Journal of Food Engineering 153 (2015) 12–19

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

Predicting functional properties of milk powder based on manufacturing data in an industrial-scale powder plant

Ville Rimpiläinen [a,*], Jari P. Kaipio [a], Nick Depree [b], Brent R. Young [c], David I. Wilson [d]

[a] Department of Mathematics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[b] Light Metals Research Centre, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[c] Department of Chemical and Materials Engineering, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[d] Electrical and Electronic Engineering, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand

ARTICLE INFO

Article history:
Received 20 October 2014
Received in revised form 5 December 2014
Accepted 8 December 2014
Available online 16 December 2014

Keywords:
Dairy
Milk processing
Powders
Powder functional properties
Multivariate

ABSTRACT

The fundamental science relating key physical and functional properties of milk powder to plant operating conditions is complex and largely unknown. Consequently this paper takes a data-driven approach to relate the routinely measured plant conditions to one vital function property known as sediment in an industrial-scale powder plant. Data from four consecutive production seasons was examined, and linear regression models based on a chosen set of processing variables were used to predict the sediment values. The average prediction error was well within the range of the uncertainty of the laboratory test. The models could be used to predict the effect of each individual plant variable on the sediment values which could be beneficial in quality optimisation. In addition the choice of the training data set used to compute regression coefficients was studied and the resultant regression models were compared to alternative PLS models built on the same data.

1. Introduction

Milk powders are widely used in the food industry as in bakery products, soups and sauces, ready meals, milk based beverages, confectioneries, milk chocolates, yoghurts and cheeses (Oldfield and Singh, 2005; Sharma et al., 2012). A key motivation to transform liquid milk into powder is to increase shelf-life and reduce transportation costs. Milk powders also possess attractive physical and functional properties. The physical properties include powder structure, particle size distribution, flowability and bulk density while the functional properties describe how the powder behaves for the customer and include reconstitution properties, heat stability and foaming properties.

The functional properties depend on the raw milk composition, the degree of standardization, the processing and subsequent storage conditions, and how the powder is used in the particular food system (Oldfield and Singh, 2005). Since some of the functional properties can be built-into the powder, there are economical interests in manufacturing such tailor-made powders due to the added value (Sharma et al., 2012).

The functional properties are usually tested by sampling the final product some time after production. However this *a posteriori* testing strategy runs the risk that an out-of-specification campaign results in considerable material being downgraded or disposed leading to significant economical losses. Furthermore, because the science currently available to explain the relations between processing conditions and functional properties of powder is immature, simply knowing that a campaign is out of spec does not help to make changes in the production.

As a solution, a data-driven approach can be used to (1) establish relations between the real-time measurable processing conditions and offline tested functional properties, (2) to predict the functional properties (in some cases in real-time) based on plant data, and (3) estimate the variance caused by each processing condition. This is sometimes known as process analytical technology (PAT) or real-time quality control (RTQ) (Bakeev, 2005; FDA, 2004; Munir et al., 2014; Munson et al., 2006; Swarbrick, 2007; van den Berg et al., 2013).

The main objective of this study was to investigate strategies to be able to predict a key functional property using only operating data measured routinely from the plant thereby making the real-time quality control possible and avoiding the necessity of a time consuming, and somewhat subjective offline laboratory analysis. The functional test is one of the many sediment tests which quantify the volume of the undissolved milk (Anon, 2014). For instant whole milk powder, the less sediment, the better, although the upper acceptable limit depends on the specific product and the customer's requirements (Sharma et al., 2012).

* Corresponding author. Tel.: +64 9 923 8818; fax: +64 9 373 7457.
E-mail address: ville@math.auckland.ac.nz (V. Rimpiläinen).

http://dx.doi.org/10.1016/j.jfoodeng.2014.12.010
0260-8774/

Previously, it has been found that high evaporator preheat temperatures, long holding times and high values of total solids increase the sediment values (Oldfield et al., 2000). Moreover, it has been found that the milk homogenization settings have only marginal effect and that lecithin addition decreases sediments (Oldfield et al., 2000). Furthermore, there seems to be an optimum value for the concentrate temperature which, however, may be plant-dependent (Oldfield et al., 2000; Oldfield and Singh, 2005). Regarding the composition, milk with a low fat content (1.5–26%) is preferable than high (>26%) (Kelly, 1998; Oldfield and Singh, 2005) while varying amounts of protein (24.9–30.8%) have only a marginal influence on the sediment results (Kelly, 1998).

Generally speaking, the main contribution of this paper is to describe a data-driven approach that can predict end-point properties of industrial products based on real-time manufacturing data and estimate how much variation in the properties is caused by each plant variable. More specifically, this approach is here tested and evaluated by predicting sediment properties of milk powder with the help of real-time manufacturing data from an industrial-scale milk powder plant. The predictions were regressed using typical operating variables that are routinely logged, and approximating the joint distribution as Gaussian. Moreover, it is suggested how the nominal values of the operating variables could be adjusted in order to improve (lower) the sediment values.

2. Theory of conditional probability distributions

The aim of this work is to predict the scalar offline measured sediment values, s, given a vector of possibly correlated m routine plant observations, $\mathbf{D} \stackrel{\text{def}}{=} [d_1, d_2, \ldots, d_m]^{\mathrm{T}}$. In this paper, the joint distribution of the plant observations and laboratory measurements is approximated as Gaussian (normal). Thus, the model predictions are linear functions of the plant observations, and easily implemented in practice.

The mean and standard deviation of probability distribution of the sediments can be predicted if correlated information on processing variables is available. This information can be incorporated by conditioning the sediment distribution with the process data. Here, the Gaussian approximations and the theory how to calculate conditional probability distributions (CPDs) are briefly described.

The following notation is used in this section: vectors and matrices are denoted upright bold. The mean is denoted by a bar, $\bar{\mathbf{z}}$, and the model prediction by a hat, $\hat{s}$.

To start, the observation s and the vector of plant data $\mathbf{D}$ are concatenated to get the augmented vector

$$\mathbf{z} = \begin{bmatrix} s \\ \mathbf{D} \end{bmatrix}. \tag{1}$$

The joint-normal approximation of s and $\mathbf{D}$ has mean value $\bar{\mathbf{z}}$ and covariance matrix $\mathbf{P}$

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{s} \\ \bar{\mathbf{D}}, \end{bmatrix} \tag{2}$$

$$\mathbf{P} = \begin{bmatrix} P_s & \mathbf{P}_{sD} \\ \mathbf{P}_{Ds} & \mathbf{P}_D \end{bmatrix}. \tag{3}$$

The joint-probability distribution is of the form

$$\pi(s, \mathbf{D}) = \pi(\mathbf{z}) \propto \exp\left(-\frac{1}{2}(\mathbf{z} - \bar{\mathbf{z}})^{\mathrm{T}} \mathbf{P}^{-1} (\mathbf{z} - \bar{\mathbf{z}})\right), \tag{4}$$

and the inverse of the joint-covariance matrix can be partitioned as

$$\mathbf{P}^{-1} \stackrel{\text{def}}{=} \mathbf{B} = \begin{bmatrix} B_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}, \tag{5}$$

with dimensions $B_{11} \in \Re^{1\times 1}, \mathbf{B}_{12} = \mathbf{B}_{21}^{\mathrm{T}} \in \Re^{1\times m}$ and $\mathbf{B}_{22} \in \Re^{m\times m}$.

The joint-probability distribution can be written as

$$\pi(s, \mathbf{D}) = \pi(s|\mathbf{D})\pi(\mathbf{D}), \tag{6}$$

where $\pi(s|\mathbf{D})$ is the conditional probability distribution of s given $\mathbf{D}$ and $\pi(\mathbf{D})$ is the probability distribution of $\mathbf{D}$. The prediction of sediments is based on utilizing the above mentioned conditional probability distribution. It can be written in the form (Eaton, 2007),

$$\pi(s|\mathbf{D}) \propto \exp\left(-\frac{1}{2}(s - \bar{s}_{\cdot|\mathbf{D}})^{\mathrm{T}} \mathbf{P}_{\cdot|\mathbf{D}}^{-1} (s - \bar{s}_{\cdot|\mathbf{D}})\right), \tag{7}$$

where

$$\bar{s}_{\cdot|\mathbf{D}} = \bar{s} - B_{11}^{-1}\mathbf{B}_{12}(\mathbf{D} - \bar{\mathbf{D}}), \tag{8}$$

$$\mathbf{P}_{\cdot|\mathbf{D}} = B_{11}^{-1} = \frac{1}{B_{11}}. \tag{9}$$

This means that the most probable outcome of the sediment test when the processing data is known is given by (8) with variance (standard deviation squared) given by (9). The notation for the predicted sediment is $\hat{s} = \bar{s}_{\cdot|\mathbf{D}}$ and the corresponding standard deviation $\sigma = 1/\sqrt{B_{11}}$. For example, if the plant is operated with the nominal operating conditions $\bar{\mathbf{D}}$, it would yield sediment result $\hat{s} = \bar{s}$. Outside these conditions, the predicted sediment value is corrected by a value proportional to the difference $(\mathbf{D} - \bar{\mathbf{D}})$.

3. Materials and methods

3.1. Standard powder sediment tests

In this study, the stability of instant whole milk powder was the primary focus and it was quantified by a standard offline laboratory sediment measurement test. The testing procedure follows (Anon, 2014) which in turn is derived from Anon (1977). The exact nature of the test depends on the customer's requirements and for this specific industrial application are proprietary. However all the sediment tests follow the same basic procedure where a measured powder sample is mixed in controlled conditions in water (or other customer-specific solvents), and the resultant undissolved material is quantified. While the test is reasonably free from any operator subjectiveness, compared to other powder functional tests, the results tend to be severely quantified (see Fig. 5 for an example).

Due to confidentiality restrictions, the sediment values presented in this work have been normalised. Such normalisation does not have an effect on the method itself.

Sediment values from four consecutive production seasons consisting of 339, 300, 273 and 284 measured samples formed the basis for this work. Results that were further than 3.5 standard deviations away from the mean value were considered outliers and removed from the data set prior to processing.

Fig. 1 shows the histograms of the normalised sediment values for each production season and overlayed is the approximating Gaussian distributions. In addition, the average and standard deviation are noted for each season. It is immediately evident that seasons 1 and 2 are similar, as are seasons 3 and 4. The first two seasons have slightly lower sediment values than the latter, and the Gaussian approximation fit is better for seasons 3 and 4. This indicates that there has been operational changes between seasons 2 and 3.

3.2. Plant overview and manufacturing data

An overview of the plant layout and the different manufacturing stages are shown in Fig. 2. The raw milk from the farms is stored in mixing tanks (or silos) before passing through the preheating stage