Multiphase flowrate measurement with time series sensing data and sequential model

Multiphase flowrate measurement with time series sensing data and sequential model*

Haokun Wang[a], Delin Hu[a], Maomao Zhang[b], Yunjie Yang[a*]

[a] *Agile Tomography Group, School of Engineering, The University of Edinburgh, Edinburgh, UK*

[b] *Tsinghua Shenzhen International Graduate School, Shenzhen, China*

Abstract

Accurate multiphase flowrate measurement is challenging but vital in the energy industry to monitor the production process. Machine learning has recently emerged as a promising method for estimating multiphase flowrates based on different conventional flow meters. In this paper, we propose a Convolutional Neural Network (CNN)-Long-Short Term Memory (LSTM) model and a Temporal Convolutional Network (TCN) model to estimate the volumetric liquid flowrate of oil/gas/water three-phase flow based on the Venturi tube. The volumetric flowrates of the liquid and gas phase vary from 0.1 - 10 m^3/h and 7.6137 - 86.7506 m^3/h, respectively. We collected time series sensing data from a Venturi tube installed in a pilot-scale multiphase flow facility and utilized single-phase flowmeters to acquire reference data before mixing. Experimental results suggest that the proposed CNN-LSTM and TCN models can effectively deal with the time series sensing data from the Venturi tube and achieve a good accuracy of multiphase flowrate estimation under different flow conditions. TCN achieves a better accuracy for both liquid and phase flowrate estimation than CNN-LSTM. The results indicate the possibility of leveraging conventional flow meters for multiphase flowrate estimation under various flow conditions.

Keywords: Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), Temporal Convolutional Network (TCN), multiphase flowrate measurement, time series data.
2010 MSC: 00-01, 99-00

1. Introduction

Measurement of multiphase flow has for long been a challenge in the energy industry [1]. Accurate measurement of oil-water-gas three phase flow is considered a key factor for achieving efficient, safe and economical production. Depending on different application perspectives, a variety of multiphase flow measurement techniques have been developed, which can generally be summarized into four major categories, i.e. flow pattern recognition [2], flow visualization [3], void fraction measurement [4] and flowrate measurement [5]. To date, several studies have investigated the multiphase flow measurement by using different techniques. For instance, in Pan et al.'s work, gas flow rate is predicted using Venturi and gamma ray technique [6]; the visualisation of a multiphase bubbly flow was achieved in a non-intrusive way in Zhou et al.'s work [7]. In recent years, the emerging machine learning technique has become prevailing in multiphase flow characterization, especially in identifying flow patterns and extracting flow features [8, 9, 10]. For instance, Artificial Neural Network (ANN) is employed to identify the multiphase flow regime by analysing pressure signals [11]. The performance of ANN on two-phase flow pattern recognition under slug flow conditions is studied in [12]. The accuracy of the hydraulic calculation is dramatically improved by implementing the data-driven hydraulic calculation method [13]. High accuracy is achieved on multiphase flow regimes identification in [14] and [15]. Other representative work includes distinguishing multiphase flow patterns for given flow conditions by analysing time-series signals [16], implementing Long-Short Term Memory (LSTM) on flow regime prediction [17], and applying the cluster-based reduced-order model to extract the flow features of slug flow [18].

This work focuses on accurate multiphase flowrate measurement. Due to the limitation of sensors and computational resources, conventional methods for multiphase flowrate measurement mainly rely upon flow separators and Single Phase Flow Meters (SPFMs), such as electromagnetic flowmeters [19]. These SPFMs are implemented to perform flowrate measurements after the separators separate the multiphase flow into single-phase flows [20]. Although the separation method is commonly utilized in energy industries, its drawbacks are also obvious; for

*Fully documented templates are available in the elsarticle package on CTAN.

*Corresponding author
Email address: y.yang@ed.ac.uk; Phone: 44(0)1316517112
(Haokun Wang[a], Delin Hu[a], Maomao Zhang[b], Yunjie Yang[a*])

example, separators are costly and cumbersome, and a long separation period of up to several hours or even days dis-enables the real-time and in-situ measurement capability of multiphase flow.

To date, various attempts have been made to overcome the disadvantages of conventional multiphase flowrate measurement methods. For instance, a modified separation strategy was proposed by Thorn et al., which implemented a vortex to separate the gas phase first and then measure the liquid phase by using single or two-phase flowmeters [21]. Other attempts include the use of differential pressure [22, 23], radiation [24, 25] and acoustic based flow meters [26] to directly measure the multiphase flowrate without separation. Venturi tube as a differential-pressure single phase flow meter is widely applied in the energy industry to perform single phase flowrate measurement. However, the over reading phenomenon usually occurs when performing the gas liquid two phase flowrate measurement due to the instantaneous rapid change of the liquid phase [27]. Meanwhile, these methods can only operate within a narrow range of Water-Liquid Ratio (WLR) and/or Gas-Volume Fraction (GVF) if only one type of flow meters is utilized. To deal with a wide range of flow conditions, multiple sensors are typically required to achieve satisfactory multiphase flowrate measurement.

The prosperity of data-driven methods has also introduced new alternatives for multiphase flowrate measurement by enabling more effective sensing data analysis. Some pioneer work includes flowrate estimation by using Flow Adversarial Networks (FANs) [5], Neural Networks (NN) [28], Deep Neural Network (DNN) [29] , Support Vector Machine (SVM) [30], virtual metering technique [31, 32], LSTM [33, 34], and CNN model [35]. Although the recent progress for multiphase flowrate measurement is substantial, accurate and simultaneous flowrate measurement of different phases under complicated flow conditions (e.g. dynamic three-phase flow) in real time remains very challenging in practice.

This paper therefore proposes a method for estimating the liquid and gas volumetric flowrate of oil/gas/water three-phase flows by combining a differential pressure based flow meter (i.e. the Venturi tube) with advanced machine learning techniques (i.e. Temporal Convolution Networks (TCNs [37]), and the combination of the Conventional Neural Network (CNN) and the Long-Short Term Memory (LSTM) model). We consider the characteristics of the multi-modal time-series sensing data obtained from the Venturi tube, i.e. the differential-pressure data, pressure data and temperature data. To overcome the limitations of convolutional architecture in dealing with the time-sequence data, the LSTM, which is a modified structure of the Recursive Neural Network (RNN) [8], is introduced to avoid the influence of short-term memory. Then, TCN, as a specifically designed model for solving time-series data forecast problems, is firstly applied on multiphase flowrate prediction in this study. We train, validate and test the proposed CNN-LSTM and TCN model based on real-world multiphase flow sensing data collected from a pilot-scale multiphase flow facility and demonstrate its effectiveness and potential for accurate liquid and gas flowrate prediction. Evaluations of the estimated results are performed, and the performance of TCN and the combination of convolutional and recurrent architectures (i.e.CNN-LSTM) is compared under various multiphase flow scenarios.

The novelty of this paper can be summarised as: 1) The successful attempt is made on multiphase flowrate estimation by only leveraging single-phase flow meters; 2) The state-of-the-art deep learning model for time series signals prediction is applied on in-situ multiphase phase flowrate estimation for the first time; 3) Volumetric gas and liquid flowrates are simultaneously estimated with satisfactory accuracy with the proposed deep learning model, which is a long-standing challenge in the energy industry.

2. Methodology

2.1. Pilot-scale multiphase flow facility

The multiphase flow experiments were conducted in the Multiphase Flow Laboratory at Tsinghua Shenzhen International Graduate School. The flow sensing data were collected from the pilot-scale multiphase flow testing facility (see the schematic in Fig. 1 and the picture in Fig. 2). The pilot-scale multiphase flow facility comprises a multiphase flow separator tank, oil, water and gas single-phase flow sections and the mixing sections, and control units. At the commencement of the experiment, three single-phase flows (i.e. oil, water and gas flows) were separately supplied and pumped into the single-phase flow pipes. The oil and water were first commingled and then blended with gas. The final admixture was transported through the multiphase flow testing section and returned to the separator for circulating utilization.

SPFMs were installed on oil, water and gas pipes to measure the flowrate of each phase, which was utilized to calculate the mixture flowrate reference values for training, validating and testing of the proposed machine learning models. A Venturi tube having a transport diameter D of 50mm and a throat diameter d of 25mm was installed in the mixture conveyance pipe to measure the flowrate. The diameter ratio of the Venturi tube is 0.5 and the sampling frequency is 10Hz. The selection of the Venturi tube is mainly based on the consideration of the maximum multiphase flowrate range to be measured, the working condition pressure during the experiment and the specification of the multiphase flow facility in the laboratory. A diagrammatic illustration of the Venturi tube is given in Fig. 3. The Venturi tube generates three pressure signals that are associated with the oil/water/gas flow dynamics, namely the dynamic pressure P and the former (ΔP_1) and posterior (ΔP_2) differential pressures. In addition, we collected the temperature T, which is

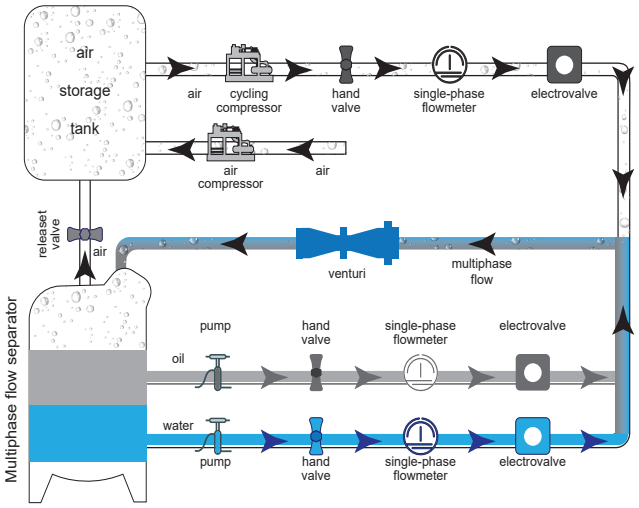


Figure 1: Diagrammatic illustration of the multiphase flow testing facility.

Figure 2: The pilot-scale multiphase flow testing facility at Tsinghua University multiphase flow laboratory.

employed as the fourth input parameter of the learning models for multiphase flowrate prediction.

2.2. Flow data acquisition

The reference data, i.e. the ground truth of the volumetric flowrate of the liquid phase, were the sum

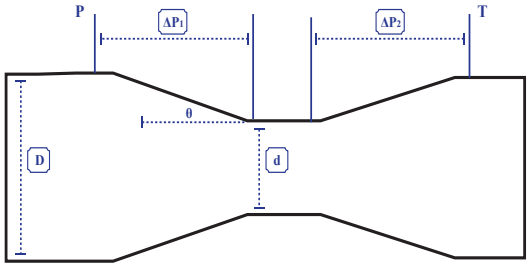


Figure 3: Schematic of Venturi tube with its sensing parameters annotated.

Table 1: Multiphase flow experiment matrix

Objects	Liquid volumetric flowrate (m^3/h)	Gas volumetric flowrate (m^3/h)	WLR / GVF
Water	0.9978 – 4.9650	-	0 – 94.5%
Oil	0.0216 – 7.0367	-	-
Gas	-	7.6134–86.7506	0 – 90%

of the oil and water volumetric flowrate. Based on the volumetric flowrate obtained from single-phase flowmeters, the flowrate of the liquid mixture could be calculated by adding the volumetric flowrate of the oil and water phase. The Venturi tube measures the instantaneous flowrate two to five times every second. Considering the difference in flow conditions at the location of the SPFMs before mixing and the Venturi tube (see Fig. 1), instead of directly using the instantaneous sensor measurements, the reference, pressure and temperature data were averaged over a period of time. The averaged flowrate could more accurately approach the true flowrate of the oil/water mixture.

Table 1 presents the experimental matrix. To mimic multiphase flows in real industrial scenarios, the dynamic flows of water, oil and gas were separately controlled to cover a large diversity of flowrate ranges. The volumetric flowrate of water changed from 0.9978 to 4.9650 m^3/h and the water-in-liquid ratio (WLR) varied from 0 to 94.5%. The volumetric flowrate of gas was 7.6134 - 86.7506 m^3/h. The GVF was settled start from 0% and increases by 10% each time until it reaches 90%. For each GVF settlement, WLR was controlled varies from 0% to 100%.

Interpolation was employed to pre-process the measurement data to compensate for the difference between the sampling rates of the Venturi tube and the SPFMs. To improve the training efficiency, for the data collected from the Venturi tube, five data points at equal time intervals of every second, which represented 0, 0.2, 0.4, 0.6 and 0.8s, were selected to achieve uniform distribution. The nearest interpolation method was applied to estimate the value at the fixed time points. In detail, the value of the five data points could be estimated by the nearest actual value that we obtained from the Venturi tube during the experiments, and the detailed procedure is shown in Fig. 4. As the Venturi tube generates three kinds of pressure data, after interpolation, a 5×3 matrix, which contains P, ΔP_1

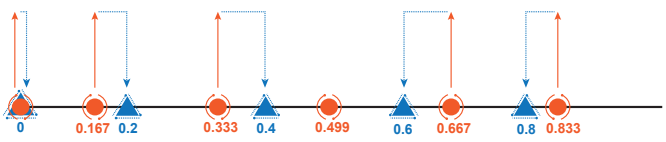


Figure 4: The principle of nearest interpolation method applied on data collected from Venturi tube.

and ΔP_2 three parameters with five data points for each parameter, was obtained every second. Therefore, each pressure parameter contains 98,570 pre-processed data for a duration of 328.57 minutes. By implementing the same procedure on the temperature data, we finally obtain the sequential data in the format of a 98,570 $\times$ 4 matrix, which is shown in Fig. 5. The final step is to calculate the average value of the interpolated sequential data every five points to match with the reference obtained from the SPFMs, the sampling rate of which is 1 Hz. Eventually, a 19714 $\times$ 4 matrix is created.

2.3. Moving average on the acquired flow data

Pursuant to the previous study, it is found that by taking the moving average of the obtained instantaneous data, both the liquid and gas flowrates demonstrate more accurate prediction results with a smaller absolute error and deviation than the prediction results based on instantaneous data [30]. The reason is that errors and non-alignment phenomena occurs between the sensory data (differential pressure and temperature data) and the reference (real liquid and gas flowrate measured by using the SPFMs). The different spatial position of the Venturi tube and SPFMs (see Fig. 1) and the highly dynamic flow status of the mixture when passing through the Venturi tube can cause and strengthen such errors and non-alignment. To limit and inhibit such occurrences, moving average was introduced to harmonize the training and target data.

In contrast to the previous study [30], rather than performing a simple moving average with an identical length "window" from the beginning to the end of the sequential data, this study employed a varying window length for the first 100 data points. In other words, the window length gradually increases from the beginning of the sequential data, and the average of the first 'n' data observations ($n \leq 100$) was calculated. Then, the traditional moving average method with a window

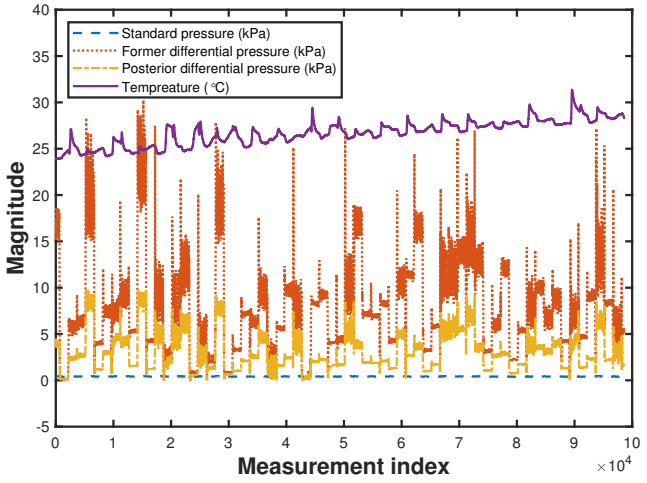


Figure 5: Visualization of the pre-processed training data.

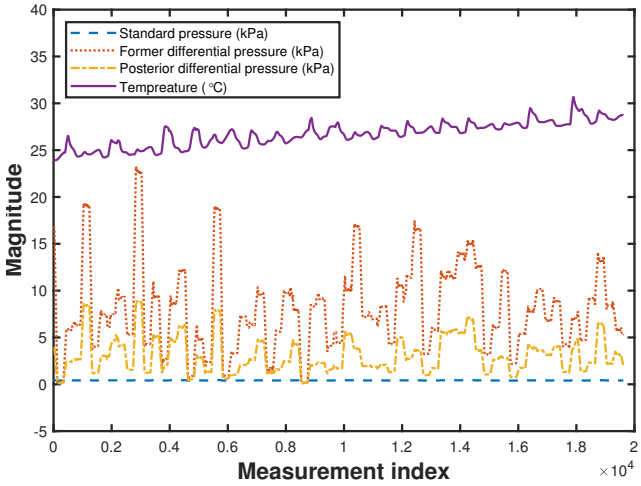


Figure 6: Visualization of the pre-processed training data after moving average operation.

length of 100 was applied on the remaining data until the end, which is shown in Fig. 6. The purpose of adapting the stated moving average strategy was to retain as much information from the raw data as possible during the training process. In detail, there were 100 average calculations by performing varying window length whereas only 1 average calculation by performing simple moving average for the first 100 data. The window length was selected as 100 mainly based on the training and testing experience in our previous work [30]. Different window length (0, 50, 100, 150, 200) was also attempted in this study and the minimum training error occurs when the window length equals to 100 for both CNN-LSTM and TCN model with the value of 0.415 and 0.218, respectively.

2.4. Reference calculation

Three single-phase flowmeters were implemented to measure the volumetric flowrate of oil, gas and water before they were mixed, respectively (see Fig. 1). The reference liquid phase volumetric flowrate (Q_v^{liquid}) could then be deducted by adding the volumetric flowrate of oil (V_o) and water (V_w), which can be formulated as:

$$Q_v^{liquid} = V_o + V_w \tag{1}$$

The calculated liquid volumetric flowrate is taken as the reference during network training and testing. The single-phase flows were then mixed and passed through the Venturi tube, and the corresponding differential pressure and temperature data were measured and recorded as the input of the network.

2.5. CNN-LSTM for flowrate estimation

Convolutional Neural Network (CNN) has been widely applied to deal with image processing problems [36]. Attempts have also been made to adapt CNN for multiphase flowrate estimation [5]. We here first propose a network (named CNN-LSTM) that combines the one-dimensional

CNN (1D CNN) and LSTM model for multiphase flowrate estimation.

2.5.1. Network structure

The proposed CNN-LSTM model is a combination of a 1D CNN and three-layer stacked LSTM model (see Fig. 7). In Fig. 7, parameters k, s and p stand for the kernel size, stride and padding, respectively. The input of the CNN-LSTM model was obtained from the Venturi tube, which contained four multiphase flow parameters: P, ΔP_1, ΔP_2 and T. Each flow parameter was manipulated by using the interpolation method and comprises 33,198 data. The reference flowrate was calculated based on SPFMs measurements before mixing. The reference of the liquid phase was the summation of the volumetric flowrate of water and oil. The output of the CNN-LSTM model is the estimated volumetric flowrate of the liquid phase.

The input layer contains four measured multiphase flow parameters and the length of one sample is set to be 50s, which includes 250 instantaneous sensory data. The sequential data reflect the variation in the multiphase flow parameters within a specific period. Instead of using instantaneous measurement data as input, this kind of input reduces the mismatch with the reference caused by the highly dynamic flows after mixing. The padding layer is designed to follow the input layer to avoid information loss with a kernel size of 3 and stride of 1. A zero padding strategy was adopted to ensure that the central Kernel was located on the data points as much as possible. The batch norm layer was designed to be followed by the convolution layer to prevent the vanishing or exploding gradient problem.

The function of the pooling layer is to downsample and reduce the dimensionality of the data. It is designed to run in parallel with the conv1D, batch norm and ReLU layers. Such a two-channel sub-sampling structure is designed to extract more advanced features of the multiphase flow parameters while retaining the main features. The concatenation layer plays a role in combining the outputs of the two channels in the previous layer, and it is supposed to connect two or more feature maps with the same dimensionality together. It is a utility layer that links the multiple input blob onto an output blob.

Once the feature extraction process of differential pressure and temperature signals is complete, the transpose convolution layer is connected. It is referred to as the fractionally strided convolution layer, and it reverts the size of the output map back to the size of the input map while keeping the connection status unchanged. Together with the convolution process, they are regarded as a complete symmetrical process in the 1D CNN-LSTM model. The fully connected layer ordinarily appears at the end of the CNN and plays the role of regression, which is the predicted flowrate of the liquid phase. The core operation of the fully connected layer is the production of the matrix. To improve the performance of the CNN network, the activation function of each neuron in the fully connected layer is set as the ReLU function in this study.

We cascaded the stacked LSTM before the fully connected layer. The LSTM, as a special RNN model, is designed to resolve the long-term dependencies problem. The structure of LSTM is shown in Fig. 8. The dimension of the input (X_t) from the transpose layer is 16*256, which contains 16 LSTM units with an input dimension (X_i) of = 1*256 ($1 \leq i \leq 16$). The output $h_i, h_i^{'}$ and $h_i^{''}$ ($1 \leq i \leq 16$) represents the output of the LSTM in each layer with the dimension of 1*64, 1*64 and 1*16, respectively. Layer 3 is also known as the output layer of the stack LSTM, where the set of each single unit's output from Layer 3 ($h_i^{''}$) has the dimension of 16*64. Meanwhile, the last-moment output of stack LSTM (h_n) can be obtained from the combination of the final output of each layer, which is the set of $\{h_{16}, h_{16}^{'}, h_{16}^{''}\}$ with the dimension of 3*64. Therefore, the final output of the stack LSTM is obtained by extracting the last element from the set h_n, which is $h_{16}^{''}$ with the dimension of 1*64. It is also consistent with the dimension of the input to the full connection layer in Fig .7 during the forward propagation process. In this study, different layers of LSTM were attempted, and the minimum training error was obtained for the CNN-LSTM model when LSTM had three layers. Meanwhile, the computation time significantly increased when the LSTM layer increased, which is not efficient in real applications.

2.6. TCN for multiphase flowrate estimation

An enhanced Temporal Convolutional Network (TCN) was designed to solve the sequential data prediction problem by combining the generic convolutional and causal convolutional architecture together in 2018 [37]. To overcome the vanishing/exploration gradient drawbacks of the traditional recurrent networks (such as RNN and LSTM) in the sequential model, a generic convolutional architecture was considered as the starting point for building the TCN model.

The structure of the TCN-based multiphase flowrate estimation model is illustrated in Fig. 9. From the input layer, the length of the input data was set to 128 which contains four training parameters with a batch size 'b' of 64. Therefore, the input dimension of the network is defined as 64*128*4. Before the input data propagate through the temporal block, it is initially modified by 1D convolutional and chomp 1D section to ensure that the network will only analyze the past and current data and that the predictions are not influenced by 'future' sequential data.

The temporal block with its input dimension of (batch*in-channel*sequence length) from the previous layer was then connected into the network. It consisted of seven semi-blocks with the same architecture but with different parameter settings. Details of the architecture of the semi-block are shown in Fig. 10.

All the parameter settings are the same for the seven semi-blocks except the out-channel, dilation ratio (d) and

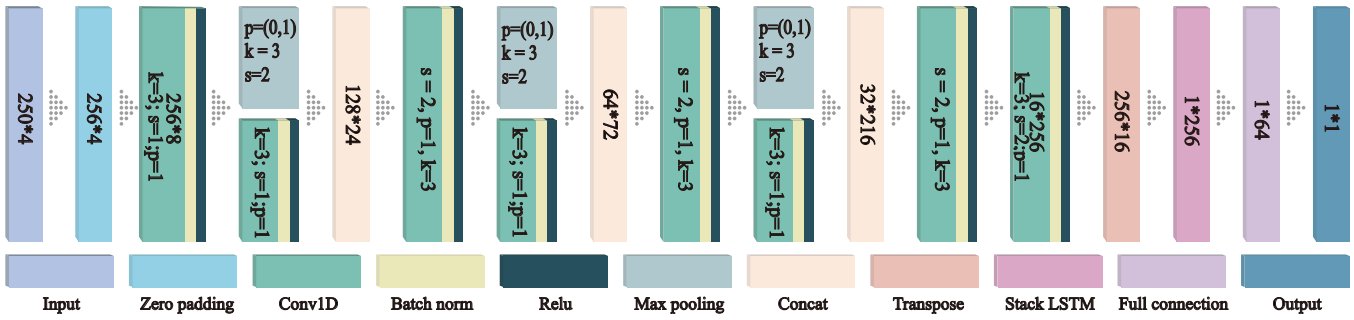


Figure 7: Schematic of the CNN-LSTM model.

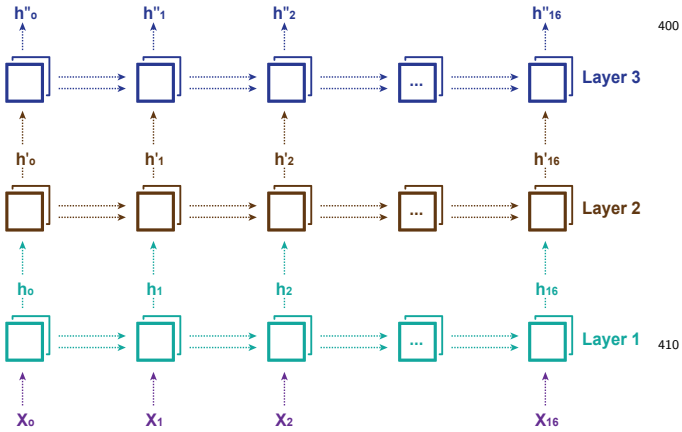


Figure 8: Schematic of the stack LSTM model.

the void rate. These three parameters of each semi-block are sequentially shown in the 'temporal block' in Fig. 9, which will not be stated again in the remainder of this paper. According to Fig. 10, the input (x) in layer ① has the dimension of 1*128*128, which indicates that the batch number is set to 1. The in-channel and sequence length are both consistent with the previous layer of the entire network, which is 128. After the input layer, the sequential data passes through a residual block for the designed TCN, which contains layers ② to ⑫. The residual contains two dilated causal convolution layers with the parameter settings of k =3, p=2 and s=1, and a dilation ratio = 2^i, where i indicates the position of the semi-block. Weight normalization was applied to the convolutional filters in Bai's work [37]. During the training process, we compared the performance of the weight and batch normalization methods. The results indicated that weight normalization has a lower computational cost and can be immune to the influence of the noise that is caused by the random property of the mini-batch. However, the robustness of the weight normalization method for parameter initialization was not sufficiently strong. This issue was solved by adding a spatial dropout layer for regularization. Specifically, the initialization was achieved by ensuring that the entire channel was zeroed out at each training step.

There is an optional route (layer ⑫ & ⑬) in the residual block, that performs the judgement operation. For the conventional residual block structure, the summarization operation can be directly performed between the input and output of the residual function. Nonetheless, the input to the TCN can have a different width than its output; addition cannot be performed if the element width is not matched. Therefore, a convolution layer with a kernel size of 1 was utilized to ensure that the widths of the in- and out-channels were aligned. In our designed residual block, the output of layer ① is added to the output of the layer ⑪ or ⑬, depending on the output channel width.

The dimension of the final output of the TCN model is 1*128, which has the same length as the input data. It is one of the clearly different characteristics of the TCN compared with CNN, which has an output dimension of 1*1. In other words, the output data length of the TCN can be adjusted by controlling the input data length and it represents the sequential forecasting results. In this study, the last digit of the output sequential data was selected as the final result because it represents the most current estimation. Therefore, for 1971 test sequential data group (each sequential data group contains 128 sequential data), we are able to obtain 1971 estimation results.

2.7. Network parameters

In Fig. 7 and Fig. 9, the kernel size of the CNN-LSTM and TCN network is noted as 'k'; 'p' stands for the padding process, and zero padding was chosen in this study. The stride, which indicates the length of the convolution step, was noted as 's'. The correct size of 's' is crucial because repeated calculations will occur, and the training efficiency of the network is reduced with a smaller stride number; however, key information may be lost, and the data features may not be extracted with a large stride. Therefore, in this study, the stride was set to 2 for the CNN-LSTM model and 1 for the TCN model.

The Rectified Linear Unit (ReLU) function was chosen as the excitation function to prevent the vanishing gradient problem during the training process. It can be expressed as:

$$ReLU(x) = \begin{cases} x & (x>0) \\ 0 & (x \leq 0) \end{cases} \quad (2)$$

The loss function was chosen to be the Mean Square Error (MSE) during the training process, which is a com-

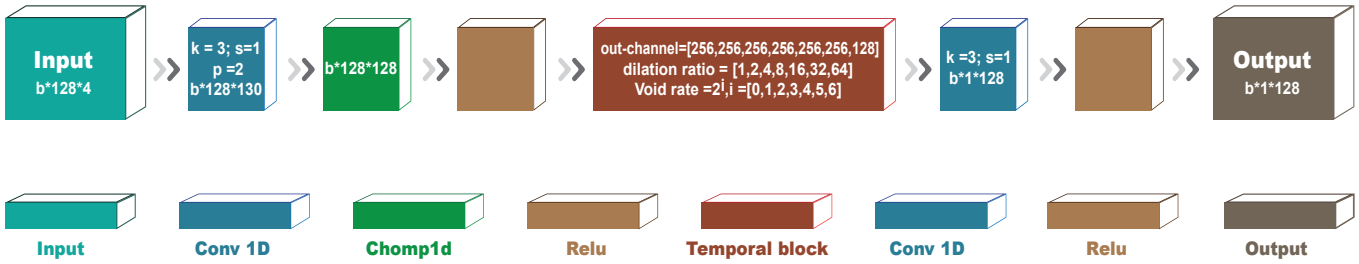


Figure 9: Schematic of the TCN model.

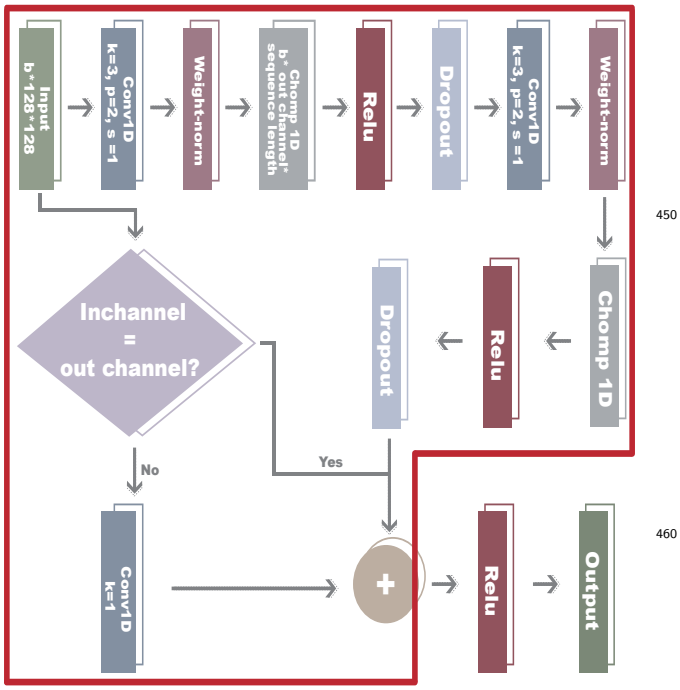


Figure 10: Schematic of the temporal block in TCN.

monly used regression loss function in machine learning. It calculates the MSE between the estimated ($\hat{y}$) and the target values (y), which can be mathematically expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[\hat{y}_i - y_i]^2 \quad (3)$$

where n is the number of samples.

3. Results and Discussion

This Section presents the liquid- and gas-phase flowrate estimation results based on the proposed CNN-LSTM and TCN models. A comparison was also made between the two models.

3.1. Network training

As stated in Section 2.2, the differential pressure, standard pressure and temperature data were pre-processed to facilitate the training process. Specifically, the number of measurements per second was unified by using linear interpolation, after which a uniform sampling rate of 5Hz was obtained. Since the summation of the instantaneous flowrate of single-phase flows before mixing did not comprehensively correspond to the instantaneous measurement data obtained from the Venturi tube due to spatial differences and the inherent characteristics of multiphase flow (such as the rapid change of flow pattern and flowrate). Therefore, it is more meaningful to predict the average flowrate over a period to mitigate the influence of such phenomena. Hence, in training, we adopted a 50 seconds period, and the average flowrate was calculated every 50s as a reference. After pre-processing, 19,714 samples were obtained. The experimental samples were randomly divided into the training, validation and testing data set in the proportions of 80%, 10% and 10%, respectively. The epoch number was set as 230 for the CNN-LSTM model and 400 for the TCN model for the best prediction results (minimum validation error occurs), respectively. The learning rate was 0.001 for both models.

3.2. Results of CNN-LSTM

Fig. 11 presents the estimated liquid and gas phase volumetric flowrate by using the proposed CNN-LSTM model. The instantaneous (un-smoothed) and pre-processed (smoothed) data were separately utilized during the multiphase flowrate estimation process, which corresponds to the results shown in Figs.11 and 12, respectively.

By comparing the flowrate estimation results for different flow phases, it is clear that a positive correlation is found between the predicted volumetric flowrate of the liquid phase and the reference flowrate in Figs. 11a and 12a. In detail, the majority of the estimation results (88.3% for smoothed data and 79.1% for un-smoothed data) are located within an acceptable error range (i.e. the $\pm 10\%$ of the regression line), with only a few points outside the boundary. This suggests the effectiveness of utilizing CNN-LSTM to manage the time-series sensing data to estimate the multiphase flowrate under more complex scenarios. A similar trend can also be obtained in the volumetric flowrate estimation of the gas phase when the gas flowrate is less than $40(m^3/h)$ in Fig. 12b. However, when the gas flowrate exceeds $40(m^3/h)$, no obvious correlation appears, and the estimation results seem to be randomly distributed. Although most of the

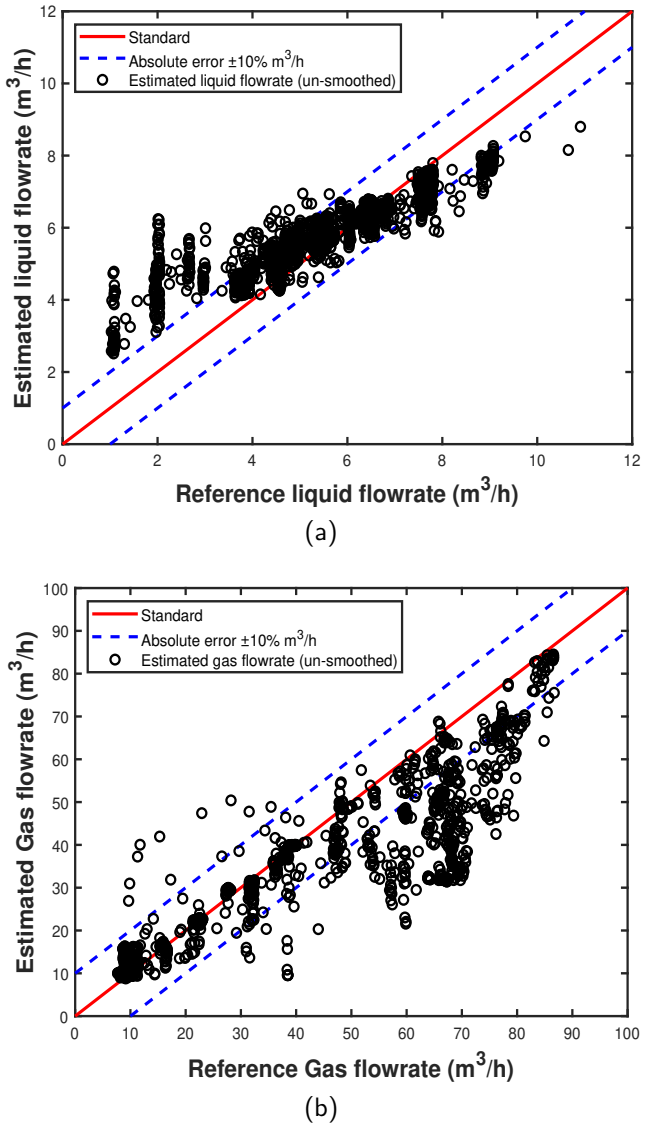


Figure 11: Multiphase flowrate estimation results of (a) liquid phase and (b) gas phase by using CNN-LSTM with un-smoothed training data.

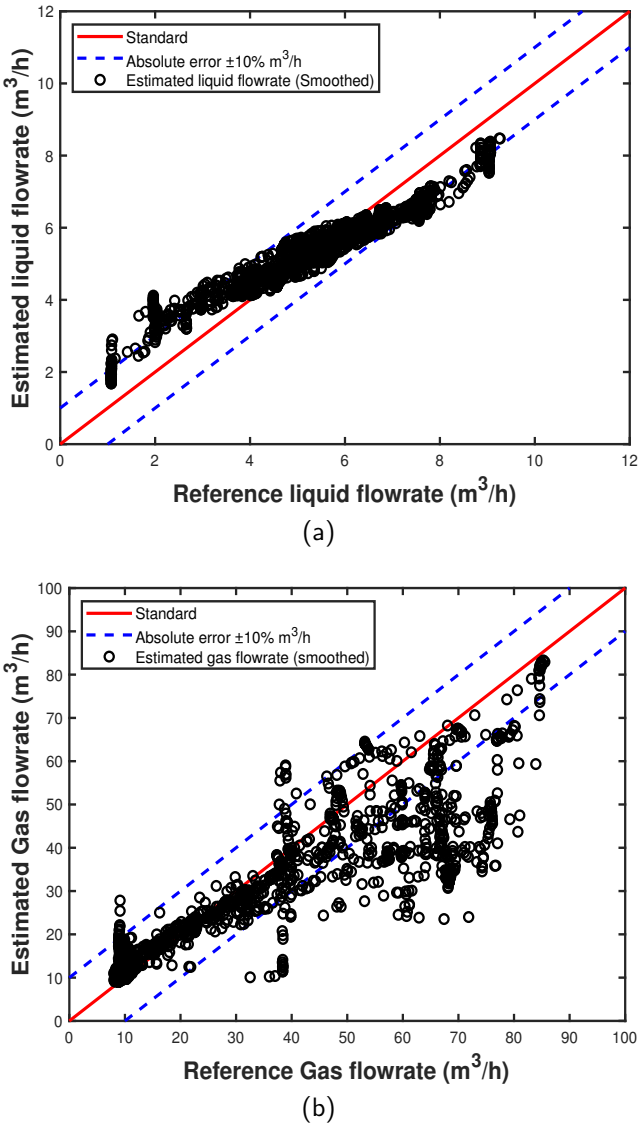


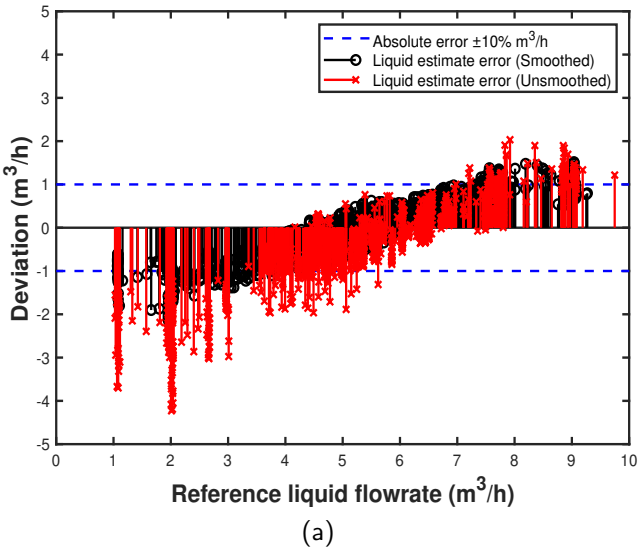
Figure 12: Multiphase flowrate estimation results of (a) liquid phase and (b) gas phase by using CNN-LSTM with smoothed training data.

estimation results (80.3% for smoothed and 80.5% for un-smoothed) for the gas phase are still located within the ±10% of the regression line, we cannot conclude that the results are acceptable due to the non-linear distribution of the estimation results.

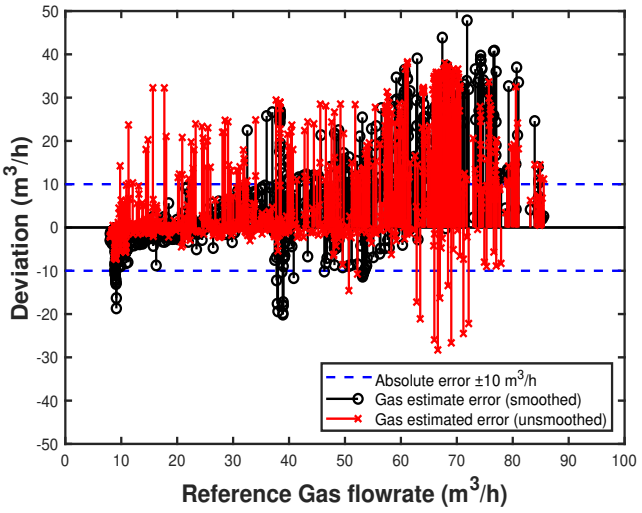
In this study, we employ absolute error instead of relative error as a performance indicator, because it is a common practice in this field, and it is extremely challenging to accurately predict the small flowrates. The deviation of the estimated liquid flowrate is presented in Fig. 13a. Most of estimation errors fall inside the acceptable range, and several points exceeds the plus or minus ten percent line. It further confirms that the CNN-LSTM model can be considered as an effective method for liquid phase flowrate estimation with time-series sensing data as the input. Meanwhile, the limitation on gas phase flowrate estimation is also non-negligible in the 1D CNN-LSTM model, which is shown in Fig. 13b. Therefore, the TCN model is developed to further deal with such issues and as a comparison to the CNN-LSTM model.

3.3. Results of TCN

The proposed TCN model was also separately trained by un-smoothed and smoothed parameters and the multiphase flowrate estimation results were shown in Fig. 14 and Fig. 15, respectively. Fig. 14 presents the multiphase flowrate estimation results obtained from the instantaneous training data of the TCN model. In Fig. 14a, a rough linear relationship can be observed between the estimated liquid flowrate and the reference. Still, 5.9% of the estimation results are not located in the ±10% range with obvious deviations. From analyzing only Fig. 14a, it could be argued that such a phenomenon is

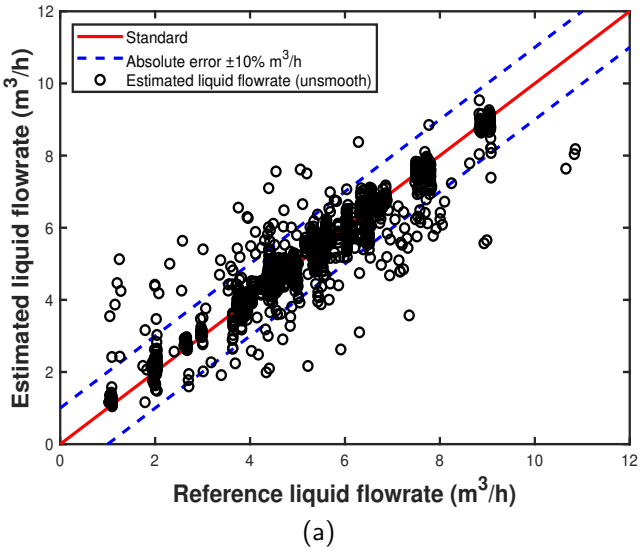


(a)

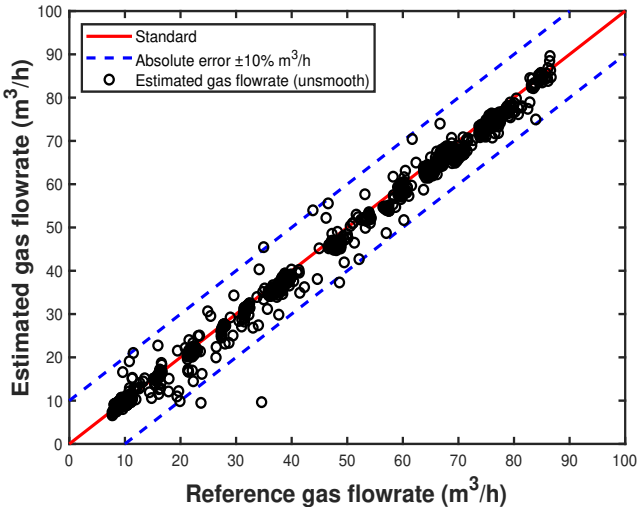


(b)

Figure 13: Deviation of (a) liquid phase and (b) gas phase by using CNN-LSTM with smoothed and un-smoothed training data by implementing CNN-LSTM model.



(a)



(b)

Figure 14: Multiphase flowrate estimation results of (a) liquid phase and (b) gas phase by using TCN with un-smoothed training data.

probably caused by the mismatch between the training and the target data, which is unavoidable when implementing instantaneous training data as input. From the data in Fig. 14b, it is apparent that the linear regression of the predicted gas flowrate is observed from the TCN model. Even the input training data is instantaneous data without any pre-processing procedure. Such a regression phenomenon on the estimated gas results initially confirms that the TCN model can resolve time-series data from multiphase flow area. In other words, the non-perfect linearity of the predicted liquid phase flowrate may be due to the characteristics of the training data rather than the inherent limitations of the TCN model. To further investigate the inherent ability of the TCN model, smoothed training data were applied, and the multiphase flowrate estimation results are presented in Fig. 15. A perfect linear regression phenomenon appears for both liquid and gas phase estimation results. Nearly all the forecasting points are located within an acceptable range (±10%) with only one exception in the gas estimation results. Fig. 15 indicates that TCN model can manage the pre-processed sequential differential pressure and temperature data for the multiphase flowrate estimation problem.

The more accurate flowrate estimation results obtained using smoothed data further confirms that the characteristic (or the smoothness process) of the input data plays a non-negligible role in the multiphase flowrate prediction problem. In particular, the alignment between the training and the reference data are more accurate with the smoothed data according to the comparison between estimation results given in Figs. 14 and 15. It could be expounded that the flow status of the three-phase mixture rapidly varies when it passes through the Venturi

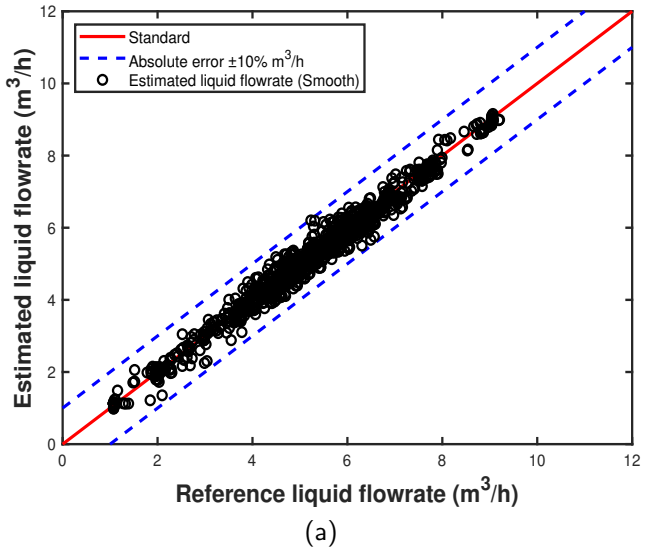


Figure 15: Multiphase flowrate estimation results of (a) liquid phase and (b) gas phase by using TCN with smoothed training data.

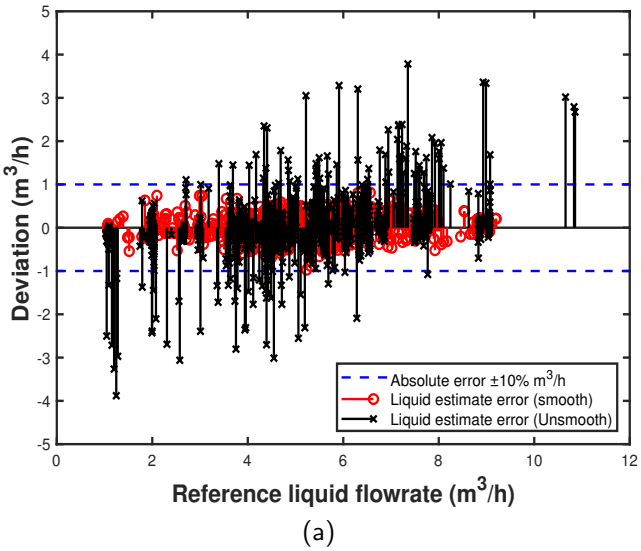


Figure 16: Deviation of (a) liquid phase and (b) gas phase by using TCN with smoothed and un-smoothed training data by implementing TCN model.

tube, which causes the instantaneous recording of the training parameters to be misaligned with the reference data, which are recorded under a more stable flow status in a single-phase flow pipe. The moving average method overcomes such inherent drawbacks of the multiphase flowrate measurement by aligning the training and reference data within a period, and the estimation results demonstrate that the pre-process is considerable when dealing with multiphase flowrate estimation problems.

The deviation demonstrated in Fig. 16 further confirms the conclusions that the relatively large estimation deviation is caused by the characteristic of the instantaneous multiphase flow parameters rather than the inherent ability of the TCN model. Both the liquid and gas estimation results demonstrate smaller deviations with the pre-processed training and reference data in Fig. 16. Larger deviations and more outside-range points are observed in Fig. 16 with the instantaneous (un-smooth) data, which indicates that non-alignment phenomena occur during the multiphase flowrate estimation process. We can further conclude that the pre-processing of instantaneous data plays an important role in multiphase flowrate estimation by reducing the affect caused by the non-alignment between the training and reference data.

In industry applications, different Venturi structures may provide various differential pressure signals. The TCN model can still be applied on multiphase flowrate estimation when differential pressure signal changes but re-training and parameter tuning of the TCN are possibly required in order to achieve the best prediction results for different scenarios.

3.4. Comparison between CNN-LSTM and TCN

To acquire comprehensive knowledge of the performance of the proposed machine learning methods on multiphase flowrate estimation, the MSE and the correlation between the reference and estimated multiphase flowrate were also considered as complementary metrics for evaluating the performance of the CNN-LSTM and TCN.

The correlation index can be calculated by:

$$\rho_{x,y} = \frac{Cov(x,y)}{\sqrt{\sigma_x \sigma_y}} \tag{4}$$

where x and y represent the estimated and reference multiphase flowrate, respectively; σ stands for the standard deviation. The calculated MSE and correlation index for the estimated multiphase flowrate of CNN-LSTM and TCN under different scenarios are shown in Table. 2, where the subscript un and s stands for the 'un-smoothed' and 'smoothed' data, respectively.

TCN results have larger correlation index and smaller MSE for liquid and gas flowrate estimation with un-smoothed and smoothed data compared with the CNN-LSTM results. This observation indicates that TCN has an increased ability for multiphase flowrate prediction under various flow conditions. The effect of the pre-processing of the training data on the prediction accuracy can be evaluated by comparing the same evaluation index of the same method but with different types of data. In detail, using CNN-LSTM as an example, $\rho_{un}^{CNN-LSTM}$ and $\rho_{s}^{CNN-LSTM}$ are 0.9190 and 0.9688 for the liquid phase flowrate estimation, respectively. Hence, the moving average pre-processing improves the liquid phase prediction accuracy when the CNN-LSTM model is implemented. Similarly, by performing such a comparison between the remaining groups in Table. 2, we can conclude that by using a moving average manipulation on the instantaneous flow parameters, higher accuracy for the predicted multiphase flowrate can be expected. The only exception is when predicting the gas phase flowrate by implementing the CNN-LSTM model. It could be explained that the CNN-LSTM model is not sufficiently sensitive to the input data to the multiphase flowrate estimation application. Specifically, although the CNN-LSTM model has the information storage ability during the decision making process, it still cannot memorize all the past information, and the final decision may also partially depend on the future information. The latter process is not achievable in an actual application because the future information is unavailable under a time-series forecasting scenario. Meanwhile, the structure of the TCN can be summarized as the addition of a 1D Fully-Convolutional Network (FCN) and the causal convolutions. Such a structure affords the TCN model two unique characteristics for the time-series data forecasting problem. The first is that the causal convolution block ensures that the output of the TCN model relies on the past information only, and in contrast to the CNN-LSTM model, all the past information contributes to the output of the TCN model. The second characteristic of the TCN model is that the output layer has the same length and width as the input layer, which is achieved by using the zero-padding method in the 1D FCN block. Therefore, the manipulation of the input data influences the output when the TCN model is implemented for the multiphase flowrate estimation problem.

According to Table. 2, TCN achieves the highest correlation index and smallest MSE for both liquid and gas phases when accepting smoothed data as input. This demonstrates the strong ability of the TCN for multiphase flowrate estimation and the necessity for manipulating the raw multiphase flow parameters to obtain more accurate estimated flowrates.

4. Conclusion

This paper proposed CNN-LSTM and TCN models for multiphase flowrate estimation of dynamic three-phase flows with multiple time-series sensing data obtained from a Venturi tube. We trained, validated and tested the proposed CNN-LSTM and TCN models using real-world flow sensing data acquired from a pilot-scale multiphase flow facility. The results validated the feasibility and effectiveness of the TCN model to estimate multiphase flowrate and the CNN-LSTM model to predict liquid phase flowrate with times-series sensing data collected from the Venturi tube. Estimation performance evaluation revealed that compared with CNN-LSTM, TCN demonstrates a stronger ability for multiphase flowrate estimation on both liquid and gas phases, and achieves more accurate estimation results. The characteristics of the input data (instantaneous or moving averaged data) also plays a non-negligible role in multiphase flowrate estimation, and a pre-processing procedure is proposed to overcome the problem caused by the mis-alignment between the training and reference data. The comparison of the liquid and gas phase flowrate estimation results revealed that TCN performs better than CNN-LSTM on time-series data in a multiphase flow area.

Future work will be conducted in the near future to establish a multi-modality data fusion platform with multi-sensors to accurately estimate the multiphase flowrate by using the proposed machine learning based method.

References

[1] L. S. Hansen, S. Pedersen, P. Durdevic, Multi-phase flow metering in offshore oil and gas transportation pipelines: Trends and perspectives, Sensors 19 (9) (2019) 2184.

[2] J. Ye, L. Guo, Multiphase flow pattern recognition in pipeline–riser system by statistical feature clustering of pressure fluctuations, Chemical Engineering Science 102 (2013) 486–501.

[3] J. Yao, M. Takei, Application of process tomography to multiphase flow measurement in industrial and biomedical fields: A review, IEEE Sensors Journal 17 (24) (2017) 8196–8205.

Table 2: Correlation index and MSE of flowrate estimation using CNN-LSTM and TCN

Index	CNN-LSTM		TCN	
	Liquid	Gas	Liquid	Gas
ρ_{un}	0.9190 (Fig. 11a)	0.9340 (Fig. 11b)	0.9476 (Fig. 14a)	0.9980 (Fig. 14b)
ρ_s	0.9688 (Fig. 12a)	0.9033 (Fig. 12b)	0.9934 (Fig. 15a)	0.9986 (Fig. 15b)
MSE_{un}	0.9586 (Fig. 11a)	109.6736 (Fig. 11b)	0.3121 (Fig. 14a)	3.1700 (Fig. 14b)
MSE_s	0.4003 (Fig. 12a)	126.2088 (Fig. 12b)	0.0363 (Fig. 15a)	2.6969 (Fig. 15b)

[4] Y. Xue, H. Li, C. Hao, C. Yao, Investigation on the void fraction of gas–liquid two-phase flows in vertically-downward pipes, International Communications in Heat and Mass Transfer 77 (2016) 1–8.

[5] D. Hu, J. Li, Y. Liu, Y. Li, Flow adversarial networks: Flowrate prediction for gas–liquid multiphase flows across different domains, IEEE transactions on neural networks and learning systems 31 (2) (2019) 475–487.

[6] Y. Pan, C. Li, Y. Ma, S. Huang, D. Wang, Gas flow rate measurement in low-quality multiphase flows using venturi and gamma ray, Experimental Thermal and Fluid Science 100 (2019) 319–327.

[7] H. Zhou, X. Niu, An image processing algorithm for the measurement of multiphase bubbly flow using predictor-corrector method, International Journal of Multiphase Flow 128 (2020) 103277.

[8] W. Dang, Z. Gao, L. Hou, D. Lv, S. Qiu, G. Chen, A novel deep learning framework for industrial multiphase flow characterization, IEEE Transactions on Industrial Informatics 15 (11) (2019) 5954–5962.

[9] E. Kanin, A. Osiptsov, A. Vainshtein, E. Burnaev, A predictive model for steady-state multiphase pipe flow: Machine learning on lab data, Journal of Petroleum Science and Engineering 180 (2019) 727–746.

[10] T. Barbariol, E. Feltresi, G. A. Susto, D. Tescaro, S. Galvanin, Sensor fusion and machine learning techniques to improve water cut measurements accuracy in multiphase application, in: SPE Annual Technical Conference and Exhibition, OnePetro, 2020.

[11] T. Xie, S. Ghiaasiaan, S. Karrila, Artificial neural network approach for flow regime classification in gas–liquid–fiber flows based on frequency domain analysis of pressure signals, Chemical Engineering Science 59 (11) (2004) 2241–2251.

[12] C. Cozin, F. E. Vicencio, F. A. de Almeida Barbuto, R. E. Morales, M. J. Da Silva, L. V. R. Arruda, Two-phase slug flow characterization using artificial neural networks, IEEE Transactions on Instrumentation and Measurement 65 (3) (2016) 494–501.

[13] G. Cui, Q.-S. Jia, X. Guan, Q. Liu, Data-driven computation of natural gas pipeline network hydraulics, Results in Control and Optimization 1 (2020) 100004.

[14] J. S. Hernandez, C. Valencia, N. Ratkovich, C. F. Torres, F. Muñoz, Data driven methodology for model selection in flow pattern prediction, Heliyon 5 (11) (2019) e02718.

[15] N. Ali, B. Viggiano, M. Tutkun, R. B. Cal, Data-driven machine learning for accurate prediction and statistical quantification of two phase flow regimes, Journal of Petroleum Science and Engineering 202 (2021) 108488.

[16] Z.-K. Gao, N.-D. Jin, Characterization of chaotic dynamic behavior in the gas–liquid slug flow using directed weighted complex network analysis, Physica A: Statistical Mechanics and its Applications 391 (10) (2012) 3005–3016.

[17] Z. Dang, M. Ishii, Two-phase flow regime prediction using lstm based deep recurrent neural network, arXiv preprint arXiv:1904.00291 (2019).

[18] N. Ali, B. Viggiano, M. Tutkun, R. B. Cal, Cluster-based reduced-order descriptions of two phase flows, Chemical Engineering Science 222 (2020) 115660.

[19] T. Leeungculsatien, G. Lucas, Measurement of velocity profiles in multiphase flow using a multi-electrode electromagnetic flow meter, Flow Measurement and Instrumentation 31 (2013) 86–95.

[20] S. S. Mofunlewi, J. A. Ajienka, Economic evaluation of multiphase meters, Leonardo J. Sci 11 (2007) 2.

[21] R. Thorn, G. A. Johansen, B. T. Hjertaker, Three-phase flow measurement in the petroleum industry, Measurement Science and Technology 24 (1) (2012) 012003.

[22] L. Xu, W. Zhou, X. Li, M. Wang, Wet-gas flow modeling for the straight section of throat-extended venturi meter, IEEE transactions on Instrumentation and Measurement 60 (6) (2011) 2080–2087.

[23] Z. Hong-jian, Y. Wei-ting, H. Zhi-yao, Investigation of oil-air two-phase mass flow rate measurement using venturi and void fraction sensor, Journal of Zhejiang University-SCIENCE A 6 (6) (2005) 601–606.

[24] S.-H. Jung, J.-S. Kim, J.-B. Kim, T.-Y. Kwon, Flow-rate measurements of a dual-phase pipe flow by cross-correlation technique of transmitted radiation signals, Applied Radiation and Isotopes 67 (7-8) (2009) 1254–1258.

[25] S. Blaney, Gamma radiation methods for clamp-on multiphase flow metering (2008).

[26] S. Al-Lababidi, D. Mba, A. Addali, Upstream multiphase flow assurance monitoring using acoustic emission, Acoustic Emission 27 (2012) 217–250.

[27] Y. Xu, Q. Zhang, T. Zhang, X. Ba, An overreading model for nonstandard venturi meters based on h correction factor, Measurement 61 (2015) 100–106.

[28] J. Han, F. Dong, Mass flow rate measurement of gas/liquid two-phase flow in horizontal pipe based on v-cone flow meter and adaptive wavelet network, in: 2009 IEEE Instrumentation and Measurement Technology Conference, IEEE, 2009, pp. 1391–1396.

[29] P. Temirchev, M. Simonov, R. Kostoev, E. Burnaev, I. Oseledets, A. Akhmetov, A. Margarit, A. Sitnikov, D. Koroteev, Deep neural networks predicting oil movement in a development unit, Journal of Petroleum Science and Engineering 184 (2020) 106513.

[30] H. Wang, M. Zhang, Y. Yang, Machine learning for multiphase flowrate estimation with time series sensing data, Measurement: Sensors (2020) 100025.

[31] T. A. AL-Qutami, R. Ibrahim, I. Ismail, M. A. Ishak, Virtual multiphase flow metering using diverse neural network ensemble and adaptive simulated annealing, Expert Systems with Applications 93 (2018) 72–85.

[32] S. Sanzo, M. Montini, L. Cadei, M. Giuliani, A. Bianco, Virtual metering and allocation using machine learning algorithms, in: International Petroleum Technology Conference, OnePetro, 2020.

[33] N. Andrianov, A machine learning approach for virtual flow metering and forecasting, IFAC-PapersOnLine 51 (8) (2018) 191–196.

[34] K. Loh, P. S. Omrani, R. van der Linden, Deep learning and data assimilation for real-time production prediction in natural gas wells, arXiv preprint arXiv:1802.05141 (2018).

[35] H. Zhang, Y. Yang, M. Yang, L. Min, Y. Li, X. Zheng, A novel cnn modeling algorithm for the instantaneous flow rate measurement of gas-liquid multiphase flow, in: Proceedings of

the 2020 12th International Conference on Machine Learning and Computing, 2020, pp. 182–187.

[36] R. Vinayakumar, K. Soman, P. Poornachandran, Applying convolutional neural network for network intrusion detection, in: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2017, pp. 1222–1228.

[37] S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv:1803.01271 (2018).