Sheffield Hallam University

Neural Brewmeister: Modelling Beer Fermentation Dynamics Using LSTM Networks

O'BRIEN, Alexander <http://orcid.org/0009-0007-0643-9144>, ZHANG, Hongwei <http://orcid.org/0000-0002-7718-021X> and ALLWOOD, Daniel <http://orcid.org/0000-0002-3735-3198>

Citation:

Article

Neural Brewmeister: Modelling Beer Fermentation Dynamics Using LSTM Networks

Alexander O'Brien [1], Hongwei Zhang [1,*] and Daniel Allwood [2]

[1] Advanced Food Innovation Centre, Sheffield Hallam University, Sheffield S9 2AA, UK; a.obrien@shu.ac.uk
[2] Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield S1 1WB, UK; d.allwood@shu.ac.uk
* Correspondence: h.zhang@shu.ac.uk

Abstract

Fermentation is a complex biochemical process that transforms brewer's wort into beer. Beer fermentation is driven by yeast and is influenced by process parameters such as the content of fermentable sugars in wort, temperature, and pH. Traditional methods of modelling this process rely heavily on empirically tuned kinetic models. However, these models tend to be recipe-specific and often require retuning when processes change. This paper proposes a data-driven approach using a Long Short-Term Memory (LSTM) network, a type of recurrent neural network, to model beer fermentation dynamics. By training the LSTM model on real-world fermentation data (1305 fermentations across ales, IPAs, lagers, and mixed-culture beers), including variables such as apparent extract (derived from specific gravity), temperature, and pH, we demonstrate that this technique can accurately predict key fermentation trajectories and support process monitoring and optimisation. When evaluated on representative medoid fermentations as one-step-ahead roll-outs over 0–300 h, the model produces accurate predictions with low errors and minimal residuals. These results show that the LSTM-based model provides accurate and robust predictions across beer styles and operating conditions, offering a practical alternative to traditional mechanistic kinetic models. This work highlights the potential of LSTM networks to enhance our understanding, monitoring, and control of fermentation processes, providing a scalable and efficient tool for both research and industrial applications. The findings suggest that LSTM models can be effectively adapted to model other fermentation processes in beverage production, opening new possibilities for advancing food science and engineering.

Keywords: beer fermentation; artificial intelligence; data-driven modelling; LSTM networks; fermentation dynamics

1. Introduction

Fermentation sits at the heart of brewing. During fermentation, fermentable sugars in brewer's wort are converted into ethanol, carbon dioxide, and a wide range of flavour-active secondary metabolites that define the organoleptic profile of the finished beer [1–3]. In production vessels, fermentation dynamics are shaped by yeast physiology, recipe formulation, temperature control, and fermentation-vessel geometry. The growth of the modern craft sector has further expanded this design space through the exploration of novel yeasts and recipes, and the use of mixed-culture fermentations [4–6]. Further, many breweries now operate with continuous inline sensing and automated data logging of

apparent extract (°P) and fluid temperature, creating an opportunity for digital tools that anticipate fermentation behaviour rather than simply recording it.

Mechanistic models of beer fermentation typically describe changes in yeast biomass, fermentable sugars, and ethanol through coupled ordinary differential equations (ODEs), often incorporating substrate inhibition or oxygen effects [7–9]. Such models have been used successfully for simulation and optimisation in specific breweries. However, such models are not robust to process changes, requiring key parameters to be tuned to empirical data, meaning recipe or yeast changes require reparametrisation for the model to perform optimally [10]. In everyday operation, many breweries primarily measure the following parameters inline: fluid temperature, apparent extract (expressed in degrees Plato), and pH (less commonly measured inline) rather than the full state vector assumed in most kinetic formulations [8]. In particular, biomass is rarely monitored inline, leading to a discrepancy between the data collected in practice and the values forecast by conventional models.

Data-driven approaches offer an alternative route. For fermentation and bioprocessing, neural networks and related methods have been applied to soft sensing, fault detection, and the prediction of biomass or product concentrations from routinely measured variables [11–13]. Recent work has also explored hybrid models that combine mechanistic structure with flexible empirical components [14,15]. More broadly, there has been increasing interest in data-driven dynamical systems and the use of digital twins within process systems engineering [16,17]. Recurrent neural networks, and in particular long short-term memory (LSTM) architectures [18–20], are also well suited to this context, as they can represent non-linear temporal dependencies without requiring ODEs derived from first principles.

In brewing, most reported neural-network models have been trained on relatively small datasets, often focusing on monoculture fermentations under tightly controlled conditions [11]. There are fewer examples of a single data-driven model trained on an extensive, heterogeneous collection of fermentations spanning multiple beer styles and mixed-culture inocula. Yet this is precisely the situation encountered in practice: breweries accumulate archives of hundreds or thousands of fermentations from a diverse product line, collected under slightly different process conditions, and an effective plant model must cope with this diversity while still providing reliable predictions for individual fermentations. Studies such as "From Data to Draught" have shown that diverse fermentation datasets can, in principle, be leveraged to develop models robust to a range of temperature and inoculum configurations [21]. Still, a robust process model trained on a large, diverse fermentation dataset from real-world commercial breweries remains needed.

The present work addresses this need by developing a data-driven plant model for beer fermentation using an LSTM-based architecture. The model is trained on a large dataset of inline-monitored fermentations recorded using a Sennos (US) BrewIQ IoT sensor array. This dataset is primarily comprised of fermentations carried out in commercial breweries and is augmented by some pilot-scale fermentations conducted in a laboratory setting. After filtering to include only ales, India pale ales (IPAs), lagers, and mixed-culture beers, 1305 fermentations were available for use in modelling. The dataset encompasses a broad range of recipes, yeast strains, and temperature profiles, resulting in diverse apparent extract and pH trajectories.

Using this dataset, we develop a multi-output, sequence-to-sequence LSTM plant model that forecasts apparent extract and pH over short horizons, conditioned on the intended future temperature schedule; full architectural and training details are provided later in the paper.

Model performance is assessed on rolling short-horizon forecast windows and through open-loop rollouts on representative fermentations in each beverage category, providing an interpretable view of fidelity in brewer-relevant terms.

The primary objectives of this work are to describe a large, stylistically diverse beer fermentation dataset suitable for data-driven modelling, to detail the design, training, and validation of a multi-output LSTM plant model that uses inline measurements of temperature, apparent extract, and pH as key modelling parameters, and to evaluate the model's capability to replicate typical fermentations for ales, IPAs, lagers, and mixed-culture beers. The findings demonstrate that the LSTM plant model achieves root-mean-square prediction errors for apparent extract and pH that are comparable to standard sensor tolerances, particularly for pH, indicating that LSTM-based neural networks of this type can provide a practical foundation for monitoring and short-term forecasting of beer fermentation in modern breweries.

2. Materials and Methods

2.1. Experimental Facilities and Instrumentation

All pilot-scale fermentations were carried out at the Advanced Food Innovation Centre (AFIC) at Sheffield Hallam University (Sheffield, UK). Wort was produced using a 200 L Speidel's Braumeister brewhouse (Figure 1) and transferred to a 240 L fermenter (Figure 2), which has been cleaned in place (CIP) and sanitised thoroughly prior to use (see Figure 3). This in-house pilot brewery enabled the production of fermentations analogous to those in small craft breweries. However, the majority of the data are sourced from data collected in commercial breweries at a range of scales using a wide variety of recipes and equipment.

Figure 1. The 200 L Speidel's Braumeister brewhouse used to produce the IPA-style wort.

Figure 2. The 240 L fermenter used for pilot-scale fermentations.

Figure 3. The Speidel clean-in-place (CIP) apparatus used to clean and sanitise the 240 L fermenter (Figure 2).

All inline measurements within the dataset were recorded using the Sennos (Durham, NC, USA) BrewIQ multi-sensor fermentation monitor, shown in Figure 4. The device measures fluid temperature, specific gravity and apparent extract (°P), pH, dissolved oxygen, and conductivity, and streams data to a secure cloud platform at regular intervals.

In this study, temperature, apparent extract, and pH are used directly in the modelling pipeline, while the remaining channels are omitted.

Figure 4. BrewIQ fermentation sensor array used to collect inline temperature, Plato, and pH measurements.

2.2. Wort Recipe, Yeast Inocula, and Operating Conditions

All pilot-scale batches were based on a single, IPA-style wort designed to provide a consistent substrate across fermentations. The grain bill comprised predominantly pale ale malt and Carapils, and hop additions chosen to produce a contemporary IPA aroma and bitterness profile (Cascade and Centennial). Using a standardised recipe isolates the influence of yeast strain, inoculum configuration, and temperature programme on the observed dynamics. This recipe was particularly important as these pilot-scale experiments constitute the overwhelming majority of data within the mixed-culture beers category.

Pilot-scale fermentations were performed by directly pitching active dry brewing yeast produced by Fermentis (Marquette-lez-Lille, France). Ale fermentations used a *Saccharomyces cerevisiae* strain suited to American pale ales and IPAs (Fermentis (Marquette-lez-Lille, France) SafAle US-05), while lager fermentations used *Saccharomyces pastorianus* (Fermentis (Marquette-lez-Lille, France) SafLager S-23). Mixed-culture beers combined ale and lager strains to create more complex flavour profiles.

Fermentations were grouped into four beverage categories: ales, IPAs, lagers, and mixed-culture beers. They were categorised based on the recorded style descriptor within the metadata of each fermentation batch.

Figure 5 illustrates three representative beers produced from the IPA-style wort using different yeast configurations at 22 °C. Even with identical wort, differences in yeast

physiology and fermentation temperature yield distinct appearance and flavour profiles, reflected in the corresponding apparent extract and pH trajectories.

(a)   (b)   (c)

Figure 5. Example beers produced from the IPA-style wort using different inocula at a primary fermentation temperature of 22 °C, photographed after two weeks of conditioning: (**a**) ale monoculture (SafAle US-05 at 80 g/hL); (**b**) mixed ale and lager culture (SafAle US-05 at 40 g/hL and SafLager S-23 at 40 g/hL); (**c**) lager monoculture (SafLager S-23 at 80 g/hL).

2.3. Dataset Construction and Preprocessing

Raw data were exported from the BrewIQ platform as comma-separated value files, each corresponding to a single fermentation. For every batch, the data included the following required attributes: the elapsed time since yeast pitch (`hours_from_pitch`, in hours), apparent extract (`plato`, in °P), fluid temperature (`fluid_temp`, in °C), and a style descriptor (`style_name`). Additional variables such as pH, dissolved oxygen, and conductivity were recorded for many fermentations, depending on the configuration used at each brewery.

The data were preprocessed and resampled before modelling. Batches lacking finite values for any of the required columns (time, Plato, temperature, style name) were discarded and two fermentations were removed on this basis. The remaining fermentations came from a mixture of pilot-scale trials and routine production in commercial breweries. The commercial data covered a wide variety of monoculture yeast inocula, temperature profiles, and recipes, including a broad range of ale and lager styles. Style names were normalised by stripping accents and diacritics, standardising capitalisation and spelling, and then mapped to beverage categories using rule-based pattern matching against curated lists. The following categories were identified: ales, IPAs, lagers, mixed-culture beers, stouts and porters, Belgian styles, sour beers, wheat beers, wines, seasonal and historical beers, flavoured and experimental beers, spirits, and uncategorised entries.

The LSTM plant model was trained on the four most populous categories: ales, IPAs, lagers, and mixed-culture beers. Restricting to these categories yielded 1305 fermentations in total. Within this modelling subset, the sampling period was 30 min, the time since pitch was capped at 300 h, and the median duration ranged from 141 h for ales to 300 h for lagers and mixed-culture fermentations. For each batch, original gravity descriptors were computed from the apparent extract (°P) time series. The initial gravity OG_{init} was defined as the first finite degrees Plato value at $t = 0$ h where available. A reference original gravity, OG_{ref}, was then obtained as the median Plato over the first 2 h of the fermentation; this approach was chosen to mitigate against measurement noise.

Temperature setpoints were reconstructed from the fluid temperature time series to provide a smoothed representation of the brewer's intended temperature programme. The

recorded temperatures were first smoothed using a three-point centred moving average, then rounded to the nearest 0.5 °C. Changes smaller than 0.5 °C were treated as noise, while sequences of approximately constant temperature were treated as plateaux. The resulting piecewise-constant trajectory, denoted $u_{\text{set}}(t)$, captures the main steps of the temperature schedule while filtering sensor noise and short-lived perturbations.

Table 1 summarises the modelling dataset by beverage category. Ales and IPAs dominate in terms of batch count and time points, reflecting their prevalence in contemporary production. The number of lager and mixed-culture fermentations is smaller but still sufficient to provide meaningful coverage for modelling. Almost all ale, IPA, and lager fermentations included pH measurements, whereas pH coverage was lower in mixed-culture fermentations.

Table 1. Summary of the modelling dataset used to train the LSTM plant model, grouped by beverage category. Durations are measured from yeast pitch to the last BrewIQ observation. Median values are reported for OG_{ref} and fluid temperature.

Metric	Ales	IPAs	Lagers	Mixed-Culture
Batches	575	503	193	34
Median duration (h)	141.0	212.0	300.0	300.0
Median OG_{ref} (°P)	11.62	14.27	12.05	12.37
Median T_{fluid} (°C)	19.6	20.1	13.9	13.2
Fermentations with pH (%)	99.8	97.4	96.4	55.9

2.4. LSTM Plant Model Formulation

The plant model (see Figure 6) is formulated as a multi-output, sequence-to-sequence LSTM network implemented in PyTorch (version 2.9.1). Its purpose is to predict short-horizon trajectories of degrees Plato and pH given a recent history of process measurements and a planned future temperature schedule. This model structure was chosen as, in combination with an optimiser, it is suitable for direct integration within a model predictive control (MPC) controller.

The model operates at a fixed sampling interval of 0.5 h. For each fermentation, a history window of length $L_{\text{hist}} = 12.0$ h and a forecast horizon of length $H_{\text{fore}} = 12.0$ h are considered, yielding $L = H = 24$ time steps per history and forecast segment. At each time step in the history window, the inputs to the model comprise the reconstructed temperature setpoint $u_{\text{set}}(t)$, the time since pitch, the original gravity descriptor (OG_{ref}), the current apparent extract value (°P), and a categorical beer style label. The style label is mapped to an integer index and embedded into a low-dimensional continuous vector through a learnable embedding layer. All continuous inputs are standardised using dataset-wide means and standard deviations, so that the network operates on inputs with approximately zero mean and unit variance per channel. The original gravity descriptor (OG_{ref}) is broadcast across time to give a constant sequence for each batch.

The recurrent backbone is a two-layer LSTM with 128 hidden units per layer. For the history window, the input sequence is formed by concatenating the scaled temperature, time, original gravity (OG_{ref}), apparent extract (°P), and style embedding at each time step, giving an input dimension of $4 + E$, where E is the 16-dimensional embedding vector. The LSTM processes this sequence and returns the sequence of hidden states, together with the final hidden and cell states. Two output heads, implemented as small feed-forward networks, map the hidden state at each forecast step to apparent extract and pH predictions, respectively. Both heads consist of a linear layer, a rectified linear unit activation, and a final linear layer that outputs a single scalar per time step.
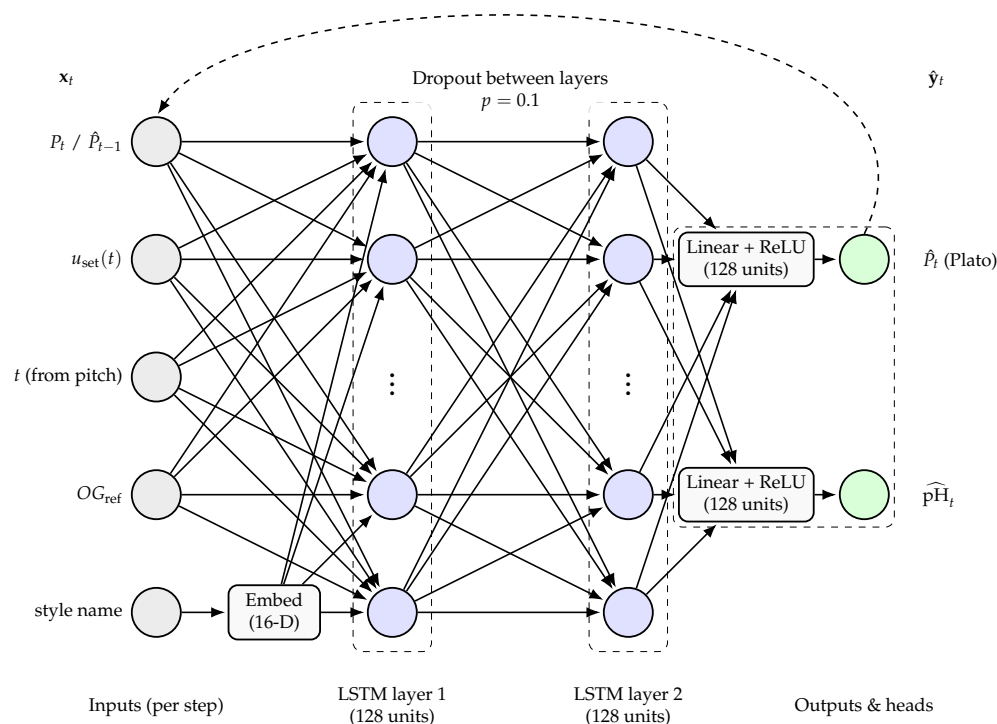
Figure 6. LSTM plant model: Per-step inputs pass through a 16-D style embedding, a two-layer LSTM backbone (128 units per layer with dropout between layers), and separate output heads. Each head consists of a hidden Linear–ReLU layer (128 units) followed by a final linear layer that outputs degrees Plato or pH. Predicted apparent extract is fed back autoregressively during forecasting.

Inference proceeds in two phases. First, the history window is passed through the LSTM to initialise the hidden and cell states. The last apparent extract value (°P) in the history segment is retained as the starting point for the autoregressive decoder. Second, the model rolls forward step by step over the forecast horizon. At each forecast step, the decoder forms an input vector by concatenating the scaled future temperature setpoint, future time, original gravity (OG_{ref}), the previously predicted degrees Plato value, and the style embedding. This input is passed through the same LSTM, using the hidden and cell states carried over from the previous step. The apparent extract (°P) and pH heads then produce predictions for the current forecast step, and the degrees Plato prediction is fed back as the apparent extract (°P) input for the next step. In this way, predicted apparent extract acts as a surrogate state variable in the decoder, allowing the network to propagate its own predictions forward in time.

2.5. Training Procedure and Loss Function

Before training, global scaling factors are estimated for each continuous variable. For every fermentation in the modelling subset, the sequences of temperature setpoint, time since pitch, original gravity (OG_{ref}), degrees Plato, and pH are concatenated, and a mean and standard deviation are computed across the combined dataset for each channel. These statistics are used to transform all inputs and targets to zero mean and unit variance. Scaling is applied identically to training and evaluation data.

Overlapping training examples are constructed by sliding a window across each fermentation. For each valid time index, a history segment of duration $L_{hist} = 12$ h and an immediately following forecast segment of duration $H_{fore} = 12$ h are extracted at 0.5 h resolution. The history contains scaled sequences of temperature setpoint, time since pitch, original gravity (OG_{ref}), degrees Plato, and the style embedding. The forecast contains scaled temperature, time, original gravity (OG_{ref}), and the corresponding degrees Plato

and pH targets. Windows containing any missing values in the input series or in the Plato targets are discarded. Windows with missing pH values are retained, but the missing entries are marked in a binary pH mask. This approach allows the model to use incomplete pH information without discarding otherwise useful windows.

Noise augmentation is applied to the training inputs to improve robustness to measurement variation. During training, the reconstructed temperature setpoint and apparent extract sequences are perturbed by truncated Gaussian noise. For each of these variables x, a zero-mean Gaussian noise term is added with a standard deviation chosen so that three standard deviations match the nominal sensor tolerance (approximately $\pm 1.0\,^{\circ}\mathrm{C}$ for temperature and $\pm 0.2\,^{\circ}\mathrm{P}$ for Plato). The perturbation is clipped to lie within the corresponding tolerance band, and the result is further restricted to physically plausible limits of 0–$30\,^{\circ}\mathrm{C}$ for temperature and 0–$20\,^{\circ}\mathrm{P}$ for apparent extract. The perturbed sequences are then scaled and used as inputs, while the original, unperturbed apparent extract and pH sequences form the targets. This procedure mimics realistic sensor noise without corrupting the ground truth.

To construct training and validation sets, all admissible windows are first enumerated without augmentation. A random subset comprising 10% of these windows is assigned to the validation set, and the remaining 90% form the training window pool. For validation, the windows are used in their unaugmented form. For training, windows are regenerated with noise augmentation, using independent random draws at each epoch. Because the split is performed at the window level rather than the batch level, windows from the same fermentation can appear in both the training and validation sets. The resulting training and validation losses therefore quantify generalisation across overlapping windows drawn from the same population of fermentations, rather than strictly held-out fermentations.

The model is trained using the Adam optimiser with a learning rate of 10^{-3}, mini-batch size of 32, and 20 epochs. Gradients are clipped to a global norm of 1.0 to improve stability, and a dropout rate of 0.1 is applied between the two LSTM layers. These hyperparameters were chosen to balance model capacity, training stability, and computational cost, and are consistent with standard practice for recurrent neural networks in process modelling [17,22].

The training objective is a masked Huber loss. For a set of predictions $\hat{z}$, targets z, and a binary mask m, we denote each prediction–target pair by $(x, y) = (\hat{z}, z)$. With a threshold parameter $\delta > 0$, the per-element Huber loss is defined as

$$\ell(x, y) = \begin{cases} \frac{1}{2}(x-y)^2, & \text{if } |x-y| < \delta, \\ \delta\left(|x-y| - \frac{1}{2}\delta\right), & \text{otherwise,} \end{cases} \tag{1}$$

and the masked Huber loss is the mean of $\ell(x, y)$ over all entries where $m = 1$. In this work, $\delta = 1.0$ is used for both apparent extract and pH in scaled units. During training, the total loss for a mini-batch is

$$\mathcal{L} = \mathcal{L}_{\text{Huber}}(\hat{P}, P; M_P) + \tfrac{1}{2}\,\mathcal{L}_{\text{Huber}}(\widehat{\text{pH}}, \text{pH}; M_{\text{pH}}), \tag{2}$$

where $\mathcal{L}_{\text{Huber}}$ denotes the masked Huber loss with $\delta = 1.0$, M_P is a mask of ones for the degrees Plato targets, and M_{pH} is the pH mask. Apparent extract (degrees Plato), therefore, acts as the primary objective, with pH contributing a secondary but non-negligible term.

To encourage robustness to missing pH measurements, an additional dropout mechanism is applied to the pH mask during training. At each batch, a random binary matrix D is sampled with entries drawn independently from a Bernoulli distribution with parameter $p = 0.30$. The effective training mask for pH is then

$$M_{\text{pH}}^{\text{train}} = M_{\text{pH}} \odot (1 - D), \tag{3}$$

where $\odot$ denotes element-wise multiplication. This mechanism effectively turns off a proportion of pH targets during training, even where measurements are available, and encourages the network to infer pH dynamics from context rather than relying solely on densely labelled trajectories. At the end of each epoch, mean training and validation losses are recorded, and the model with the lowest validation loss is retained for subsequent analysis.

2.6. Medoid Selection and Full-Run Evaluation

The training objective is defined on short forecast windows. To assess how well the LSTM plant model reproduces complete fermentations, it is evaluated in a one-step-ahead, open-loop configuration over the full 0–300 h horizon for a representative fermentation in each beverage category.

Representative fermentations are selected using a medoid approach. For each of the four categories (ales, IPAs, lagers, mixed-culture beers), only fermentations with BrewIQ data covering the entire 0–300 h interval are considered. Their degrees Plato trajectories are interpolated onto a uniform 1 h grid, and pairwise distances are computed between these interpolated trajectories as root-mean-square differences over time. For each candidate trajectory, the mean distance to all others in the same category is calculated. The medoid is defined as the fermentation whose mean distance is minimal. Where possible, the candidate pool is restricted to fermentations with pH measurements throughout most of the horizon, so that both degrees Plato and pH can be assessed.

For a chosen medoid, the trained LSTM plant model is rolled forward as follows: at each time step t in 0.5 h increments, a 12 h history window ending at t is extracted, scaled, and passed through the encoder to update the hidden state. The decoder then receives the next planned temperature setpoint and time index, together with original gravity (OG_{ref}), the style embedding, and the most recent degrees Plato value, and returns a one-step-ahead forecast for apparent extract (°P) and pH at time $t + 0.5$ h. Predictions are transformed back to physical units using the stored scalers. This procedure is repeated across the entire 0–300 h horizon. Root-mean-square error (RMSE) and mean absolute error (MAE) are computed for each medoid, using only those time points for which the relevant ground truth measurements are available. These medoid trajectories and metrics provide an intuitive picture of model performance across complete fermentations, complementing the window-based training and validation losses.

3. Results

3.1. Training and Validation Behaviour

Figure 7 shows the evolution of the training and validation losses over 20 epochs. The quantities plotted are the mean mini-batch losses at the end of each epoch, computed using the masked Huber objective on the scaled apparent extract and pH variables, as described in Section 2.5. The loss is therefore dimensionless and reflects the average discrepancy between predictions and measurements in standardised units.

Training begins from mean losses of 0.0775 on the training windows and 0.0556 on the validation windows. Both curves decrease steadily and remain close to each other throughout. By epoch 20, the mean training and validation losses have converged to 0.0128 and 0.0124, respectively. This behaviour indicates that the chosen architecture and regularisation provide a good fit to the windowed data without obvious overfitting at the window level.

Because the training vs validation split is constructed at the window level, with substantial overlap between windows drawn from the same fermentations, these losses should be interpreted as measures of within-dataset consistency rather than as fully independent

batch-level generalisation. Batch-level performance is explored qualitatively through the medoid trajectories presented in the following subsections.
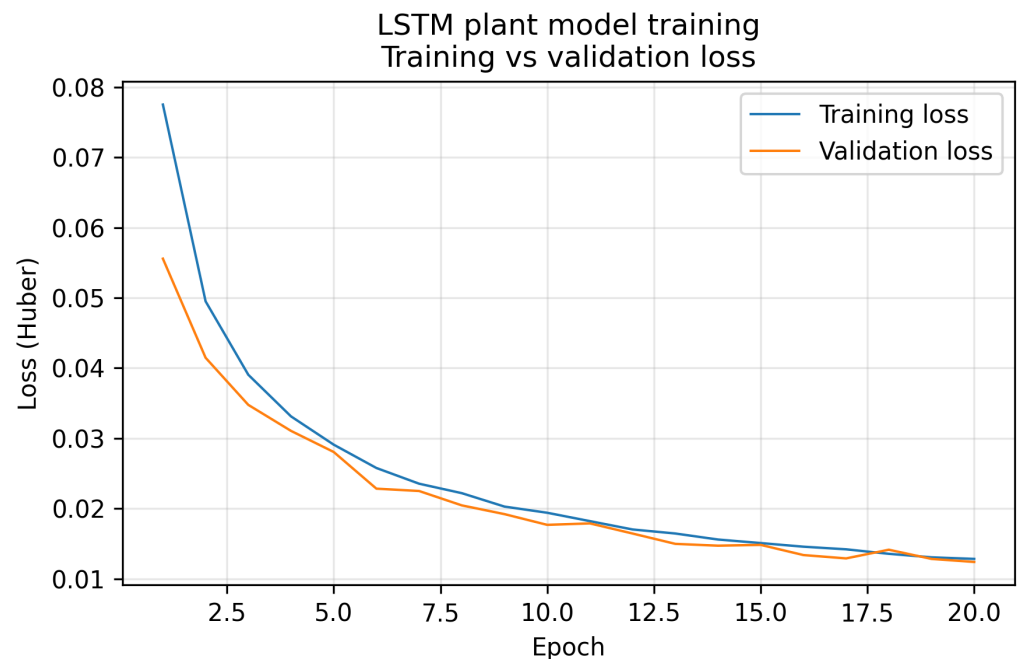

Figure 7. Training and validation losses for the LSTM plant model over 20 epochs, computed using the masked Huber objective on scaled degrees Plato and pH values.

3.2. Medoid Fermentations: Apparent Extract and pH Trajectories

The performance of the LSTM plant model across complete fermentations is illustrated by medoid trajectories for each beverage category. For each medoid, apparent extract and pH predictions are generated in a one-step-ahead, open-loop mode over the full 0–300 h horizon, as described in Section 2.6. The corresponding plots (Figures 8–15) show measured and predicted trajectories, together with residuals and tolerance bands of ±0.2 °P for apparent extract and ±0.1 for pH.

Figures 8 and 9 present the medoid ale fermentation. The ale medoid corresponds to an American pale ale with an original gravity (OG_{ref}) of 11.6 °P and a typical ale temperature schedule. The LSTM plant model follows the steep attenuation phase between roughly 25 and 100 h, reproducing both the rapid drop in degrees Plato and the subsequent approach to terminal gravity. The pH trajectory is also captured well, including the characteristic decline during active fermentation and the gradual recovery during conditioning. Across the full horizon, the degrees Plato predictions yield an RMSE of 0.304 °P and an MAE of 0.200 °P, while pH predictions achieve an RMSE of 0.096 and an MAE of 0.056.

Figures 10 and 11 show the medoid IPA fermentation, representing a higher gravity, more heavily hopped beer with an original gravity (OG_{ref}) of approximately 15.2 °P. The model again reproduces the overall shape of the degrees Plato trajectory, including the faster attenuation and slightly extended conditioning compared with the ale medoid. Errors peak at the end of the lag stage and during the most active phase of fermentation, where steeper gradients and potentially greater sensor noise are present, but remain within acceptable limits for process supervision in many settings. The apparent extract RMSE and MAE for the IPA medoid are 0.467 and 0.262 °P, respectively. For pH, the RMSE and MAE are 0.082 and 0.056, respectively.
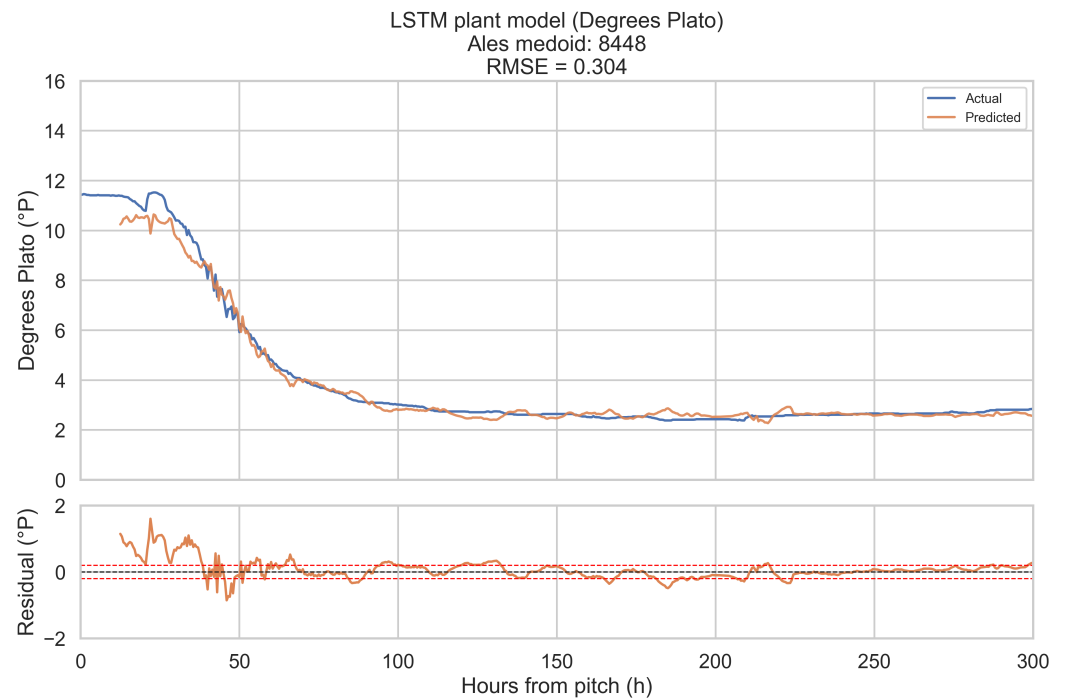
Figure 8. Medoid ale fermentation: LSTM plant model predictions versus BrewIQ observations for degrees Plato over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.
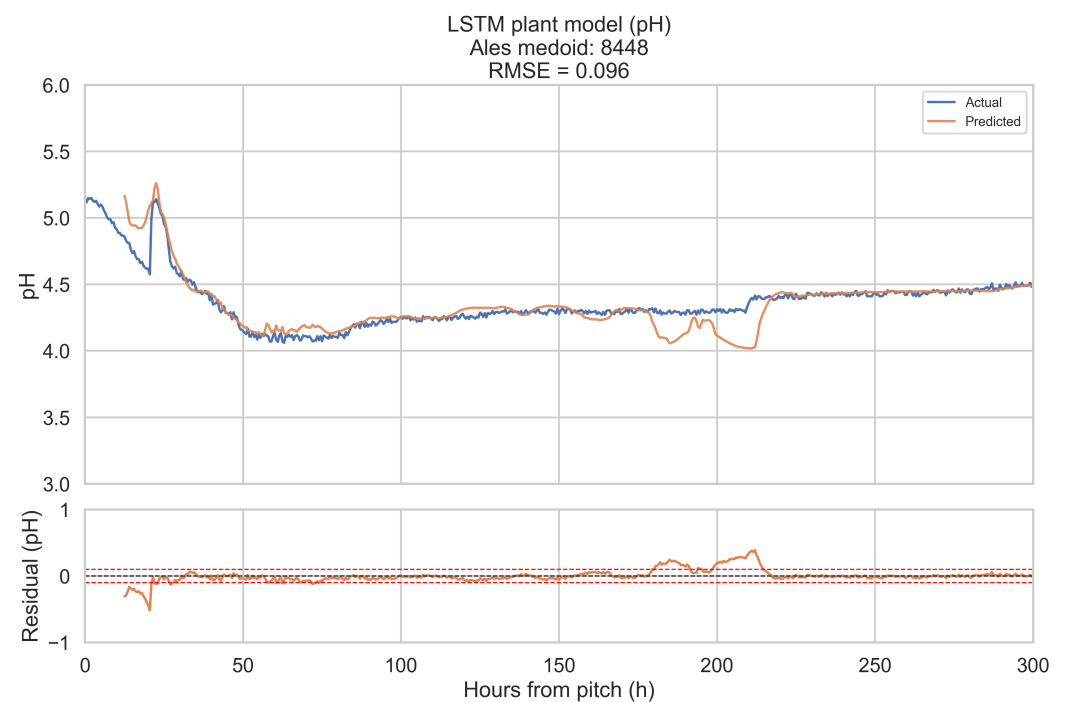

Figure 9. Medoid ale fermentation: LSTM plant model predictions versus BrewIQ observations for pH over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.
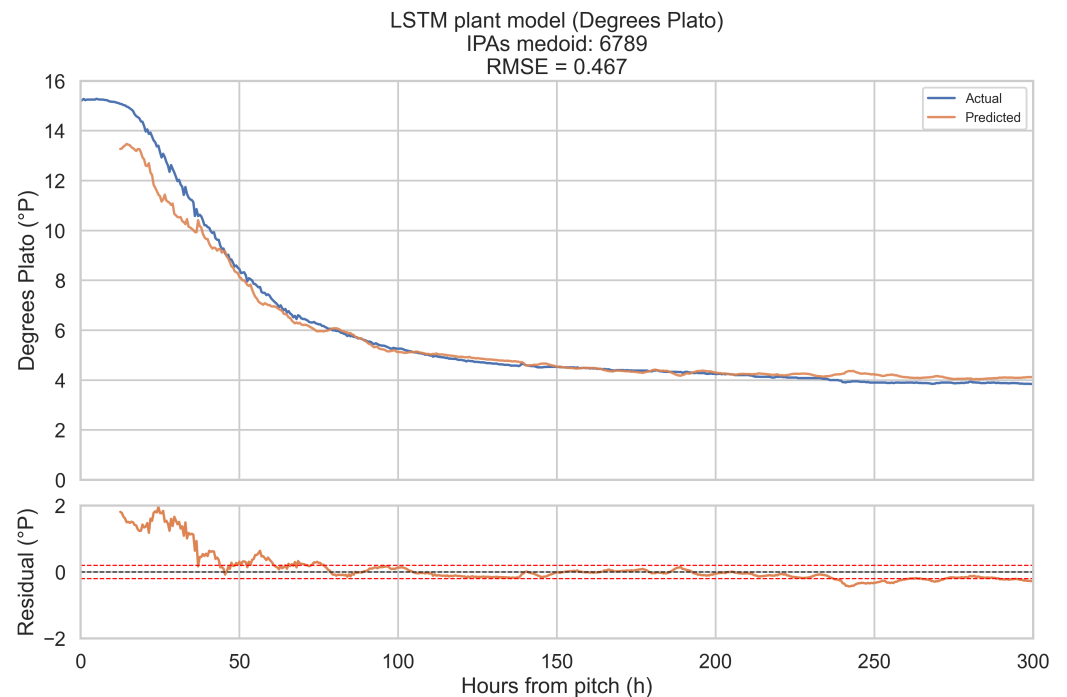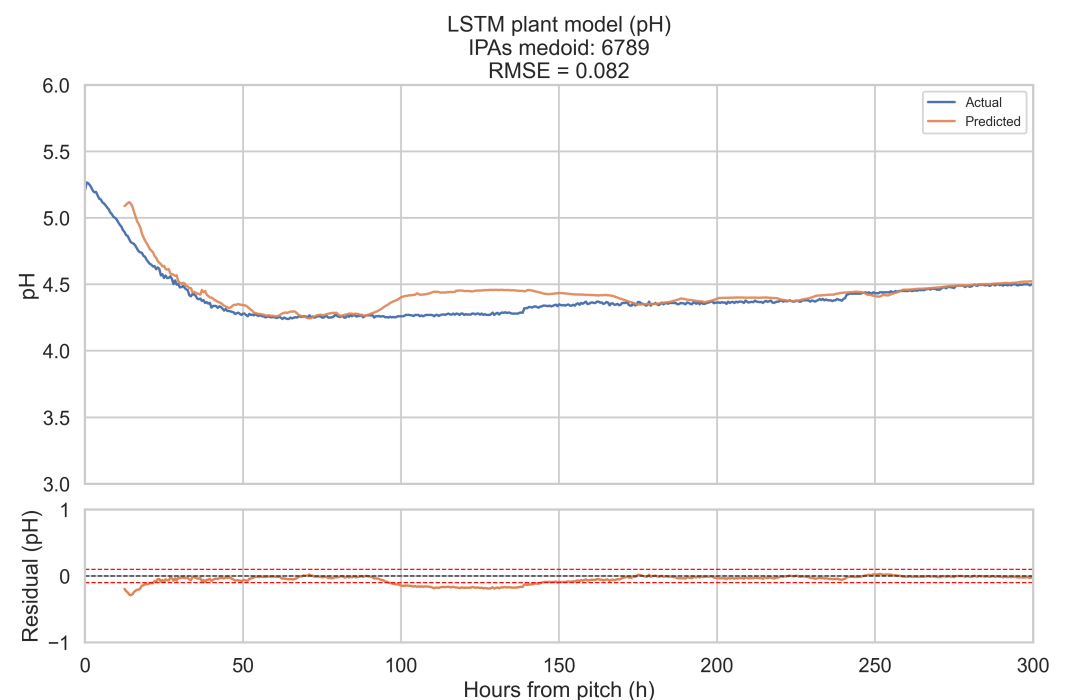
Figure 10. Medoid IPA fermentation: LSTM plant model predictions versus BrewIQ observations for degrees Plato over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.

Figure 11. Medoid IPA fermentation: LSTM plant model predictions versus BrewIQ observations for pH over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.

The medoid lager fermentation is presented in Figures 12 and 13. Lager fermentations typically exhibit slower attenuation (reflected in the degrees Plato curve), a diacetyl rest, and an extended cold conditioning period. Despite the longer duration and lower temperatures, the LSTM-based plant model closely tracks the trajectory. The apparent extract RMSE and MAE are 0.355 and 0.236 °P, and the pH RMSE and MAE are 0.046 and 0.030. The pH residuals are small and largely unbiased, suggesting that the model has learnt consistent relationships between temperature, attenuation, and pH evolution in lager fermentations.
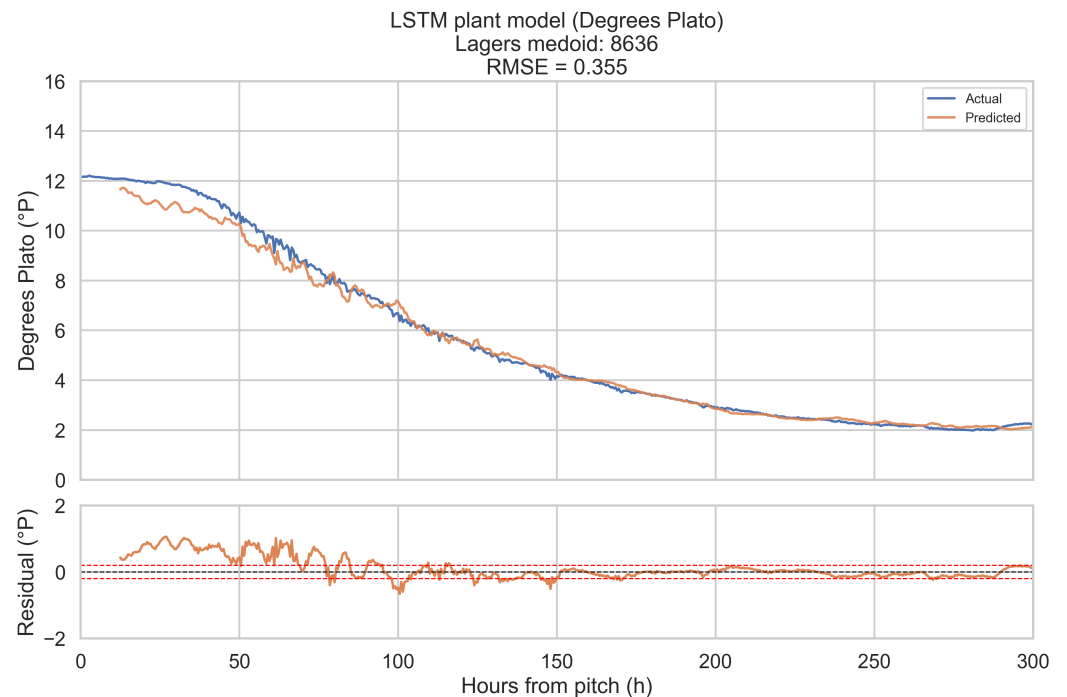
Figure 12. Medoid lager fermentation: LSTM plant model predictions versus BrewIQ observations for degrees Plato over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.
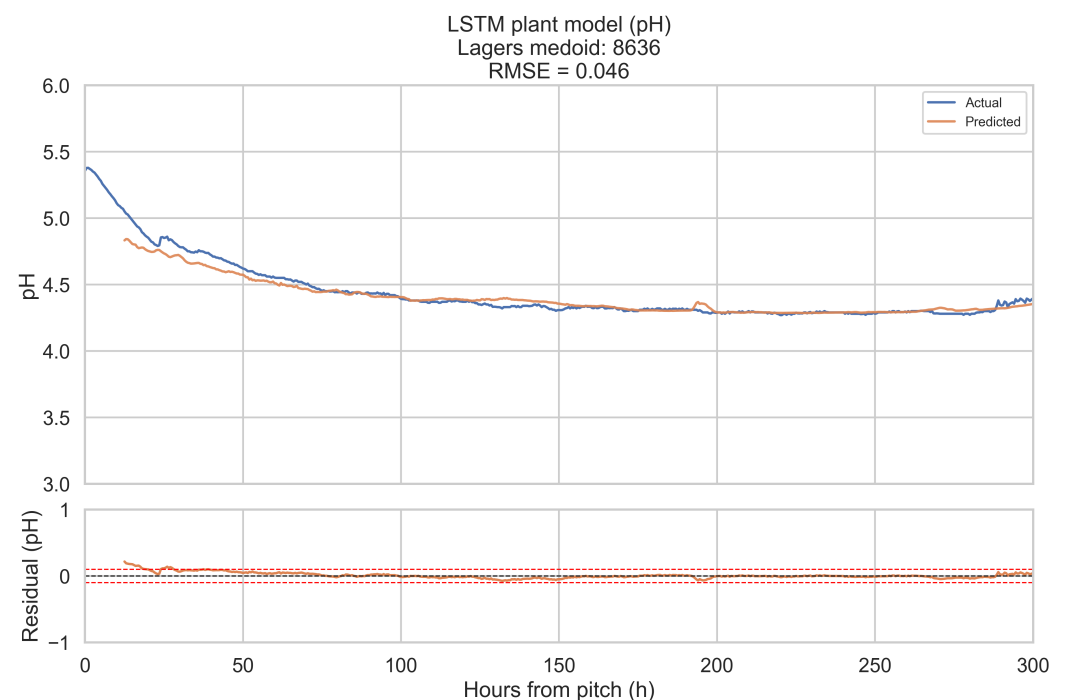

Figure 13. Medoid lager fermentation: LSTM plant model predictions versus BrewIQ observations for pH over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.

Mixed-culture fermentations pose a more challenging modelling problem because they involve interactions among multiple yeast species. Figures 14 and 15 show the medoid of mixed-culture fermentations present within the dataset. The LSTM plant model captures the decline in degrees Plato and the overall pH change associated with acidification. Over 0–300 h, the Plato RMSE and MAE are 0.228 and 0.128 °P, respectively. The pH RMSE and MAE are 0.071 and 0.033.
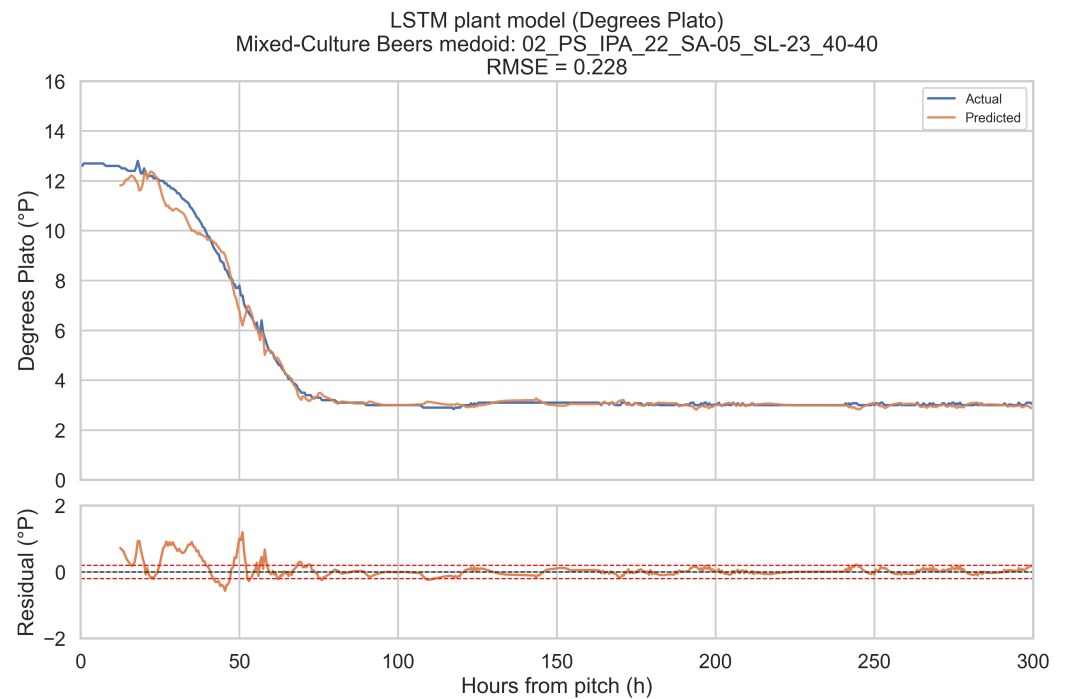
LSTM plant model (Degrees Plato)
Mixed-Culture Beers medoid: 02_PS_IPA_22_SA-05_SL-23_40-40
RMSE = 0.228

Figure 14. Medoid mixed-culture fermentation: LSTM plant model predictions versus BrewIQ observations for degrees Plato over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.
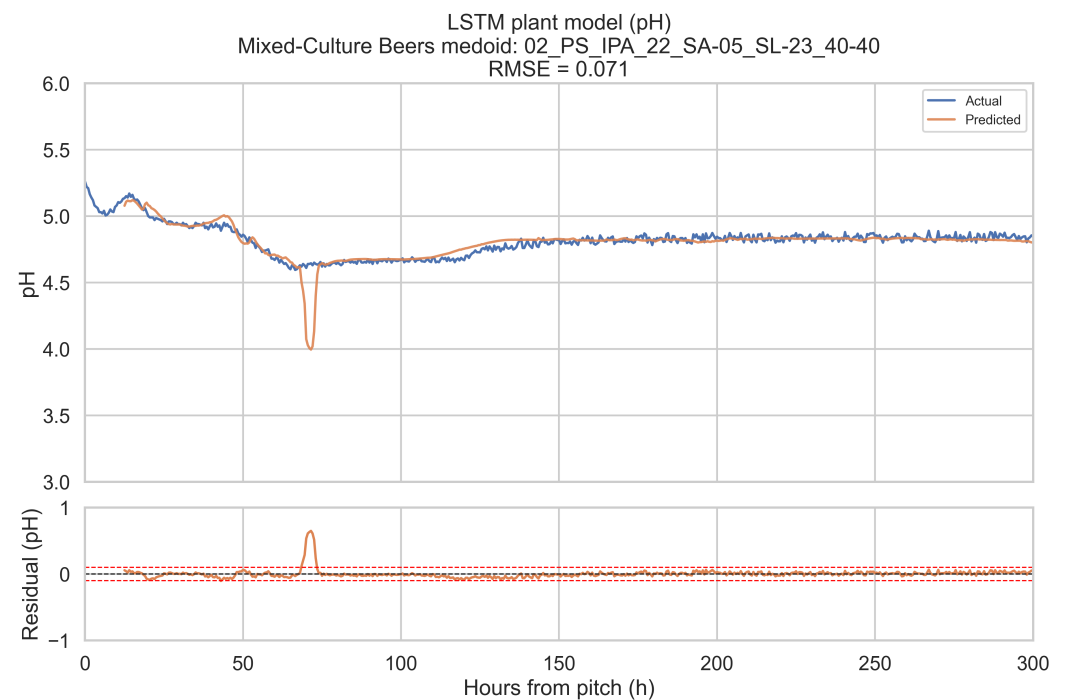
LSTM plant model (pH)
Mixed-Culture Beers medoid: 02_PS_IPA_22_SA-05_SL-23_40-40
RMSE = 0.071

Figure 15. Medoid mixed-culture fermentation: LSTM plant model predictions versus BrewIQ observations for pH over the 0–300 h horizon. Residuals and tolerance bands are shown in the lower panel.

These medoid trajectories should be viewed as representative examples rather than exhaustive summaries of performance across all batches. They highlight how the LSTM plant model behaves during typical fermentations in each category, but do not capture the full distribution of errors across the entire dataset.

3.3. Aggregate Medoid Metrics

Table 2 summarises the one-step-ahead, open-loop errors for the medoid fermentations in each category. Averaged over the four medoids, the mean Plato RMSE is 0.339 °P and the mean Plato MAE is 0.207 °P. For pH, the mean RMSE and MAE are 0.074 and 0.044, respectively. For pH, these values are similar in magnitude to the 0.1 pH unit tolerance assumed in the noise model. For Plato, the errors are somewhat larger than the 0.2 °P tolerance used in augmentation, but still of the same order of magnitude.

Table 2. LSTM plant model performance on medoid fermentations for each beverage category, evaluated as one-step-ahead, open-loop predictions over 0–300 h.

Category	Plato RMSE (°P)	Plato MAE (°P)	pH RMSE	pH MAE
Ales	0.304	0.200	0.096	0.056
IPAs	0.467	0.262	0.082	0.056
Lagers	0.355	0.236	0.046	0.030
Mixed-culture beers	0.228	0.128	0.071	0.033
Mean over categories	0.339	0.207	0.074	0.044

The pattern of errors differs between categories. IPAs show the largest apparent extract errors, consistent with their higher original gravities and fast attenuation, which magnify any mismatch between the model and the process. Conversely, mixed-culture fermentations yield the smallest apparent extract errors, as these batches share a common IPA-style wort and a mixed *S. cerevisiae* and *S. pastorianus* inoculum, and such combinations of *Saccharomyces* strains with complementary sugar utilisation profiles are known to support reliable, well-attenuated fermentations rather than simply altering their speed [23–26]. Across all categories, the relatively low pH errors suggest that the multi-output design, in which degrees Plato and pH share the same recurrent backbone, has captured useful cross-correlations between attenuation and acidification.

3.4. Errors and Temporal Behaviour

Inspection of the residuals across all medoids indicates that prediction errors are not uniformly distributed in time. For degrees Plato, the largest residuals occur during rapid changes in apparent extract, particularly in the early stages of active fermentation and around temperature ramps. During lag and late conditioning phases, residuals are typically small and centred near zero. For pH, residuals tend to be slightly biased around the point of maximum acidification, with the LSTM sometimes overestimating the depth of the pH minimum before converging back towards the observed trajectory. These biases remain of limited magnitude and do not accumulate over time.

Importantly, neither the apparent extract nor the pH prediction residuals exhibit signs of long-term drift across the 0–300 h model roll-outs. Suggesting that the autoregressive decoder remains stable when applied iteratively, even though it was trained only on 12 h forecast windows. Instead, the residual patterns align with a model whose errors are primarily driven by local dynamical discrepancies and moderate input noise, rather than by systematic biases or fundamental structural flaws. A more comprehensive analysis of variability across all fermentations—such as examining the distribution of RMSE values over the entire dataset—is deferred to future work.

4. Discussion

The results indicate that a single LSTM plant model can represent beer fermentation dynamics across ales, IPAs, lagers, and mixed-culture beers with an accuracy that is broadly compatible with inline measurement tolerances, especially for pH. This is achieved without

separate parameter sets for each category and without explicit mechanistic equations for biomass, sugar, or ethanol concentrations. Instead, the model operates directly on temperature, degrees Plato, pH, original gravity descriptors, and beer style labels, learning the collective behaviour of 1305 fermentations. Several aspects of the formulation contribute to this performance.

Firstly, the input feature set is deliberately close to what brewers observe in practice. The reconstructed temperature setpoint captures the main steps of the thermal control strategy, while the original gravity (OG_{ref}) summarises wort strength in a form that is robust to measurement noise. The use of time since pitch as an explicit input allows the LSTM to locate each forecast within the broader fermentation timeline, a key consideration given that fermentation is a time-variant process. Combined with the recent degrees Plato history, these features provide a compact yet informative representation of the present state of the process.

Secondly, the multi-output design, with a shared recurrent backbone and separate degrees Plato and pH output heads, allows the model to exploit relationships between attenuation and acidification. In many fermentations, pH dynamics are strongly coupled to yeast metabolism, and changes in pH can provide indirect information about the underlying state of the culture [27,28]. By jointly predicting apparent extract and pH, the network can use pH measurements when available to refine its internal representation, even though pH is not used as an input. The masked Huber loss, together with pH-specific dropout, prevents windows with sparse pH measurements from dominating training while still contributing useful information for soft-sensing.

Moreover, the choice of an LSTM architecture is well aligned with the data's characteristics. Beer fermentations display long-range temporal dependencies, for example, decisions made in the first 24 h can influence attenuation, pH, and flavour many days later. LSTMs are designed precisely to handle such long-range time dependencies [18,19]. The present model uses a relatively modest architecture by modern standards, with only two layers of 128 units and a 16-dimensional style embedding. Yet, the training curves and medoid results indicate that this capacity is sufficient for the process modelling challenge. In this respect, the LSTM plant model extends previous data-driven work in brewing [11,29] by operating on a larger, more varied dataset and by modelling both apparent extract and pH directly, without reliance on offline biomass measurements.

From an operational perspective, prediction error magnitudes in excess of 0.3–0.5 °P in apparent extract and 0.05–0.10 pH units relative to ground truth are meaningful. For many breweries, such errors are comparable to the variability observed between replicate fermentations under nominally identical conditions. A plant model with this level of fidelity can support several tasks. It can provide short-horizon forecasts of apparent extract and pH, helping brewers to anticipate when a fermentation will reach target attenuation or pH limits. It can act as a soft sensor for pH, providing consistency checks against noisy measurements or suggesting likely trajectories when pH measurements are temporarily unavailable. It can also serve as a basis for anomaly detection by identifying fermentations whose trajectories deviate significantly from the behaviour encoded in the training data.

At the same time, there are important limitations. The LSTM plant model, at present, is trained predominantly on fermentations drawn from a variety of commercial breweries, encompassing a broad mix of monoculture yeasts, recipes, and process equipment, with the mixed-culture beers representing the category in which laboratory-scale datasets contribute most strongly. Although this provides a richer basis than a single-wort, single-site study, it still covers only a subset of the diversity found in the wider brewing industry. Generalising the model to breweries with substantially different vessel designs, control philosophies, or product portfolios may therefore require additional data volumes and attributes. Similarly,

while the style embedding allows the network to adapt to different beer styles within the dataset, it does not have explicit access to formulation variables such as hop additions, adjunct use, or yeast generation, which are important factors of fermentation behaviour in other contexts.

Another limitation is that the model is purely data-driven. It does not explicitly enforce physical constraints such as non-negative extract, limits on achievable attenuation, or monotonic relationships between certain variables. Within the range of the training data, these constraints are respected implicitly, but there is no guarantee that extrapolations will remain physically realistic. Physics-informed or hybrid approaches, which embed approximate kinetic relationships or monotonic behaviours into the network [15,30], could be explored to improve extrapolation and provide stronger guarantees on model behaviour in sparse or novel operating regimes.

The way in which training and validation sets are constructed also warrants careful interpretation. Windows are sampled at 0.5 h intervals with substantial overlap, and the training and validation split is performed at the window level, not at the batch level. This means that windows from the same fermentation can be used for both training and validation. The excellent agreement between training and validation losses, therefore, indicates consistency across overlapping windows from the same population of fermentations, but may overestimate performance on entirely unseen fermentations. A stricter evaluation, in which complete fermentations are held out at the batch level, would provide a more conservative measure of generalisation. Similarly, the medoid-based analysis, while intuitive and informative, reports full-run errors for only one representative fermentation per category. It does not capture the full distribution of errors across all batches, nor does it quantify worst-case performance. This, however, is to be treated strictly as an exemplar of "typical" model performance, as a complement to the reported error metrics.

The term "digital twin" in the wider literature encompasses a broad spectrum of models and couplings [17]. In the context of this research, the LSTM plant model is best viewed as a data-driven surrogate of the fermentation dynamics that could be embedded within a digital twin or MPC architecture, rather than as a complete digital twin in its own right. The present study focuses on establishing that a compact, recurrent model can reproduce the core dynamics of typical fermentations with useful accuracy.

Finally, there is scope to refine the representation of batch-level information and to optimise hyperparameters. In the current formulation, original gravity (OG_{ref}) and the style embedding are the only batch-level descriptors. Future models could incorporate additional descriptors, such as yeast generation, fermentation volume, or key recipe parameters, or could use learned embeddings of entire recipes or process histories. However, accessing or producing such a dataset presents a significant research challenge. Hyperparameters such as the number of LSTM layers, the hidden size, and the style embedding dimension were chosen to balance model capacity and training stability based on "best practices" and the scale of the dataset, rather than through an exhaustive search. Systematic hyperparameter optimisation, possibly using Bayesian optimisation or more sophisticated methods, could further improve performance.

Overall, the present results suggest that LSTM networks provide a practical route to building data-driven plant models of beer fermentations from routinely collected brewery data. They complement rather than replace mechanistic models and brewing expertise, and they provide a flexible foundation for developing more advanced monitoring and control strategies.

5. Conclusions

This paper has presented a long short-term memory (LSTM) plant model for beer fermentation, trained on a large collection of inline-monitored fermentations using Sennos' BrewIQ sensor array (see Figure 4). The model is built upon reconstructed temperature setpoints, time since pitch, original gravity descriptors, apparent extract, and beer style labels, and predicts short-horizon trajectories of apparent extract and pH. The training dataset comprises 1305 fermentations drawn from ales, IPAs, lagers, and mixed-culture beers.

The LSTM plant model uses a two-layer architecture with a style embedding and a multi-output head structure for apparent extract and pH. Training is carried out on overlapping 12 h history and 12 h forecast windows, using a masked Huber loss to accommodate incomplete pH data and modest noise augmentation on temperature and degrees Plato inputs.

When evaluated on representative medoid fermentations in each beverage category, the model reproduces degrees Plato and pH trajectories over the full 0–300 h horizon with errors that are on the same order as the underlying sensor tolerances. Across the four medoids, the mean apparent extract RMSE and MAE are 0.339 and 0.207 °P, and the mean pH RMSE and MAE are 0.074 and 0.044. For pH, these values are close to a 0.1 pH unit tolerance and for apparent extract, they are modestly larger than a 0.2 °P sensor tolerance, but remain within a range that is likely to be useful for monitoring and forecasting.

The research underscores the importance of careful data handling in industrial machine-learning applications. Uniform resampling at 0.5 h, robust estimation of original gravity (OG_{ref}), reconstruction of temperature setpoints, and explicit encoding of beer style all contribute to the success of the model. The masked Huber loss and pH-aware dropout provide a straightforward way to learn from incomplete pH measurements while retaining the benefits of joint apparent extract and pH modelling.

From an application standpoint, the LSTM plant model can support short-horizon forecasting of attenuation and pH (via soft sensing), and anomaly detection during fermentation. With further development, it could also be embedded within a model predictive control framework, allowing breweries to explore alternative temperature schedules and assess their likely impact on fermentation trajectories "in silico" before committing to process changes.

Key limitations are the reliance on the variables available from inline sensing (excluding biomass and other biochemical states), incomplete pH coverage, and the limited representation of mixed-culture fermentations in the training data. For future work and deployment, priorities include external validation across additional breweries and vessel geometries, uncertainty quantification and drift monitoring, and incorporation of richer operational metadata (e.g., yeast strain, pitching rate, oxygenation) where available.

In conclusion, the LSTM-based plant model introduced in this work delivers a data-driven yet computationally efficient model of beer fermentation behaviour. It demonstrates how recurrent neural networks trained on production data can serve as practical building blocks for brewery digital twins and decision-support workflows.

Author Contributions: Conceptualisation, A.O.; methodology, A.O.; software, A.O.; validation, A.O.; formal analysis, A.O.; investigation, A.O.; resources, A.O.; data curation, A.O.; writing—original draft preparation, A.O.; writing—review and editing, A.O., H.Z., and D.A.; visualisation, A.O.; supervision, H.Z. and D.A.; project administration, H.Z. and D.A. All authors have read and agreed to the published version of the manuscript.

References

1. Briggs, D.E.; Brookes, P.A.; Stevens, R.; Boulton, C.A. *Brewing: Science and Practice*; Elsevier: Amsterdam, The Netherlands, 2004.
2. Lodolo, E.J.; Kock, J.L.F.; Axcell, B.C.; Brooks, M. The yeast Saccharomyces cerevisiae- the main character in beer brewing. *Fems Yeast Res.* **2008**, *8*, 1018–1036. https://doi.org/10.1111/j.1567-1364.2008.00433.x.
3. White, C.; Zainasheff, J. *Yeast: The Practical Guide to Beer Fermentation*; Brewers Publications: Boulder, CO, USA, 2010.
4. Tonsmeire, M. *American Sour Beers*; Brewers Publications: Boulder, CO, USA, 2014.
5. Sparrow, J. *Wild Brews: Beer Beyond the Influence of Brewer's Yeast*; Brewers Publications: Boulder, CO, USA, 2005.
6. Yerolla, R.; Mehshan K M, M.; Roy, N.; Harsha, N.S.; Pavan Ganesh, M.P.; Besta, C.S. Beer fermentation modeling for optimum flavor and performance. *IFAC-PapersOnLine* **2022**, *55*, 381–386. https://doi.org/10.1016/j.ifacol.2022.04.063.
7. Gee, D.A.; Ramirez, W.F. A Flavour Model for Beer Fermentation. *J. Inst. Brew.* **1994**, *100*, 321–329. https://doi.org/10.1002/j.2050-0416.1994.tb00830.x.
8. de Andrés-Toro, B.; Girón-Sierra, J.M.; López-Orozco, J.A.; Fernández-Conde, C. Optimization of a Batch Fermentation Process by Genetic Algorithms. *IFAC Proc. Vol.* **1997**, *30*, 179–184. https://doi.org/10.1016/S1474-6670(17)43157-0.
9. de Andrés-Toro, B.; Girón-Sierra, J.; López-Orozco, J.; Fernández-Conde, C.; Peinado, J.; Garcia-Ochoa, F. A kinetic model for beer production under industrial operational conditions. *Math. Comput. Simul.* **1998**, *48*, 65–74. https://doi.org/10.1016/S0378-4754(98)00147-5.
10. de Andrés-Toro, B.; Girón-Sierra, J.M.; Fernández-Blanco, P.; López-Orozco, J.A.; Besada-Portas, E. Multiobjective optimization and multivariable control of the beer fermentation process with the use of evolutionary algorithms. *J. Zhejiang-Univ.-Sci. A* **2004**, *5*, 378–389.
11. Trelea, I.C.; Titica, M.; Landaud, S.; Latrille, E.; Corrieu, G.; Cheruy, A. Predictive modelling of brewing fermentation: From knowledge-based to black-box models. *Math. Comput. Simul.* **2001**, *56*, 405–424. https://doi.org/10.1016/S0378-4754(01)00311-1.
12. Paul, G.; Hawkins, C. The Real-Time Optimisation of an Industrial Fermentation Process. *IFAC Proc. Vol.* **2004**, *37*, 529–534. https://doi.org/10.1016/S1474-6670(17)32636-8.
13. Chai, W.Y.; Teo, K.T.K.; Tan, M.K.; Tham, H.J. Fermentation Process Control and Optimization. *Chem. Eng. Technol.* **2022**, *45*, 1731–1747. https://doi.org/10.1002/ceat.202200029.
14. Kondakci, T. Recent Applications of Advanced Control Techniques in Food Industry. *Food Bioprocess Technol.* **2017**, *10*, 522–542.
15. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168.
16. Seborg, D.E.; Edgar, T.F.; Mellichamp, D.A.; Doyle. *Process Dynamics and Control*, 3rd ed., International Student Version, ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011.
17. Brunton, S.L.; Kutz, J.N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2022.
18. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
19. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924.
20. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. https://doi.org/10.1038/nature14539.
21. O'Brien, A.; Zhang, H.; Allwood, D.M.; Rawsthorne, A. From Data to Draught: Modelling and Predicting Mixed-Culture Beer Fermentation Dynamics Using Autoregressive Recurrent Neural Networks. *Modelling* **2024**, *5*, 201–222. https://doi.org/10.3390/modelling5010011.

22. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019.

23. Gallone, B.; Steensels, J.; Prahl, T.; Soriaga, L.; Saels, V.; Herrera-Malaver, B.; Merlevede, A.; Roncoroni, M.; Voordeckers, K.; Miraglia, L.; et al. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. *Cell* **2016**, *166*, 1397–1410.e16. https://doi.org/10.1016/j.cell.2016.08.020.

24. Gibson, B.; Geertman, J.M.A.; Hittinger, C.T.; Krogerus, K.; Libkind, D.; Louis, E.J.; Magalhães, F.; Sampaio, J.P. New yeasts—New brews: Modern approaches to brewing yeast design and development. *Fems Yeast Res.* **2017**, *17*, fox038. https://doi.org/10.1093/femsyr/fox038.

25. Basso, R.F.; Alcarde, A.R.; Portugal, C.B. Could non-Saccharomyces yeasts contribute on innovative brewing fermentations? *Food Res. Int.* **2016**, *86*, 112–120. https://doi.org/10.1016/j.foodres.2016.06.002.

26. Dysvik, A.; La Rosa, S.L.; Liland, K.H.; Myhrer, K.S.; Østlie, H.M.; De Rouck, G.; Rukke, E.O.; Westereng, B.; Wicklund, T. Co-fermentation Involving Saccharomyces cerevisiae and Lactobacillus Species Tolerant to Brewing-Related Stress Factors for Controlled and Rapid Production of Sour Beer. *Front. Microbiol.* **2020**, *11*, 279–279. https://doi.org/10.3389/fmicb.2020.00279.

27. Martens, H.; Iserentant, D.; Verachtert, H. Microbiological Aspects of a Mixed Yeast—Bacterial Fermentation in the Production of a Special Belgian Acidic Ale. *J. Inst. Brew.* **1997**, *103*, 85–91. https://doi.org/10.1002/j.2050-0416.1997.tb00939.x.

28. De Roos, J.; De Vuyst, L. Microbial acidification, alcoholization, and aroma production during spontaneous lambic beer production. *J. Sci. Food Agric.* **2019**, *99*, 25–38. https://doi.org/10.1002/jsfa.9291.

29. Peleg, M. A New Look at Models of the Combined Effect of Temperature, pH, Water Activity, or Other Factors on Microbial Growth Rate. *Food Eng. Rev.* **2022**, *14*, 31–44. https://doi.org/10.1007/s12393-021-09292-x.

30. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.