Article | Received 22 November 2024; Accepted 25 January 2025; Published 14 March 2025
https://doi.org/10.55092/mt20250001

Underflow concentration soft sensing for cone thickener system based on a direct data-driven quantile regression forecasting

Yongxiang Lei * and Hamid Reza Karimi

Department of Mechanical Engineering, Politecnico di Milano, Milano, 20156, Italy

* Correspondence author; E-mail: yongxiang.lei@polimi.it.

Highlights:

- A novel prediction framework based on probabilistic models is proposed.
- Quantile regression incorporates and quantifies prediction uncertainty.
- Multi-horizon prediction is implemented in the presented framework.
- Industrial experiments have validated and verified the performance.

Abstract: Paste-filling is a crucial process in the mining industry. Traditional sensor devices often overlook model errors and struggle to measure the certainty of key quality variables. This study addresses these challenges by proposing a novel efficient data-driven quantile regression forecasting framework, DDQRF, to predict the concentration of deep cone thickeners. Conventional methods rely on normal regression models, minimizing residual mean square error to estimate underflow concentration, resulting in inaccuracies due to residual error accumulation in recursive strategies. Specifically, complex high-quality feature representation is essential for accurate prediction models, particularly for multiple horizon predictions necessary for hierarchical optimal control. The presented framework introduces direct data-driven regression prediction, leveraging temporal machine learning models to extract features effectively. Unlike probabilistic Bayesian models, our approach offers efficient implementation and deployment, utilizing prediction intervals to quantify forecast uncertainty. In addition, the proposed model directly predicts multiple horizons, contrasting with traditional recursive single-point forecasting, offering enhanced training and memory efficiency. These characteristics are validated through industrial experiments on a deep cone thickener, comprehensively comparing performance with state-of-the-art counterparts.

Keywords: machine learning; quantile regression; direct data-driven prediction; uncertainty qualification; cone thickener system (CTS)

1. Introduction

The extraction of resources is a cornerstone of modern civilization, serving as a catalyst for societal progress, with the mining industry playing a central and indispensable role in this pursuit. Raw mining materials are essential to a wide range of applications, cutting across numerous sectors [1], including critical industries such as aerospace and medicine, further emphasizing the multifaceted importance of mining [2–4]. However, despite its pivotal contribution, the mining sector faces significant challenges. Chief among these is the precise regulation of underflow concentration—a critical quality parameter that directly influences the paste-filling process [5]. Accurate measurement of this concentration is hindered by several factors, including prohibitive device costs, complex coupling effects, and substantial time delays [6]. Maintaining a consistent and accurate underflow concentration is crucial for the robustness of paste-filling operations, as deviations can lead to imbalances and mud oscillations within thickeners. As a result, effective control of the cone thickener system (CTS) is vital to monitor paste extraction concentration and mitigate potential safety risks [7]. Given the mechanical complexities of CTS, reliable prediction of underflow concentration serves as a prerequisite for implementing effective control measures. In addition, keeping the underflow concentration within a narrowly defined dynamic range is essential to prevent blockages or crushing while meeting the stringent quality demands of the paste-filling process.

The accurate underflow concentration is of critical importance in CTS. However, traditional time series forecasting models primarily focus on y_{t+1} subject to recent history $y : t = (y_t, , y_0)$ [8]. Among these, the Box-Jenkins methodology, including ARIMA models [9], is a widely used point-estimation approach that employs historical data points in a recursive autoregressive framework for multi-horizon forecasting. Nonetheless, real-world forecasting often entails higher complexity, with time series exhibiting long-term dependencies between inputs and outputs [10]. Additionally, accurate estimation of prediction intervals is crucial for quantifying forecast uncertainties and optimizing control performance in decision-making, particularly for multistep, long-horizon predictions [11]. To tackle these challenges, a variety of modern techniques have been proposed, each addressing specific aspects of the forecasting problem [12].

Quantile uncertainty prediction serves as a pivotal approach in time series forecasting, particularly for applications requiring precise estimation of prediction intervals and robust decision-making. Unlike traditional point prediction methods that estimate the conditional mean of a target variable, quantile prediction focuses on estimating specific conditional quantiles of the target distribution. This approach provides a comprehensive representation of forecast uncertainties by generating prediction intervals at desired confidence levels. By learning to predict these quantiles directly, quantile regression eliminates the need for strong distributional assumptions, offering greater flexibility and applicability across diverse datasets [13]. Such predictions are particularly beneficial in scenarios with asymmetric costs associated with under- and over-prediction, as they enable more informed and risk-aware control strategies. Moreover, the ability to quantify uncertainty through prediction intervals enhances the reliability of forecasts in industrial systems, where maintaining operational consistency and safety is paramount. Consequently, quantile-based uncertainty prediction has emerged as a robust tool for advancing predictive analytics in complex, data-driven environments [14].

Most neural network-based approaches for time series forecasting adopt the recursive strategy [15], where the model predicts y_{t+1} based on history $y_{:t}$ and iteratively updates this estimate to forecast longer horizons. However, this recursive approach can suffer from error accumulation, particularly when employing Recurrent Neural Networks (RNNs), as the model relies on its own predictions rather than actual data during inference [16]. By contrast, the direct strategy predicts y_{t+k} directly from $y_{:t}$ for each k, reducing bias, improving stability, and enhancing robustness against model misspecification [17]. Recent studies comparing multi-step forecasting strategies with neural networks highlight the efficacy of the direct multi-horizon approach, where the model is trained to predict a multivariate target $(y_{t+1};\cdot;y_{t+k})$, thereby mitigating error accumulation while ensuring parameter efficiency. In decision-making contexts with asymmetric costs for over- and under-prediction, probabilistic forecast models that estimate the full conditional distribution $p(y_{t+k}|y_{:t})$ are preferable to point forecast models that predict only the conditional mean $E(y_{t+k}|y_{:t})$. Traditionally, for real-valued time series, probabilistic models assume Gaussian error distributions or stochastic processes for residual series $\varepsilon_t = y_t - \hat{y}_t$. However, quantile regression has gained prominence in scenarios where the prediction of specific quantiles is necessary for minimizing losses and quantifying uncertainties [13, 14]. This method learns to predict conditional quantiles $y_{t+k}^{(q)}|y_{:t}$ of the target distribution without assuming specific distributional forms, enabling accurate probabilistic forecasts with precise prediction intervals. As a result, quantile regression is increasingly recognized as a robust approach for prediction calibration.

Direct data-driven control algorithms utilize input-output data to design control strategies, integrating parameter identification with model-based control. In industrial cone thickener systems, several challenges underscore the need for a multi-horizon quantile strategy driven by direct data [7]. These challenges include:

(1) The difficulty of measuring historical underflow concentration distributions, which limits the applicability of traditional Bayesian probabilistic predictors.

(2) The necessity for consistent operation and production in the paste-filling process, which demands accurate underflow concentration forecasts. Reliable forecasts enable precise control, enhance efficiency, and mitigate costs and safety hazards.

(3) The challenge of obtaining a comprehensive mathematical model of the CTS for implementing model-based methods, making data-driven approaches more viable. Recent advancements in direct data-driven prediction techniques offer promising solutions to address these challenges [18–20].

Building on the previous analysis, the time series prediction of the CTS underflow concentration is critical yet remains insufficiently explored. In this work, a consistent underflow concentration prediction method based on deep LSTM and quantile regression is proposed. The motivation is that the underflow concentration prediction is generated using a data-driven multi-horizon forecasting mechanism with uncertainty quantification. The key contributions of this study are as follows:

- **Novel architecture:** Unlike the traditional time series prediction models [21, 22], which predict y_t given recent history $y_{t-1}, y_{t-2}, \ldots, y_0$, this study proposes a new prediction model architecture that implements direct data-driven multi-horizon prediction.

- **Quantified uncertainty:** Distinct from previous deterministic prediction methods [7, 23–25], this quantile regression model incorporates and quantifies prediction uncertainty.
- **Learning capacity and robustness:** The presented framework fully leverages the learning capacity of temporal machine learning models and the robustness of probabilistic models to uncertainty. Unlike traditional Bayesian probabilistic models, this study introduces a novel Gating Recurrent Network (GRN) for encoder representation, characterized by its ease of use and implementation.
- **Industrial application:** This study represents the inaugural effort to tackle the challenge of predicting underflow concentration within an industrial CTS system. The proposed model provides substantial insights, serving as a critical reference for control strategies.

The proposed method's feasibility and effectiveness are demonstrated through a real-world industrial application for CTS, supported by core model verification and validation. The paper is structured as follows: Section 2 presents the foundational work, covering problem formulation, multi-horizon forecasting, and underflow probabilistic prediction. Section 3, Section 4, and Section 5 discuss the prediction approach, convergence analysis, and experimental studies, respectively. Finally, Section 6 provides the conclusion.

2. Preliminary foundations

The deep thickener system is the central element of the industrial paste-filling process. Furthermore, it supplies the essential components required to construct the proposed probabilistic, multi-horizon prediction framework. Notably, these components include the LSTM unit and the recurrent gating mechanism.

2.1. CTS description

As shown in Figure 1, A deep understanding of the CTS is crucial for subsurface paste filling, ensuring consistent concentration levels for underground mining operations. Maintaining an appropriate subterranean concentration is essential, as excessive concentration can result in pipe blockages, while insufficient concentration compromises the quality of the backfill paste, posing significant safety hazards. Consequently, developing an efficient model to predict underflow concentrations for CTS is vitally important. The CTS processes a low-concentration crude slurry flow (about 20%–30%), combining it with a flocculant to improve settling and aid in collecting dissolved particles in a mud bed. The system generates a properly concentrated feed flow from the bottom, while clean water from the overflow pipe is recycled for reuse. The primary control objective of the CTS is to maintain a consistent and precise underflow concentration, a critical parameter for evaluating the effectiveness and efficiency of the underground paste-filling process. To predict the underflow concentration, time-series architectures can leverage prior knowledge and historical data, enabling accurate and reliable forecasting to support operational safety and efficiency.

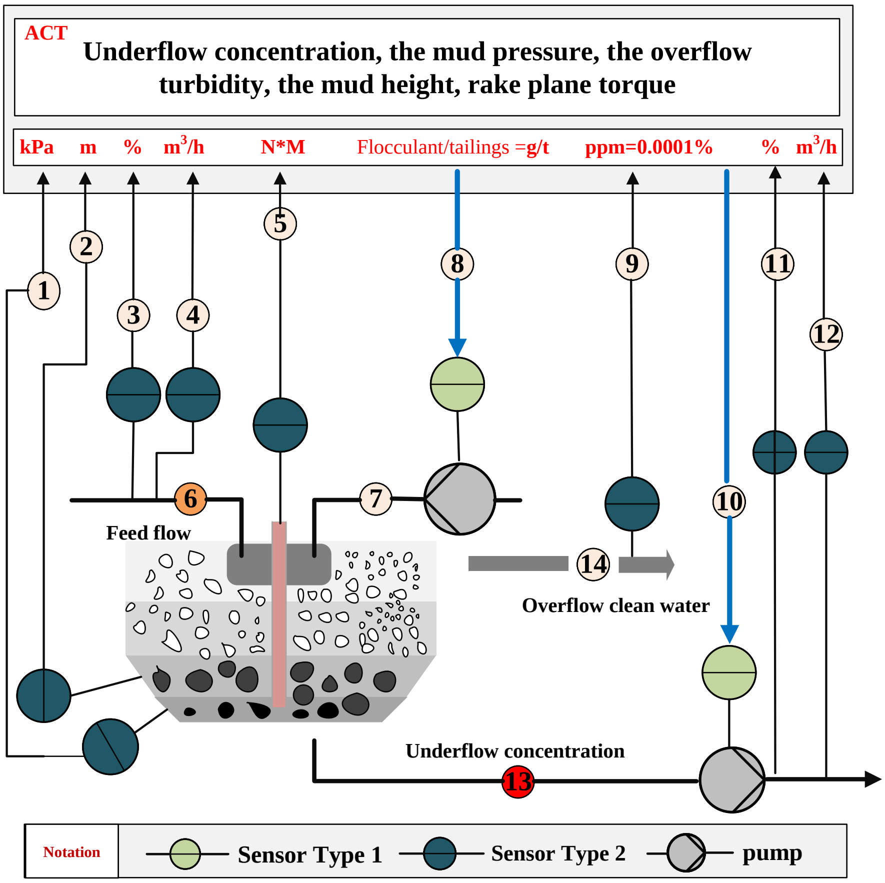

Figure 1. Deep Cone Thickener Unit.

ID	Process Parameter	Measurement Unit
1	Mud well pressure	kPa
2	Mud column height	m
3	Feed slurry concentration	wt%
4	Feed volumetric flow	m^3/h
5	Rake torque	N·m
6	Tailings production rate	t/h
7	Flocculant dosage	g/t
8	Flocculant feed velocity	m/s
9	Overflow turbidity	ppm
10	Rake rotational speed	rpm
11	Historical underflow conc. (t-interval)	wt%
12	Underflow volumetric flow	m^3/h
13	Underflow line pressure	kPa
14	Overflow channel concentration	wt%

Figure 2. Specification of Process Variables

2.2. Multi-horizon forecasting problem statement

In this section, a new sequence-to-sequence direct data-driven framework that generates multiple horizon quantile forecasts with uncertainty consideration for the industrial CTS is presented. In addition, the traditional multi-horizon forecast problem is shown in Figure 3. It is represented as:

$$p(y_{t+k,i}, \cdots, y_{t+1,i} \mid y_{:t,i}, x^{(h)}_{:t,i}, x^{(f)}_{t:,i}, x^{(s)}_{i}), \tag{1}$$

where $y_{\cdot,i}$ represents the ith time series to be predicted, while $x^{(h)}_{:t,i}$ denotes historical temporal covariates, $x^{(f)}_{t:,i}$ contains future information, and $x^{(s)}_{i}$ includes static, time-invariant features. Each time series serves as an individual input for a sequential neural network, such as an RNN or CNN.

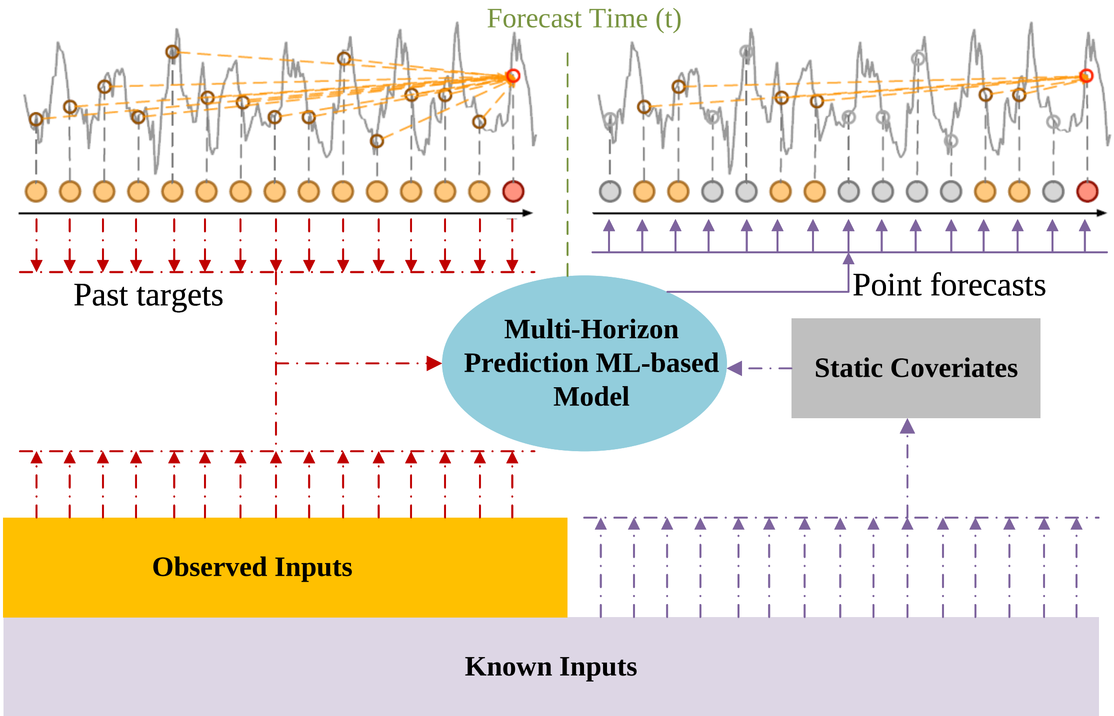

Figure 3. Illustration of multi-horizon forecasting.

2.3. Related counterparts

All the previous time series prediction methods can be summarized and classified into the following two categories, deterministic models, probabilistic models with a statistical viewpoint, or recursive single-point prediction, and multi-horizon strategy.

Deterministic models: RNNs and CNNs are two kinds of basic units for the time-series models [26]. Quantile time series modeling with deep learning has been applied in [13]. Zhang *et al.* discussed the comparative study for the time-series RNN model [27]. Some other univariate and multivariate attention models are also discussed in [23, 28, 29]. Recently, Huang *et al.* evaluated various multi-step strategies applied to a multi-layer perceptron, highlighting the effectiveness of the direct data-driven approach [30]. *Probabilistic models:* Regarding probabilistic prediction with encoder-decoder models, a popular prediction method, DeepAR, a Seq2Seq architecture, was proposed [31]. DeepAR is similar to a neural network that predicts Gaussian parameters and such a strategy. Zhang *et al.* investigated the latent adversarial regularized autoencoder for the high-dimensional probabilistic time prediction, however, it also requires computing the output distribution [32]. Koo *et al.* designed a quantile autoregressive neural net for the stretching application [33]. The quantile idea is used into these presented models.

LSTM is a type of model designed to retain historical data, effectively circumventing the long-term dependency challenges often encountered in RNNs [34]. This capability is facilitated by a memory gate that preserves information from prior contexts. At its core, the basic LSTM unit features three gates—input, output, and forget gates—that collectively protect and manage the cell state, ensuring the selective retention and updating of information. An advanced LSTM configuration encompasses four interacting layers, including repeating modules, which further refine its processing capabilities. These layers enable the LSTM to adeptly toggle between gates, establishing a durable temporal memory that mitigates the issue of vanishing gradients, thus enhancing the model's ability to learn from long sequence data.

Denote the historical vector $x_{(t)}$ and the hidden state $h_{(t-1)}$, and the previous inputs $c_{(t-1)}$. Then, the forget gate is:

$$f(t) = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

The gate of input and new candidate vectors are formulated:

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right), \\ \tilde{C}_t &= \tanh\left(W_{(C)} \cdot [h_{t-1}, x_t] + b_C\right). \end{aligned} \tag{3}$$

The new state of an LSTM cell is updated by

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \tag{4}$$

The output can be given by

$$\begin{aligned} o_t &= \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right), \\ h_t &= o_t * \tanh\left(C_t\right). \end{aligned}$$

Here, σ denotes the activation function, typically the Sigmoid function, while tanh represents the nonlinear tangent activation. The symbol $*$ indicates point-wise multiplication. Parameters W_c, W_o, and W_i correspond to the associated weights, whereas b_i, b_c, and b_o represent the respective biases. A variant of bidirectional LSTM is presented in Figure 4.

2.4. Underflow concentration multi-horizon prediction

In CTS, concentration is intertwined with numerous other variables [35]. Figure 2 presents all potential variables within the CTS system. In practice, the myriad of interdependent variables and the often opaque nature of their relationships with the target variable complicate the identification of relevant predictors. Furthermore, establishing the appropriate prediction horizon and the extent of historical observations required poses additional challenges. To address these complexities, a coefficient corelation analysis was conducted to identify key variables in the cone thickener production process. The selection of vital variables is a pivotal step in constructing a data-driven prediction model. Through mechanical analysis, mud height and pressure were identified as significant factors due to their nonlinear correlations with the target variable and were thus incorporated into the model. Building upon the foundation laid by [35], the final input variables for the model include mud pressure, mud height, the most recent feed flow rate, the flow rate of the highest volume, underflow density, and historical underflow concentration. Consequently, this sets the stage for multi-horizon prediction of underflow concentration in CTS systems.

$$\begin{aligned} & p(y_{t+k,i},\cdots,y_{t+1,i} \mid y_{:t,i}, x^{(h)}_{:t,i}, x^{(f)}_{t:,i}, x^{(s)}_{i}) \\ & = p(y_{t+k,i},\cdots,y_{t+1,i} \mid \langle \mathbf{C}_{i(t)}, \mathbf{Q}_{i(t)}, \mathbf{F}_{(t)}, \mathbf{C}_{o(t)}, \mathbf{Q}_{o(t)} \rangle) \\ & = p(y_{t+k,i},\cdots,y_{t+1,i} \mid \mathbf{W} \cdot NN \sum_{i=1}^{L} \langle y_i, y'_i \rangle \end{aligned} \tag{6}$$

where NN represents the neural network-based predictor, $\mathbf{W}$ is the corresponding weights. The input variables are denoted by the matrix $[C_i, Q_i, F, C_o, Q_o]$. However, prediction intervals is very insightful for the control and operation. So, Equation (6) with quantile addition can be rewritten as:

$$\begin{aligned} & p(y_{t+k,i},\cdots,y_{t+1,i} \mid y_{:t,i}, x^{(h)}_{:t,i}, x^{(f)}_{t:,i}, x^{(s)}_{i}, q) \\ & = p(y_{t+k,i},\cdots,y_{t+1,i} \mid \langle \mathbf{C}_{i(t)}, \mathbf{Q}_{i(t)}, \mathbf{F}_{(t)}, \mathbf{C}_{o(t)}, \mathbf{Q}_{o(t)} \rangle, q) \\ & = p(y_{t+k,i},\cdots,y_{t+1,i} \mid \mathbf{W} \cdot NN \sum_{i=1}^{L} \langle y_i, y'_i \rangle \end{aligned} \tag{7}$$

which implements the predicted q^{th} quantile of the k-step-ahead forecast at instant t.

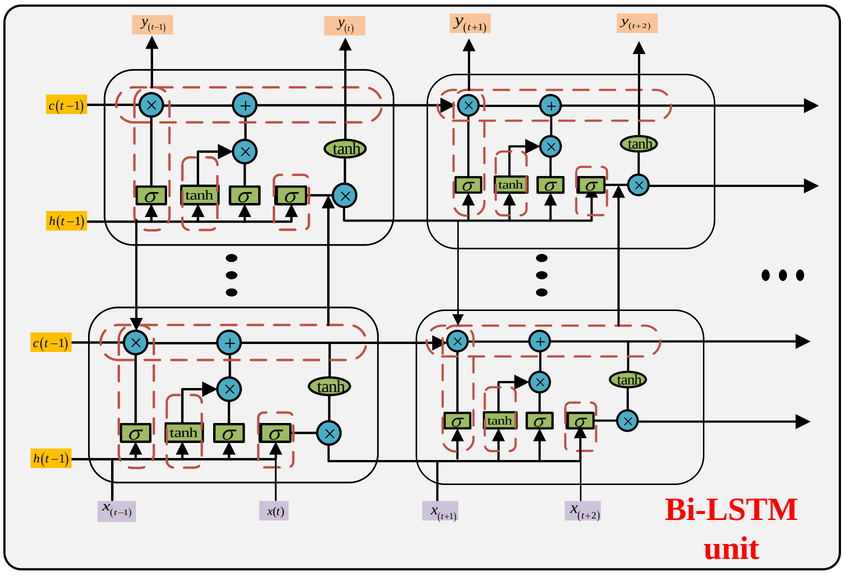

Figure 4. The basic illustration of BiLSTM unit.

3. Methodology

The primary objective of this segment is to address the challenge outlined in (7), which entails determining the optimal weights for the data-driven model through the training process, given all the input data. This process aims to accurately generate the quantile underflow concentration output. Detailed descriptions of the proposed model are given as following sections, encompassing the encoder representation, the specially designed Gated Recurrent Network (GRN), and the decoder output, respectively.

3.1. Encoder representation

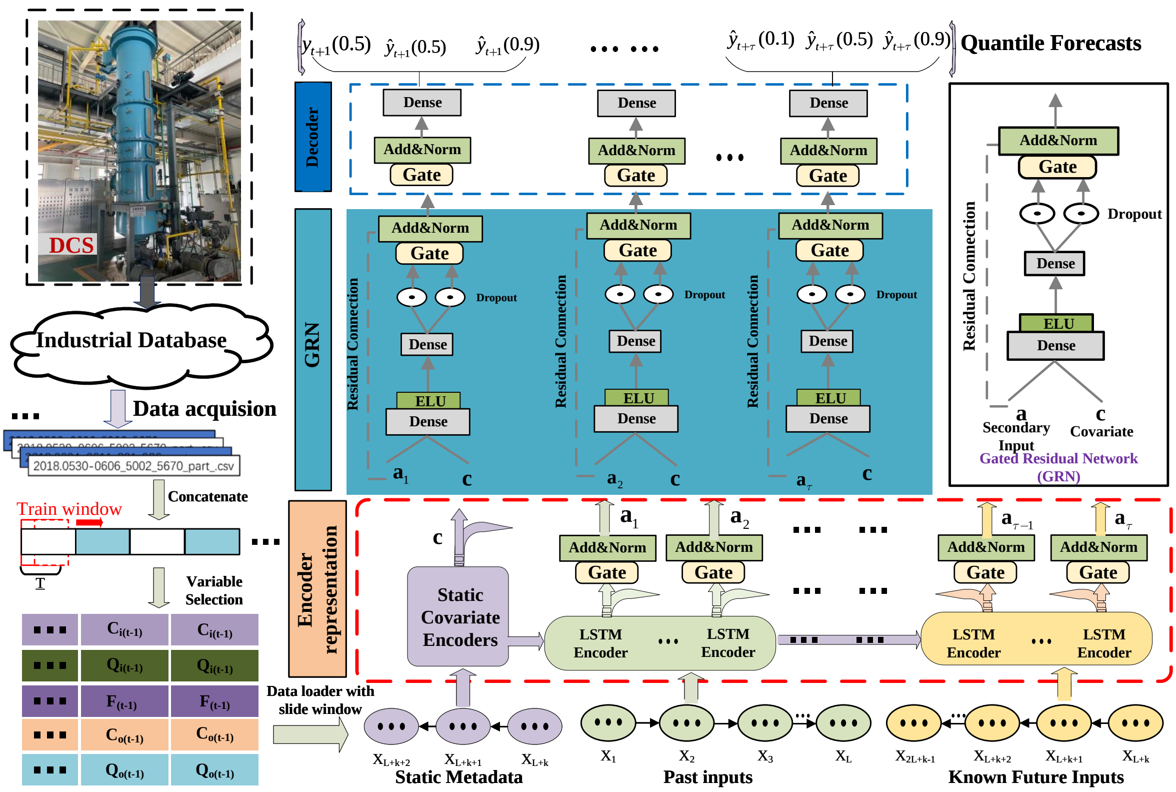

Figure 5. Flowchart of the proposed DDQRF.

As shown in Figure 5, the original data is preprocessed into three subsequent data, the static metadata, past inputs and the known future inputs. The static metadata is calculated by

$$\mathbf{c} = Cov(X_{L+k}, \ldots, X_{2L+k}) = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})], i \neq j \tag{8}$$

where μ_{X_i} and μ_{X_j} are the different expectations for the static metadata.

The past inputs and future inputs are both transformed to the LSTM, inherited by the Gate mechanism and normalization term, then the output of the first LSTM encoder would be

$$\phi(X, c) = LSTM(W, X, c) \tag{9}$$

Then the final output after the gate mechanism and layer normalization is described as

$$\mathbf{a} = LayerNorm(GLU(\phi(X,c))) \tag{10}$$

3.2. Gated recurrent network

Motivated by the nonlinear representability ability of neural networks, a new architecture is proposed as the building block of the proposed framework. The GRN takes the secondary inputs that extract from the LSTM encoders **a** and static covariate **c** as the inputs, then yields:

$$\begin{aligned} GRN(\mathbf{a},\mathbf{c}) &= Norm(\mathbf{a} + GLU(\eta_1)) \\ \eta_1 &= \mathbf{W_1}\eta_2 + \mathbf{b_1} \\ \eta_2 &= ELU(\mathbf{W_2 a} + \mathbf{W_3 c} + \mathbf{b_2}) \\ GLU(\gamma) &= \sigma(\mathbf{W_4}\gamma + \mathbf{b_4}) \odot (\mathbf{W_5}\gamma + \mathbf{b_5}) \end{aligned} \tag{11}$$

where ELU is the exponential kernel. **b** is the bias of the ELU.

3.3. Quantile decoder

As presented in previous work, the underflow concentration is continuous and is linked to the previous representation of the characteristics. In the designed framework, the decoder representation is given as:

$$\varphi(X,a,c) = LayerNorm(GLU(\xi)) \tag{12}$$

The decoder representation gives the high-level features of the underflow concentration prediction. In the top layer, the model generates the prediction intervals by generating various simultaneous quantiles. Probabilistic quantile forecasters are activated by the temporal decoder representation:

$$\hat{y}(q,t,k) = W_q \hat{\varphi}(X,a,c) + b_q \tag{13}$$

where k is the horizons in the future, $W_q \in \mathbb{R}^{1\times L}$, $b_q \in \mathbb{R}$ is the bias for the different quantile q. In the training, the objective is to minimize the total quantile loss $\mathscr{L}_q$:

$$\mathscr{L}_q(y,\hat{y},q) = q(y-\hat{y})_+ + (1-q)(\hat{y}-y)_+ \tag{14}$$

where $(\cdot)_+ = \max(0,\cdot)$. When $q = 0.5$, the $\mathscr{L}_q$ is simply the mean absolute error (MAE). Let τ be the maximum number of forecast horizons, $\mathscr{M}$ be the matrix of quantiles of interest, then the $\tau \times \mathscr{M}$ matrix $\hat{\mathbf{Y}} = \left[\hat{y}^{(q)}_{t+k}\right]_{k,q}$ is the output of a parametric model $g(y_{:t}, x, \theta)$, for example an RNN. $\sum_t \sum_q \sum_k \mathscr{L}_q(y_{t+k}, \hat{y}^{(q)}_{t+k})$ is the model loss, where t iterates through all forecast creation times (FCTs). In the underflow concentration prediction, the loss function would be the joint quantile loss, which is summed across all the outputs:

$$\mathscr{L}(\mathscr{N},\mathbf{W})=\sum_{y_t\in\mathscr{N}}\sum_{q\in\mathscr{M}}\sum_{k=1}^{k=\tau}\frac{\mathscr{L}_q(y,\hat{y},q)}{\mathscr{N}k} \tag{15}$$

where $\mathscr{N}$ is the domain field of all the data-driven training data. $\mathbf{W}$ is the corresponding weights. $\mathscr{M}$ is the quantile hyperparameter list that is manually denoted, in our experiment, $\mathscr{M}=[0.1,0.5,0.9]$. The definition of $(.)_+$ is given by $(.)_+=\max(0,.)$. Finally, the prediction framework presented is formulated and an intuitive demonstration is given in Figure 3.

4. Convergence analysis

The mean squared error (MSE) loss function is:

$$\mathbb{MSE}(y,\hat{y}):=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{16}$$

When the residuals' mean $\varepsilon:=y-\hat{y}$ is zero, minimizing this loss function results in estimating the conditional mean $\hat{Y}=E(Y|X)$.

Quantile regression modifies the MSE loss function to estimate conditional quantiles instead of means. The core idea is to transform quantile estimation from a sorting problem into a probability estimation. For example, for $q\in(0,1)$, the check function is defined as follows:

$$\rho_q(x):=\begin{cases}x(q-1) & \text{if } x<0\\ xq & \text{otherwise}\end{cases} \tag{17}$$

the associated mean quantile loss is given by

$$\mathbb{MQL}(y,\hat{y}):=\frac{1}{n}\sum_{i=1}^{n}\rho_q(y_i-\hat{y}_i) \tag{18}$$

The aim is to find $\hat{y}$ that minimizes

$$\begin{aligned}\mathbb{E}\left[\rho_q(Y-\hat{y})\right]&=(q-1)\int_{-\infty}^{\hat{y}}f(t)(t-\hat{y})dt+q\int_{\hat{y}}^{\infty}f(t)(t-\hat{y})dt\\&=q\int_{-\infty}^{\hat{y}}f(t)(t-\hat{y})dt-\int_{-\infty}^{\hat{y}}f(t)(t-\hat{y})dt\\&=q\int_{-\infty}^{\hat{y}}f(t)tdt-\hat{y}q\int_{-\infty}^{\hat{y}}f(t)dt-\int_{-\infty}^{\hat{y}}f(t)tdt\\+\hat{y}\int_{-\infty}^{\hat{y}}f(t)dt&=q\int_{-\infty}^{\hat{y}}f(t)tdt-\hat{y}q-\int_{-\infty}^{\hat{y}}f(t)tdt+\hat{y}F(\hat{y})\end{aligned}$$

where f is the PDF of Y. Differentiate with respect to $\hat{y}$ and set the expression equal to zero then

$$0 = -q - \hat{y}f(\hat{y}) + F(\hat{y}) - \hat{y}f(\hat{y}) = F(\hat{y}) - q \tag{19}$$

showing that $\hat{y} \in F^{-1}(q)$, i.e. that it is indeed a qth quantile.

5. Case study

In the evaluation design, different ablation studies are conducted. The first-class competitors are those deterministic recursive methods, which include the ARIMA, LSTM, and RNN. The other counterparts are those probabilistic strategy methods, such as DeepAR, and the Bayesian methods. The evaluation index RMSE can be given as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_t^{\,i} - y_t^i)^2} \tag{20}$$

In addition, MAE is identified as another evaluation index that yields:

$$MAE = \frac{1}{N_u}\sum_{i=1}^{N_u}|\widehat{y}_t - y_t| \tag{21}$$

The experimental setup mirrors that described in the reference by [21] with respect to the DualLSTM framework. The dropout rate is configured at 0.3, the state size is determined to be 160, and the prediction horizon is established at 20 time steps. For model training, the Adamdelta optimizer is employed for weight adjustments and for calculating the gradient descent of the loss function. The network's architecture encompasses an input layer, which can be comprised of RNN, LSTM, or bidirectional LSTM (BiLSTM) units, each with 128 hidden neurons. This layer is followed by a fully-connected layer featuring 64 neurons, and culminates in an output layer with a single neuron. A Gated Recurrent Network (GRN) model is also developed, featuring a dense architecture. Recognizing the impact of training epochs on prediction accuracy, the ablation study adjusts the number of iterative epochs in a progressively increasing sequence: [1, 5, 10, 50, 100, 150, 200, 500, 800, 1000], while maintaining all other hyperparameters at their original settings.

5.1. Validation and verification result analysis

To fulfill the validation and verification process, the LSTM and GRN are selected as the comparison benchmarks. The ablation study results are summarized in Figure 6. Figure 6a compares the different regressors with different quantiles. The observation underflow concentration is fully bounded by the upper and lower boundary, the Regressor 1 (green dash line) and Regressor 3 (the dotted line). It demonstrates that the proposed framework correctly predicts the underflow concentration in multiple horizons. On the other hand, Regressor 2 (the blue dot line), also the mean probabilistic prediction follows the actual underflow concentration value timidly, which means that the proposed method achieves a great performance. Similar characteristics are also discovered in the subsequent series, the presented method implements a satisfactory prediction performance. Compared to the presented approach, the LSTM, the black line in Figure 6a, shows an apparent residual error between the prediction and observation. Figure 6b show that the GRN and LSTM usually underestimate the underflow concentration, while the proposed Regressor 2 got the superior performance. Figure 6c also shows that in time instant $t = 1000$, the actual underflow concentration is around 65%, while the LSTM predicted value is 60%, such a prediction residual degrades the LSTM model's performance. Figure 6b verified the initial performance of those different methods, as shown in interval $[880, 980]$, the proposed three quantile regressors achieve accurate prediction, while the LSTM and GRN caused inaccurate underflow concentration prediction, with almost 15% and 10% residual errors, respectively.

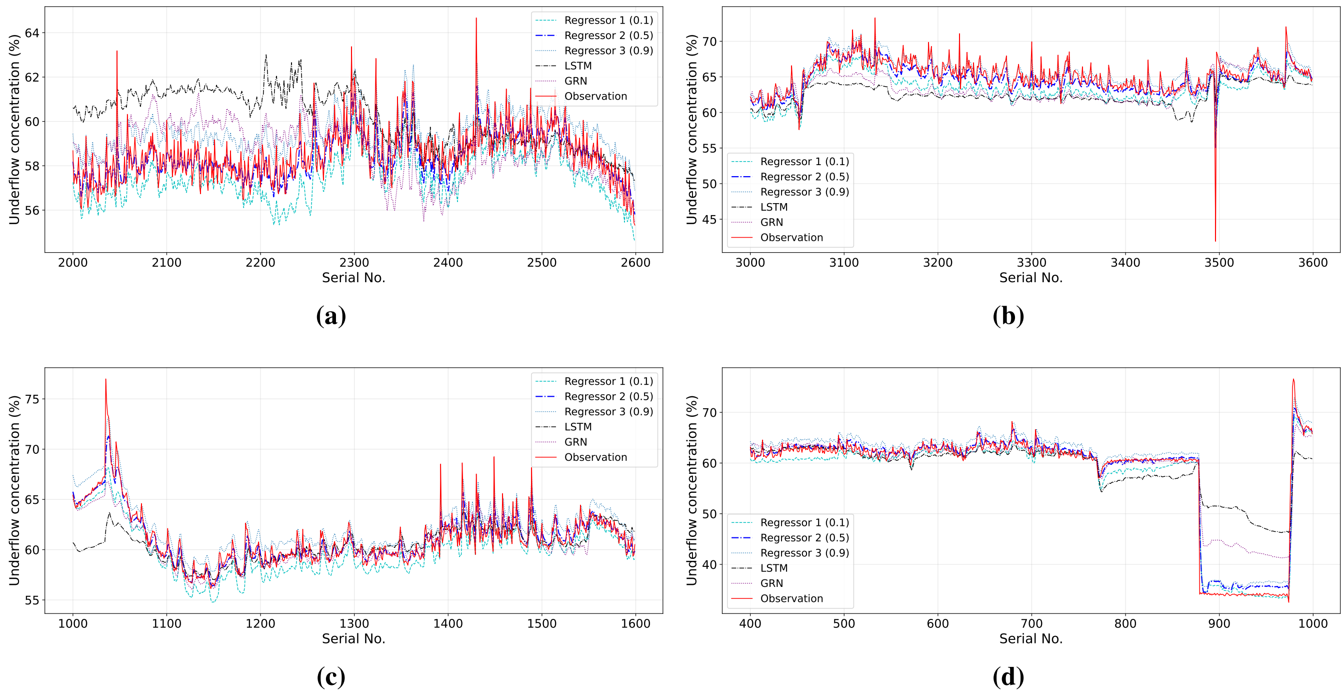

Figure 6. The proposed model for industrial underflow prediction results; (a)-(d) are different prediction intervals, respectively.

Regarding verification, the method introduced has been implemented in the CTS system. Figure 7 presents the application performance of the proposed quantile forecaster. Figures 7a–7d describe 4 different industrial applicative series. The main feature is that all the observations are bounded by an 80% prediction interval, which is computed by the prediction upper boundary ($q = 0.9$) and lower boundary ($q = 0.1$). The proposed regressor covers the observation very well, even in some abrupt oscillation intervals, such as in an interval $[900, 1000]$. However, the proposed method also shows a weakness, which is the ability to cucumber the situation of a single outlier, the intuitive point is shown in Figure 7a, around instant $t = 220$, and Figure 7d, instant $t = 3500$. This outcome is justified given that the accuracy surpasses that of other contemporary approaches, which demonstrate inferior capabilities in handling outliers. Moreover, within practical applications, a singular outlier does not constitute the primary concern that requires prioritization.

The more detailed ablation study results are given in Figure 7. An increasing epoch with different quantiles is conducted while keeping other hyper-parameter fixed. Figure 7 gives the intuitive result presentation of the MSE, associated with the bar, and RMSE with lines. Compared to $q = 0.1$ and $q = 0.9$, $q = 0.5$'s performance outperforms others, with the lowest MSE and RMSE. It also shows that when the epoch is set as 100, associated with quantile coefficient $q = 0.5$, the lowest MSE 0.662, is obtained. Therefore, as fully explored, for the deployment verification, the optimal epoch is set as 100. When the epoch is set as 50, the $q = 0.1$ and $q = 0.9$ achieves the highest RMSE, 0.824 and 0.827, respectively, while the $q = 0.5$ is 0.814. The result explores the optimal hyper-parameters for the proposed framework.

Table 1 gives the comparison of the different competitors, which covers the deterministic methods SVR [23], MPA-RNN [28], SS-PdeepFM [29], DSTP-RNN [36] and other probabilistic approaches DualLSTM [21], DeepAR [31]. The recursive point strategy DA-RNN [37], DA-LSTM [26]. The proposed method is scope of multi-horizon, and probabilistic. Compared to the basic LSTM, the GRN improves the prediction ability. While the GRN's feature learning ability is also limited. Therefore, the proposed method has a lower MSE and RMSE, compared to its own basic benchmarks. Despite relying on a singular feature, the proposed method successfully incorporates these benefits.

Regarding verification, the method introduced has been implemented in the CTS system. Figure 7 presents the application performance of the proposed quantile forecaster. Figures 7a–7d describe 4 different industrial applicative series. The main feature is that all the observations are bounded by an 80% prediction interval, which is computed by the prediction upper boundary ($q = 0.9$) and lower boundary ($q = 0.1$). The proposed regressor covers the observation very well, even in some abrupt oscillation intervals, such as in an interval $[900, 1000]$. However, the proposed method also shows a weakness, which is the ability to cucumber the situation of a single outlier, the intuitive point is shown in Figure 7a, around instant $t = 220$, and Figure 7d, instant $t = 3500$. This outcome is justified given that the prediction accuracy of the proposed method surpasses that of other contemporary approaches, which demonstrate inferior capabilities in handling outliers. Moreover, within practical applications, a singular outlier does not constitute the primary concern that requires prioritization.

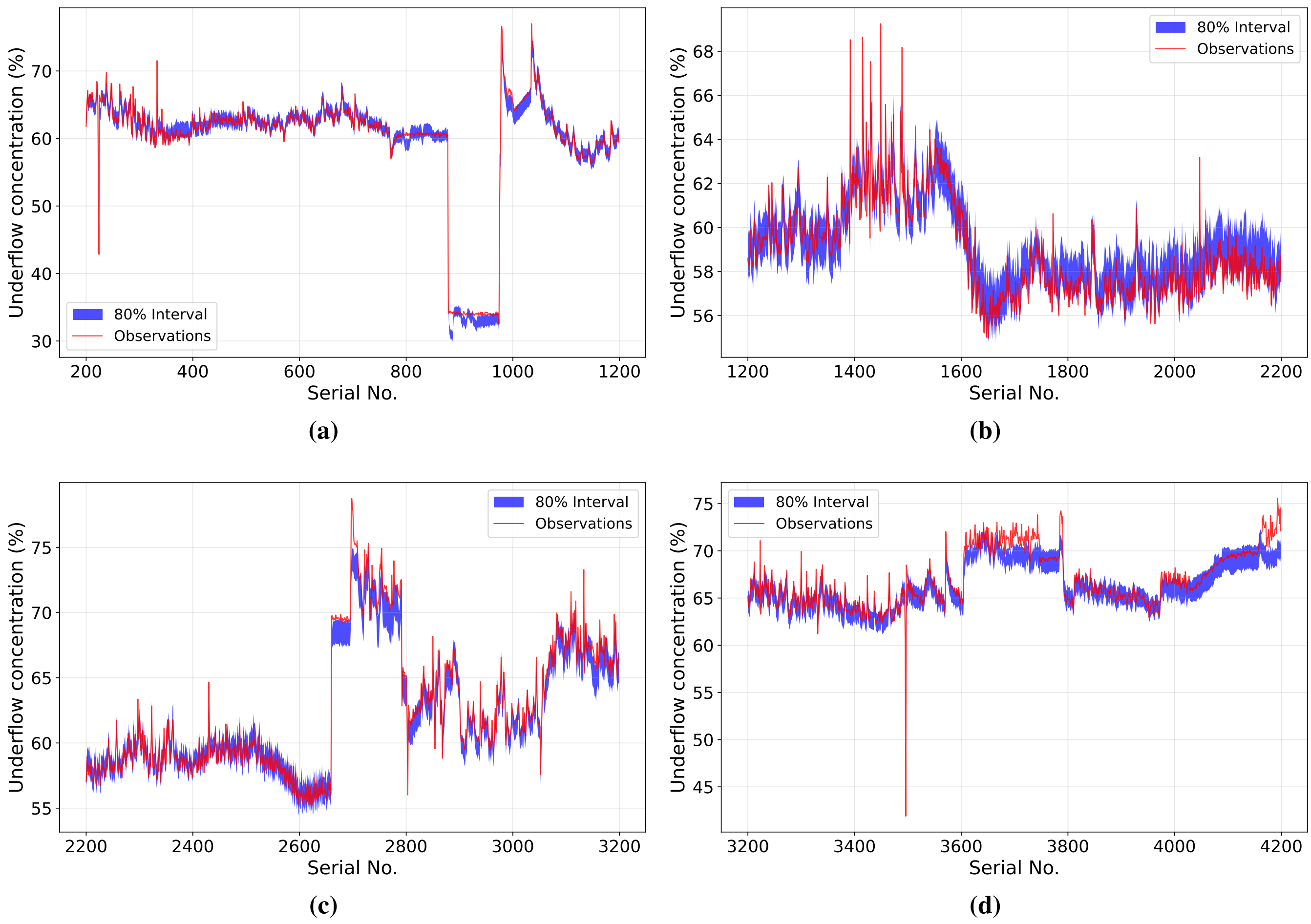

Figure 7. The proposed DDQRF model for industrial underflow prediction results. The subfigure is the different time series and all the results are conducted on the Testing metadata.

The more detailed ablation study results are given in Figure 7. An increasing epoch with different quantiles is conducted while keeping other hyper-parameter fixed. Figure 8 gives the intuitive result presentation of the MSE, associated with the bar, and RMSE with lines. Compared to $q = 0.1$ and $q = 0.9$, $q = 0.5$'s performance outperforms others, with the lowest MSE and RMSE. It also shows that when the epoch is set as 100, associated with quantile coefficient $q = 0.5$, the lowest MSE 0.662, is obtained. Therefore, as fully explored, for the deployment verification, the optimal epoch is set as 100. When the epoch is set as 50, the $q = 0.1$ and $q = 0.9$ achieves the highest RMSE, 0.824 and 0.827, respectively, while the $q = 0.5$ is 0.814. The result explores the optimal hyper-parameters for the proposed framework.

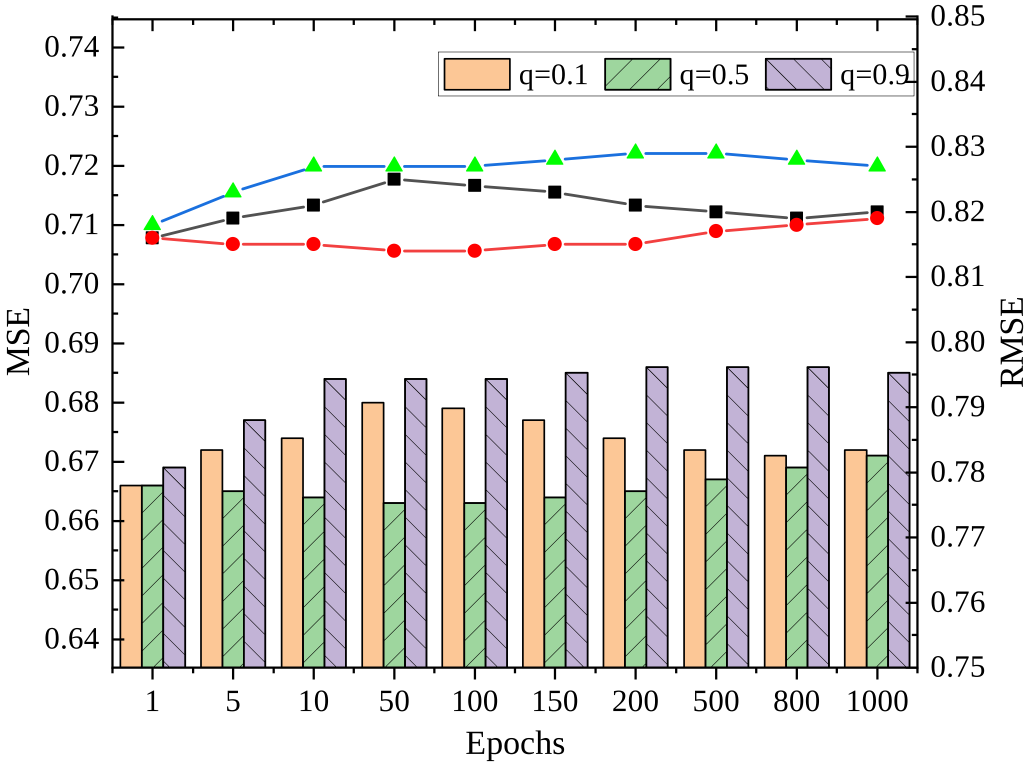

Figure 8. The error performance comparison (RMSE and MAE) for the different methods in an increasing epoch. Specifically, the quantiles of our proposed method are based on 0.1, 0.5, and 0.9, respectively. The rmse legend for quantile q = 0.5 is marked with red, q = 0.9 is marked in a blue triangle, and q = 0.1 is marked in a black rectangle.

Table 1 gives the comparison of the different competitors, which covers the deterministic methods SVR [23], MPA-RNN [28], SS-PdeepFM [29], DSTP-RNN [36] and other probabilistic approaches DualLSTM [21], DeepAR [31]. The recursive point strategy DA-RNN [37], DA-LSTM [26]. The proposed method is scope of multi-horizon, and probabilistic. Compared to the basic LSTM, the GRN improves the prediction ability. While the GRN's feature learning ability is also limited. Therefore, the proposed method has a lower MSE and RMSE, compared to its own basic benchmarks. Despite relying on a singular feature, the proposed method successfully incorporates these benefits.

Table 1. Evaluation indexes for the different competitive prediction methods. Compared to the existing state-of-the-art deterministic approaches, our proposed benchmark implements probabilistic data-driven forecasting, and also multiple horizon prediction.

Categories	Methods	RMSE	MAE
Deterministic	RNN	0.6712 ± 0.0124	0.3488 ± 0.0082
	GRN	0.4634 ± 0.0078	0.2499 ± 0.0345
	LSTM	0.7533 ± 0.0030	0.5119 ± 0.0079
	SVR	0.8133 ± 0.0002	0.6003 ± 0.0013
	BiLSTM	0.8091 ± 0.0092	0.6083 ± 0.0118
	GSTAE	0.7735 ± 0.0068	0.5296 ± 0.0097
	SS-PdeepFM	0.7650 ± 0.0036	0.5206 ± 0.0073
Probabilistic	DualLSTM	0.7985 ± 0.0116	0.5847 ± 0.0164
	DeepAR	0.7571 ± 0.0033	0.5235 ± 0.0092
	Suggested	**0.6654 ± 0.0021**	**0.631 ± 0.0013**
Recursive point strategy	ARIMA	0.7515 ± 0.0034	0.2074 ± 0.0072
	LSTM-DeepFM	0.7497 ± 0.0026	0.5091 ± 0.0071
	MPA-RNN	0.7515 ± 0.0034	0.2074 ± 0.0072
	DA-RNN	0.851 ± 0.0029	0.2322 ± 0.0052
	DA-LSTM	0.851 ± 0.0029	0.2322 ± 0.0052
	Attention	0.9086 ± 0.0021	0.2445 ± 0.0031
Multiple horizon strategy	DeepAR	0.9086 ± 0.0021	0.2445 ± 0.0031
	DSTP-RNN	0.8496 ± 0.0051	0.2254 ± 0.0042
	Suggested	**0.6654 ± 0.0021**	**0.631 ± 0.0013**

6. Conclusion

This study introduces a high-efficiency, key-quality model for predicting underflow concentration in deep cone thickener systems, aiming to overcome the challenges associated with underflow prediction in industrial CTS facilities. It unveils a pioneering approach to soft sensing of underflow concentration in the industrial paste filling process through probabilistic methods. Markedly, this research is the first to apply probabilistic forecasting of underflow concentration, offering an alternative to the existing deterministic approaches. Furthermore, it distinguishes itself from the probabilistic Bayesian networks by proposing a novel quantile regression framework equipped with multi-horizon forecasting capabilities. This model sets a benchmark for similar process industries, providing a readily adaptable and utilizable solution. The proposed method leverages the nonlinear mapping capabilities of time sequence models alongside the uncertainty quantification of probabilistic models. Validation and verification experiments conducted within an industrial CTS context have demonstrated its superior performance compared to existing methods. Future research will concentrate on addressing the instability caused by outliers, as highlighted in previous discussions. Exploring and integrating algorithms that mitigate the impact of outliers in our probabilistic regression forecasting model represents a promising avenue for further investigation.

Acknowledgement

This work was partially supported by the Italian Ministry of Education, University and Research through the Project "Department of Excellence LIS4.0-Lightweight and Smart Structures for Industry 4.0" and in part by the Horizon Marie Skłodowska-Curie Actions program (101073037) and China Scholarship Council (202006370101).

Author's contribution

Conceptualization, Yongxiang Lei; methodology, Yongxiang Lei; software, Yongxiang Lei; validation, Yongxiang Lei; resources, Yongxiang Lei; data curation, Yongxiang Lei; writing—original draft preparation, Yongxiang Lei; writing—review and editing, Yongxiang Lei; visualization, Yongxiang Lei; supervision, Hamid Reza Karimi; project administration, Hamid Reza Karimi; funding acquisition, Hamid Reza Karimi. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Bahrami O, Wang W, Hou R, Lynch JP. A sequence-to-sequence model for joint bridge response forecasting. *Mech. Syst. Signal Process.* 2023, 203:110690.

[2] Huan W, Ting L, Yuning C, Aixiang W. Underflow concentration prediction model of deep-cone thickener based on data-driven. *J. China Univ. Posts Telecomm.* 2019, 26(6):63.

[3] Tan CK, Setiawan R, Bao J, Bickert G. Studies on parameter estimation and model predictive control of paste thickeners. *J. Process Control* 2015, 28:1–8.

[4] Fang C, He D, Li K, Liu Y, Wang F. Image-based thickener mud layer height prediction with attention mechanism-based CNN. *ISA Trans.* 2022, 128:677–689.

[5] Takács I, Patry GG, Nolasco D. A dynamic model of the clarification-thickening process. *Water. Res.* 1991, 25(10):1263–1271.

[6] Xiao D, Xie H, Jiang L, Le BT, Wang J, *et al.* Research on a method for predicting the underflow concentration of a thickener based on the hybrid model. *Eng. Appl. Comput. Fluid Mech.* 2020, 14(1):13–26.

[7] Yuan Z, Hu J, Wu D, Ban X. A dual-attention recurrent neural network method for deep cone thickener underflow concentration prediction. *Sensors* 2020, 20(5):1260.

[8] Valcamonico D, Baraldi P, Zio E, Decarli L, Crivellari A, *et al.* Combining natural language processing and bayesian networks for the probabilistic estimation of the severity of process safety events in hydrocarbon production assets. *Reliab. Eng. Syst. Saf.* 2024, 241:109638.

[9] Box GE, Jenkins GM, Reinsel GC, Ljung GM. *Time series analysis: Forecasting and control*, John Wiley & Sons,2015.

[10] Jaderberg M, Simonyan K, Zisserman A, *et al.* Spatial transformer networks. *Advances in neural information processing systems* 2015, 28.

[11] Torossian L, Picheny V, Faivre R, Garivier A. A review on quantile regression for stochastic computer experiments. *Reliab. Eng. Syst. Saf.* 2020, 201:106858.

[12] Mei LF, Yan WJ, Yuen KV, Ren WX, Beer M. Transmissibility-based damage detection with hierarchical clustering enhanced by multivariate probabilistic distance accommodating uncertainty and correlation. *Mech. Syst. Signal Process.* 2023, 203:110702.

[13] Taylor JW. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *J. Forecasting* 2000, 19(4):299–311.

[14] Xiang S, Qin Y, Luo J, Wu F, Gryllias K. A concise self-adapting deep learning network for machine remaining useful life prediction. *Mech. Syst. Signal Process.* 2023, 191:110187.

[15] Yang D, Karimi HR, Sun K. Residual wide-kernel deep convolutional auto-encoder for intelligent rotating machinery fault diagnosis with limited samples. *Neural Networks* 2021, 141:133–144.

[16] Zhang Y, Zhang M, Liu C, Feng Z, Xu Y. Reliability enhancement of state of health assessment model of lithium-ion battery considering the uncertainty with quantile distribution of deep features. *Reliab. Eng. Syst. Saf.* 2024, 245:110002.

[17] Marrel A, Iooss B. Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation. *Reliab. Eng. Syst. Saf.* 2024, 247:110094.

[18] Lei Y. Intelligent Optimal Framework for the industrial mining plant-wide prediction control. *IEEE Trans. Instrum. Meas.* 2023, 72:1–12.

[19] Mosallam A, Medjaher K, Zerhouni N. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *J. Intell. Manuf.* 2016, 27:1037–1048.

[20] Liu K, Shang Y, Ouyang Q, Widanage WD. A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery. *IEEE Trans. Ind. Electron.* 2020, 68(4):3170–3180.

[21] Lei Y, Karimi HR. DualLSTM: A novel key-quality prediction for a hierarchical cone thickener. *Control Eng. Pract.* 2023, 137:105566.

[22] Lei Y, Karimi HR. Underflow concentration prediction based on improved dual bidirectional LSTM for hierarchical cone thickener system. *Int. J. Adv. Manuf. Technol.* 2023, pp. 1–12.

[23] Yuan X, Qi S, Wang Y, Xia H. A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data. *Control Eng. Pract.* 2020, 104:104614.

[24] Yuan X, Li L, Shardt YA, Wang Y, Yang C. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Trans. Ind. Electron.* 2020, 68(5):4404–4414.

[25] Xie W, Wang J, Xing C, Guo S, Guo M, *et al.* Variational autoencoder bidirectional long and short-term memory neural network soft-sensor model based on batch training strategy. *IEEE Trans. Ind. Inf.* 2020, 17(8):5325–5334.

[26] Zhu Y, Xie S, Xie Y, Chen X. Temperature prediction of aluminum reduction cell based on integration of dual attention LSTM for non-stationary sub-sequence and ARMA for stationary sub-sequences.

Control Eng. Pract. 2023, 138:105567.

[27] Zhang W, Quan H, Srinivasan D. An improved quantile regression neural network for probabilistic load forecasting. *IEEE Trans. Smart Grid* 2018, 10(4):4425–4434.

[28] Geng J, Yang C, Li Y, Lan L, Luo Q. MPA-RNN: a novel attention-based recurrent neural networks for total nitrogen prediction. *IEEE Trans. Ind. Inf.* 2022, 18(10):6516–6525.

[29] Ren L, Wang T, Laili Y, Zhang L. A data-driven self-supervised LSTM-DeepFM model for industrial soft sensor. *IEEE Trans. Ind. Inf.* 2021, 18(9):5859–5869.

[30] Huang F, Deng Y. A spatiotemporal deep neural network for fine-grained multi-horizon wind prediction. *Data Min. Knowl. Discovery* 2023, 37:1441–1472.

[31] Salinas D, Flunkert V, Gasthaus J, Januschowski T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecasting* 2020, 36(3):1181–1191.

[32] Zhang J, Dai Q. Latent adversarial regularized autoencoder for high-dimensional probabilistic time series prediction. *Neural Networks* 2022, 155:383–397.

[33] Koo E, Kim H. Empirical strategy for stretching probability distribution in neural-network-based regression. *Neural Networks* 2021, 140:113–120.

[34] Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. *IEEE Trans. Neural Networks Learn. Syst.* 2016, 28(10):2222–2232.

[35] Lei Y, Karimi HR, Cen L, Chen X, Xie Y. Processes soft modeling based on stacked autoencoders and wavelet extreme learning machine for aluminum plant-wide application. *Control Eng. Pract.* 2021, 108:104706.

[36] Liu Y, Gong C, Yang L, Chen Y. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Syst. Appl.* 2020, 143:113082.

[37] Qin Y, Song D, Chen H, Cheng W, Jiang G, *et al.* A dual-stage attention-based recurrent neural network for time series prediction. *arXiv* 2017, arXiv:1704.02971.