Available online at website: https://journal.bcrec.id/index.php/jcerp

Journal of Chemical Engineering Research Progress, 2 (2) 2025, 232-245

Research Article

Hybrid Approaches for Steam Demand Forecasting: Combining First Principles, Box and Jenkins, and Neural Network Models

Davies Agommuoh[1]*, Antony Higginson[1], Kevin Brooks[1], Philip de Vaal[2]

[1]Data Analytics and Numerics in Control Engineering Unit, Department of Chemical and Metallurgical Engineering, University of the Witwatersrand, Johannesburg, South Africa.
[2]Department of Chemical Engineering, University of Pretoria, Pretoria, South Africa.

Received: 9th September 2025; Revised: 20th September 2025; Accepted: 23th September 2025
Available online: 26th September 2025; Published regularly: December 2025

Abstract

The pulp and paper industry relies heavily on batch sulphite digesters for chemical cellulose production, where steam is a critical utility – lack of steam prediction results in venting losses, increased costs, and a negative environmental impact. Accurate prediction of steam demand is therefore essential for optimising digester cooking cycles and resource allocation. This study aims to develop and compare predictive models for steam demand in batch pulp digesters using magnesium bisulphite cooking liquor. Three years of production data were pre-processed to extract digester temperature profiles and batch steam demands. Seven modelling approaches were evaluated: a mechanistic first-principles energy balance model, Box–Jenkins ARIMA, two neural network models (LSTM and CNN), and three hybrid models combining first-principles with ARIMA, LSTM, and CNN. The hybrid frameworks employed dimensionless parameters from the mechanistic model as exogenous variables to compensate for unavailable process data. Model accuracy was assessed using RMSE and MAE metrics. The results show that hybrid models consistently outperformed their standalone counterparts. In particular, the hybrid first-principles–CNN model achieved the highest predictive accuracy, demonstrating the CNN's ability to extract features and capture nonlinear temporal dependencies in steam demand. The hybrid first-principles–ARIMA model also surpassed both the standalone ARIMA and mechanistic models. Integrating mechanistic insights with data-driven methods significantly enhances prediction accuracy in complex batch processes. The findings highlight the value of hybrid modelling strategies for improving steam demand forecasting, with potential benefits for process optimisation, energy efficiency, and batch scheduling in the pulp and paper industry.

Keywords: Steam demand forecast; Batch pulp digester; Hybrid Modelling; Neural Networks; First-principles models

How to Cite: Agommuoh, D., Higginson, A., Brooks, K., de Vaal, P. (2025). Hybrid Approaches for Steam Demand Forecasting: Combining First Principles, Box and Jenkins, and Neural Network Models. *Journal of Chemical Engineering Research Progress*, 2 (2), 232-245 (doi: 10.9767/jcerp.20485)

Permalink/DOI: https://doi.org/10.9767/jcerp.20485

1. Introduction

The pulp and paper industry is a key player in global manufacturing, producing essential materials such as chemical cellulose (dissolving pulp), which is used in various products, ranging from paper to food additives [1]. This research addresses the challenge of predicting steam demand in batch pulp digesters used in the production of chemical cellulose, a key process in the pulp and paper industry. The production involves cooking wood chips in digesters using magnesium bisulphite and steam. The efficient use of steam is crucial, but the complexity of the batch process, which includes sequential phases across multiple digesters, makes it difficult to predict steam requirements accurately. Excess steam generation leads to inefficient venting, which is both economically and environmentally detrimental. Accurate steam demand prediction would enable optimal resource allocation, reduce waste, and improve plant scheduling.

To address this challenge, the study proposes the development of hybrid predictive models that

* Corresponding Author.
Email: davieschuks.a@gmail.com (D. Agommuoh)

combine first-principles (also called mechanistic models) with data-driven approaches, including Box-Jenkins ARIMA (Autoregressive Integrated Moving Average) model and neural networks (Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN)). While first-principles models use foundational knowledge of physics and chemistry to fill data gaps, black-box models like neural networks are well-suited for handling the complex, poorly understood variables in steam demand prediction [2]. ARIMA models, which are adept at capturing linear dependencies in time series data, complement neural networks, which excel at non-linear relationships [3].

Predictive modelling in chemical engineering has traditionally relied on first-principles (mechanistic) models, which provide valuable physical insights but are often limited by missing process data and their inability to capture nonlinear dynamics [4,5]. Data-driven methods, particularly neural networks, have emerged as strong alternatives, consistently outperforming mechanistic models across domains such as electrodialysis [6], supercritical solubility [7], fuel cells [8], and catalytic cracking [9], owing to their ability to handle nonlinearities and reduce computational complexity. More recently, hybrid models that integrate mechanistic models with neural networks or statistical approaches such as ARIMA have shown superior performance, combining mechanistic interpretability with data-driven flexibility. These hybrid approaches have improved forecasting in diverse applications, from wastewater treatment emissions [10] to time series problems such as water quality [11] and commodity prices [12], demonstrating their ability to capture both linear and nonlinear patterns. Despite these advances, the application of hybrid models remains limited in the pulp and paper sector. Most existing work has focused on predicting the Kappa number in kraft pulping processes, with little attention to batch sulphite digesters [13-15]. Furthermore, while recurrent neural networks (RNNs), particularly long short-term memory (LSTM) variants, have been explored for fault diagnosis [16,17], convolutional neural networks (CNNs) remain largely underutilised in time series regression for process engineering, as their applications have predominantly focused on image recognition tasks [18].

The study evaluates eight models: standalone first-principles, ARIMA, and neural network models (LSTM & CNN), along with hybrid models that couple first-principles with ARIMA, LSTM, and CNN. The first-principles model is grounded in the energy balance equation and is parameterised using temperature profile data, solved via optimisation techniques. The hybrid models integrate these first-principles insights to provide exogenous variables for the ARIMA and neural network models. Initial findings suggest that the hybrid models outperform their standalone counterparts.

There have been limited studies concerned with sulphite pulping processes, as most existing research has focused on kraft pulping and its well-established reaction kinetics. Therefore, this research intends to advance predictive modeling in the pulp and paper sector by investigating sulphite batch digesters and introducing convolutional neural networks (CNNs) for time-series regression, a technique rarely applied in process engineering. The novelty of this study lies in addressing the challenge of insufficient process data inherent in batch processes by rendering the governing energy balance equation into a dimensionless form, thereby generating key parameters that serve as exogenous inputs for data-driven models. This approach creates a form of "latent data" – parameters that cannot be directly measured in real time due to process complexity but can be inferred through hybrid modeling. The objectives of this research are: (i) to develop and implement a dimensionless first-principles framework to derive these latent parameters; (ii) to integrate these parameters as exogenous variables in the hybrid models; and (iii) to evaluate and identify the most effective model structure for forecasting steam demand in batch sulphite digesters under conditions of limited process data and unobservable reaction variables. This approach provides a systematic pathway for modeling complex chemical systems where process measurements are sparse or incomplete.

2. Methods

In this study, model development relies heavily on plant data for regression and training, distinguishing mechanistic from black-box models. Historical data was extracted using OsiSoft PI in Microsoft Excel, creating a detailed three-year dataset with 10-minute interval readings. Key parameters include total steam consumption per batch, steam temperature, and the time required. Due to operational similarities, the analysis focuses on a single digester. Data preprocessing, conducted in Python, addressed noise, missing values, and outliers, improving the dataset's reliability. The data, covering the period from January 2018 to June 2021, was collected from a pulp mill in CSV format and included variables such as temperature, steam consumption, and timestamps. Given shutdowns and equipment malfunctions, only periods of normal operation were analysed, with a continuous data stretch from June 2020 to January 2021 selected for model training. The dataset was divided into 462 batches using Python, defined by zero steam consumption before and after each batch, with each batch containing temperature and steam demand profiles.

2.1 First-Principles Model

The first-principles model begins with the energy balance for a batch digester in Equation (1) [19].

$$\frac{dT}{dt} = \frac{\dot{Q} - \dot{W}_S + (-\Delta H_{rx})(-r_A V)}{\sum N_i C_{Pi}} \tag{1}$$

Where, T is the temperature in Kelvin, t is time in seconds, $\dot{Q}$ is the rate of heat flow in J/s, $\dot{W}_s$ is the shaft work in J/s, ΔH_{rx} is the heat of reaction in the digester in J/mol, $-r_A$ is the rate of reaction of the limiting reagent, A (wood) in mol/L.s, V is the digester volume in litres, N_i is the number of moles of species i in moles, C_{Pi} is the specific heat of species i in J/mol.K. A schematic diagram of the batch digester illustrating the main reaction constituents (wood, liquor and pulp) is presented in Figure 1.

The assumptions made in the development of the first principles models are listed below: (1). First-order reaction kinetics in the digester; (2). Wood is simplified to its major component, cellulose, and treated as the limiting reagent; (3). The heat capacity of the mixture is the weighted sum of the components, assuming negligible heat capacity change during the reaction ($\Delta C_p \approx 0$); (4). $\Delta C_p \approx 0$ as the reaction happens in the solid-liquid phase; (5). The reaction involves heat transfer, with no shaft work, and the system operates in batch mode.

The denominator in Equation (1) can be expanded from the fourth assumption, $\Delta C_p \approx 0$

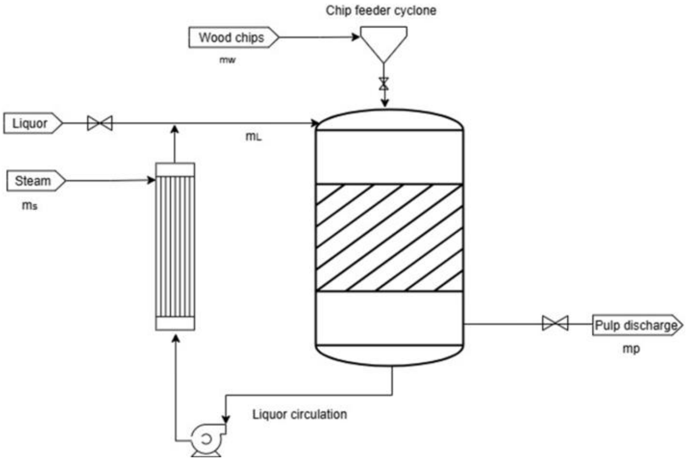

Figure 1. Schematic diagram of a batch pulp digester.

$$\sum N_i C_{Pi} = \sum N_{i0} C_{Pi} = N_{W0} \sum \theta_i C_{Pi} = N_{W0} C_{pm} \tag{2}$$

In Equation (2), subscripts W and L represent the reagents wood and liquor, respectively, while subscript 0 represents the variables at the input stage. Subscript m refers to the property of the mixture.

$$\theta_i = \frac{N_i}{N_{i0}} \text{ [19]}$$
$$C_{pm} = \sum_{i=1}^{n} C_{pi} y_i = C_{pW} y_W + C_{PL} y_L + C_{PP} y_P \tag{3}$$

Subscript P refers to the physical property of pulp, and y_i is the mole fraction of i in the mixture. The reaction rate can also be expanded from the first assumption.

$$r_A = -k[A] = -k[A]_0 e^{-kt} \tag{4}$$

k is the rate constant for a first-order reaction in 1/second, $[A]$ is the concentration of species A.

A stoichiometric table of the reaction in the digester determines the molar fractions, y_i in terms of the conversion. The resulting expression for the heat capacity of the mixture was incorporated into C_{Pi}, and the first-order reaction rate was substituted for $-r_A$. This resulted in Equation [5].

$$\frac{dT}{dt} = \frac{\dot{Q} + (-\Delta H_{rx})(N_{W0} k e^{-kt})}{N_{W0}(C_{P0} + y_{W0} X (C_{PP} - C_{PW}))} \tag{5}$$

Equation (5) was rendered dimensionless by introducing dimensionless variables for temperature (θ) and time (τ), resulting in Equation (6) with three dimensionless parameters σ, δ, ρ.

$$\frac{d\theta}{d\tau} = \sigma + \delta\rho \tag{6}$$
$$\theta = \frac{T - T_0}{T_0} \tag{7}$$
$$\tau = \frac{t}{t_b} \tag{8}$$
$$\sigma = \frac{(\dot{Q} t_b)}{C_{P0} T_0 N_{W0}} \tag{9}$$
$$\delta = \frac{Da(-\Delta H_{rx})}{C_{P0} T_0} \tag{10}$$
$$\rho = e^{-Da\tau} \tag{11}$$

Equation (6) was integrated to produce Equation (12).

$$\theta = \sigma\tau - Da\delta(e^{-Da\tau} - 1) \tag{12}$$

Essentially, Equation (5) presents a challenge due to the lack of measurable values for several terms (e.g., $\dot{Q}$, ΔH_{rx}, C_{PP}, k and others). To address this, the system was non-dimensionalised, grouping the unmeasured terms into three new variables. This transformation reduced the number of unknowns from nine to three, leading to Equation (6). In Equation (6), the three dimensionless parameters, σ, δ, and ρ can be interpreted as follows:

σ: Represents the efficiency of energy transfer from steam to the batch (energy required to increase the batch content by 1 kJ).

δ: Represents the heat released or absorbed during the reaction relative to the reagent's heat content (efficiency of heat transfer in the reaction).

ρ: A dimensionless factor representing the process decay over time, based on the Damkohler number.

This non-dimensionalisation simplifies the system, making it more manageable for analysis. The final, dimensionless form of the model highlights the balance between external heat input and the heat generated by the reaction. Using cellulose as a proxy for wood simplifies the model's complexity while maintaining accuracy in predicting steam demand in the digester. The plant data, which included temperature and steam consumption for each batch, was processed using Equations (7) and (8) to render the temperature and steam columns dimensionless. The dimensionless parameters in Equation (12) were then determined by fitting the data using the L-BFGS optimisation algorithm in SciPy. The L-BFGS-B algorithm is an efficient optimisation method that approximates function curvature with limited memory, making it ideal for large-scale problems [20]. The first-principles model concentrates on the parameter σ, which represents the steam flow rate to the digester – a value we aim to forecast as outlined in Equation (9). A significant issue arises with the assumption that σ remains constant throughout a batch. In reality, σ fluctuates due to continuous changes in the heat flow rate, $\dot{Q}$. To address this variability and maintain σ as a constant for each batch, we utilise the average heat transfer across a batch. This approach simplifies the representation of σ as:

$$\sigma = \frac{(\dot{Q}t_b)}{C_{P0}T_0N_{W0}} = \frac{Q}{C_{P0}T_0N_{W0}} \tag{13}$$

$$Q = m * C_p * \Delta T = m_s * H_{fg} \tag{14}$$

$$\sigma = \frac{(m_s H_{fg})}{C_{P0}T_0N_{W0}} \tag{15}$$

$$m_s = \frac{(\sigma C_{P0}T_0N_{W0})}{H_{fg}} \tag{16}$$

The value of m_s can be obtained from Equation (16), provided we can estimate the values of C_{P0}, T_0, N_{W0} and H_{fg} .

Another challenge with the first-principles model is its requirement for prior knowledge of the batch-specific parameter σ to forecast future steam flow rates. The parameter σ is derived from the batch's temperature profile, which is unavailable before the batch begins. Consequently, operators cannot determine σ in advance when predicting steam demand for a future batch, rendering Equation (16) unusable without this value. To address this issue, a black-box model becomes essential to forecast future values of σ based on historical data. Since σ has no exogenous variables, a univariate model suffices. A Convolutional Neural Network (CNN) was found to be the most effective approach for predicting σ, and it was used to generate future values of σ for computing the mass of steam.

2.2 Box and Jenkins' ARIMA Model

For over fifty years, Box-Jenkins ARIMA linear models have been widely used in time series forecasting. A nonseasonal time series is typically modelled as a combination of past values and errors, represented as ARIMA (p, d, q), where p and q refer to the orders of the autoregressive and moving average components. The general form is:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{17}$$

Here, ϕ and θ are coefficients, X_t and ε_t are the values of the modelled variable and residual at time t, respectively. The ARIMA model combines three components to capture patterns in time series data. The autoregressive (AR) terms incorporate the influence of past observations on the current value, while the moving average (MA) terms account for the impact of past forecast errors. Differencing (I) is applied to remove trends and ensure stationarity, a key requirement for effective modelling. Finally, the error term represents the random noise not explained by the model. Together, these components allow ARIMA to model both systematic dependencies and stochastic variations in the data [21]. The Box and Jenkins' methodology for developing the ARIMA model is depicted in Figure 2. The methodology outlined in Figure 2 begins with importing the pre-processed data and checking the stationarity of the steam demand series.

If the data is not stationary, first-order differencing is applied by calculating the difference between each observation and its preceding value, and this process is repeated until stationarity is achieved. Once the data is stationary, correlation plots are generated to identify significant lag values. These lag values guide the development of the ARIMA model, although Auto-ARIMA may also be used to automatically determine the optimal order of each model component. The batch temperature and steam data are utilised to develop the ARIMA model in a Python Jupyter notebook. To ensure reliable evaluation, 90% of the dataset is used for training the model, while the remaining 10% is reserved for testing its ability to predict steam demand for future batches.

2.3 Neural Network Models

The two neural networks employed in this study, LSTM and CNN, were implemented using the Keras library in Python. The neural network models were trained using historical steam demand data to forecast future demand and tested multiple times across lags 1 to 6. Lag 1 used the previous batch's steam demand on the test dataset, while lag 6 incorporated the last six batches' steam demand to predict the next batch. The hybrid first-principles models used an equal number of lagged exogenous variables (from the first principles model) and steam demand variables, also tested multiple times across six lags to find the most stable configuration.

The models were trained using the stochastic gradient descent (SGD) backpropagation algorithm paired with the Adam optimiser, aiming to minimise the root mean squared error (RMSE) over 95 epochs. After 95 epochs, the RMSE remained stable across both the training and validation datasets, indicating consistent model performance. Empirical evidence also shows that the Adam optimiser accelerates training compared to other optimisers like RMSprop and SGD."

2.3.1 LSTM model

Deep neural networks, inspired by neurobiology, have become powerful tools for function approximation and pattern recognition. They are broadly categorised into feedforward networks, where data flows unidirectionally, and recurrent neural networks (RNNs), which incorporate feedback loops to retain information from previous inputs. Although RNNs can model sequential data, they are limited in capturing long-term dependencies. To overcome this, Hochreiter and Schmidhuber developed Long Short-Term Memory (LSTM) networks, which introduce memory cells regulated by forget, input, and output gates. These mechanisms allow LSTMs to preserve or discard information over extended sequences, enabling effective learning in tasks such as language modelling, prediction, and sequence analysis [22].

The LSTM model has a three-layer architecture with a depth of 3 and a width of 64. Increasing the layers or neurons adds complexity without improving performance, while removing a layer slightly degrades it. The LSTM architecture is illustrated in Figure 3. The LSTM model consists of an input layer, defined with a shape of (k, x), where k represents the number of lagged time steps and x denotes the number of forecasting variables for steam demand. This is followed by an LSTM layer containing 64 memory cells, which process the input sequences while managing the input, forget, and output gates, as well as the cell states. The output from the LSTM layer is passed through a fully connected dense layer with 8 units and ReLU activation, introducing non-linearity and transforming the features. Finally, a dense layer with a single unit and linear activation produces the final forecasted output.

2.3.2 CNN model

Convolutional Neural Networks (CNNs) are specialised neural networks that use convolution

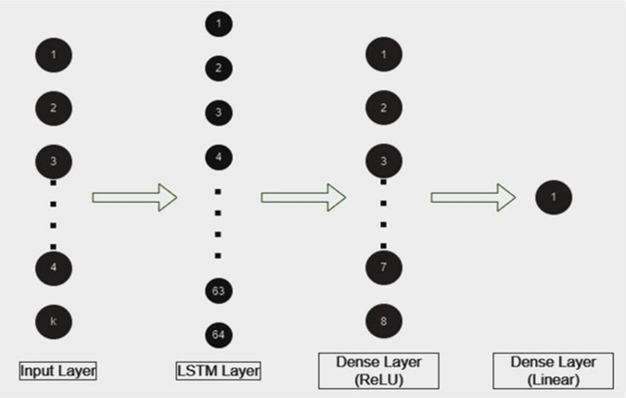

Figure 3. LSTM model architecture development.

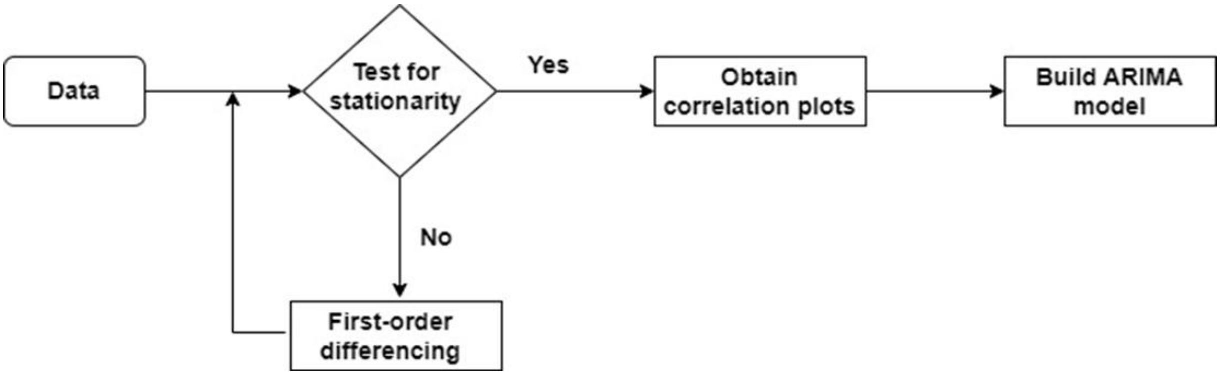

Figure 2. Box and Jenkins ARIMA model development methodology.

and pooling operations to extract deep features, making them highly effective for recognising patterns in complex data such as images, text, and video. Their architecture, as seen in Figure 4, is also well-suited for analysing seasonal time series with trends.

In this study, CNNs, typically applied to image classification, are adapted for time series regression. The model begins with convolutional layers that capture time-lag correlations from past observations, effectively extracting predictive patterns from sequences of previous data points. These convolutional layers are followed by pooling layers, which merge similar features and reduce dimensionality by selecting the maximum value from neighbouring neurons, ensuring robustness to input shifts. Finally, fully connected layers summarise the extracted features and model both linear and nonlinear relationships, generating the final output for the time series forecast [23].

The CNN model employs a four-layer architecture designed to balance complexity and performance. The input layer is shaped as (kc, xc), where kc represents the number of lagged time steps and xc denotes the forecasting variables for steam demand. A Conv1D layer with 64 filters and a kernel size of 1, using ReLU activation, extracts features from the input sequence. The output is then flattened into a one-dimensional array, which is passed through a dense layer with 8 neurons and ReLU activation to capture complex relationships. Finally, a dense layer with a single neuron and linear activation produces the forecasted output. This architecture effectively meets the objectives of the study without unnecessary complexity.

2.4 Hybrid First Principles-ARIMA Models

In the hybrid first-principles-ARIMA model, the first-principles approach calculates the dimensionless parameters, which are then incorporated as input variables alongside the plant's temperature data in the ARIMA model to forecast steam demand. This framework can be viewed as an ARIMAX model, where the dimensionless parameters serve as exogenous variables in the ARIMA structure. This dynamic is illustrated in Figure 5.

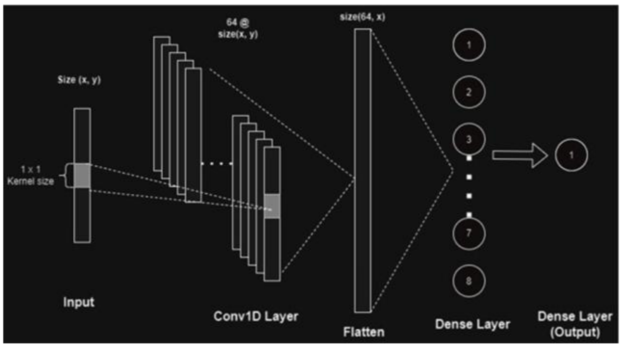

Figure 4. CNN model architecture.

2.5 Hybrid First Principles-Artificial Neural Network Models

In the hybrid first principles-neural network model, the dimensionless parameters estimated from the first principles models alongside the temperature variable are used as inputs for the LSTM or CNN models to predict steam demand. As a result, the hybrid first principles-LSTM model functions as a multivariate LSTM, and the hybrid first principles-CNN model operates as a multivariate CNN. Their structure parallels that of the hybrid first-principles–ARIMA model shown in Figure 5, with the ARIMA model component replaced by either an LSTM or CNN.

2.6 Model Evaluation Metrics

To evaluate the forecast accuracy of all models, two approaches were considered: analysing the entire test dataset versus specific subsets. Given the need to predict steam demand for upcoming batches, the focus was on the next batch. Evaluating the entire test set could skew the results due to error accumulation; therefore, accuracy was assessed over a five-batch forecast horizon using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). This horizon strikes a balance between relevance and reduces error propagation. Training metrics were computed across the full dataset.

The RMSE is the square root of the average of the squares of the differences between the predicted values and the actual values, as formulated in Equation (18).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (18)$$

The MAE is the average of the absolute differences between the predicted values and the

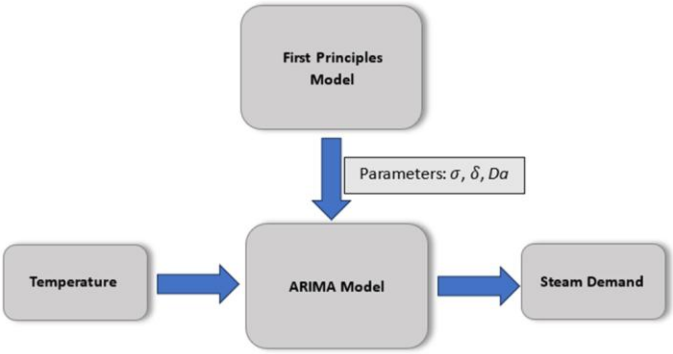

Figure 5. Layout of the hybrid first principles-ARIMA model.

actual values. Unlike RMSE, MAE uses the L1 norm ($|y_i - \hat{y}_i|$) to compute the average absolute error, making it less sensitive to outliers since deviations are not squared. This robustness makes MAE a more reliable performance metric for regression models when outliers are present [24].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{19}$$

3. Results and Discussion

This section presents the results of applying the models to the test dataset. Training plots generated from the training data are provided as Supporting Information. This approach is taken because a model's performance is best evaluated on the test dataset, which consists of new, unseen data.

3.1. First-Principles Model Performance

Figure 6 compares the actual steam consumption (blue) with the predictions from the first principles model (red). The model significantly underestimates the actual steam demand and fails to capture the variability in the data, resulting in high error metrics: an RMSE of 11.32 and an MAE of 11.23. This discrepancy arises from two main assumptions in the model. First, the assumption of a constant σ across a batch leads to an underestimation of the actual heat transfer Q, which in turn underestimates the steam mass m_s. Second, by considering the wood load as constant, the model cannot account for the actual variability in steam demand. As a result, the model predicts an average steam demand for every batch (about 16-19 tonnes), as illustrated in Figures 7 and 8, and also fails to account for fluctuations in the data.

To improve accuracy, a correction term m_c was introduced. The refined first-principles model incorporates mc to adjust the predicted steam demand, resulting in better alignment with the actual data. However, it still does not fully capture the inherent variability due to the constant wood load assumption. Resolving this would require precise data on wood and liquor loads for each batch. The modified equation becomes:

$$\sigma = \frac{((m_s + m_c)\, H_{fg})}{C_{P0} T_0 N_{W0}} \tag{20}$$

$$\sigma = \frac{(m_s\, H_{fg})}{C_{P0} T_0 N_{W0}} + \frac{(m_c\, H_{fg})}{C_{P0} T_0 N_{W0}} \tag{21}$$

This leads to an adjusted σ, denoted by σ+σ' in Equation (22).

$$\frac{d\theta}{d\tau} = \sigma + \sigma' + \delta\rho \tag{22}$$

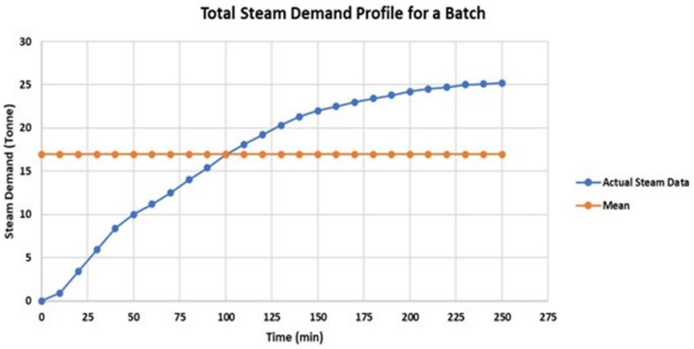

Figure 7. Total steam consumption per minute for a typical batch.

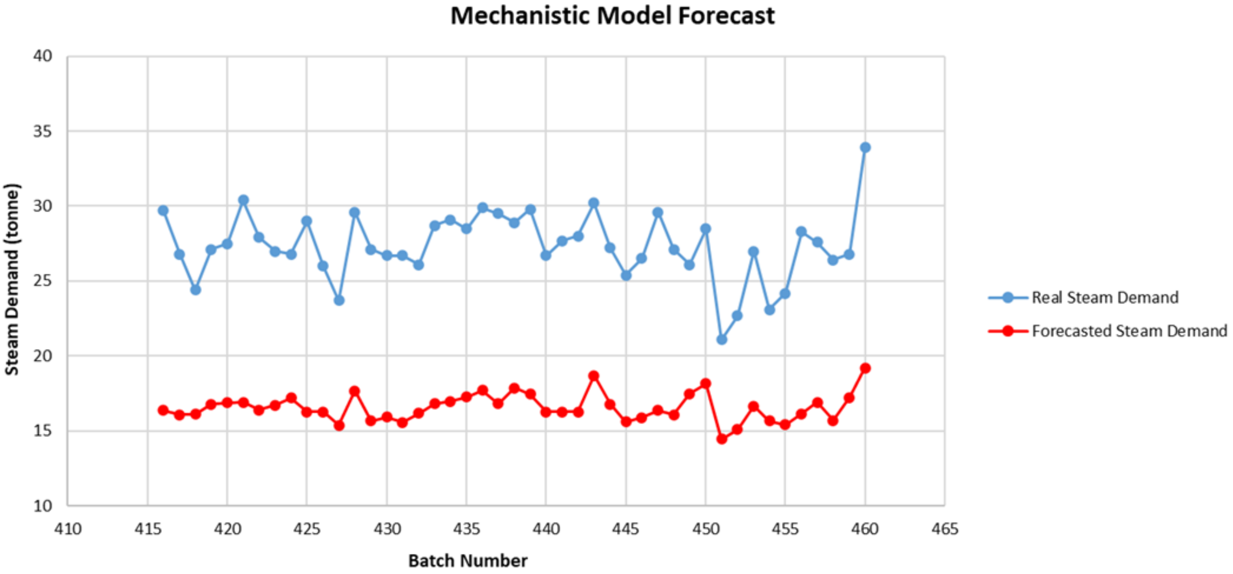

Figure 6. Comparison of first-principles model forecasts with actual data.

Here, m_c is calculated as the Mean Absolute Error (MAE) between the actual and forecasted average steam demand across the test dataset:

$$m_c = \frac{|\sum_{i=1}^{n} actual,m - \sum_{i=1}^{n} forecasted,m\,|}{n} \qquad (23)$$

Figure 9 displays the actual steam consumption (blue), the original model's prediction (red), and the adjusted prediction (orange). The adjusted model shows improved performance with an RMSE of 2.03 and an MAE of 1.63, indicating a better fit to the actual data. The refined first-principles model outperforms the original by incorporating Mean Absolute Error (MAE) for more accurate predictions within the actual data range. However, it still falls short in capturing data variability due to constant wood load. Addressing this requires precise data on both wood and liquor loads for each batch.

3.2. Box and Jenkins' ARIMA Model Performance

The ARIMA model is developed by identifying the optimal combination of autoregressive, integrated, and moving average components. Since stationarity is a prerequisite, the analysis begins with testing and ensuring that the steam demand data is stationary. Autocorrelation (ACF) and partial autocorrelation (PACF) plots are then used to examine temporal dependencies and guide model specification. In addition, the auto-ARIMA function is employed to automatically determine the most suitable model parameters. This section presents the stationarity results, ACF and PACF analyses, and the final auto-ARIMA model specification, followed by its application to the steam demand data.

3.2.1 Stationarity test

The Augmented Dickey-Fuller (ADF) test, using Python's 'adfuller' method, confirms stationarity in the steam usage data with a p-value of 0.011% – well below the 5% threshold – and a test statistic lower than the 5% critical value. This allows us to reject the null hypothesis and proceed with ARIMA modelling.

3.2.2 ACF and PACF

The autocorrelation (ACF) and partial autocorrelation (PACF) functions are visualised using Python's statsmodels graphics tools in Figures 10 and 11. Figure 10 shows moderate correlations between the current series value and its preceding values up to 10 lags, with the highest correlation at lag 1 being 0.35. This suggests that future values can be reasonably predicted based on the past 10 observations. Figure 11 indicates that the first three lags, especially the first, have a direct impact on the current value, highlighting the significant influence of recent data even after accounting for intermediate lags. However, the strength of these correlations remains moderate, capped at 0.35. The blue regions in both plots denote correlations insignificant at the 5% significance level (95% confidence). The ACF plot reveals significant autocorrelations up to lag 13, while the PACF plot shows significant autocorrelations up to lag 3.

3.2.3 Optimal ARIMA configuration

The optimal ARIMA model configuration was determined using ACF and PACF plots, suggesting an initial ARMA(3,10) model. However, to streamline the process, the 'auto-ARIMA' function identified ARIMA(3,1,1) as the best fit, balancing model complexity. Despite the

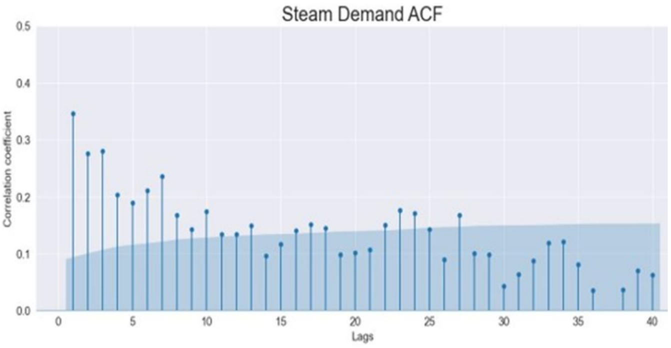

Figure 10. ACF plot of steam demand.

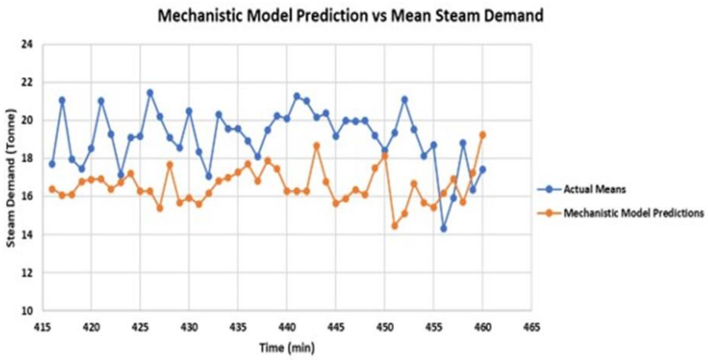

Figure 8. Comparison of first-principles model forecasts with the means of actual batches.

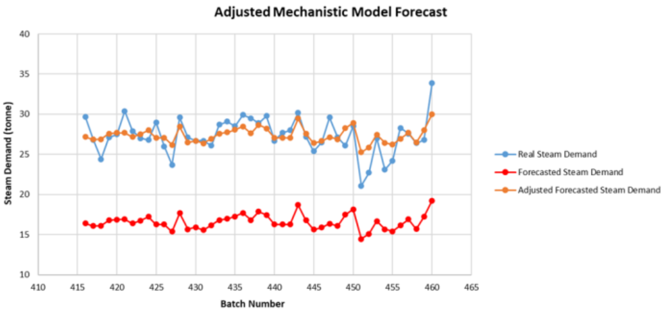

Figure 9. Adjusted mechanistic model test forecast.

data's stationarity, a single order of differencing was employed, which improved the log-likelihood and Akaike Information Criteria, possibly due to seasonality in some data segments. Thus, ARIMA(3,1,1) is selected.

3.2.4 ARIMA(3,1,1) model performance

Figure 12 illustrates the forecasting performance of the ARIMA(3,1,1) model on the test dataset. As expected, the model performs better on the training data than on new test scenarios. The forecast appears as a near-straight line, failing to capture the variability in future steam values due to the model's simplicity and the low correlation between past and future steam demand. Incorporating process variables, such as wood loading and liquor loading, could enhance its performance. The test dataset forecast has an RMSE of 3.24 and an MAE of 2.90. The straight-line forecast results from the model's integrated component of 1, which models the first-order difference in steam demand data. With limited information to explain changes in steam demand, the model predicts minimal variations after the initial points. Including exogenous variables could better capture the variation in steam demand. Table 1 presents the coefficients and p-values of the ARIMA(3,1,1) model. Except for the lag 3 coefficient, all p-values exceed the 5% significance level, indicating that most coefficients are not individually significant, although the overall model fit is more important. The lag 3 coefficient, despite being insignificant, contributes to the model's goodness of fit. The MA component has a large negative coefficient compared to others, suggesting a compounding effect that negates positive changes in steam demand as more data points are included. This explains the slight variation at the beginning of the test forecasts and the subsequent straight-line predictions as the forecasts extend further.

3.3. Neural Network Model Performance

The neural network models were evaluated using 6 different lags, with lag 3 proving to be the most stable and yielding the lowest error metrics. This result aligns with the PACF plot, which indicates that the preceding three values primarily influence steam demand.

3.3.1 LSTM model performance

The LSTM model performs similarly on both the training and test datasets, indicating no overfitting; however, its accuracy is suboptimal, with predictions deviating by approximately 2 tonnes of steam. It struggles to predict steam demand variability due to relying solely on past steam data, resulting in an RMSE of 2.24 ± 0.062 and an MAE of 2.06 ± 0.056 (uncertainties arise from the stochastic nature of neural networks). Despite this, its numerous parameters enable it to model steam variability better than a univariate statistical time series model. Incorporating exogenous variables could potentially improve its performance. Figure 13 displays the LSTM test forecasts.

3.3.2 CNN model performance

The CNN model achieved an RMSE of 2.00 ± 0.02 and an MAE of 1.63 ± 0.023, exhibiting lower error metrics on the test dataset compared to the ARIMA and LSTM models, although it performed slightly worse on the training data. This difference arises because the LSTM model, with more parameters, fits the training data better but generalises less effectively. In contrast, the CNN's simpler structure facilitates generalisation and enables faster training, while effectively capturing short-term fluctuations. Figure 14 displays the CNN test forecasts.

Table 1. ARIMA(3,1,1) model coefficient.

	Coefficient	p-Value
ar.Lag1	0.1874	0.000
ar.Lag2	0.0970	0.030
ar.Lag3	0.0879	0.092
ma.Lag1	-0.9726	0.000
Log-Likelihood: -546.53		
AIC: 1103.07		

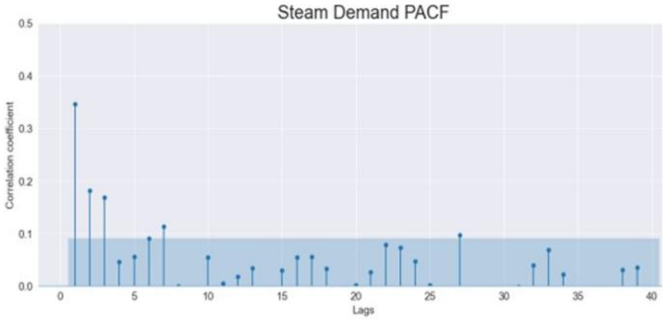

Figure 11. PACF plot of steam demand.

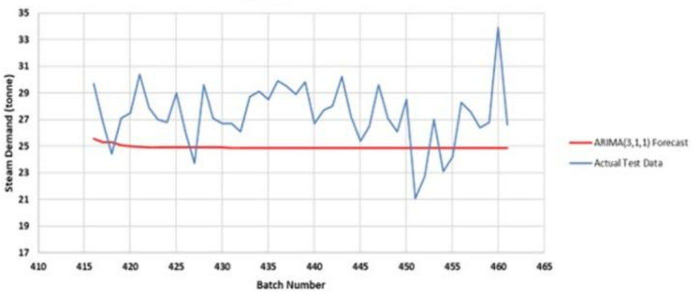

Figure 12. ARIMA(3,1,1) model forecast on test set vs. actual test data.

3.4. Hybrid Model Performance

3.4.1 Hybrid first principles-ARIMA model performance

While the ARIMA model handles training data adequately, it produces constant predictions for the test data, indicating limitations due to missing batch-specific plant features. To improve accuracy, a hybrid first-principles ARIMA model incorporates derived dimensionless parameters (σ, δ, Da) as exogenous variables, allowing it to capture both historical demand and key influencing factors, enhancing predictive performance over a standard ARIMA model. The hybrid model, which incorporates exogenous variables, outperforms the ARIMA model without them. Figure 15 demonstrates that the hybrid model continues to surpass the ARIMA model in forecasting future steam demand, with an RMSE of 2.29 and an MAE of 1.93 – about 1 tonne of steam lower than the ARIMA model's errors. This consistent performance indicates that overfitting is unlikely. Table 2 presents the coefficients of the Hybrid ARIMA(3,1,1) model, which boasts a significantly higher log-likelihood and lower AIC values than the ARIMA model, indicating a better fit. Even without data on wood loading, wood type, and liquor loading at the start of the batch, the exogenous variables effectively capture steam demand patterns. These dimensionless variables, derived from the energy balance equation, provide valuable insights into the chemical and physical behaviours of the pulping process, enabling the model to make accurate predictions.

3.4.2 Hybrid first principles-LSTM model performance

The hybrid LSTM model was trained using historical steam demand data and three parameters (σ, δ, Da) from the first principles model, with equal lags for both steam demand and exogenous variables. This model exhibits significantly lower RMSE and MAE (0.95 ± 0.013 and 0.83 ± 0.022) compared to the standalone LSTM, due to the inclusion of exogenous variables that enhance forecasting accuracy. As shown in Figure 16, the hybrid LSTM model captures steam demand variability more effectively than both the ARIMA and LSTM models, highlighting the potential of ANNs in accurately modelling steam demand.

3.4.3 Hybrid first principles-CNN model performance

Figure 17 indicates that it outperforms all other models in predicting both datasets. The hybrid CNN model has an RMSE of 0.89 ± 0.027

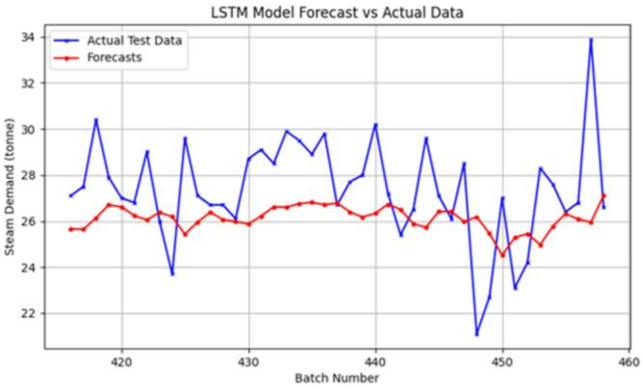

Figure 13. LSTM model forecast on test set vs. actual test data.

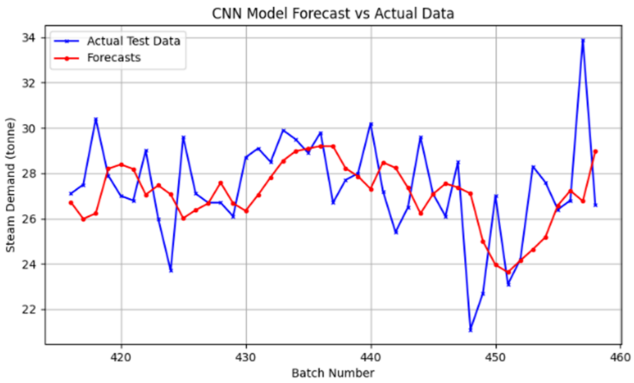

Figure 14. CNN model forecast on test set vs. actual test data.

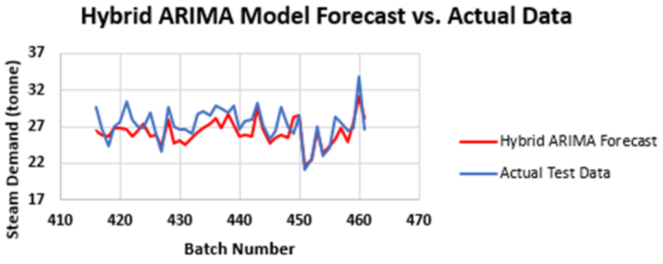

Figure 15. Hybrid ARIMA model forecast on test set vs. actual test data.

Table 2. Hybrid first principles-ARIMA(3,1,1) model coefficients.

	Coefficient	p-Value
ar.Lag1	0.1596	0.003
ar.Lag2	0.0882	0.062
ar.Lag3	0.0109	0.829
ma.Lag1	-0.9808	0.000
δ	-0.0979	0.010
Da	-0.1264	0.006
σ	0.5840	0.000
Log-Likelihood: -546.53		
AIC: 1103.07		

and an MSE of 0.71±0.024. This superior performance is due to the Hybrid CNN model's ability to create additional features through convolution operations. Alongside the dimensionless parameters derived from the first-principles dataset, the Hybrid CNN model utilises convolutional kernels to generate additional features that effectively capture local patterns within the data.

3.5. Comparative Analysis of the Performance of all Models

Table 3 presents the RMSE, MAE, and degrees of freedom (DOF) for each model, facilitating the identification of the simplest and most effective model by comparing their complexity. Figure 18 visualises these results, showing that hybrid neural network models perform best, while the ARIMA model performs the worst. The ARIMA model underperforms due to a low autocorrelation coefficient (below 0.5) between steam demand and its past values, indicating significant influence from external factors not captured by ARIMA's linear, temporal approach. Incorporating exogenous variables enhances performance in the hybrid ARIMA model. The first-principles model, based on fundamental physical and chemical processes of digester reactions, is expected to outperform time-series models, such as Box and Jenkins. However, its predictive accuracy is limited by the absence of crucial initial variables such as wood and liquor content and the actual heat capacity of the digester content. Despite these limitations, it still

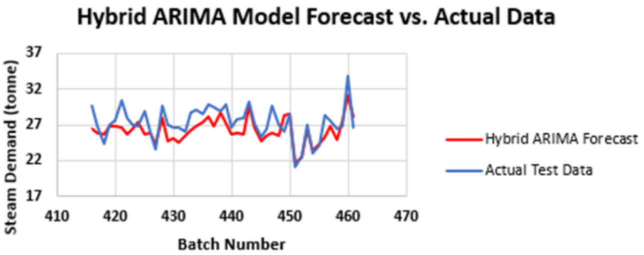

Figure 15. Hybrid ARIMA model forecast on test set vs. actual test data.

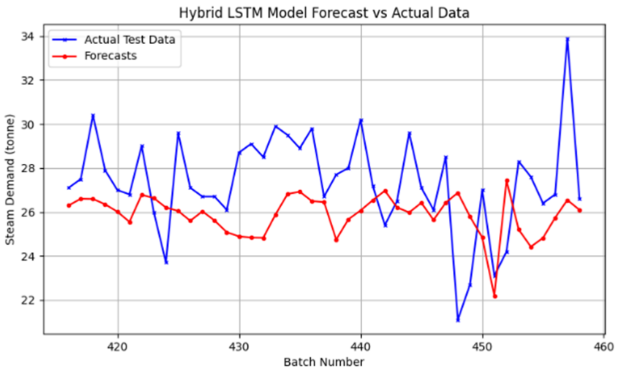

Figure 16. Hybrid LSTM model forecast on test set vs. actual test data.

Table 3. Error metrics and degrees of freedom of the models.

Model	RMSE	MAE	DOF
First Principles	2.03	1.63	3
ARIMA	3.24	2.90	5
Hybrid ARIMA	2.29	1.93	8
LSTM	2.24 ± 0.062	2.06 ± 0.056	17425
Hybrid LSTM	0.95 ± 0.013	0.83 ± 0.022	17681
CNN	2.00 ± 0.020	1.63 ± 0.023	1681
Hybrid CNN	0.89 ± 0.027	0.71 ± 0.024	1745

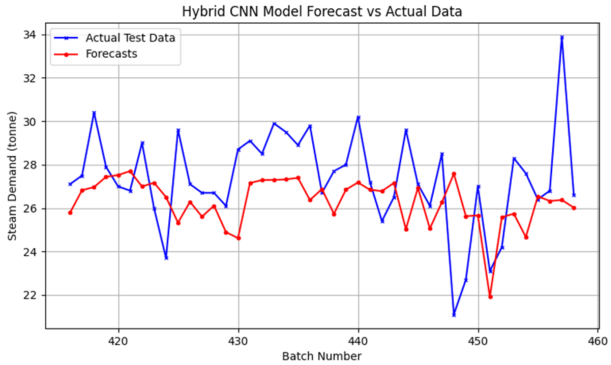

Figure 17. Hybrid CNN model forecast on test set vs. actual test data.

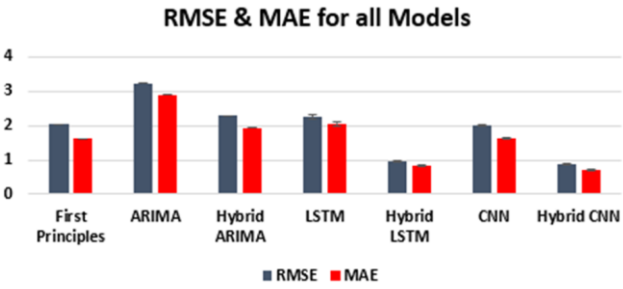

Figure 18. RMSE and MAE values for all models.

surpasses the traditional time series models. Complex neural network models, such as LSTM and CNN, have high degrees of freedom, with LSTM being the most intricate. While they perform comparably, their reliance solely on steam demand data makes them less optimal than their hybrid versions. Nevertheless, their nonlinear complexity only allows them to perform on par with the hybrid ARIMA model.

Among neural networks, CNN generally outperforms LSTM while using fewer parameters due to its utilisation of shared weights. This means that as the number of time steps increases, the parameter count in CNN models remains constant if the number of filters stays the same. Hybrid models like the hybrid LSTM and hybrid CNN significantly outperform their standalone counterparts by incorporating exogenous variables that better explain steam demand variations. Therefore, the hybrid first principles-CNN model is the most effective structure for modelling steam demand. The hybrid first principles-ARIMA model performs significantly better than the standard ARIMA. Even without detailed process data from the mill, the dimensionless first-principles model generated features that neural networks used to accurately model steam demand.

The findings of this study align with previous applications of hybrid mechanistic–data-driven models in chemical engineering, as outlined in the Introduction. Similar trends are observed in the pulp and paper industry, where most research has focused on kappa number prediction and other kraft-specific variables. For example, a hybrid kraft digester model coupling kinetics and diffusion with ANN achieved more accurate delignification and kappa number predictions than either approach individually [25]. Even within control theory, hybridisation has proven valuable, with a Koopman MPC framework enhancing regulation of Kappa number and fibre properties under feed fluctuations relative to single-model controllers [26]. Distinct from these studies, which emphasise digester chemistry or process control, our work extends hybrid modelling to steam demand forecasting – an equally critical but underexplored aspect of pulp mill operations. Consistent with prior findings, our hybrid mechanistic–data-driven models outperformed standalone mechanistic, ARIMA, and neural network models, highlighting the effectiveness of hybrid strategies for improving both accuracy and reliability in complex industrial processes.

4. Conclusions

This study addressed the challenge of predicting steam demand in batch sulphite digesters by developing and comparing first-principles, ARIMA, neural network, and hybrid models. The hybrid frameworks, which incorporated dimensionless parameters from the mechanistic model as exogenous variables, consistently outperformed standalone models. Among them, the hybrid first-principles–CNN model achieved the highest predictive accuracy, followed by the hybrid LSTM and ARIMA models. These results confirm that integrating mechanistic insights with data-driven techniques provides a more reliable approach for forecasting steam demand in complex batch processes. Future research should explore incorporating higher-order reaction kinetics to more accurately capture underlying chemical dynamics and investigate alternative time series methods, such as ARDL or Kalman filtering, to address non-stationary data. In addition, extending neural network architectures to advanced models like transformers could further improve predictive performance.

Acknowledgement

The authors gratefully acknowledge the support of the Data Analytics and Numerics in Control Engineering (DANCE) research group at the University of the Witwatersrand. We also extend our appreciation to Sappi for providing data and valuable insights that contributed to this research.

CRedit Author Statement

Author Contributions: D. Agommuoh: Methodology, Investigation, Writing, Software, Editing; A. Higginson: Methodology, Data Curation, Supervision, Project Administration; K. Brooks: Validation, Methodology, Review and Editing, Data Curation, Supervision; P. de Vaal: Supervision, Review and Editing, Validation. All authors have read and agreed to the published version of the manuscript.

References

[1] Bajpai, P. (2015). Pulp and Paper Industry: Chemicals. 1st ed. Elsevier.

[2] May, R., Maier, H., Dandy, G., Fernando, T. (2008). Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modelling & Software*, 23(10-11). 1312-1326. DOI: 10.1016/j.envsoft.2008.03.007

[3] Khashei, M., Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 11(2), 2664-2675. DOI: 10.1016/j.asoc.2010.10.015

[4] Kwon, Y., Kim, S., Choi, Y., Kang, S. (2022). Generative modeling to predict multiple suitable conditions for chemical reactions. *Journal of Chemical Information and Modeling*, 62(23), 5952-5960. DOI: 10.1021/acs.jcim.2c01085

[5] Gallegos, L., Luchini, G., St. John, P., Kim, S., Paton, R. (2021). Importance of engineered and learned molecular representations in predicting organic reactivity, selectivity, and chemical properties. *Accounts of Chemical Research*, 54(4), 827-836. DOI: 10.1021/acs.accounts.0c00745

[6] Sadrzadeh, M., Mohammadi, T., Ivakpour, J., Kasiri, N. (2008). Separation of lead ions from wastewater using electrodialysis: comparing mathematical and neural network modeling. *Chemical Engineering Journal*, 144(3), 431-441. DOI: 10.1016/j.cej.2008.02.023

[7] Lashkarbolooki, M., Vaferi, B., Rahimpour, M.R. (2011). Comparing the capability of artificial neural network (ANN) and EOS for prediction of solid solubilities in supercritical carbon dioxide. *Fluid Phase Equilibria*, 308(1-2), 35-43. DOI: 10.1016/j.fluid.2011.06.002

[8] Grondin-Perez, B., Roche, S., Lebreton, C., Benne, M., Damour, C., Kadjo, A.J.J. (2014). Mechanistic model versus artificial neural network model of a single-cell pemfc. *Engineering*, 6(8), 418-426. DOI: 10.4236/eng.2014.68044

[9] Miheț, M., Cristea, V.M., Agachi, P.Ș. (2009). FCCU simulation based on first principle and artificial neural network models. *Asia-Pacific Journal of Chemical Engineering*, 4(6), 878-884. DOI: 10.1002/apj.312

[10] Li, K., Duan, H., Liu, L., Qiu, R., van den Akker, B., Ni, BJ., Chen, T., Yin, H., Yuan, Z., Ye, L. (2022). An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. *Environmental Science & Technology*, 56(4), 2816-2826. DOI: 10.1021/acs.est.1c05020

[11] Faruk, D.Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23(4), 586-594. DOI: 10.1016/j.engappai.2009.09.015

[12] Ghahnavieh, A.E. (2019). Time series forecasting of styrene price using a hybrid ARIMA and neural network model. Independent *Journal of Management & Production*, 10(3), 915-933. DOI: 10.14807/ijmp.v10i3.877

[13] Phatwong, A., Koolpiruck, D. (2019). Kappa number prediction of pulp digester using LSTM neural network. *16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 151-154.

[14] Shah, P., Choi, H.K., Kwon, J.S.I. (2023). Achieving optimal paper properties: A layered multiscale kMC and LSTM-ANN-based control approach for kraft pulping. *Processes*, 11(3), 809. DOI: 10.3390/pr11030809

[15] Correia, F.M., d'Angelo, J.V.H., Zemp, R.J., Mingoti, S.A. (2014). Prediction of kappa number in eucalyptus kraft pulp continuous digester using the Box & Jenkins methodology. *Advances in Chemical Engineering and Science*, 4(4), 539-547. DOI: 10.4236/aces.2014.44055

[16] Shao, B., Hu, X., Bian, G., Zhao, Y. (2019). A multichannel LSTM-CNN method for fault diagnosis of chemical process. *Mathematical Problems in Engineering*, 2019(1). DOI: 10.1155/2019/1032480.

[17] Chen, H., Cen, J., Yang, Z., Si, W., Cheng, H. (2022). Fault diagnosis of the dynamic chemical process based on the optimized CNN-LSTM network. *ACS Omega*, 7(38), 34389-34400. DOI: 10.1021/acsomega.2c04017

[18] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, 13(22), 4712. DOI: 10.3390/rs13224712

[19] Fogler, H.S. (2006). Elements of Chemical Reaction Engineering. 4th ed. Prentice Hall.

[20] Lai, K., Mishra, S., Panda, G., Chakraborty, S., Samei, M., Ram, B. (2020). A limited memory q-BFGS algorithm for unconstrained optimization problems. *Journal of Applied Mathematics and Computing*, 2021(66), 183-202. DOI: 10.1007/s12190-020-01432-6

[21] Ho, S., Xie, M., Goh, T. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, 42(2-4), 371-375. DOI: 10.1016/S0360-8352(02)00036-0.

[22] Vani, S., Rao, T., Naidu, C. (2019). Comparative analysis on variants of neural networks: an experimental study. *5th International Conference on Advanced Computing & Communication Systems*, 429-434. DOI: 10.1109/ICACCS.2019.8728327

[23] Liu, S., Ji, H., Wang, M. (2020). Nonpooling convolutional neural network forecasting for seasonal time series with trends. *IEEE Transactions on Neural Networks and Learning Systems*, 31(8), 2879-2888. DOI: 10.1109/TNNLS.2019.2934110

[24] Chicco, D., Warrens, M., Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7. DOI: 10.7717/peerj-cs.623

[25] Filhoa, R., Aguiar, H., Polowskia, N. (2005). Hybrid modelling development for a continuous industrial kraft pulping digester. *European symposium on computer aided process engineering: ESCAPE-15 conference paper*.

[26] Son, S., Choi, H., Moon, J., Kwon, J. (2022). Hybrid Koopman model predictive control of nonlinear systems using multiple EDMD models: An application to a batch pulp digester with feed fluctuation. *Control Engineering Practice*, 118(14), 104956. DOI: 10.1016/j.conengprac.2021.104956.