Available online at www.sciencedirect.com

ScienceDirect

Journal of Food Engineering 83 (2007) 167–172

JOURNAL OF FOOD ENGINEERING

www.elsevier.com/locate/jfoodeng

Modelling beer fermentation variability

M. Defernez [a], R.J. Foxall [a,*], C.J. O'Malley [b], G. Montague [b], S.M. Ring [a], E.K. Kemsley [a]

[a] Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK

[b] School of Chemical Engineering and Advanced Materials, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

Available online 22 February 2007

Abstract

In this paper we present the outcomes of three different approaches to characterising beer fermentations, with the particular aim of predicting the likelihood of the target 'present gravity' (PG) being reached within a given time window. The study uses data collated at real brewery sites, and from three different beer qualities. The approaches include: the modelling of the PG curve by a mathematical function; a nearest neighbour (NN) approach; and the generation of centile curves. We show that it is useful to combine these approaches; a software package allowing them to be integrated has been developed, which enables an informed judgement to be made as to whether a given fermentation deviates from normal behaviour.

Keywords: Beer; Centile; Fermentation; Incomplete beta-function; Nearest neighbour

1. Introduction

The fermentation of beer is an inherently variable process, affected by factors such as the raw ingredients' composition and the yeast characteristics (Hough, Briggs, & Stevens, 1971). In practice, this means that fermentation times can vary considerably between batches of the same beer quality. This leads to the current practice where the fermentation process is often continued longer than strictly necessary, to ensure that the fermentation is sufficiently progressed before proceeding to the next stage of beer production. This is remaining so despite attempts at modelling this process in the past (De Andres-Toro et al., 1998; Rousu, Elomaa, & Aarts, 1999; Trelea et al., 2001; Warnes, Glassey, Montague, & Kara, 1996; Siebert, 2001). The piece of work presented here is based on the fact that the two primary (although not necessarily only) endpoints of fermentation are that the present gravity (PG) and the diacetyl (or vicinal diketones) reach their specifications, after which the beer can in principle be processed further and packaged. Confident prediction of these fermentation endpoints would enable better planning, scheduling, and therefore better vessel utilisation, thus increasing competitiveness and profitability.

The work presented here has resulted from a collaborative project with two UK industrial partners. The project involves the collection and analysis of real brewery data, and aims to (1) better characterise fermentations, by generating centile charts for PG versus time profiles, using the statistical distribution of real data (i.e. "empirical centiles" as opposed to a centiles based on theoretical distributions); (2) investigate the relationship between various process variables and fermentation time, with a view to predicting the latter.

2. Materials and methods

Fermentation data was obtained from three industrial beer qualities: two from one UK brewery ('Brew A' and 'Brew B'), and a third ('Brew C') from another UK brewery. This data consists of PG measurements over time (~2 measurements per 24 h) for each 'batch' (fermentation), along with other potentially relevant information, for example yeast generation and vessel size. Towards the end of fermentation, diacetyl readings were also taken

* Corresponding author. Fax: +44 (0)1603 507723.
E-mail address: Rob.Foxall@bbsrc.ac.uk (R.J. Foxall).

0260-8774/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2007.02.033

Table 1
Summary of fermentation data for three different brew qualities

Brew	Number of batches	Target PG (°S)	Desired end point time range (h)
A	129	14.5	65–90
B	97	14.5	57–83
C	50	15	105–145

(typically 1–4 per batch). This paper focuses on the problem of modelling PG as a function of time. There are two related aims: firstly, to predict the precise time at which "target" PG will be reached (i.e. the endpoint of fermentation), and secondly, to predict whether or not this time will fall within a given desired time range, thus classifying each batch as either 'standard' or 'non-standard'. Brew information is summarised in Table 1.

For these predictions to be of real practical value, they must be made within 60 h of the start of fermentation, and preferably within 24–48 h. In our work, we have used three different methods for generating these predictions. The first method uses smooth analytical functions to model the observed PG data as a function of time. We have concentrated on the incomplete beta-function (IBF), previously used by Trelea, Latrille, Landaud, and Corrieu (2001) in the context of modelling fermentation data. In the present implementation, the IBF relates PG to elapsed Time via 5 parameters, as follows:

$$\mathrm{PG} = \mathrm{PG}_{\max} - (\mathrm{PG}_{\max} - \mathrm{PG}_{\min}) \times \mathrm{IBF}(\alpha, \beta, \gamma \times \mathrm{Time})$$

The minimum and maximum PG parameters ($\mathrm{PG}_{\min}$ and $\mathrm{PG}_{\max}$) can be obtained directly from the data or from expert opinion, whilst the remaining 3 parameters (α, β, and γ) must be optimised. Alternatives to the IBF include the *logistic function*, and also *neural networks* – previously used for Van Breusegem, Thibault, and Cheruy (1991) and Vlassides, Ferrier, and Block (2001).

The second method employed was a variant on "nearest neighbour" (NN) approaches (Singhal & Seborg, 2002). Here, the object is to choose from a "history" of relevant batches the particular instance where the initial time versus PG profile most closely resembles the profile of the current batch. The argument is that if this historical batch is similar to the current batch over the first 48–60 h, then it is likely to stay similar over the remainder of the fermentation, and thus the time at which the historical batch reached the target PG is a valid prediction of when the current batch will do so.

In general, each batch had its PG measured at slightly different timepoints. *Interpolating splines* were therefore employed to create a time versus PG profile for each of the historical batches, so that all historical batches were sampled at the same defined timepoints and could be matched to the observation times of the current batch. Once the times are matched in this fashion, the similarity of the current batch to each of the historical batches is measured using *Euclidean distance*. If the current batch $\mathrm{PG}^{(c)}$ has n PG measurements taken at times $t_1, t_2, \ldots, t_n$, then the Euclidean distance between $\mathrm{PG}^{(c)}$ and the ith historical batch $\mathrm{PG}^{(i)}$ is given by

$$\mathrm{Distance}_i = \sqrt{\sum_{j=1}^{n} \left(PG_{tj}^{(c)} - PG_{tj}^{(i)}\right)^2}$$

The historical batch with the smallest distance from the current batch is then identified. An extension of this idea would be to match batches not only by initial time versus PG profiles, but also taking into account other variables that may be relevant, such as yeast generation or dissolved oxygen, although this was not investigated in the present study. Initial models including the temperature profile over time have been employed; thus far with mixed results – of the three brews, including temperature information improved prediction measures for one brew, made no difference to another brew and worsened matters for the third brew.

The third alternative method of detecting non-standard batches was to develop reference centile charts which depict normal behaviour of PG during the course of fermentation. Currently such graphs are used in breweries, but generally comprise two somewhat arbitrarily determined trend-lines. We are aiming to provide a similar tool, which will be something brewers are familiar with, but based upon the statistical distribution of real data. The method is based upon fitted analytical functions (such as the IBF), again interpolated to a suitable sampling rate (e.g. once per hour). Estimates of the median and desired percentiles at each time point are then obtained by bootstrapping. We initially used data from all available batches to generate the empirical centiles, but early testing showed that the most extreme batches should be removed from the dataset if the centiles were to flag up 'borderline' batches. These were identified by a suitable outlier detection method, for example, by examining the scores plot obtained by principal component analysis (applied to the interpolated curves), or the coefficients obtained by fitting an IBF to the data.

3. The "Fermentor software package

Measurements of PG are typically obtained infrequently, and moreover are measured with limited precision. Thus, it is arguably risky to blindly trust model predictions. Our intention is rather that visual (i.e. subjective) guides should be used in conjunction with numerical predictions, to make assessments of fermentation progress. Although initially the development of models was done using the free software package "R" (R Development Core Team, 2005), and the commercial package MATLAB (The Mathworks Inc.), eventually the methods must be accessible to an end-user in the brewery. For this purpose, a simple stand-alone software package has been developed using Visual Studio C++, and is currently undergoing testing at the participating breweries.

A simple sequence of events is required to use this software. After its invocation, the initial control window of the package appears as illustrated in Fig. 1.

The first step required, rather than entering in current batch data, is to select the beer quality, which contains information relevant to the Brew to be modelled, such as Target PG. This is kept in a simple text file, an example of which is shown below:

"Historical Data = C:\Beer_Data\Historical\BrewA_Hist.txt
Centiles = C:\Beer_Data\Centiles\BrewA_Cent.txt
Minimum PG = 11.5
Target PG = 14.5
Target Range = 65 to 90 hours"

The historical data and the centile data are also kept in text files, in a tabular format; an example of a centile file is shown in Table 2.

If there is no historical or centile information (for example during the first batches of a new beer quality), these values can be set to "none", and the software will not attempt to employ these methods. In this case, only the IBF approach may be used for making predictions.

The final step is to input the data observed thus far. This takes the form of the time and PG data, and can be directly typed in to the (initially empty) data screen shown in Fig. 1, or read from an existing file in the format given in Table 3.

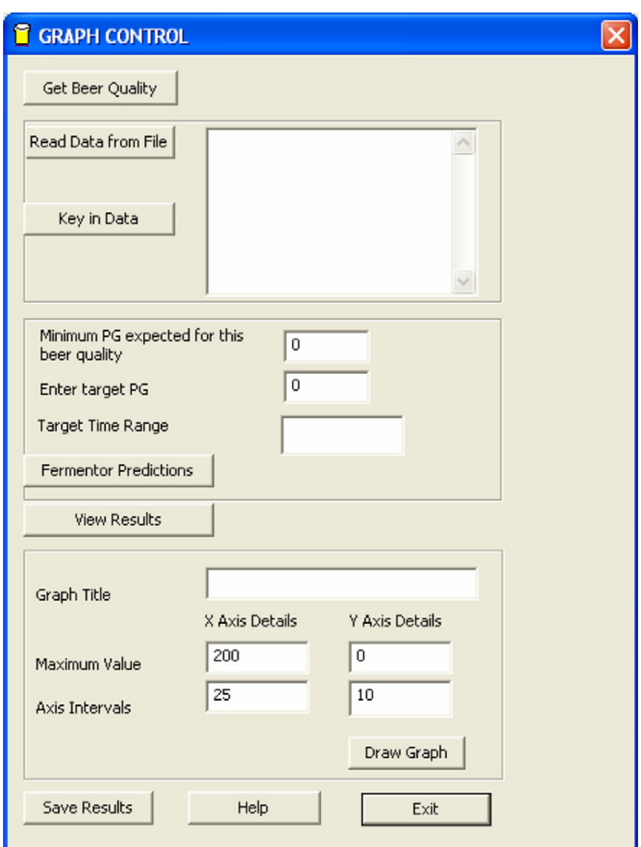

Fig. 1. Initial control window for Fermentor software.

Table 2
Example centile file

Time	1%	6%	22%	50%	78%	94%	99%
3	48.73	48.81	49.98	50.89	51.68	52.24	52.27
4	48.4	49.14	50.53	51.31	51.84	52.12	52.32
5	48.29	49.22	50.45	51.33	51.9	52.1	52.32
6	48.01	49.2	50.19	51.06	51.79	52.08	52.36
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3
An example Time versus PG data file

Time	PG
9.25	51.0
24.0	47.5
32.25	43.0
42.0	36.4
48.0	31.8

Clicking the button "Fermentor Predictions" fits the various models to the data (as appropriate), and outputs summary information pertaining to the predicted time to target PG, with an assessment of the likelihood that this is within the defined standard time range to target PG (see Fig. 2).

Finally, so that the user can get a visual guide to the fit of the data, the predicted profile is plotted on a graph, along with the observed data (Fig. 3).

4. Results

4.1. Incomplete beta-function (IBF)

From fitting the IBF to the initial time versus PG profiles for many batches, it became apparent that in many cases the fit was inappropriate for predicting the course of the fermentation. There were several possible reasons for this. In some instances there were only two observations made in the allotted initial time period (48 or 60 h, depending on Brew) – clearly not enough for fitting a non-linear function. Other cases had several observations, but these happened to lie essentially on a straight line which again does not provide sufficient information for model fitting – this usually leads to the fitted IBF curve either having a sharp corner, or a too flat profile. Finally, in some cases the observations demonstrated a rise in PG with time, which is biochemically impossible, and so strongly suggests that the particular data is unreliable. In Fig. 4 some examples of implausible model fits are shown. Note that these situations are all situations in which there is a "problem" with the initial data, and so having an implausible IBF fit, despite not providing a usable predicted time to reach target PG, is a useful warning of potential problems.

Table 4 displays summaries of how well the IBF predicts the time to target PG. *Root Mean Squared Error* (*RMSE*) is used to indicate the accuracy of predicting the actual time. The probabilities of correctly predicting both a standard and a non-standard batch are also given. Summary figures are based upon all models from all available data, and in

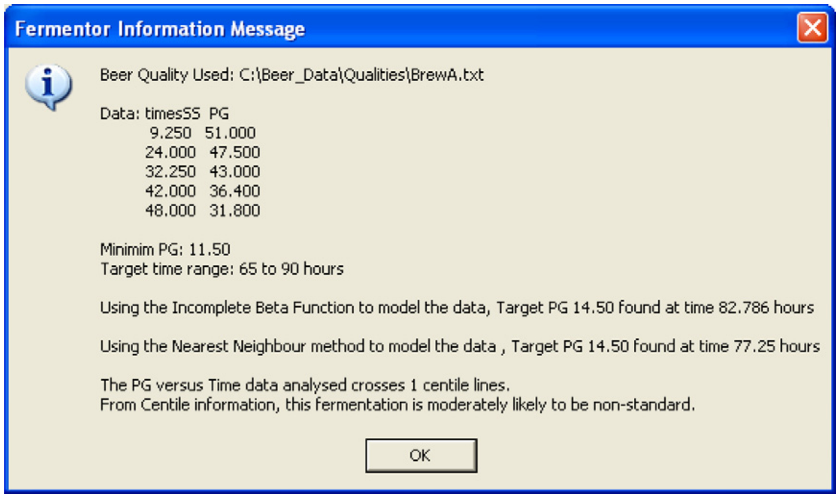

Fig. 2. Summary of model fitting by Fermentor software.

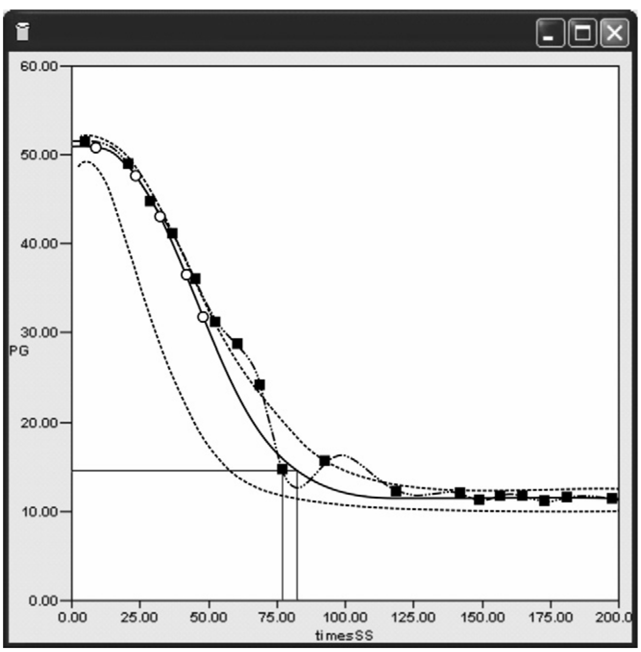

Fig. 3. Plot of model fitting by Fermentor software. Lines (---) are 6% and 94% centiles. Line (—) is fitted IBF curve, with points (○) the observed data before 48 h. Line (-··-) is spline fit to NN data, given by the points (■). Lines (—) denote target PG of 14.5, and predicted values from IBF and NN approaches.

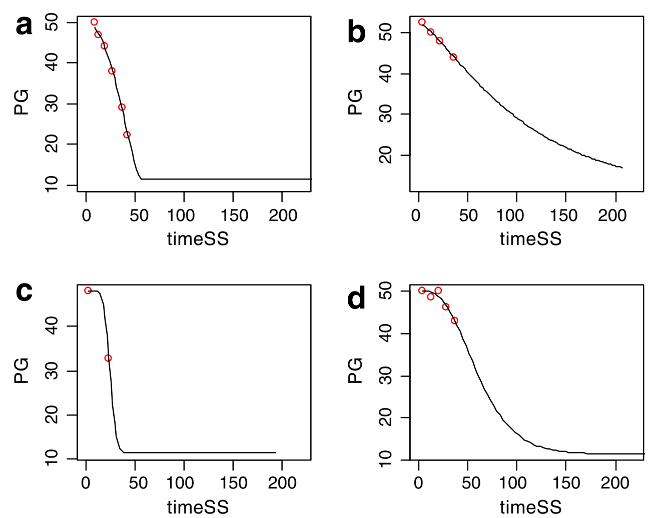

Fig. 4. Some real examples. (a) Points on straight line – sharp corner. (b) Points on straight line – too flat. (c) Only two points available. (d) PG increases between the second and third observations.

square brackets, from a selection of models only. In the latter case, the fitted curves were filtered by an initial assessment, made visually, of whether the curve (or the data) was acceptable, rejecting cases of the types presented in Fig. 4. For each of the three brew qualities, this assessment resulted in approximately half of the model fits being discarded.

It is clear that filtering by eye improves the performance as assessed by both the RMSE, and by the probability of correctly identifying standard batches. However, there is a slight decrease in the probability of correctly identifying non-standard batches. The explanation for this could be that batches giving unsatisfactory models tend also to be those that initially exhibit non-standard behaviour, and are thus more likely to reach Target PG outside the target time range, and thus become non-standard batches. Removing all such batches effectively removes a higher proportion of those which were, in fact, destined to be correctly classified as non-standard.

Also of note is the relatively poor fit to Brew C data compared to both Brews A and B. Brew C takes considerably longer for target PG to be reached, and the 60 h used as the initial measurement period is proportionally less than the 48 h used for Brews A and B. It seems likely that for this particular beer quality, 60 h of initial data is insufficient to adequately fit the IBF. Fig. 5 contrasts a typical (and good) model fit to Brew A data with a typical (and bad) model fit to Brew C data.

Table 4
Results for IBF method, all the model fits [only model fits deemed "plausible"], based on first 48 h (A,B), or 60 h (C)

Brew	RMSE (h)	Pr (correct 'standard')	Pr (correct 'non-standard')
A	31.7 [17.6]	47.4% [64.9%]	75.5% [67.9%]
B	27.9 [13.2]	42.9% [77.8%]	63.4% [48.0%]
C	63.9 [54.9]	15.6% [31.3%]	73.3% [62.5%]

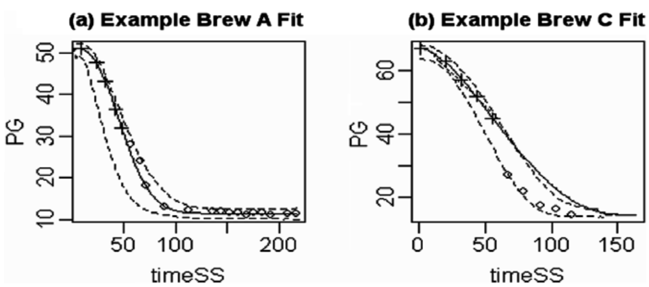

Fig. 5. Fitted IBF (—) and outer centiles (---) plotted with data. (a) IBF fits the data well. (b) IBF fails to fit well; the initial data lies close to straight line. Data used for training (+); subsequent data (○).

5. Empirical centiles

The centiles method does not give a predicted time to target PG; instead it indicates whether or not target PG is likely to be obtained during the target time range for the current batch. This prediction is based on where the initial observations lie with respect to the centiles generated for the appropriate beer quality. If they arise beyond around the lower (less than 6th) or upper (above the 94th) centile, the prediction is that the batch will exhibit non-standard behaviour and thus be a non-standard batch. Results for the centiles method are shown in Table 5.

The classification performance of the centiles method is generally good, and quite consistent across the three Brew qualities. The results for all three Brews tend to be superior to those from the IBF method; this is particularly true for Brew C.

6. Nearest neighbour (NN)

To implement the NN method, it is necessary to have available a set of "historical" data from the relevant brew quality, from which the historical batch most closely matching the current batch can be identified. For Brews A and B, the first 30 batches were designated the historical data (leaving 99 and 67 batches respectively for assessment); for Brew C, due to the smaller number of available batches, only the first 20 were chosen (leaving 30 batches for assessment).

The assessment of performance was done using two slightly different approaches. In the first, the historical data was defined as respectively the initial 30 or 20 batches only. In the second, once a test batch had been analysed, it was then used to augment the set of historical batches so that subsequent batches would have all the preceding batches to provide historical data. This was done in date order so that it would realistically reflect how the method would operate under actual brewing conditions. Results from both approaches are shown in Table 6.

The RMSE results are somewhat better than those of the IBF, especially for Brew C. The probability of predicting a standard batch is very high, better than for both the IBF and the centiles methods, although the difference between the NN with updated histories and centiles methods is smaller. Conversely, the probabilities of correctly predicting non-standard batches are relatively poor for the NN method, although for predicting non-standard batches, the performance is improved by having an updated history compared to the fixed history.

It is interesting to have a closer look at the poor performance at predicting non-standard batches. For example for Brew C, using a fixed history, none of the 11 non-standard batches were correctly classified. On examining the 20 historical batches, we find that none of them reached target PG *before* the target time range. So, for the subsequent 6 out of the 11 batches that were non-standard due to reaching target PG too soon, there was no candidate batch in the history to correctly predict this outcome. Furthermore, there were only four historical batches that reached target PG *after* the target time range. Clearly, for the NN method to work, it is necessary that the historical data set is large enough to contain representatives of all likely profiles. For Brew C, with only 50 available batches in total, it is difficult to accurately assess the performance of the NN method, even using the updated history approach. Only once enough historical data has been collected so that it is representative of all brew outcomes, will the NN method provide reliable estimates of brew quality.

In practice, when using the NN method as each brew is completed its data is added to the historical data set. Thus for any brew quality the historical data set will in time grow to be representative of all likely fermentation profiles. If such historical data was allowed to grow unchecked this may negatively impact on performance, as there will be increasing chance of brews dissimilar at

Table 5
Classification results for centiles method, based on first 48 h (A,B), or 60 h (C)

Brew	Pr (correct 'standard') (%)	Pr (correct 'non-standard') (%)
A	78.9	70.7
B	83.3	73.0
C	56.7	76.5

Table 6
Results for NN method, based on first 48 h (A,B), or 60 h (C), and fixed histories of size 30 batches (A,B) or 20 batches (C), [updated histories, initially 30 batches (A,B) or 20 batches (C)]

Brew	RMSE (h)	Pr (correct 'standard')	Pr (correct 'non-standard')
A	12.4 [12.4]	87.7% [75.4%]	38.1% [50.0%]
B	14.6 [11.9]	84.2% [78.9%]	44.8% [48.3%]
C	25.6 [30.3]	94.1% [76.5%]	0% [27.3%]

the end of fermentation having similar initial profiles. To prevent this the size of the historical data set has a fixed limit, currently set to be 120 batches. This is set up so that the most recent 120 batches are those kept – in this way, if the properties of a particular brew are in any way changing over time, the historical data set will evolve to reflect these changes.

7. Discussion

Of the three methods examined in this work, fitting the IBF to the initial data is perhaps the simplest way to produce predictions of target PG. This method does not rely on already having many batches from the same quality before it can be implemented, unlike using centiles or the NN methods. Performance of the IBF however, in terms of RMSE and the probability of correctly predicting standard batches, was relatively poor compared to the other methods. This assessment was improved by using simple visual examination to reject implausible model fits. However, this resulted in making no predictions at all for approximately 50% of the batches. Nevertheless, the probability of correctly predicting non-standard batches was nearly as good as the centiles and better than the NN method.

Of the two methods dependent upon already having a history of relevant batches, the centiles method was better at correctly identifying standard and non-standard batches; however it does not predict actual time to target PG. Whether this is a necessary requirement depends upon the needs of the brewery end-user. The NN method shows promise in comparison to the IBF in terms of RMSE. It remains to be seen whether the probability of predicting non-standard batches will continue to increase as more historical data becomes available.

It seems likely that using all three methods together (once sufficient histories exist) will offer benefits. Confidence in prediction increases if all three methods agree: for example, for Brew A, when all three methods predict a non-standard batch, this is correct 81.3% of the time – a better performance than by any of the methods individually. Finally, the brewery end-user needs to assess the costs associated with misclassifying standard batches as non-standard compared with non-standard batches as standard. This will assist in determining which combination of methods (any or all) will best suit the particular brewery requirements.

8. Conclusions

No fully-automated system to reliably predict the future course of a fermentation based on early readings currently exists. Rather, predictive models need to be used in conjunction with experienced opinion. Software to facilitate this has been developed, with an emphasis on visual as well as numerical guides. Of the numerical methods presented here, the performance of the centiles method is superior (with the lowest false positive rate), although the comparison is not entirely direct, as the same data was used to generate the centiles as to assess the method's performance. Future data (currently being collected) will provide a fairer comparison of the methods.

The empirical centiles method demonstrates that batches exhibiting non-typical behaviour at the beginning of a fermentation tend to stay non-typical throughout the whole fermentation. Improving the performance of the IBF method might be possible by increasing the sampling rate at the beginning of fermentation, but this would only be the case if the profile is already exhibiting the desired non-linear behaviour; it would be of limited help for Brew C, for instance, where the issue is that the behaviour is rather linear during the initial measurement period, and therefore difficult to model accurately.

Acknowledgements

This study is carried out in collaboration with Charles Wells Ltd. and Scottish Courage Brewing Ltd., and is supported by DEFRA LINK Project AFM183.

References

De Andres-Toro, B., Giron-Sierra, J. M., Lopez-Orosco, J. A., Fernandez-Conde, C., Peinado, J. M., & Garcia-Ochoa, F. (1998). A kinetic model for beer production under industrial operational conditions. *Mathematics and Computers in Simulation, 48*(1), 65–74.

Hough, J. S., Briggs, D. E., & Stevens, R. (1971). Brewery fermentations. In *Malting and brewing science* (pp. 505–536). London: Chapman and Hall Ltd.

R Development Core Team (2005). *R: Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rousu, J., Elomaa, T., & Aarts, R. (1999). Predicting the speed of beer fermentation in laboratory and industrial scale. *Engineering Applications of Bio-Inspired Artificial Neural Networks, 1607*, 893–901.

Siebert, K. J. (2001). Chemometrics in brewing – A review. *Journal of the American Society of Brewing Chemists, 59*(4), 147–156.

Singhal, A., & Seborg, D. E. (2002). Pattern matching in multivariate time series databases using a moving-window approach. *Industrial and Engineering Chemistry Research, 41*, 3822–3838.

Trelea, I. C., Latrille, E., Landaud, S., & Corrieu, G. (2001). Reliable estimation of the key variables and of their rates of change in the alcoholic fermentation. *Bioprocess and Biosystems Engineering, 24*, 227–237.

Trelea, I. C., Titica, M., Landaud, S., Latrille, E., Corrieu, G., & Cheruy, A. (2001). Predictive modelling of brewing fermentation: from knowledge-based to black-box models. *Mathematics and Computers in Simulation, 56*, 405–424.

Van Breusegem, V., Thibault, J., & Cheruy, A. (1991). Adaptative neural models for on-line prediction in fermentation. *The Canadian Journal of Chemical Engineering, 69*, 481–487.

Vlassides, S., Ferrier, J. G., & Block, D. E. (2001). Using historical data for bioprocess optimization: Modelling wine characteristics using artificial neural networks and archived process information. *Biotechnology and Bioengineering, 73*(1), 55–68.

Warnes, M. R., Glassey, J., Montague, G. A., & Kara, B. (1996). On data-based modelling techniques for fermentation processes. *Process Biochemistry, 31*, 147–155.