integration of machine learning (ML) methodologies within this study is both conceptually and technically warranted due to the inherently nonlinear and interdependent behavior of wet gas flow dynamics. Conventional empirical models, including linear regression or physics-based formulations, often fall short in capturing the complex interactions between parameters such as Liquid Volume Fraction (LVF), Lockhart-Martinelli parameter (X_{LM}), Froude number (Fr_g), Weber number (We), Density Ratio (DR), and Reynolds numbers for both gas (Re_g) and liquid (Re_l). These flow descriptors do not exhibit straightforward relationships with the ultrasonic flow meter's measured output due to the multi-phase-induced signal distortions, acoustic impedance mismatches, and phase velocity shifts that give rise to over-reading errors.

The necessity of ML in this context arises from its capacity to learn and represent these intricate relationships without assuming a fixed functional form. This aligns with prior metrology research; for example [15], reported RFR-based ultrasonic deviation models with RMSE ≤ 0.06 % and uncertainty ≤ 0.12 %. Likewise [4], emphasized the diagnostic reliability of hybrid ML models under changing salinity, flow regimes, and pressure conditions, reinforcing their industrial viability. Furthermore, the deployment of ML is not merely for predictive enhancement as it represents a paradigmatic shift toward adaptive, data-driven metering systems capable of compensating for flow regime transitions, measurement uncertainty, and environmental variability. As supported by literature [16], the adoption of ML facilitates dynamic calibration frameworks, contextual sensor fusion, and improved metrological traceability in real-time. Thus, the experimental validation and cross-referenced evidence strongly affirm that machine learning is not only a viable solution but a necessary advancement for modern wet gas flow metering system

Building on this review, the present research aimed to compare machine learning (ML) models such as Random Forest Regression and Scikit Neural Network with regression models, including Multilinear Regression, Ridge Regression, Lasso Regression, and Scipy Optimization. Based on experimental ultrasonic flow meter data. The study intended to establish the most accurate and efficient model for OR correction to predict wet gas flow rate based on USM. This comparison addressed the efficiency and robustness of the USM data-driven over-reading correction model (USMOR), particularly in handling high-dimensional data and improving model interpretability. The findings of this study hold significant potential for advancing the precision of wet gas flow rate measurement and enhancing the reliability of industrial operations.

3. Materials and methods

3.1. Transit time ultrasonic flow meter

In this study, wet gas flow rates were measured using a clamp-on transit-time ultrasonic flow meter (TT USM), specifically the Fluxus G800 from Flexim GmbH (Germany). The TT USM uses ultrasonic transit time for single-phase flow measurement, with transducers positioned at an angle to the gas flow, alternating as transmitters and receivers to generate ultrasonic pulses.

The TT-USM operates on the transit-time differential principle, whereby paired transducers alternately transmit and receive ultrasonic pulses across the pipe. A pulse travelling in the direction of flow is accelerated, while one travelling against the flow is decelerated, producing a measurable difference in transit times. This difference is directly related to the average flow velocity along the acoustic path, which, when multiplied by the known cross-sectional area of the pipe, yields the volumetric flow rate. Mathematically, the average velocity is expressed as in Fig. 1 below:

Referring to Fig. 1 above, and considering the transit times between two transducers t_1 and t_2 the distance L between the transmitting and receiving transducer, and the inclination angle θ, the transit times t_1 and

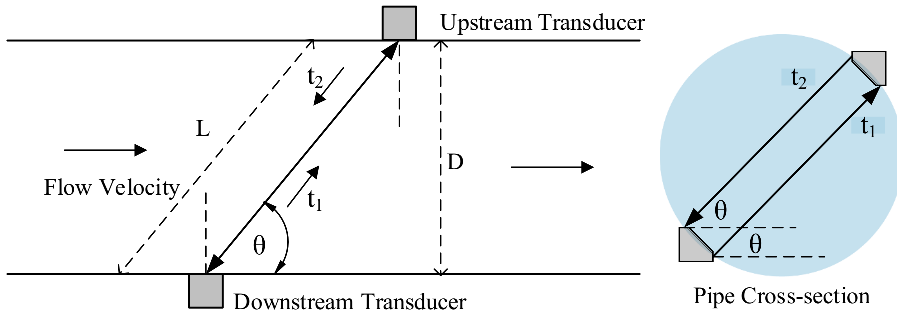

Fig. 1. Operating principle of transit time ultrasonic flow meter.

t_2 can be expressed in Eq. (1) and (2), respectively.

$$t_1 = \frac{L}{c - v\cos\theta} \tag{1}$$

$$t_2 = \frac{L}{c + v\cos\theta} \tag{2}$$

The difference in transit times, t_1 and t_2, is expressed in Eq. (3), under the assumption that t_2 is achieved when the ultrasonic pulse travels downstream, against the direction of flow, while t_1 corresponds to the ultrasonic pulse travelling upstream, along the flow. The equation for the average velocity is derived from Eq. (3) and is presented in Eq. (4). Furthermore, when the cross-sectional area of the pipe, A is pre-determined or known, the actual gas volumetric flow rate Q_g of the fluid can be calculated using Eq. (5), where $Q = AV$.

$$t_2 - t_1 = \frac{L}{c + v\cos\theta} - \frac{L}{c - v\cos\theta} \tag{3}$$

$$v = \frac{L}{2\cos\theta}\cdot\left(\frac{t_2 - t_1}{t_2 t_1}\right) \tag{4}$$

$$Q_g = A\cdot\frac{L}{2\cos\theta}\left(\frac{t_2 - t_1}{t_2 t_1}\right) \tag{5}$$

3.2. TT-USM over-reading in wet gas measurement

The presence of liquid droplets or films in wet gas reduces the effective cross-sectional area available for gas transport (A'_g) and alters the velocity distribution, leading to systematic over-reading in transit-time ultrasonic flow meters (TT-USM). These effects are particularly pronounced in mist and stratified flow regimes, where multipath interference and signal attenuation distort the transit-time differential used to compute flow velocity [1,2].

In stratified flow, the liquid phase typically occupies the lower portion of the pipe, displacing the gas upward and shifting the velocity maximum toward the upper wall. This asymmetry increases shear at the gas–liquid interface and modifies the velocity profile from the near-parabolic distribution typical of single-phase flow to a more flattened or skewed profile. In mist flow, entrained droplets increase turbulence intensity and disrupt uniformity along the acoustic path, introducing localised velocity fluctuations. Both regimes reduce A'_g relative to the dry-gas area A_g, forcing the gas phase to accelerate through the remaining open area [7].

For transit time ultrasonic metering systems (TT USM), two main implications arise. First, there is acoustic path distortion, which occurs when the ultrasonic beam encounters varying acoustic impedance in the liquid phase, causing refraction and scattering that affect the effective path length (L′) and beam angle (θ′), leading to biased transit-time differentials. Second is velocity bias, which happens when higher local velocities in the gas core result in an overestimation of average velocity, particularly if the meter assumes a single-phase profile.

The theoretical basis for this behaviour can be expressed through Eqs. (6) and (7) adapted from the formulation presented by Xu et al. [2] and van Putten & Dsouza [1]. Eq. (8) defines the measured wet gas flow rate Q_{tp} recorded by the TT-USM:

$$Q_{tp} = Ag\cdot\frac{L'}{2\cos\theta'}\left(\frac{t_2 - t_1}{t_1 t_2}\right) \tag{6}$$

Here, A'_g is the effective cross-sectional area for gas flow in the presence of liquid, reduced from the dry-gas area A_g due to liquid holdup [7]; L' is the effective acoustic path length between transducers in wet gas; θ' is the ultrasonic beam propagation angle in wet gas, and t_1 and t_2 are the upstream and downstream transit times, respectively, modified by the altered velocity profile [2].

Eq. (7) in this paper defines the over-reading factor (OR) as the ratio of the measured wet gas flow rate to the actual dry gas flow rate:

$$OR = \frac{Q_{tp}}{Q_g} = \frac{A}{A_g}\cdot\frac{L}{L'}\cdot\frac{\cos\theta'}{\cos\theta} \tag{7}$$

The area ratio $\frac{A'_g}{A_g}$ captures the reduction in effective gas flow area due to liquid holdup, the path length ratio $\frac{L}{L'}$ accounts for changes in acoustic path geometry, and the angle ratio $\frac{\cos\theta'}{\cos\theta}$ reflects deviations in beam alignment or refraction. Together, these terms quantify the geometric and acoustic deviations between dry and wet gas conditions, forming the foundation for the correction models developed in this study [7].

3.3. USM over-reading parameters

The data collected served as input for developing the over-reading (OR) model, enabling prediction of the corrected flow rate for the ultrasonic flow meter (USM). In this study, a commercial Flexim clamp-on TT-USM was used to measure the total wet gas flow rate, while the liquid volume fraction (LVF) was calculated from the actual gas flow rate Q_g and liquid flow rate Q_l as expressed in Eq. (8):

$$LVF = 1 - GVF = \frac{\frac{Q_l}{Q_g}}{1 + \frac{Q_l}{Q_g}} \tag{8}$$

The Lockhart–Martinelli parameter (XLM), representing the ratio of liquid to gas mass flow rates and accounting for density differences between phases, is given in Eq. (8) [1,2].

Literature reports indicate that LVF and liquid film thickness at the ultrasonic wave reflection point are primary contributors to OR in USMs [7,16]. The over-reading phenomenon arises due to velocity and density disparities between the gas and liquid phases, leading to deviations in the measured gas flow rate [2]. Dimensional analysis, following Buckingham's Pi theorem [18], was applied to derive dimensionless parameters from multiphase flow continuity and momentum equations (Eqs. (9), 10). These parameters include the Froude number of gas (Fr_g), density ratio (DR), slip ratio (S), Weber number (We), and Reynolds numbers for gas (Re_g) and liquid (Re_l).

$$\frac{\partial(\alpha_g\rho_g\mathbf{v}_g)}{\partial \mathbf{t}} + \nabla\cdot(\alpha_g\rho_g\mathbf{v}_g\mathbf{v}_g) = -\alpha_g\nabla\mathbf{P} + \alpha_g\rho_g\mathbf{g} + \mathbf{F}_{g-l} \quad (9)$$

$$\frac{\partial(\alpha_l\rho_l\mathbf{v}_l)}{\partial t} + \nabla\cdot(\alpha_l\rho_l\mathbf{v}_l\mathbf{v}_l) = -\alpha_l\nabla P + \alpha_l\rho_l\mathbf{g} - \mathbf{F}_{g-l} \quad (10)$$

In agreement with Buckingham's Pi theorem, the over-reading correction model was developed using a set of Pi groups that preserve the same dimensionless relationships as the underlying physical variables. The relevant groups are: Group 1: Pipe diameter and inner pipe wall roughness (D, ε), Group 2: Viscosity and densities of gas and liquid (μ_g,ρ_g, μ_l, ρ_l), Group 3: Superficial velocity of gas and liquid (V_{SG}, V_{SL}), Group 4: Reynolds number of gas and liquid (R_{eg} , R_{el}) and Group 5: Gravitational and interfacial tension (g, σ_{gl}).

From Groups 1 and 2, key dimensionless parameters were obtained, including the Froude number of gas (Fr_g), the Lockhart–Martinelli parameter (X_{LM}), and the density ratio (DR). Applying dimensional analysis and scaling the gas void fraction (α) yielded a total of eight dimensionless parameters. These parameters form the basis of the over-reading model, reducing the original eleven independent variables to eight, in accordance with the Pi theorem, which states that the number of dimensionless parameters equals the number of variables minus the number of fundamental units (m, kg, s) [17].

The general functional form of the over-reading correction model is expressed in Eq. (17), while the Froude number Fr_g, used to characterise multiphase flow under gravitational effects, is calculated from Eq. (11):

$$Fr_g = \frac{V_{Sg}{}^*}{\sqrt{gD}}\sqrt{\frac{DR}{1-DR}}, \quad (11)$$

where V_{Sg}^* is the superficial gas velocity, g is gravitational acceleration, and D is the pipe diameter. The critical Froude number $Fr_g^* = \frac{V_{Sg}^*}{\sqrt{gD}}$ is approximately unity at the transition from stratified to dispersed flow [1]. Flows with $Fr_g > Fr_g^*$ are fully dispersed [2].

The density ratio is defined as:

$$DR = \frac{\rho_l}{\rho_g} \quad (12)$$

where ρ_l, ρ_g are the densities of liquid and gas, respectively.

To complete the development of the wet gas over-reading correction model for the ultrasonic flow meter, additional dimensionless variables were derived from the relevant parameters in Groups 3, 4, and 5. These include the slip ratio (S), defined as the ratio of superficial gas velocity to superficial liquid velocity, which is used in determining the kinematic viscosity of the wet gas, and the Weber number (We), which characterises the influence of inertial forces relative to surface tension and is employed to evaluate the gas–liquid interfacial tension (σ_{gl}), as given in Eq. (16). Furthermore, the Reynolds numbers for the gas and liquid phases (Re_g, Re_l) are determined to obtain the corresponding dynamic viscosities of gas (μ_g) *and liquid* (μ_l), as expressed in Eqs. (14) and (15).

$$S = \frac{V_{Sg}}{V_{Sl}} \quad (13)$$

$$Re_g = \frac{\rho_g V_{Sg} D}{\mu_g}, \quad (14)$$

$$\mathrm{Re}_l = \frac{\rho_l V_{Sl} D}{\mu_l}, \quad (15)$$

$$W_e = \frac{\rho_g V_{Sg}{}^2 D}{\sigma_{tp}}, \quad (16)$$

Here, $\mu_g$and μ_l are the dynamic viscosities of gas and liquid, V_{Sg} and V_{Sl} are the superficial velocities of gas and liquid, and σ_{tp} is the gas–liquid interfacial tension. The final set of influencing parameters for the regression-based USM over-reading model comprises: LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR. The general OR correction model is expressed as;

$$OR = \frac{Q_{tp}}{Q_g} = f(LVF,\ X_{LM}, S,\ Fr_g,\ DR, we,\ Re_g,\ Re_l) \quad (17)$$

where, Q_{tp} is the calculated or measured wet gas flow rate by USM, Q_g indicates the actual gas flow rate.

3.4. Experimental setup and data collection

The experimental campaign was conducted at the PSE Laboratory at Cranfield University, employing a fully instrumented multiphase flow loop specifically designed to replicate wet gas conditions. The selected working fluids comprised compressed air, serving as the gas phase, and tap water, representing the liquid phase. These fluids were chosen for their accessibility, safety, and well-established thermophysical properties, which are essential for accurately mimicking critical flow characteristics such as density contrast, phase slip, and interfacial behavior, as established in prior studies [1,2].

The flow loop featured a 50.8 mm (2-inch) stainless steel pipeline with a wall thickness of 3 mm and an extensive length exceeding 50 m, ensuring fully developed flow conditions. The ultrasonic flow meter utilized was a commercial Flexim clamp-on transit-time ultrasonic flow meter (TT-USM), configured in a V-path arrangement at a 45° angle to optimize measurement accuracy (Fluxus G800). The transducers were externally mounted using manufacturer-recommended coupling gel to enhance the acoustic signal. Furthermore, the pipe material and wall thickness parameters were meticulously configured in the meter's software to ensure precise acoustic path calibration. The TT-USM was factory-calibrated and subsequently verified against a reference Coriolis flowmeter, which possesses a certified uncertainty of ±0.25 %. This reference device was installed within the dry gas loop, providing reliable ground truth measurements of the actual gas flow rate.

The experimental matrix encompassed Gas Volume Fractions (GVF) ranging from 95 % to 100 %, with flow rates varying from 10 to 550 m^3/h. The temperature of the gas-liquid mixture was maintained between 20 °C and 25 °C, while the pressure varied from 1.12 to 3.03 bar. Flow regime transitions were visually confirmed using a sight glass, where stratified flow was observed at lower flow rates (<5 m^3/h) and mist flow dominated at higher flow rates (greater than 10 m^3/h). The Liquid Volume Fraction (LVF) was adjusted between 0.01 and 0.05, corresponding to GVFs of 95–99.99 %. Whilst in this test Re_l and Re_g varied ranging from (5270 to 25,400) and (1330 to 2800) respectively. An inline sight glass with a 50.8 mm diameter was installed in the metering section to facilitate monitoring of flow regimes in close proximity to the ultrasonic flow meter, thereby enabling real-time observation of flow behavior. The experimental procedure involved the following steps:

(a) The dry gas flow rate (Q_g) was measured using the Coriolis meter, while water was injected at controlled rates to achieve the desired LVF.
(b) The total wet gas flow rate (Q_{tp}) was measured using the TT-USM.
(c) The actual liquid flow rate (Q_l) was recorded using a calibrated liquid flowmeter.
(d) Over-reading (OR) was calculated as the ratio Q_{tp}/Q_g for each test condition.
(e) Dimensionless parameters including X_{LM}, S, Fr_g, DR, we, Re_g and Re_l were computed using measured fluid properties and flow conditions based on Eq. (11) through (16).

The schematic of the setup is provided in Fig. 2, and all relevant symbols and units are summarized in Appendix A, Table A1

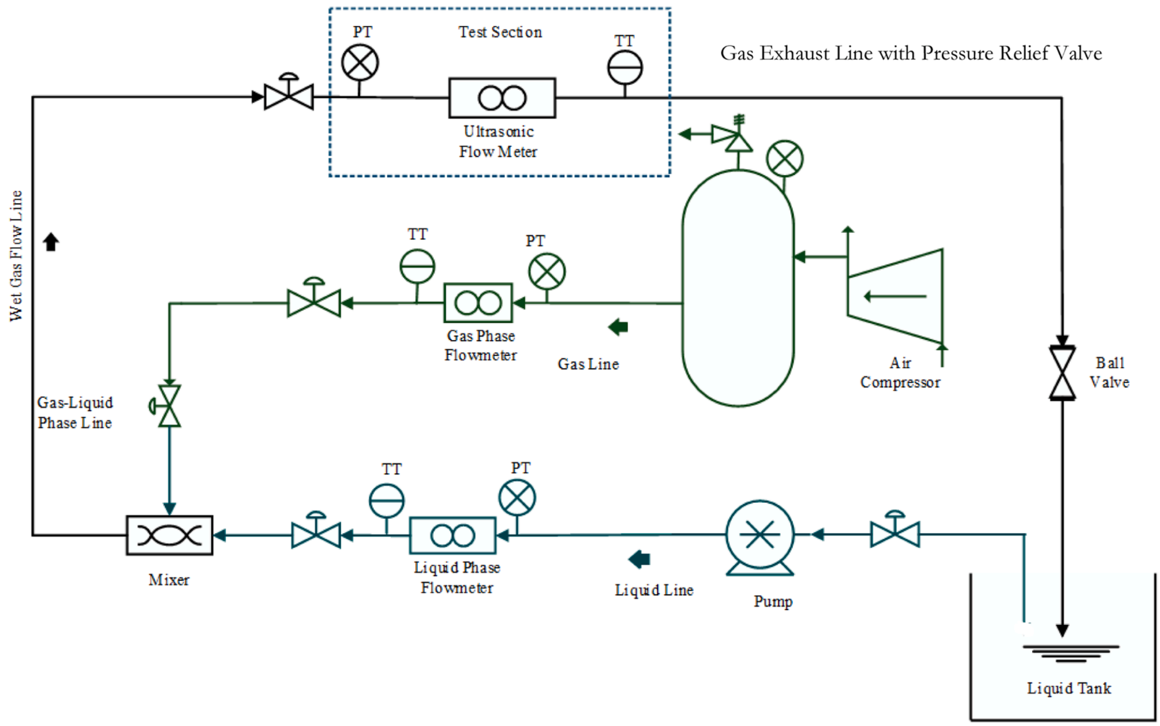

Fig. 2. Experimental setup of ultrasonic flow meter for wet gas measurement in a multiphase flow rig.

3.5. Modeling of USM over-reading correction

USM over-reading is defined as the positive deviation between the wet gas measured flow rate and the actual flow rate caused by the presence of liquid phases. The correction models aim to mitigate this deviation by accurately predicting the over-reading coefficient (OR). The experimental dataset was fitted to models using computational regression modelling techniques to develop and validate USM over-reading models. The selected modelling techniques included: Polynomial Regression (PLR), Random Forest Regression (RFR), Scikit Neural Network (SNN), Lasso regression (LR), Ridge Regression (RR), and Multilinear Regression (MLR).

3.5.1. Polynomial regression (PR)

The Polynomial Regression (PL) technique models the nonlinear relationship between the over-reading factor (OR) and the independent variables, namely LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR by fitting a polynomial equation to the experimental data. This approach captures complex interactions between predictors and provides flexibility in representing curved relationships. The modelling process involved transforming the independent variables into polynomial features and applying the least-squares method to fit the model. Then the model's performance is assessed using metrics like R-squared and Mean Squared Error (MSE). The general form of the polynomial regression model is expressed in Eq. (18)

$$\boldsymbol{OR} = \beta_0 + \beta_1 \boldsymbol{LVF} + \beta_2 \boldsymbol{X_{LM}}^2 + \beta_3 \boldsymbol{xFr_g}^3 + \beta_4 \boldsymbol{xS}^4 + \beta_5 \boldsymbol{xWe}^5 + \beta_6 \boldsymbol{xDR}^6 \cdots + \beta_7 (\boldsymbol{Re_g}/\boldsymbol{Re_l})^7 \tag{18}$$

Here; OR represents the dependent variable, LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR are the independent variable, $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ indicates the coefficients determined from the regression fit.

3.5.2. Random forest regression (RFR)

RFR is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and control overfitting. In this case, RFR is selected for its effectiveness in handling high-dimensional data and capturing complex interactions between the independent variables.LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR) and the over-reading factor OR. In the process of modelling OR, the experimental dataset is well prepared then the data is trained in multiple decision trees on different subsets of the data. The predictions were aggregated from all trees to get the final prediction. The model was evaluated using metrics like R-squared and Mean Absolute Error (MAE). The prediction $\widehat{OR}$ is the average of predictions from all individual trees presented in Eq. (19) below:

$$\widehat{\boldsymbol{OR}} = \frac{1}{\boldsymbol{N}} \sum_{i=1}^{\boldsymbol{N}} \boldsymbol{T_i} \left(\boldsymbol{LVF},\ \boldsymbol{X_{LM}},\ \boldsymbol{Fr_g},\ \boldsymbol{We},\ \boldsymbol{S},\ \boldsymbol{Re_g},\ \boldsymbol{Re_l},\ \boldsymbol{DR} \right) \tag{19}$$

where: N is the number of trees, $T_{\,i}(x)$ is the prediction from the i^{th} tree

3.5.3. Scikit neural network (SNN)

A Multi-Layer Perceptron (MLP) neural network consists of an input layer, multiple hidden layers, and an output layer. The neural network is trained using backpropagation and gradient descent to learn the complex patterns in the data and accurately predict OR. In the modeling process data are normalized and split it into training and testing sets. The model is configured by defining the architecture of the neural network (number of layers, neurons per layer) as illustrated in Fig. 3 below. The model's performance is assessed using metrics like R-squared and Root Mean Squared Error (RMSE). The fundamental equation of the output of a neuron is given by:

$$\boldsymbol{OR} = \boldsymbol{f}\left(\sum_{i=1}^{n} \boldsymbol{w_i x_i} + \boldsymbol{b} \right) \tag{20}$$

where: f is the activation function, w_i are the weights, x_i are the inputs i.e.; LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR and b is the bias term.

3.5.4. Multilinear regression (MLR)

The MLR technique models the linear relationship between OR and the independent variables LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR by fitting

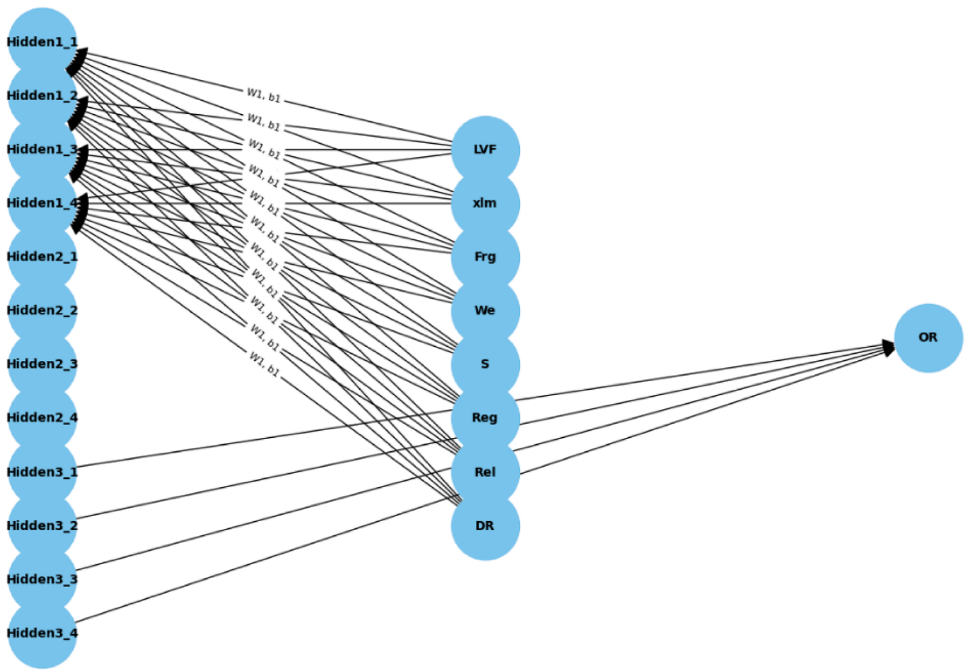

Fig. 3. A multi-layer perceptron (MLP) neural network for over-reading modelling.

a linear equation to the data. Multilinear regression provides a straightforward and interpretable approach to modelling over-reading, although it may struggle with capturing nonlinearities and interactions between variables. In this study, the OR was regarded as the function of LVF.(X_{LM}. Fr_g. We. S. Re_l/ Re_g) other than the general linear regression function. The MLR model function is expressed in Eq. (21) below.

$$\boldsymbol{ln}\ (\boldsymbol{OR}) = \beta_0 + \beta_1 \cdot \boldsymbol{ln}\ (\boldsymbol{LVF}) + \beta_2 \cdot \boldsymbol{ln}\ (\boldsymbol{X_{LM}}) + \beta_3 \cdot \boldsymbol{ln}\ (\boldsymbol{Fr_g}) + \beta_4 \cdot \boldsymbol{ln}\ (\boldsymbol{We}) + \beta_5 . \boldsymbol{ln}\ (\boldsymbol{S}) + \beta_6 \cdot \boldsymbol{ln}\ (\boldsymbol{Re_l} /\ \boldsymbol{Re_g}) \tag{21}$$

Where, β_0 is the interception and β_1- β_7 are the coefficients for each independent variable.

3.5.5. SciPy optimization (SO)

In this method, the SciPy optimization was performed to determine the constant values of the general over-reading equation derived from the general Eq. (22), which was then fitted to the experimental dataset. The parameters in this case were reduced to allow the iteration to converge based on the selected samples. Eq. (23) was comprehended as the result of the optimization

$$\boldsymbol{OR} = 1 + \boldsymbol{C} \cdot (\boldsymbol{LVF})^{\boldsymbol{n}} \cdot (\boldsymbol{X_{LM}})^{\boldsymbol{m}} \cdot (\boldsymbol{Fr_g})^{\boldsymbol{a}} \cdot (\boldsymbol{We})^{\boldsymbol{b}} \cdot \left(\frac{\boldsymbol{Rel}}{\boldsymbol{Reg}}\right)^{\boldsymbol{c}} \tag{22}$$

$$\boldsymbol{OR} = 1 + 1.7169 \cdot (\boldsymbol{LVF})^{1.2860} \cdot (\boldsymbol{X_{LM}})^{2.7452 \times 10^{-10}} \cdot (\boldsymbol{Fr_g})^{2.3106 \times 10^{-6}} \cdot (\boldsymbol{We})^{0.1948} \cdot \left(\frac{\boldsymbol{Re_l}}{\boldsymbol{Re_g}}\right)^{0.5962} \tag{23}$$

3.6. Model training, cross-validation, generalization and reliability

A structured model training and validation framework was developed to ensure predictive reliability and statistical rigor in correcting ultrasonic flow meter over-reading (OR) under mist flow conditions. The experimental dataset used in this study comprised 1400 random samples captured at a logging frequency of 60 readings per minute. Each sample reflected distinct wet gas behavior characterized by Gas Volume Fractions (GVF) between 95 % and 100 %, suitable for mist regime analysis. To preserve statistical balance and allow unbiased model evaluation, the data were randomly stratified into 70 % training, 20 % validation, and 10 % testing subsets, in accordance with standard machine learning practice [18].

Model development focused on predicting the over-reading correction factor (OR), defined as the ratio between the ultrasonic-measured wet gas flow rate (Q_{tp}) and the actual dry gas flow rate (Q_g), as per Eq. (17). The input variables included critical flow descriptors and dimensionless parameters: Liquid Volume Fraction (LVF), Lockhart–Martinelli parameter (X_{LM}), Froude number (Fr_g), Weber number (We), slip ratio (S), density ratio (DR), and Reynolds numbers for gas and liquid phases (Re_g, Re_l). These features were selected for their proven significance in multiphase flow modeling and their scalability across varying flow regimes [1].

To assess model generalization and mitigate overfitting risks, the study employed a Repeated Holdout Cross-Validation strategy. This approach involved iterative resampling and performance averaging over randomized splits of the training and validation subsets, offering computational efficiency and facilitating iterative model refinement, particularly critical for high-dimensional regressors like Random Forest Regression (RFR) and Neural Networks [19]. Although K-Fold Cross-Validation is widely recognized, repeated holdout was preferable here, given the moderate dataset size and the model complexity involved.

Each model was trained and evaluated using performance metrics including Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), computed over both the validation and held-out testing sets. To standardize feature scales and improve convergence rates during training, normalization was applied to all input variables, ensuring consistent pre-processing across cross-validation iterations.

Hyperparameter tuning for each model was performed using a Grid Search strategy tailored to the unique architectural requirements of each algorithm. For Random Forest Regression (RFR), key parameters such as the number of trees (n_estimators), maximum tree depth (max_depth), minimum number of samples required for a split (min_samples_split), and the number of features considered per split (max_features) were optimized to balance model complexity and overfitting control. The Multilinear Regression (MLR) model was refined through the selection of appropriate regularization schemes, application of log-transformed predictors, and the inclusion of interaction terms to enhance the linear approximation of nonlinear behavior. Polynomial Regression (PR) required calibration of the polynomial degree and feature expansion scope to capture higher-order relationships between variables. For the Scikit-Learn Neural Network (SNN), hyperparameter tuning encompassed adjustments to the network architecture, including the number of hidden layers, neuron count per layer, learning rate, and activation functions. Lastly, the SciPy Optimization (SO) model was fine-tuned by setting appropriate convergence thresholds and bounding functional parameters to ensure stable curve fitting and minimization of residuals. This targeted approach to hyperparameter optimization enhanced model flexibility, predictive accuracy, and generalization capability across the mist flow dataset.

4. Results

4.1. Performance evaluation of USM over-reading correction models

The performance metrics of the empirical over-reading correction models were derived from a controlled experimental dataset and evaluated using a suite of statistical regression techniques. The Random Forest Regression model (RFR) was fitted using ensemble decision tree methods and represents one of the most robust correction models in this study. The SciPy Optimization model (SO) applied nonlinear curve fitting based on least squares techniques, calibrated in accordance with Eq. (22). The Multilinear Regression model (MLR), formulated using log-transformed inputs and referenced in Eq. (21), was particularly effective in addressing linear dependencies. Polynomial Regression (PR), Ridge Regression (RR), and Lasso Regression (LR) models were also developed to capture nonlinearity and regularization effects. The Scikit-Learn Neural Network model (SNN), structured as a multi-layer perceptron, offered a flexible architecture for representing complex data patterns and flow conditions. All seven models successfully predicted over-readings in the ultrasonic flow meter under varying wet gas conditions. Whereas, the neural network model (SNN) demonstrated a loss metric of 0.7356 and an R^2 value of 4.56×10^{-29}, reflecting its iterative training performance and limitations in over-reading estimation for the experimental dataset.

A comparative evaluation of the regression models is illustrated in Fig. 4, which presents R^2 and adjusted R^2 values for MLR, RFR, PR, RR, and LR. Models MLR, RFR, and PR exhibited notably high R^2 values, indicating strong fits to the empirical data. Their adjusted R^2 scores further affirmed their reliability and generalization capability. In contrast, RR and LR yielded substantially lower R^2 and adjusted R^2 values, signifying weak fits and reduced model robustness. These findings suggest that MLR, RFR, and PR are the most reliable modeling techniques, primarily due to their ability to accommodate both linear and nonlinear relationships, handle outliers effectively, and adapt to complex datasets. Conversely, the diminished performance of RR and LR may stem from constraints introduced by their regularization methods, which are less suited to highly nonlinear flows or noisy data.

Cross-validation results presented in Table 1, based on MSE, MAE, and RMSE metrics, reinforce these observations. Models RFR, MLR, and PR achieved superior scores across all validation criteria. The high performance of MLR can be attributed to its effectiveness in capturing linear trends with clear structural assumptions. RFR excelled due to its robustness against outliers and capacity to model nonlinear interactions, while PR offered enhanced flexibility in fitting curved relationships within the data. By contrast, RR and LR were comparatively limited by

Table 1
The summarized cross-validation analysis.

Statistical Analysis Model	R^2	MSE	RMSE	MAE	Durbin Watson
Multilinear Regression (MLR)	0.9991	0.000293	0.01712	0.0082	1.7238
Polynomial Regression (PR)	0.9961	0.000218	0.01475	0.0104	1.6570
Random Forest Regression (RFR)	0.9963	0.000037	0.00608	0.0026	2.1440
Ridge Regression (RR)	0.3990	0.035704	0.18895	0.1324	0.2030
Lasso Regression (LR)	0.1198	0.522830	0.22866	0.1680	0.0434
Scikit-Lean Neural Network (SNN)	–	–	–	–	–
SciPy Optimization (OS)	–	–	–	–	–

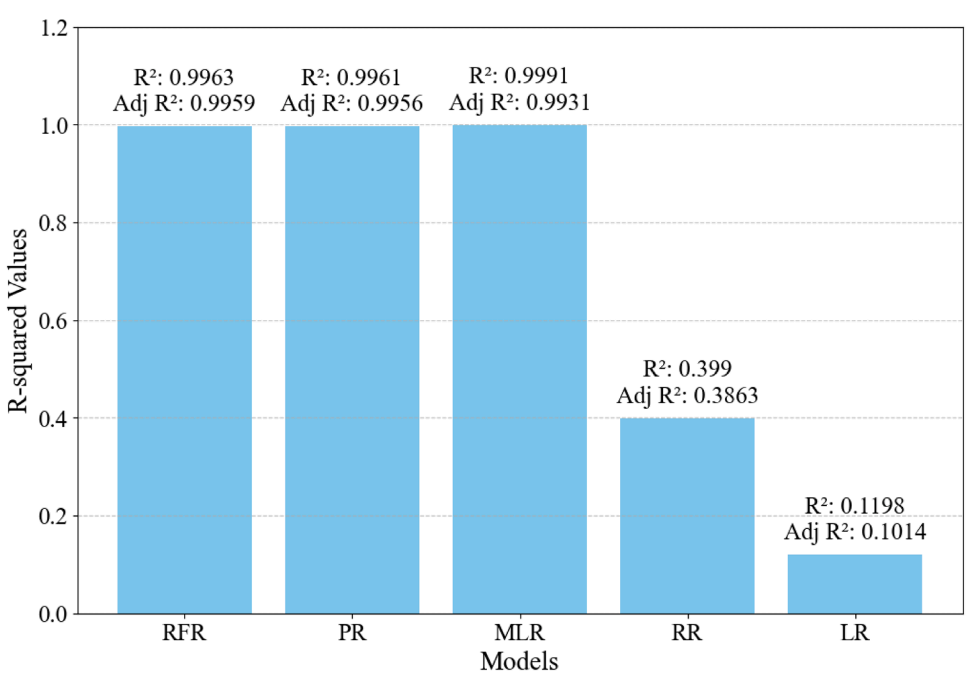

Fig. 4. R^2 and Adjusted R^2 to measure performance of regression models.

their regularization mechanisms, which can suppress model adaptability in the presence of substantial nonlinearity or experimental variability.

Furthermore,

Table 1 presents a comparative analysis of Durbin Watson (DW) statistics for models RFR, MLR, PR, RR, and LR, providing insight into the degree of autocorrelation present in their residuals. The RFR displays minimal negative autocorrelation, indicating a robust model fit and well-behaved residuals. Similarly, the PR and MLR exhibit slight positive autocorrelation, which remains within acceptable limits and reflects relatively strong predictive performance. In contrast, RR and LR demonstrate pronounced positive autocorrelation, suggesting that their residuals are less independent and pointing to potential issues in model specification or structural inadequacy. These findings support the conclusion that RFR, PR, and MLR offer superior statistical integrity in capturing the correction factor behavior, while RR and LR may require further refinement to mitigate autocorrelation effects. It is worth noting that DW statistics ranging between 1.5 and 2.5 are generally considered acceptable, signifying that autocorrelation is unlikely to compromise model reliability.

4.2. Residual diagnostics and model reliability

Residual analysis was conducted to further evaluate the reliability and performance of the over-reading correction models. Using the test

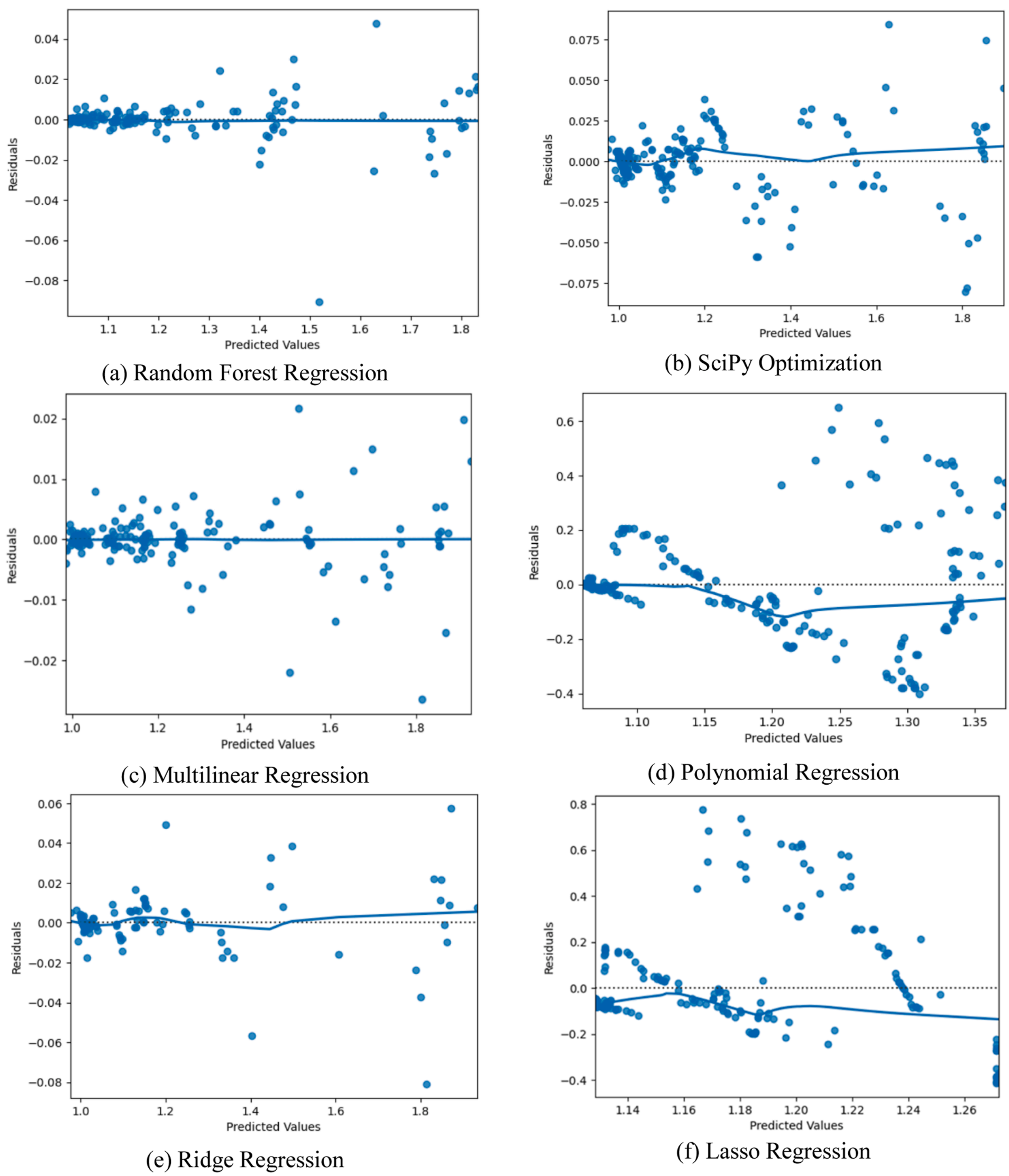

Fig. 5. Residual analysis plotting residuals vs predicted values.

dataset, which included the predicted over-reading values (OR), Liquid Volume Fraction (LVF), and relevant dimensionless parameters, residual plots were generated to assess model behavior and diagnostic validity. Fig. 5 illustrates the residuals plotted against the predicted OR values for each regression model, with Fig. 5(a), (c), and (e) highlighting the performance of RFR, MLR and RR models, respectively.

The distribution of residuals for RFR, MLR, and RR models revealed desirable statistical characteristics: randomness, symmetry, and lack of discernible patterns. These properties suggest that the models sufficiently capture the underlying relationships between predictor variables and the correction factor, without systemic bias. Moreover, the residuals' adherence to normality and their independence from predicted values further reinforce the appropriateness of these models for wet gas over-reading estimation. The analysis confirms that RFR, MLR, and RR provide reliable regression fits, while other models may require additional validation to ensure similar robustness in their predictive structure.

4.3. Feature importance analysis in over-reading prediction

Feature importance analysis is a crucial aspect in evaluating the performance of predictive models. In this study, the Random Forest Regression (RFR) model uniquely succeeded in predicting the future importance of input features, with values of 0.5051, 0.4915, 0.00015, 0.0011, 0.0004, and 0.0003 for X_{LM}, LVF, Fr_g, Re_l/Re_g, S and We, respectively. These results highlighted the significant influence of the Lockhart-Martinelli parameter (X_{LM}) and liquid volume fraction (LVF) in the modeling of ultrasonic flow meter (USM) over-reading (OR). In contrast, the remaining parameters, including DR, S, We, Re_l, Re_g and Fr_g demonstrated comparatively minimal contributions to the model's predictive performance. While Random Forest Regression (RFR) is commonly recognized for its predictive accuracy, particularly in nonlinear and multivariate datasets, its interpretability is equally critical when considering industrial deployment. In this study, feature importance analysis revealed that LVF and the X_{LM} contributed approximately 99 % of the total model influence, underscoring their dominant role in ultrasonic flow meter over-reading correction. This interpretability offers several practical benefits.

First, understanding feature rankings enables targeted sensor integration. For example, recognizing that LVF is the most influential predictor suggests that installing real-time phase fraction sensors can substantially improve model input fidelity and correction accuracy. Second, feature attribution guides operational diagnostics: by quantifying how dimensionless parameters (e.g., DR, S, We, Re_l, Re_g and Fr_g) predictions, field engineers can interpret deviations in flow rate as linked to specific flow dynamics, which is valuable in troubleshooting and optimizing process conditions.

In industrial environments such as oil and gas, chemical processing, and reservoir management, models must not only produce accurate results but also offer insight into measurement behaviors. The ability of RFR to attribute prediction variance to concrete physical variables ensures it can be embedded within decision-support systems. For instance, its feature importance framework can be leveraged to design adaptive flow metering protocols, where operators prioritize calibration or maintenance based on dominant predictors.

Moreover, the interpretability of RFR in this study supports regulatory compliance and model auditing, particularly when deployed in custody transfer systems. By offering a transparent breakdown of model decision pathways, stakeholders can validate the model's reliability under varied flow regimes, mist, dispersed, and transitional, thus enhancing trust in automated metering systems.

4.4. Experimental validation of correction models

The comparative analysis is discussed in this section based on the statistical analysis of the developed USMOR models. Whereas, in this study, a test dataset was selected as detailed in Section 3.5 and statistically analyzed to test the performance of USM over-reading models. Consequently, a comparative analysis is performed to govern the selection of the precise modelling techniques and validate the accuracy of the respective over-reading correction models.

4.4.1. Calibration over-reading vs predicted over-reading

In this scenario, all ML and regression predicted over-reading models were compared with the experimentally measured over-reading correction OR (also referred to as calibration OR) as calculated from Eq. (18), which defines OR as the ratio of the ultrasonic-measured wet gas flow rate Q_{tp} to the actual dry gas flow rate Q_g.

Fig. 6 below signifies the 3D surface diagram that presents the correlation between the predicted ultrasonic meter over-reading models (RFR, SO, MLR, PR, RR, LR and SNN) and the experimentally measured over-reading values (OR). These visual comparisons assess each model's predictive accuracy against empirical observations across mist flow conditions. Notably, the models developed using RFR and MLR demonstrated the highest fidelity, with their corrected estimates falling within $a \pm 5$ % absolute error margin relative to experimental over-reading measurements (see Fig. 7). This performance highlights their reliability and alignment with empirical trends under varying wet gas conditions. From a theoretical standpoint, the strengths of Multilinear Regression and Random Forest Regression, such as their ability to accommodate multiple predictors, mitigate multicollinearity, prevent overfitting, and yield interpretable outcomes, render them particularly well-suited for modeling over-reading correction factor in ultrasonic flow meters (USMs) under complex multiphase scenarios [8].

Moreover, RFR demonstrate the ability to capture nonlinear relationships, its flexibility in OR modeling intricate interactions, and its robustness against noisy datasets, which contribute to its enhanced predictive power and fit quality [15]. These attributes support its consistent performance across test cases and confirm its utility for accurate correction modeling. Taken together, the results affirm that RFR and MLR offer compelling solutions for estimating over-reading correction factors in wet gas measurement systems, achieving performance margins within acceptable error thresholds.

4.4.2. Actual gas flow rate vs. corrected wet gas flow rate

A comparative analysis was undertaken to validate the predictive accuracy and reliability of each over-reading correction model by evaluating the correlation between the actual gas flow rate (Q_g) and the corrected flow rate (Q_{corr}) produced by the respective modeling techniques (RFR, SO, MLR, PR, RR, LR and SNN). These ML and regression-based models derived across the different modeling techniques were assessed for their ability to replicate calibration flow conditions. Fig. 8 depict the comparison of actual gas flow rate measured by USM (Q_g) against the corrected flow rates (Q_{corr}) by OR of multiple models. Within the operational flow range (10–550m^3/h), most approaches demonstrated acceptable correlation to the experimental benchmark. However, the predictions generated by RFR, MLR and PR showed particularly strong alignment with the actual gas flow rate. Their output values consistently matched empirical measurements, reinforcing their suitability for compensating for over-reading in ultrasonic wet gas measurement.

Further insights are illustrated in Fig. 9, which highlights the top-performing models in this study. Specifically, the machine learning approach using Random Forest and the regression model based on Multilinear techniques achieved superior correlation accuracy. While Random Forest maintained relative absolute errors within ±10 %, Multilinear Regression delivered enhanced precision with a relative error within ±5 % of the true gas flow rate. These findings emphasize the strength of these methods in translating ultrasonic sensor data into accurate flow estimates, making them reliable candidates for real-time

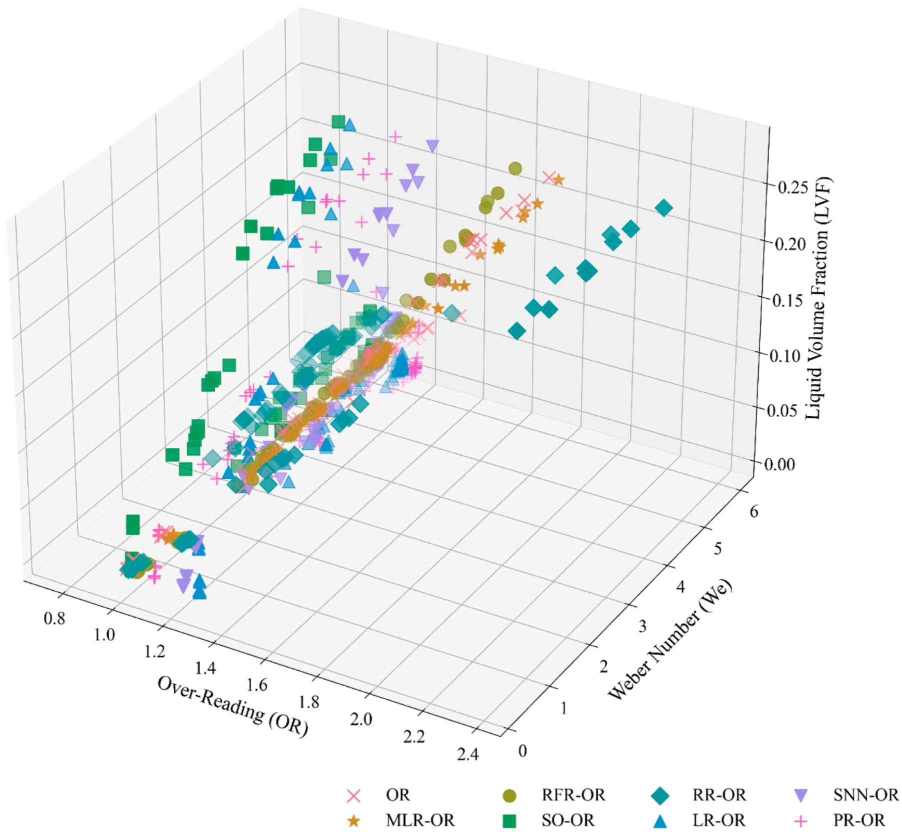

Fig. 6. Correlation of over-reading correction models with the experimental over-reading correction factor under the influence of LVF and We.

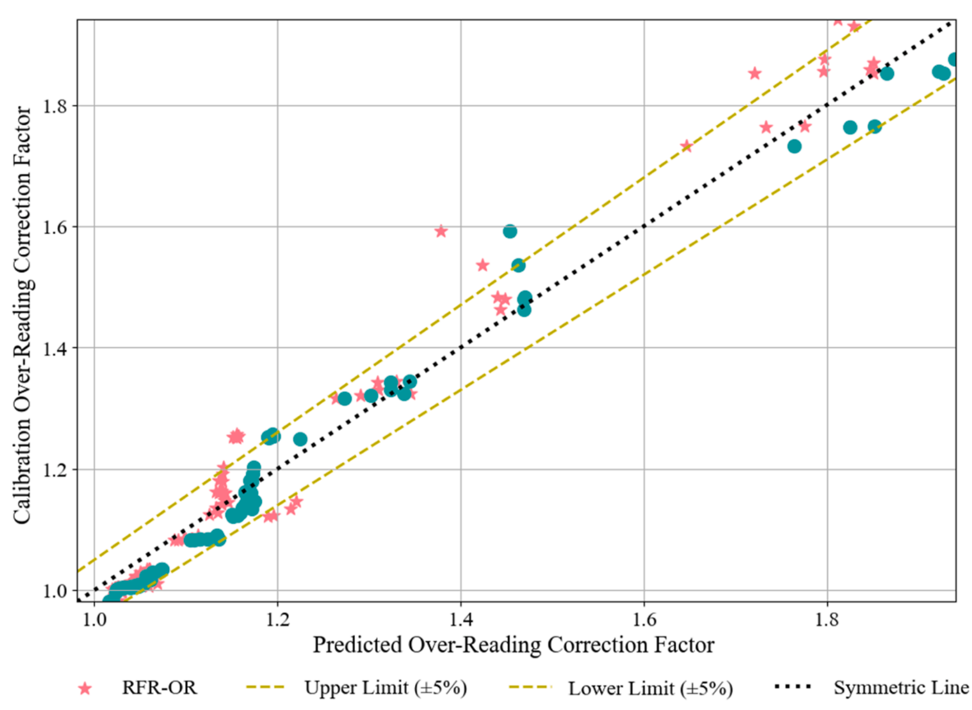

Fig. 7. The correlation of experimental over-reading OR against predicted USM over-reading corrections models by MLR and RFR.

correction modeling in complex multiphase environments.

4.3.3. Relative error analysis across models

In this scenario, all regression models were evaluated by determining the relative error as a measure of their robustness in predicting the over-reading factor. Provided that over-reading of the flow rate by USM may be influenced by different factors, in this case relative error of each model was analyzed while reflecting all influencing predictors, which

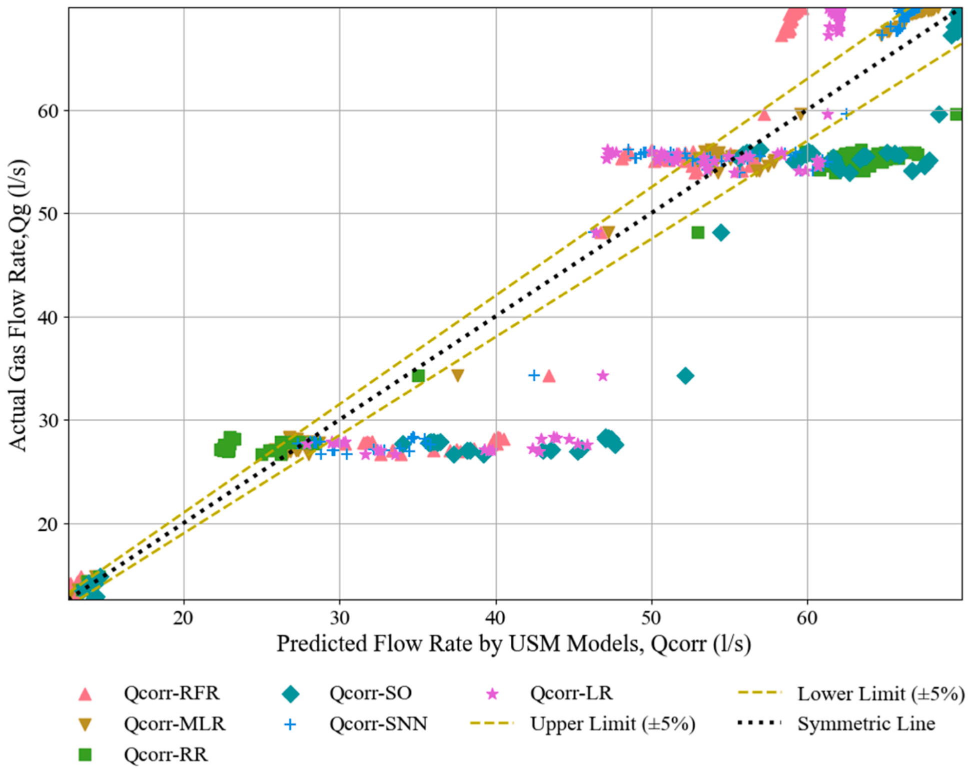

Fig. 8. Correlation of the corrected gas flowrate determined by USM over-reading models with the actual gas flowrate.

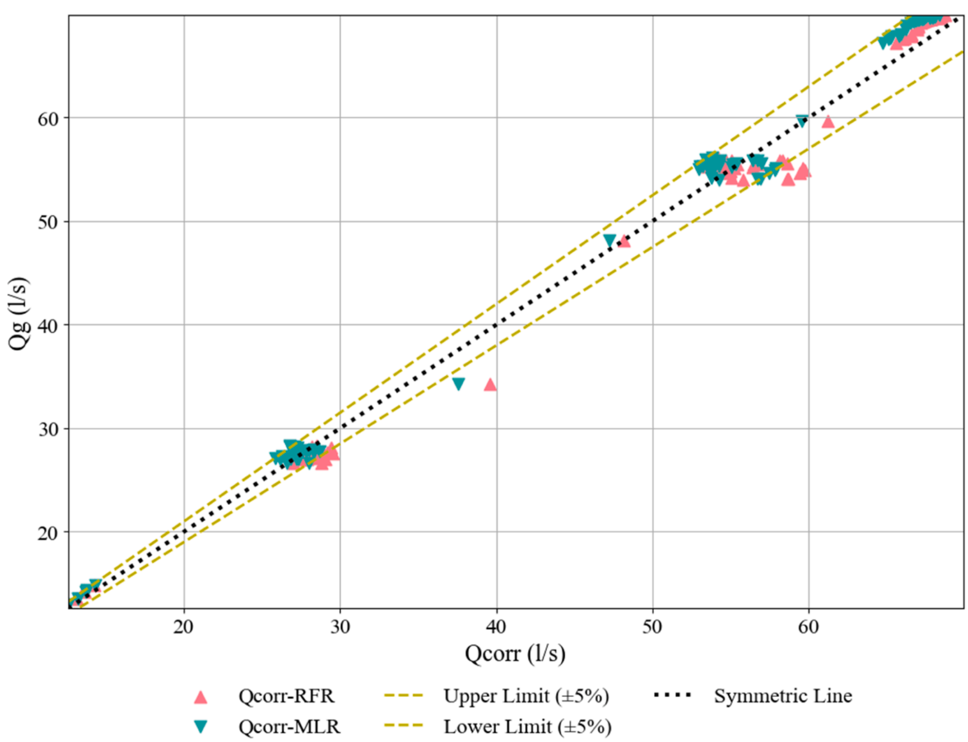

Fig. 9. Correlation of the corrected wet gas flow rate determined by USM over-reading models (MLR and RFR) with the actual gas flow rate.

include LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR as stated in Eq. (17). The OR error can be calculated as the difference between the corrected flow rate and the actual gas flow rate while considering the influencing factors, as suggested by van Putten et al. [1], that the liquid holds to be the critical predictor of relative error as in Eq. (24) below.

$$\varepsilon \mathbf{OR} = \mathbf{OR}\alpha_g - 1 \tag{24}$$

Where εOR denotes over-reading correction error, OR represents the over-reading correction factor, and αg signifies the void fraction.

Fig. 10 illustrates the relative absolute error (RAE) exhibited by each

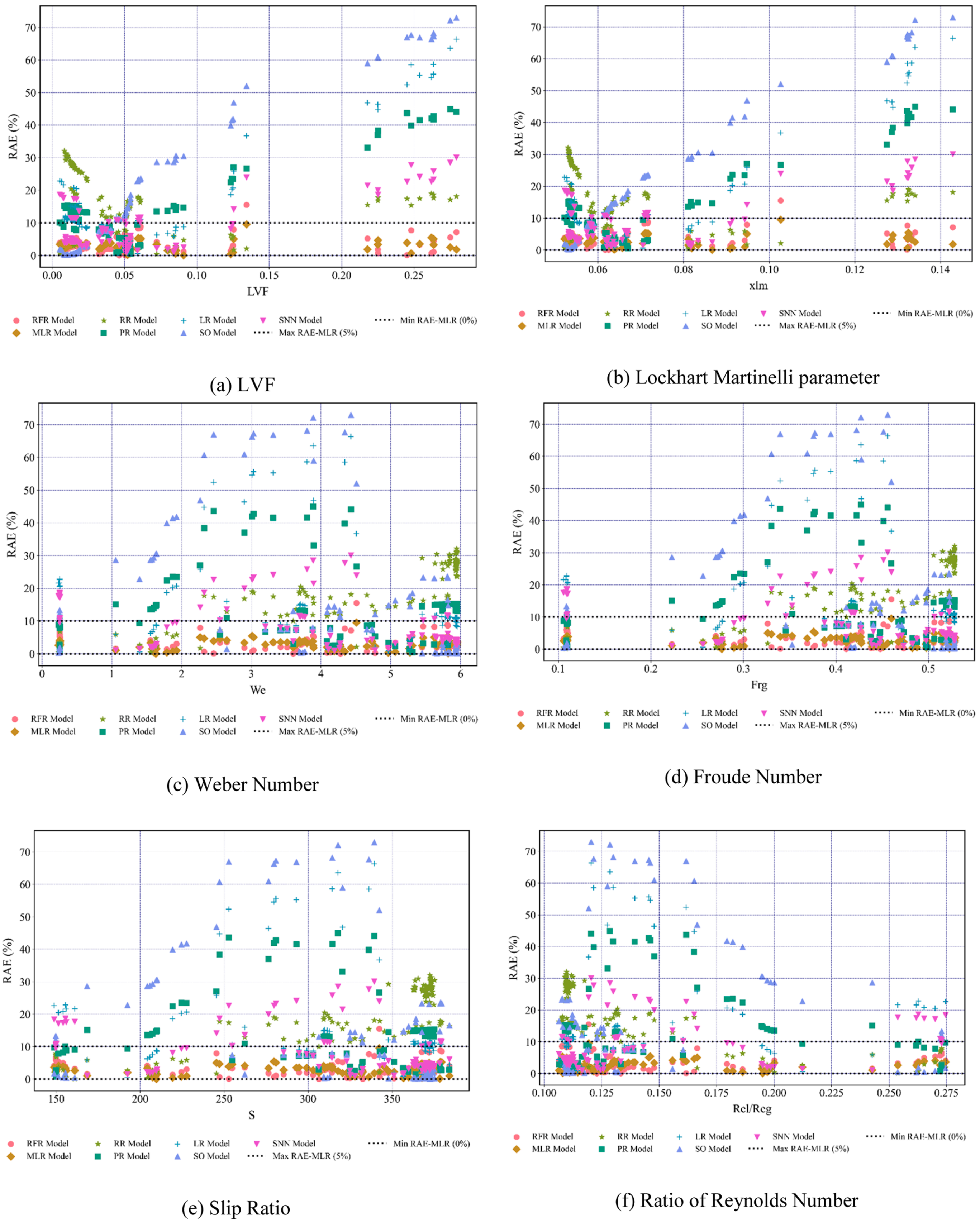

Fig. 10. Relative error for multiple USM Over-reading models as related to the influencing parameters.

regression model in response to the liquid volume fraction and various selected dimensionless parameters, including; X_{LM}, Fr_g, We, S, Re_g, Re_l and DR. Multiple models assessed exhibited varying sensitivities to these parameters. These data-based results signify that RFR, MLR, and RR consistently achieve lower relative errors in estimating corrected flow rates compared to the other modeling approaches. Their stability across diverse flow conditions indicates superior generalization capabilities, particularly in managing the complex dynamics of wet gas systems. In contrast, models such as LR, PR, and SNN demonstrate larger deviations, suggesting reduced reliability when confronted with the multidimensional variability of the underlying flow parameters.

Fig. 11 further validates these findings through scatter plots that highlight the robustness of RFR and MLR with respect to the X_{LM} parameter. Over the test dataset, both models maintain correction accuracy within a maximum relative error margin of ±10 %. Specifically, the average RAE for MLR is evaluated as 3.02 %, while RFR achieves a closely comparable figure of 3.20 %, as analyzed in Fig. 11 below. These low error magnitudes underscore the effectiveness of both techniques, MLR for its linear interpretability and RFR for its nonlinear adaptability, in accurately estimating over-reading correction factors for ultrasonic flow meters in wet gas applications.

4.4.4. *Comparison with existing DP over-reading models*

Comparative analysis of USM single-phase correlation models with traditional DP meter (including Murdock [20], Chisholm [21], Homogeneous, and Smith and Leang model [22]) has been carried out using experimental data, revealing that the USMOR models demonstrate superior performance in wet gas measurement applications. Fig. 12 highlights the correlation between the predicted flow rate derived from existing differential pressure (DP) over-reading models, the MLR-USM over-reading model, and experimental over-reading (OR). The predicted flow rate determined by the USM over-reading model exhibits a stronger positive correlation with the calibration flow rate compared to the predictions made by models such as Homogeneous, De Leeuw, Chisholm, Smith & Leang, and Murdock, as implied in Fig. 12.

The inferiority of DP over-reading models in comparison to the USM over-reading model can be attributed to their reliance on physical modeling rather than data-driven approaches. Additionally, these models are constrained by their dependence on the Lockhart-Martinelli parameter, which is derived solely from mass flow rates and the densities of the flow phases. Other limitations associated with conventional DP over-reading models stem from their measurement methodologies. These methods tend to promote flow pressure drops and swirl or eddy formations due to their intrusive nature, which disrupts flow fields and consequently increases measurement uncertainties [23,16]. Moreover, the integration of radioactive gamma-ray techniques for density and composition measurement raises significant health and safety concerns [23]. DP sensors also face challenges in accurately capturing critical phase distribution parameters, such as void fractions and film thicknesses [2]. Furthermore, over-reading correction models often lack experimental validation and obtaining the initial liquid flow rate required for gas-phase flow rate calculations remains practically unfeasible in industrial applications [2]. These findings further reinforce the suitability of Multilinear Regression techniques and machine learning approaches, such as Random Forest Regression, for accurately modeling USM over-reading corrections in wet gas measurement applications.

5. Discussion

The above comparative analysis results established that RFR and MLR achieved the best performance in modelling the ultrasonic flow meter over-reading correction factor (USMOR) for wet gas measurement in the mist flow regime. The superior performance of these two regression techniques in predicting over-reading using the experimental dataset was attributed to several key factors. Firstly, the multilinear regression model excels in handling multiple predictors simultaneously, allowing it to capture the combined effect of various factors on the dependent variable. This capability is crucial in complex datasets where multiple variables interact to influence the outcome [5]. In this case, a complex dataset of seven parameters (LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR) provided in Eq. (17). attributes the modelling of USMOR. Furthermore, the interpretability of the multilinear regression model is a significant advantage, as it provides clear coefficients for each predictor; (LVF, X_{LM}, Fr_g, We, S, Re_g, Re_l and DR), making it easy to understand the influence of each variable on the over-reading model as illustrated in

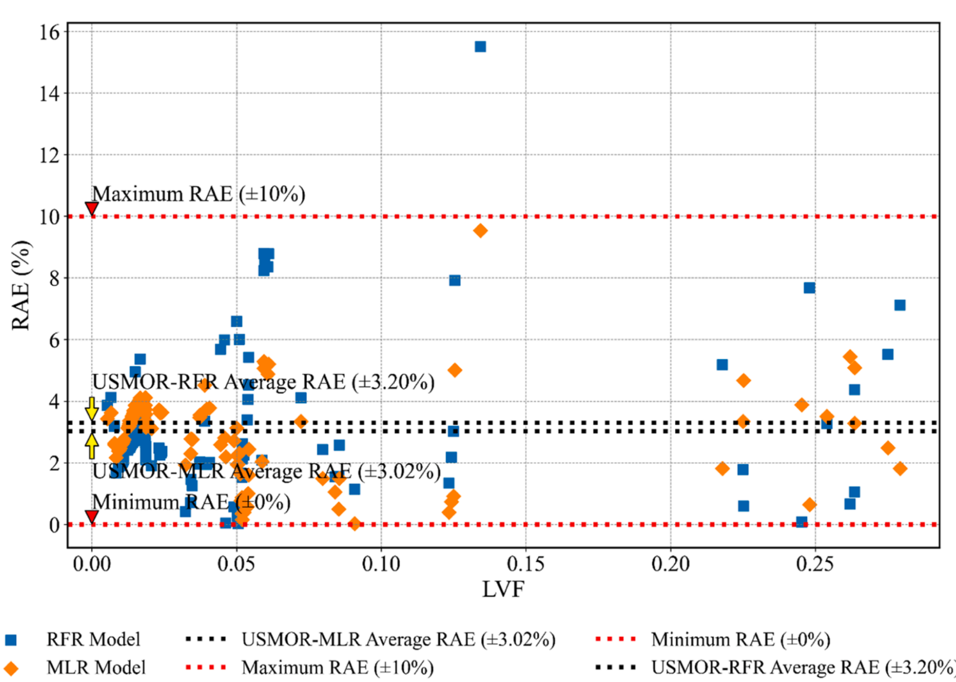

Fig. 11. Relative error of over-reading correction model fitted using random forest regression and multiple regression in respect to Lockhart-Martinelli parameter (X_{LM}).

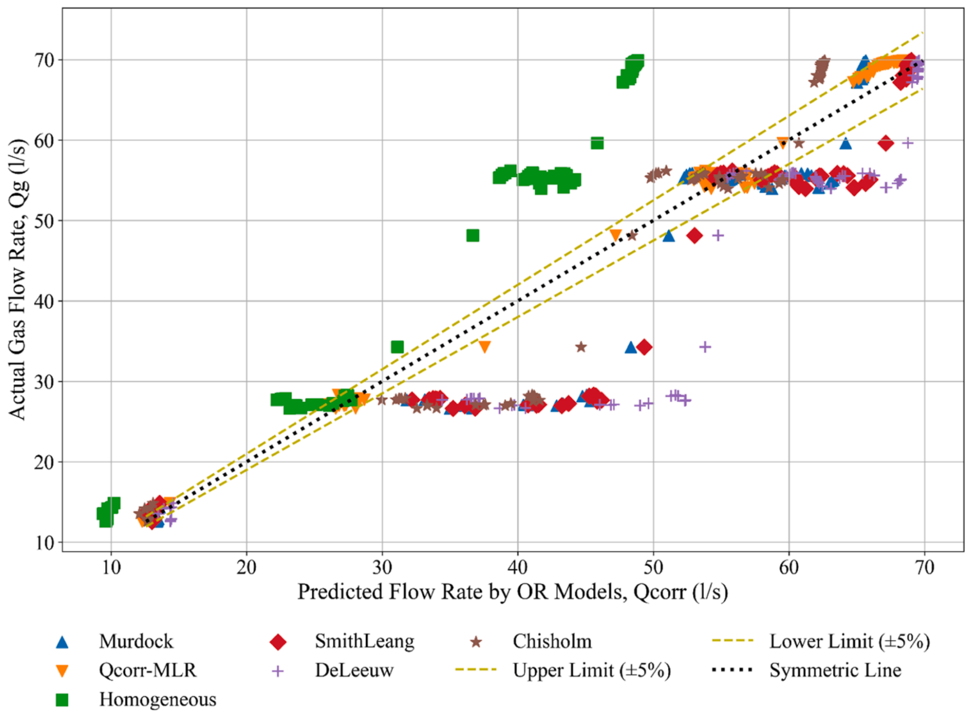

Fig. 12. Correlation plot of actual flow rate with predicted flow rate by multilinear regression USM over-reading (USM3) and existing DP over-reading correction.

Fig. 10 above. MLR also offered flexibility in including interaction terms and polynomial terms, further enhancing its ability to model complex relationships [13], as OR in Fig. 7, which was modelled as a function of multiple parameters instead of linear. As also reflected in [13], if the assumptions of linearity, independence, and homoscedasticity are met, multilinear regression can provide efficient and unbiased estimates, contributing to its strong performance. While Random Forest Regression (RFR) is often praised for its predictive performance in high-dimensional and nonlinear datasets, its interpretability is particularly valuable in industrial applications where model transparency is essential for diagnostics, auditing, and real-time control. In this study, feature importance analysis revealed that LVF and X_{LM}, accounted for approximately 99 % of the model total influence, indicating a physically meaningful correlation with the over-reading correction factor.

In industrial application, this interpretability offers substantial utility in the oil and gas sector, where real-time measurement accuracy is critical. For instance, understanding that LVF is the dominant predictor informs instrumentation engineers about the value of deploying inline phase fraction sensors to enhance flow meter calibration. Moreover, attributing prediction variance to specific flow descriptors enables targeted maintenance and calibration protocols based on dominant parameter sensitivity. Such feature-driven interpretability supports adaptive metering strategies, as discussed by Mofunlewi et al. [10] who emphasized the robustness of RFR in predicting oilfield flow rates, particularly under conditions of dimensional complexity and phase asymmetry. Additionally, Jiang et al. [11] demonstrated the practical viability of RFR in comparison to other machine learning models for gas-liquid metering, noting its favorable performance when generalizing to industrial datasets. The present study extends those findings, showing that RFR not only achieves low RAE (3.02 %) but also offers transparent and explainable predictions, essential for deployment in allocation metering and regulatory environments. By embedding RFR into metering systems with documented variable sensitivity, engineers can confidently interpret outputs and justify measurement decisions across various operating regimes

Conversely, the RFR algorithm demonstrated outstanding performance in modeling the USMOR based on experimental data attributed to its ability to capture complex nonlinear relationships between independent variables such as LVF and dimensionless flow parameters and the dependent variable, the over-reading factor (OR), as documented in prior studies [5,6]. In wet gas measurement, the interdependencies among variables are inherently nonlinear and multidimensional. While Multilinear Regression (MLR) traditionally offers linear modeling capabilities, its extension through higher-order polynomial terms enabled it to capture underlying patterns with increased fidelity, resulting in improved USMOR estimations. The capacity to model intricate nonlinear behaviors is essential for flow diagnostic applications, as reinforced by recent findings that underscore the limitations of linear assumptions in ultrasonic flow metering [7].

RFR offers greater flexibility compared to linear regression by allowing the inclusion of interaction terms and higher-order terms [6]. This flexibility enables the model to fit a wider range of data patterns, making it more adaptable to the complexities of the experimental data. RFR also excels in handling complex interactions between variables [7]. In the context of ultrasonic flow meter over-reading correction, multiple factors amounting to dimensionless numbers in Eq. (11) through (16), not limited to flow rates, pressure, velocities, pipe diameter, densities, and temperature, may interact in non-linear ways. However, RFR managed to effectively capture these interactions, leading to more reliable predictions of USMOR, and hence ensuring robustness of models as presented in Fig. 9 above. Moreover, as highlighted in Section 4.3 above, about feature importance analysis in RFR, the Lockhart Martinelli parameter (X_{LM}) and liquid volume fraction (LVF) were contributed to the USMOR model by 50 % and 49 %, respectively, while the remaining parameter contributed 1 %.

Notably, the Random Forest Regression (RFR) model demonstrated a remarkable ability to handle heterogeneous data types, thereby enhancing the applicability and reliability of the over-reading correction framework in mist flow regimes. This versatility underscores the model's generalized performance across the wet gas flow spectrum, particularly when combined with Multilinear Regression (MLR), which further supports consistent prediction under fully developed mist flow conditions. Together, these data-driven approaches exhibit strong adaptability in modeling over-reading phenomena, leveraging robust statistical and learning-based techniques to accommodate the complexities of ultrasonic wet gas measurements.

However, the scope of the study was limited to mist and dispersed flow regimes, and did not comprehensively investigate performance under stratified, wave, or annular flow conditions. At transitional flow rates, particularly near regime boundaries, the correction models exhibited increased error, indicating reduced robustness and generalization in these more complex flow architectures. This observation highlights the need for caution when extrapolating the current model configurations to other flow regimes not explicitly represented in the training dataset. The implications of these findings are significant for the process, chemical, and oil and gas industries, where accurate wet gas flow measurement is critical for operational efficiency, production optimization, and reservoir management. Improving correction models to reduce measurement uncertainty directly impacts decision-making in flow monitoring, allocation metering, and performance diagnostics.

6. Conclusion

This study conducted a comprehensive evaluation of seven modeling techniques, namely Random Forest Regression (RFR), Multilinear Regression (MLR), Polynomial Regression, Ridge Regression, Lasso Regression, SciPy Optimization, and Scikit-Learn Neural Networks in predicting ultrasonic over-reading correction factor in mist flow regimes. Among these, RFR and MLR consistently outperformed the alternatives, achieving average relative absolute errors as low as 3 % and demonstrating strong generalization across the validation sets. Techniques RFR proved highly effective in modeling nonlinear interactions and handling multivariate complexity, while MLR techniques offered interpretable, statistically consistent outputs that reinforced its practical value. The performance of these models was substantiated through rigorous residual analysis, feature importance diagnostics, and comparative metric evaluations.

These findings affirm the viability of data-driven modeling frameworks in improving ultrasonic wet gas metering accuracy. To enhance future applicability, further research should explore model extension to stratified and annular regimes, integrate real-time LVF sensing for dynamic correction, and investigate salinity-dependent dielectric properties under varying operating conditions. Such efforts will support the deployment of robust, sensor-integrated correction systems suitable for industrial environments.

CRediT authorship contribution statement

Ishigita Lucas Shunashu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Osmund Kaunde:** Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Table A1
– Symbols and units.

Symbol	Description	Unit
Q_{tp}	Total flow rate measured by ultrasonic flow meter (USM)	m^3/h
Q_g	Actual gas flow rate (reference)	m^3/h
Q_l	Liquid flow rate	m^3/h
OR	Over-reading factor (Q_{tp}/Q_g)	–
LVF	Liquid Volume Fraction	–
X_{LM}	Lockhart-Martinelli parameter	–
Fr_g	Froude number of gas	–
We	Weber number	–
S	Slip ratio	–
DR	Density ratio (ρ_{l/ρ_g})	–
Re_g	Reynolds number of gas	–
Re_l	Reynolds number of liquid	–
ρ_g	Gas density	kg/m^3
ρ_l	Liquid density	kg/m^3
μ_g	Gas dynamic viscosity	Pa·s
μ_l	Liquid dynamic viscosity	Pa·s
V_{sg}	Superficial gas velocity	m/s
V_{sl}	Superficial liquid velocity	m/s
D	Pipe diameter	m
σ	Surface tension	N/m
g	Gravitational acceleration	m/s^2
ε_m	Dielectric constant of wet gas mixture	–
$\Delta\theta$	Phase shift from microwave sensor	degrees (°)
ΔA	Amplitude attenuation from microwave sensor	dB

References

[1] D. van Putten, B. Dsouza, Wet gas over-reading correction for ultrasonic flow meters, Exp. Fluids 60 (3) (Mar. 2019), https://doi.org/10.1007/s00348-019-2693-6.

[2] Y. Xu, P. Yu, Z. Zhu, C. Yuan, T. Zhang, Over-reading modeling of the ultrasonic flow meter in wet gas measurement, Meas. J. Int. Meas. Confed. 98 (Feb. 2017) 17–24, https://doi.org/10.1016/j.measurement.2016.11.007.

[3] K. L, X.J.M. Bao, M. Wang, Integrating machine learning with sensor technology for multiphase flow measurement, IEEE Sens. J. 15 (1) (2024) 37–48, https://doi.org/10.1109/JSEN.2024.3437292.Abstract.

[4] S. Hosseini, "Enhancing wet gas metering capabilities using machine learning : implications of data-driven models and integration with physical principles," no. October, pp. 1–11, 2024.
[5] J. Morshed, S.E. Powers, Regression and dimensional analysis for modeling two-phase flow, Transp. Porous Media 38 (1–2) (2000) 205–221, https://doi.org/10.1023/A:1006607132347.
[6] J.M. Klusowski, Complete analysis of a random forest model, ArXiv 13 (2018) 1063–1095, https://doi.org/10.48550/arXiv.1805.02587.
[7] M. Li, Z. Li, C. Li, In-use measurement of ultrasonic flowmeter based on machine learning, Meas. J. Int. Meas. Confed. 223 (1) (2023) 1–6, https://doi.org/10.1016/j.measurement.2023.113721.
[8] H. Wang, M. Zhang, Y. Yang, Machine learning for multiphase flowrate estimation with time series sensing data, Meas. Sens. 10–12 (2020) 100025, https://doi.org/10.1016/j.measen.2020.100025.
[9] A. Gryzlov, L. Mironova, S. Safonov, M. Arsalan, Evaluation of machine learning methods for prediction of multiphase production rates, Soc. Pet. Eng. - SPE Symp. Artif. Intell. - Towar. Resilient Effic. Energy Ind. 2021 (2021), https://doi.org/10.2118/208648-MS.
[10] S.S. Mofunlewi, J.A. Ajienka, D. Appah, Determination of multiphase flow meter reliability and development of correction charts for the prediction of oilfield fluid flow rates, Leonardo J. Sci. (12) (2008) 165–174. https://ljs.academicdirect.org/A12/165_174.pdf?.
[11] Z. Jiang, H. Wang, Y. Yang, Y. Li, Comparison of machine learning methods for multiphase flowrate prediction, in: IST 2019 - IEEE Int. Conf. Imaging Syst. Tech. Proc, 2019, https://doi.org/10.1109/IST48021.2019.9010450.
[12] V.-T. Nguyen, H.J. Sagar, Computational fluid dynamics modeling and experiments of two-phase flows, Fluids 9 (9) (2024) 207, https://doi.org/10.3390/fluids9090207.
[13] R. Qiu, H. Dong, J. Wang, C. Fan, Y. Wang, Modeling two-phase flows with complicated interface evolution using parallel physics-informed neural networks, Phys. Fluids 36 (9) (2024), https://doi.org/10.1063/5.0216609.
[14] Y. Chen, G. Chinello, P. Tait, J. Jia, A new correlation to determine the Lockhart-Martinelli parameter from vertical differential pressure for horizontal venturi tube over-reading correction, Flow Meas. Instrum. 88 (November 2023, 2022), https://doi.org/10.1016/j.flowmeasinst.2022.102266.
[15] L.T. Zhu, et al., Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors, Ind. Eng. Chem. Res. 61 (28) (2022) 9901–9949, https://doi.org/10.1021/acs.iecr.2c01036.
[16] L. Xing, et al., A combination method for metering gas-liquid two-phase flows of low liquid loading applying ultrasonic and coriolis flowmeters, Flow Meas. Instrum. 37 (2014) 135–143, https://doi.org/10.1016/j.flowmeasinst.2014.01.005.
[17] A. Al-Sarkhi, V. Duc, C. Sarica, E. Pereryra, Upscaling modeling using dimensional analysis in gas-liquid annular and stratified flows, J. Pet. Sci. Eng. 137 (2016) 240–249, https://doi.org/10.1016/j.petrol.2015.11.028.
[18] N.M. Latha, K. Geetha, S. Damodharan, Overview of regression models and how to determine the best model for data, J. Sci. Res. Rep. 30 (10) (2024) 250–266, https://doi.org/10.9734/jsrr/2024/v30i102452.
[19] D. Berrar, Cross-validation, Encycl. Bioinforma. Comput. Biol. ABC Bioinforma. 1–3 (January 2018) (2018) 542–545, https://doi.org/10.1016/B978-0-12-809633-8.20349-X.
[20] J.W. Murdock, Two-phase flow measurement with orifices, J. Fluids Eng. Trans. ASME 84 (4) (1962), https://doi.org/10.1115/1.3658657.
[21] D. Chisholm, Flow of compressible two-phase mixtures through sharp-edged orifices, J. Mech. Eng. Sci. 23 (1) (1981), https://doi.org/10.1243/JMES_JOUR_1981_023_008_02.
[22] R.V. Smith, J.T. Leang, Evaluations of correlations for two-phase flowmeters three current—one new, J. Eng. Gas Turbines Power 97 (4) (1975), https://doi.org/10.1115/1.3446072.
[23] I. Ismail, J.C. Gamio, S.F.A. Bukhari, W.Q. Yang, Tomography for multi-phase flow measurement in the oil industry, Flow Meas. Instrum. 16 (2–3) (Apr. 2005) 145–155, https://doi.org/10.1016/j.flowmeasinst.2005.02.017.