

# COMBINING MECHANISTIC MODELLING WITH MACHINE LEARNING IN AN INDUSTRIAL CASE STUDY: PREDICTING CREAM CHEESE FERMENTATION

W. Yu<sup>1,\*</sup>, D. I. Wilson<sup>2</sup>, Y. Lin<sup>1</sup> and B. R. Young<sup>1,\*</sup>  
<sup>1</sup>University of Auckland & <sup>2</sup>Auckland University of Technology  
Auckland New Zealand

## ***Abstract***

The modelling of the fermentation of a live culture using solely regressed dynamic models can be improved by incorporating deep learning recurrent models such as long-short term memory (LSTM). Not only do these models improve the predictability of a key component during the reaction, the very fact that they work so well challenges a key assumption that the output variable is only algebraically related to the states. This example shows that machine learning has a place even when fundamental models exist.

## ***Keywords***

LSTM, fermentation kinetics, NARX, pH prediction.

## **Introduction**

The challenge to produce a consistent, reproducible biological product such as cream cheese at industrial scale is hampered by the vagaries inherent in anything fermented from a live culture, (Munir 2015). Any sort of model-based control, or even just employing sensor fusion, requires the control designer to first select a suitable dynamic model structure, and then regress the parameters appropriate to the specific application, (Mears, 2017). However, even a cursory glance at the literature uncovers a plethora of competing biological growth models, all with parameters of widely varying orders of magnitudes (e.g. Monod (1949), Bouguetoucha (2011), and Amrane, (1999)). A natural question therefore is why is this so hard?

A more modern approach may well be to lessen the deterministic model requirement, (or even to abandon it completely), and to instead employ machine learning techniques. This has the advantage of efficiently using the logged process data that is readily available (from the historian) as opposed to the more informative, (but much more expensive and therefore rare) validation concentration/quality data, Boiarkina (2018). Of course

experience has shown that such an approach is hardly a panacea, (see for example Depree, (2019)), and given that the number of deep learning paradigms is increasing at an exponential rate it is difficult to know which strategy is appropriate for a given application. The choice of learning model structure and appropriate tuning factors is a combinatorially large problem, and sub-optimality is deadly to what is a time-consuming computational exercise.

## **Industrial cream cheese prediction**

One of the steps in the industrial production of cream cheese is knowing when to terminate the fermentation process. Traditionally pH has been used given that it is a good proxy for the extent of reaction, and sensors are available, although Fig.1 indicates the typical variation throughout the season, and the sensor drift. These variations are considered detrimental to the final quality of the food product.

---

\* To whom all correspondence should be addressed

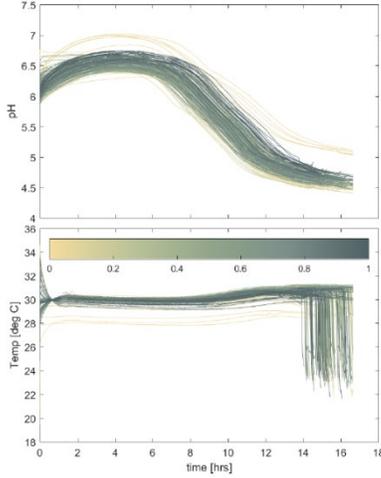


Figure 1: Multiple pH and temperature profiles showing typical variations from a single fermentation vat over the course of a season. The shading corresponds to the day throughout the production season

In this paper, we used a combination of a traditional kinetic model and a deep ANN network LSTM (long-short term memory, Hochreiter (1997)) to predict the cream cheese pH which is a key indicator as to when to stop the fed-batch fermentation. The kinetic model was validated using experiment data than those used for the network training. The overall modelling framework is shown in Fig. 2.

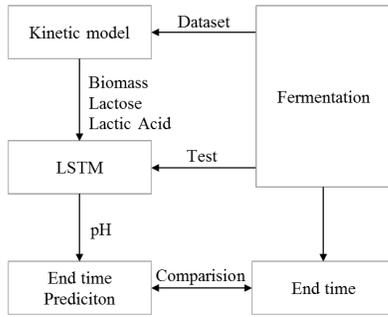


Figure 2: Model structure

### Cream cheese fermentation data

The lab-scale fermenter used in this research for model validation is a BioFlo 3000 provided by New Brunswick (Chemical and Materials Engineering, University of Auckland), where the pH and temperature could be monitored continuously. Samples are taken at every 1.5 hours for biomass, lactose and lactic acid concentration measurement.

### Kinetic model

By comparing many different kinetic models, we selected a model structure proposed by Boonmee et al. (2013). The model consists of three dynamic equations:

Cell growth rate:

$$\frac{dX}{dt} = \mu_{max} \left( \frac{S}{K_{Sx} + S} \right) \left( 1 - \frac{P - P_{ix}}{P_{mx} - P_{ix}} \right) \left( \frac{K_{ix}}{K_{ix} + S} \right) X \quad (1)$$

Lactose consumption rate:

$$\frac{dS}{dt} = q_{s,max} \left( \frac{S}{K_{Ss} + S} \right) \left( 1 - \frac{P - P_{is}}{P_{ms} - P_{is}} \right) \left( \frac{K_{is}}{K_{is} + S} \right) X \quad (2)$$

Lactic acid production rate:

$$\frac{dP}{dt} = \alpha \frac{dX}{dt} + q_{p,max} \left( \frac{S}{K_{Sp} + S} \right) \left( 1 - \frac{P - P_{ip}}{P_{mp} - P_{ip}} \right) \left( \frac{K_{ip}}{K_{ip} + S} \right) X \quad (3)$$

Where the variables  $X$ ,  $S$ ,  $P$  are the concentration of biomass, lactose and lactic acid ( $\text{g L}^{-1}$ ).  $\mu_{max}$  is the maximum specific growth rate ( $\text{h}^{-1}$ ),  $q_{p,max}$  is the maximum specific lactic acid production rate ( $\text{g g}^{-1} \text{h}^{-1}$ ) and  $q_{s,max}$  is the maximum specific lactose utilisation rate ( $\text{g g}^{-1} \text{h}^{-1}$ ).  $K_s$  is the Monod constant or substrate saturation constant ( $\text{g L}^{-1}$ ),  $P_i$  is the threshold lactic acid concentration ( $\text{g/L}$ );  $P_m$  is the maximum lactic acid concentration ( $\text{g/L}$ );  $K_i$  is the product inhibition constant ( $\text{g/L}$ ).  $\alpha$  is the growth - associated product form coefficient.

### Long-Short Term Memory (LSTM) network

The LSTM has become very popular due to its accuracy in predicting time series data and avoiding the “vanishing gradient” problem whilst learning. Compared to a simple recurrent network, the LSTM proposes the concepts of a memory cell, input gate, output gate, and the forget gate. Fig. 3 shows the (vanilla) LSTM architecture chosen for this work, (Greff et al, 2017).

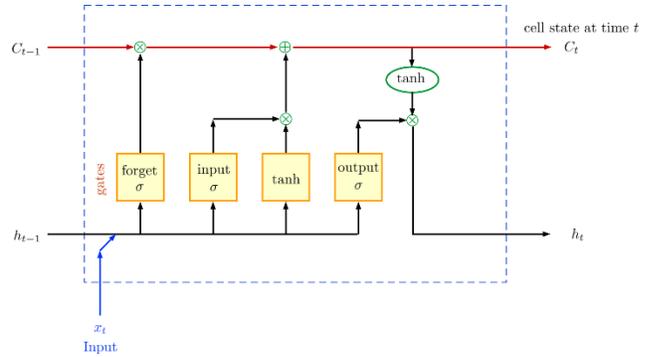


Figure 3: The LSTM architecture used for this work

### Results and Discussion

The kinetic model fitting one experiment data set is shown in Fig 4.

The model fits the experimental data reasonably well given the difficulties of a bio-system, with all the  $R^2$  values greater than 0.88. We also tested the model with different experimental initial conditions, and the model delivers an average  $R^2 > 0.84$  except for the biomass. The biomass measurements include significant measurement variations which may be from  $\pm 16\%$  to 52% reported by ISO (2003). We have to reduce the parameter weights for biomass parameters.

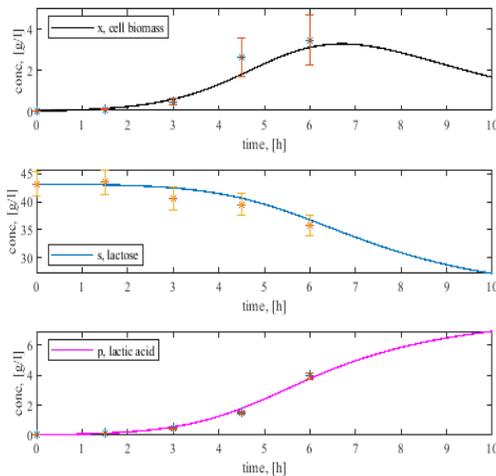


Figure 4: The fitting of the 3 states from the cream cheese kinetic model using experiment data

This industrial application shows that combining traditional regression techniques (with all the difficulties inherent in such an approach) with an LSTM neural-network model for the prediction of a key measured variable, (pH), out-performed other strategies, including black-box nonlinear models such as NARX, as shown in Fig 5.

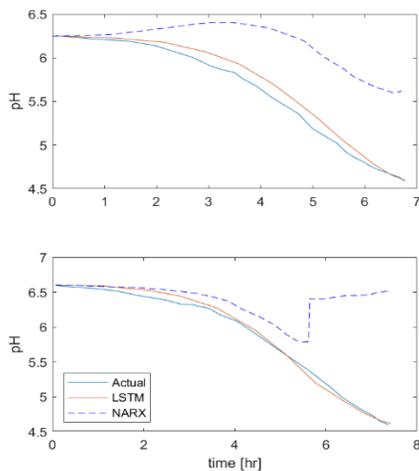


Figure 5: Comparing NARX & LSTM pH predictions.

What is interesting though is that the successful approach used a traditional dynamic model to predict the states, and an LSTM model to predict the output. From theory, it was expected that the output is only algebraically related to the states (as opposed to dynamically related), so a generic algebraic model with no internal memory should suffice. As it turned out, a dynamic model was required, (the LSTM model with considerable internal delays) and that proved substantially better than the Hammerstein type nonlinear black box model.

## Significance

Given that industrial process engineers are reluctant to embrace models with little or no underlying fundamental theory, it is interesting to note that if a dynamic model consistently out-performs a static model, then this indicates that the underlying phenomena may be more complex than originally postulated. We see this sort of challenge, created by using deep learning models that remove any preconceived bias a useful check for process engineering model-based control.

## Acknowledgments

The authors would like to acknowledge Matthew Proctor from Fonterra for their help in the development of the underlying models, and discussions throughout this project.

## References

- Amrane, A and Prigent, Y. (1999) Differentiation of pH and free lactic acid effects on the various growth and production phases of *lactobacillus helveticus*. *Journal of Chemical Technology and Biotechnology*, 74(1):33–40.
- Boiarkina, I, Depree, N, Prince-Pike, A, Yu, W, Wilson, D.I., Young B.R., (2018) Using Big Data in Industrial Milk Powder Process Systems, *Computer Aided Chemical Engineering*, (44), Pages 2293-2298.
- Boonmee, M., Leksawasdi, N., Bridge, W., Rogers, R.L., 2003. Batch and continuous culture of *Lactococcus lactis* NZ133: experimental data and model development, *Biochemical Engineering Journal*, 14, 2, 127-135.
- Bouguettoucha, A, Balannec, B, and Amrane, A. (2011). Unstructured models for lactic acid fermentation-a review. *Food Technology and Biotechnology*, 49(1):3
- Depree, N, Prince-Pike, A, Young, B.R., Wilson, D.I. (2019) Predictive modelling of instant whole milk powder functional performance across three industrial plants, *Journal of Food Engineering*, (252), 1-9.
- Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28, 2222-2232.
- Hochreiter, S., and J. Schmidhuber, J. (1997). Long short-term memory. *Neural computation*. Vol. 9, Number 8, pp.1735–1780.
- ISO, (2003). *Yogurt-Enumeration of characteristic microorganisms- Colony-count technique at 37C*, in ISO 7889. 2003, International Organization for Standardization.
- Lin, Y (2018). Cream cheese fermentation pH prediction. *Masters Thesis*, The University of Auckland, New Zealand.
- Mears, L, Stocks, S.M., Sin, G, and Gernaey, K.V. (2017) A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *Journal of Biotechnology*, (245)
- Monod, J. (1949) The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394.
- Munir, T, Yu, W, Young, B.R. and Wilson, D.I. (2015) The current status of process analytical technologies in the dairy industry, *Trends in Food Science & Technology*, 43, (2), 205-218.