Journal of Dairy Research

cambridge.org/dar

Measurement of water-holding capacity in fermented milk using near-infrared spectroscopy combined with chemometric methods

Pernille Uma Havmand[1], Lau Gustav Zachariassen[1], Richard Ipsen[1] and Vera Kuzina Poulsen[2]

[1]Department of Food Science, University of Copenhagen, Copenhagen, Denmark and [2]Microbial Screening, Discovery, Chr. Hansen A/S, Hoersholm, Denmark

Research Article

Cite this article: Havmand PU, Zachariassen LG, Ipsen R and Poulsen VK (2022). Measurement of water-holding capacity in fermented milk using near-infrared spectroscopy combined with chemometric methods. *Journal of Dairy Research* **89**, 194–200. https://doi.org/10.1017/S0022029922000279

Received: 31 May 2021
Revised: 12 January 2022
Accepted: 24 January 2022
First published online: 7 April 2022

Keywords:
Lactic acid bacteria (LAB); Near-infrared spectroscopy (NIR); Partial least-squares-discriminant analysis (PLS-DA); Water-holding capacity (WHC)

Author for correspondence:
Vera Kuzina Poulsen,
Email: verakuzina@gmail.com

Abstract

We investigated the use of near-infrared spectroscopy (NIR) for measuring water-holding capacity (WHC) in fermented milk. Increased WHC ensures improved texture and decreased syneresis in fermented dairy products and also improves cheese yield. NIR combined with partial least-squares-discriminant analysis (PLS-DA) was found to be a promising rapid and non-invasive method with no pretreatment of the samples for prediction of WHC in fermented milk samples. Analysis of the chemical bonds in the region 10 700–4500 cm^{-1} (935–2200 nm) of the electromagnetic spectrum was able to distinguish between samples with high vs. low WHC. This technique was successfully used to screen different strains of lactic acid bacteria for their ability to provide fermented milk with increased WHC, which is of great importance for use in various dairy products.

Milk fermentation by lactic acid bacteria (LAB) results in prolonged shelf-life due to acidification, as well as flavour and texture development. The latter is affected by milk composition, manufacturing process and added starter culture. Two common textural defects are syneresis and granular structure, occurring especially in low-fat products (Nguyen *et al*., 2018). LAB surface characteristics and interactions with food matrix components are known to affect the textural properties of fermented dairy products through cell chaining, clumping, formation of pili (Burgain *et al*., 2014; Tarazanova *et al*., 2018a, 2018b) or by production of exocellular polysaccharides (EPS). WHC refers to the ability to retain water. The use of EPS(+) starter cultures can help to increase the WHC of fermented milk, leading to a decreased level of whey separation (syneresis) in yoghurt and increased moisture content and yield in cheeses such as mozzarella (Low *et al*., 1998; Broadbent *et al*., 2001; Awad *et al*., 2005; Amatayakul *et al*., 2006; Dabour *et al*., 2006; Costa *et al*., 2010; Burgain *et al*., 2014).

Several methodologies have been implemented for measuring WHC in fermented milk, such as centrifugation, siphoning, drainage and nuclear magnetic resonance (NMR) (Hinrichs *et al*., 2004; Amatayakul *et al*., 2006; Lee and Lucey, 2010; Gilbert *et al*., 2020). The siphon method determines the level of spontaneous whey separated on the surface of gels such as set yogurts. Drainage measures the level of whey separated from cut gels under the influence of gravity; it is more relevant to cottage cheese than yogurt gels (Lucey and Singh, 1997). An NMR-based method (wash-out-test) has been applied to simulate syneresis in fresh cheese (Hinrichs *et al*., 2004); the method allowed quantification of the degree of immobilization of serum and could also determine whether it was retained in closed pores or located in open capillaries. Water in capillaries can be washed out, while water in pores cannot. The fraction of the serum bound or retained in closed pores in a gel structure was shown not to contribute to syneresis, while the part present in capillaries open to the surface of the sample contributed to syneresis (Hinrichs *et al*., 2004). Low-frequency nuclear magnetic resonance (^{1}H-LF-NMR) is a non-destructive technique, which has been reported to make it possible to assess and describe water retention by yogurt and spontaneous syneresis when associated with image analysis of micrographs (Gilbert *et al*., 2020). Mid-infrared (MIR) spectroscopy is widely used for rapid, real-time, and nondestructive evaluation of milk composition (fat, protein, lactose) and has also shown potential to estimate curd yield in cheese production (El Jabri *et al*., 2019). Despite its obvious shortcomings, centrifugation to estimate the water entrapped into the matrix appears to be the most commonly used method for quantifying WHC, even though centrifugation time and the force applied varies markedly amongst researchers and strongly affects the outcome. The centrifugation method measures the level of whey separated from the collapsed gel as a result of an applied high external force (centrifugal force), i.e. resistance of the gel to compaction. The amount of whey collected as a result of

Hannah Dairy Research Foundation

CAMBRIDGE UNIVERSITY PRESS

centrifugation can be influenced by other factors such as the rigidity and rheological properties of gels (Lucey and Singh, 1997).

The various methods used to determine WHC in dairy products have not been standardized, which complicates data comparison between studies. Furthermore, these methodologies are time consuming and often destructive. This increases the incentive to develop a standardized rapid method for measuring WHC. In recent years, NIR has been used as a rapid and robust method with little or no pretreatment of samples for analysing chemical bonds in the region 12 820–3959 cm^{-1} (780–2526 nm) of the electromagnetic spectrum (Franzoi *et al*., 2021). The combination bands and overtones of water are found at approx. 10 350 cm^{-1} (965 nm) (stretching, 2nd overtone), 6980 cm^{-1} (1450 nm) (stretching, 1st overtone) and 5180 cm^{-1} (1930 nm) (combination of stretching and bending) (Metrohm, 2013). NIR has been successfully tested for prediction of WHC in meat (Brøndum *et al*., 2000; Bowker *et al*., 2014) and fermented goat milk (Zhu and Ding, 2010), but has not yet been applied for cow milk, as in the present study. The reflected light is dependent on the wavelength of the light and the chemical bonds of the sample. The light exposed to the samples holds an energy dependent on the frequency (wavelength) of the light. A chemical bond between two atoms holds an energy, which is dependent on the length of the bond. The vibration of such a bond can be described by the harmonic model. If the energy of the near-infrared light exposed to a sample matches the energy of a chemical bond between two atoms, the energy is absorbed and transits to non-contiguous energy states. The vibration can no longer be described by the harmonic model, but rather the anharmonic model, where the bond can dissociate. A hydrogen bond will have a high dipole moment causing little energy required for the bond to dissociate and bonds to hydrogen atoms are therefore highly anharmonic. The transition to higher energy levels, described by the anharmonic model, can be detected in the NIR region as overtones and combination bands of stretching and bending of the chemical bond. Therefore, when the O–H bonds of water molecules are bound to protein or EPS (increasing the WHC of the sample), the vibration becomes more harmonic, and cannot be detected by NIR (Miller, 2001). Samples with bound water molecules resulting in high WHC will hence have lower absorption in the bands related to O–H and using chemometric methods, the shift in bound and free water can be detected.

Based on the above, we hypothesized that it would be possible to detect the difference between samples with high and low WHC using NIR by extracting the information in the spectra with chemometric methods. The aim of the study was to build a prediction model that can be used as a rapid method for screening fermented milk samples for high WHC and determine the feasibility of this.

Materials and methods

Strains and media

Eleven *Streptococcus thermophilus* strains originating from the Chr. Hansen culture collection were used in this work: A, B, C, D, E, F, G, H, I, J, and K. Strain A, B, D, F, H, and J were considered to be high-WHC strains, as they resulted in 3.3 to 5.0% increased yield of mozzarella cheese when tested in 150-l scale in the Application Test Center of Chr. Hansen in Hoersholm, compared to the strains C, E, G and K, which were considered as low WHC strains. Cheese yield data obtained in 150 l scale are not shown, as we do not have data from one single experiment with all the 11 strains in the same conditions. Because of the large milk volumes used for each experiment and limitations in the amount of sample that can run at once, many experiments spread in time were performed instead. The experiments were performed with two to three strains at a time, using different milk batches and sometimes also slightly different cheese making parameters. Strain I was not tested in cheese production in 150-l scale, as this strain was texturing.

The sample preparation for the applied centrifugation method and NIR method described below is shown in Figure 1. Cultures incubated overnight in M17 broth (Terzaghi and Sandine, 1975) containing 2% lactose were used for milk acidification (1% inoculum) in severely heat-treated skim milk (SHT-SM) and pasteurized low-fat milk (past-LFM), respectively. For the samples applied for NIR, pH indicator was added as described in Poulsen *et al*. (2019). The pH indicator does not contain molecules that can interfere in NIR region, and the concentration is negligible in terms of affecting the dry matter content of the samples. SHT-SM was prepared by reconstituting skim milk powder containing 38% protein, 53% lactose, <1.25 fat, and 3.9% moisture (Arla Foods amba, Denmark) to a level of dry matter of 9.5% and pasteurized at 99°C for 30 min, followed by cooling to 40°C. The high temperature treatment ensures that the fermentation of this type of milk is not affected by the indigenous enzymes or bacteria. Past-LFM from Arla containing a fat content of 1.5% and a dry matter content of 12.8%, was pasteurized at 90°C for 20 min, cooled down to 40°C, and enriched with 0.003% Na^+ formate (HCOONa) to stimulate the growth of *S. thermophilus* (Derzelle *et al*., 2005). We chose to apply both milk types in order to investigate the feasibility of the method using two types of cow milk with varying dry matter content.

Centrifugation for WHC measurement

Empty sterile plastic bottles were weighed, and 200 ml past-LFM containing 1% overnight culture and 0.003% Na^+ formate was added. The samples were incubated at 40°C until a pH of 4.55 was achieved and subsequently cooled down in ice water for 30 min and then kept overnight at 4°C. The samples were centrifuged at $2325 \times \boldsymbol{g}$ for 3 min, and the whey was carefully poured out of the plastic bottles. The samples were weighed before and after whey removal. The WHC was expressed as curd yield. Four replicates of each sample were prepared; one was used to monitor milk acidification using pH electrodes, and three remaining replicates were used for WHC analysis.

NIR measurement

A volume of 40 ml SHT-SM or past-LFM containing 1% inoculum and pH indicator was transferred to glass Petri dishes, 100 mm in diameter, and incubated at 37°C overnight anaerobically. The samples were stored at 4°C for two days, before they were measured using NIR. A FT-NIR Spectrometer (MB-160, ABB Bomem, Quebec, Canada) with 64 scans per measurement was used to record reflectance in NIR region 12 000–4000 cm^{-1} (833–2500 nm) and a resolution of 16 cm^{-1}. Purging with dry O_2 was applied. The samples were scanned through the Petri dish. A spectrum was obtained of an empty Petri dish and used to account for the glass Petri dish in the calculation of the absorbance of the samples. The samples were measured in random order. Each sample was measured three times by moving the Petri dish containing the sample after each

https://doi.org/10.1017/S0022029922000279 Published online by Cambridge University Press

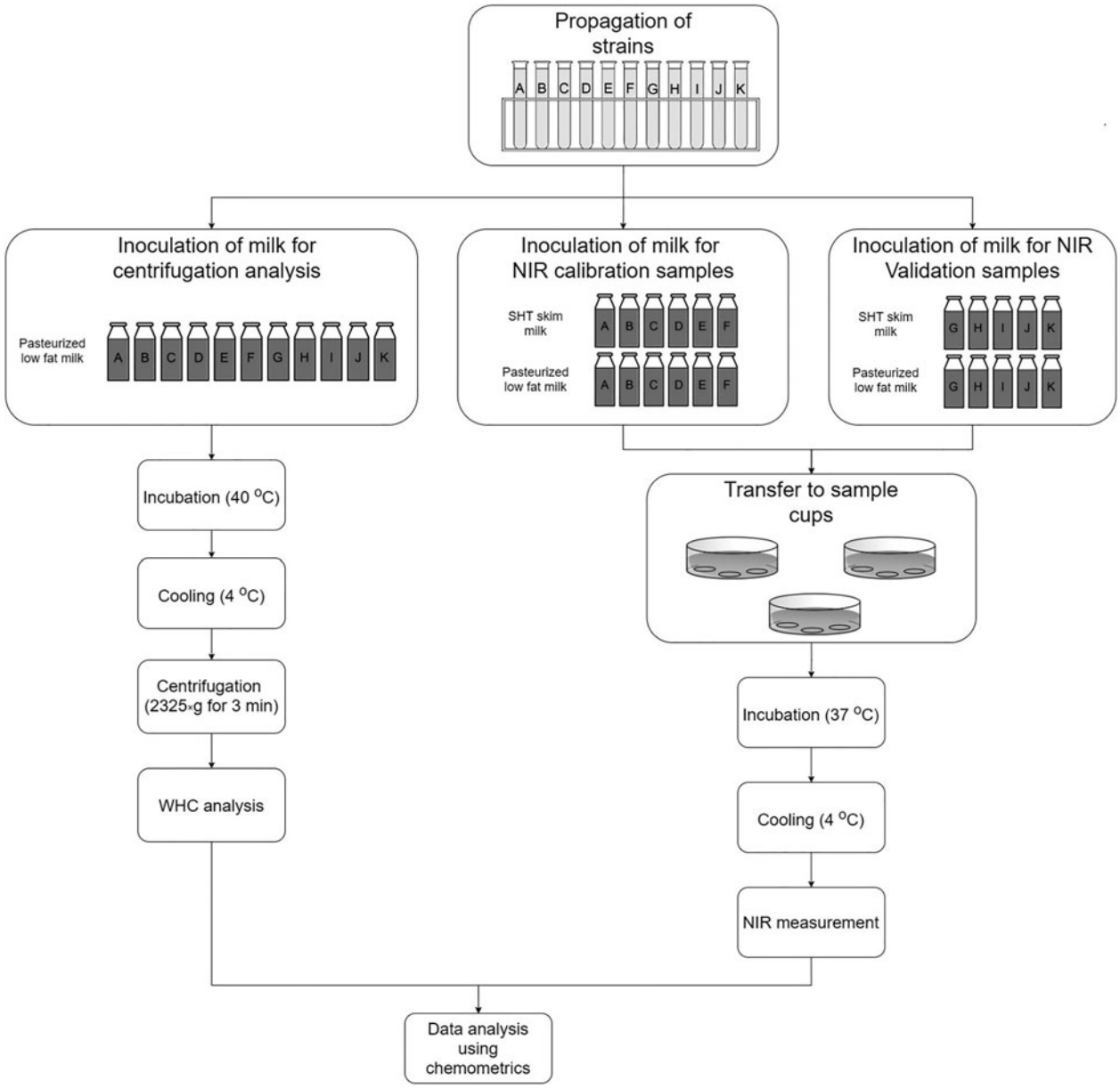

Fig. 1. Overview of sample preparation for centrifugation and NIR methods and subsequent analysis. Overnight inoculum in M17 broth added lactose was used to acidify past-LFM added Na^+ formate to pH 4.55 in plastic bottles. The samples were cooled to 4°C, centrifuged and WHC was calculated as the ratio of curd. The WHC was used to classify the strains as high and low WHC. The calibration and validation samples for NIR method were prepared by adding overnight inoculum in M17 broth and pH colour indicator to 200 ml of SHT-SM and past-LFM, respectively. The 200 mL of inoculated milk was poured into three Petri dishes and incubated overnight at 37°C and cooled down to 4°C. The Petri dishes were measured three times, as indicated with circles in the figure, with NIR. Classification of high and low WHC strains obtained by the centrifugation method and the NIR spectra were analysed using PCA and PLS-DA.

measurement, so that the measurement was performed in a new spot of the Petri dish each time to ensure suitable sampling (Fig. 1). Three independent experiments were conducted with fermented SHT-SM and past-LFM using strains A–F, measured in triplicates per strain, which were used as calibration samples. One experiment was conducted using strains G–K, measured in triplicates per strain, which were used as validation samples.

The raw NIR spectra exhibited scattering, which was accounted for by pre-processing. It was chosen to apply Standard Normal Variate (SNV) followed by Mean-Center (MC). However, other standard pre-processing methods were equally good at discerning between the high and low WHC strains.

Data analysis

A Student's *t*-test was applied for testing significance between the high and low WHC of the strains measured with centrifugation (significance level of $P < 0.05$). The spectra obtained by NIR were analysed with principal component analysis (PCA) and PLS-DA using PLS Toolbox 8.6, a toolbox for MATLAB v. R2017b (Eigenvector, 2018). PCA is tool for explorative analysis of multivariate datasets (Bro and Smilde, 2014). PLS-DA is a

https://doi.org/10.1017/S0022029922000279 Published online by Cambridge University Press

supervised classification method that uses PLS for prediction of a dummy variable. The dummy variable is a vector consisting of 1 and 0 indicating the sample is in or out of class, regarding high WHC (Lee *et al.*, 2018). A variable selection was conducted, and the range 10 700–4500 cm^{-1} (935–2200 nm) of the spectra were included in the PCA and PLS-DA models. The number of latent variables (LV) were chosen individually for each model. The PLS-DA models were cross-validated using leave-one-strain-out cross-validation (CV). The prediction of the samples was based on a discrimination line that divides the samples as belonging to either Class 1 (high WHC strain) or Class 0 (low WHC strain). Since we only had two classes, we assumed that if the prediction of samples does not belong to Class 1, it must belong to Class 0. The discrimination line was based on maximizing the sensitivity and minimizing the specificity, where the sensitivity is the number of true positive samples (should be near 1), and the specificity is the number of false negatives (should be near 0) (Lee *et al.*, 2018). The PLS-DA models were evaluated by Matthew's correlation coefficient (MCC), which is based on the true and false positive rates. The MCC can range from −1 to +1. If MCC is 1, it represents a perfect correlation between predictions and references, MCC = 0 represent an incidental correlation and MCC = −1 indicates no relation between the predictions and references (Chicco, 2017).

Results and discussion

WHC measured by centrifugation

Strain A, B, D, F, H, I, J and K have higher WHC (curd yield above 54%) compared to the other strains and are classified as high WHC strains, and strain C, E and G are classified as low WHC strains (Fig. 2). There is a statistical difference between high and low WHC strains (P-value = 0.023), when applying this separation. Milk acidified using strain I had the highest WHC compared to the other strains (Fig. 2). Strain D had the second highest WHC. The scoring of the strains and the separation into high-WHC *vs.* low-WHC groups was similar for the 150 L and in 200 mL scale (results not shown). Strains considered as high WHC strains based on the cheese yield data obtained in the 150 L experiment, had higher WHC measured by the centrifugation analysis, compared to the strains considered as low WHC except for strain K, which was found to have low WHC in large scale production (Fig. 2). This indicates that measuring WHC is not an in- or out-class case, however due to the high error on this method, it was not possible to obtain accurate results using PLS. Therefore, we chose to divide the samples into high and low WHC samples and differentiate them using PLS-DA.

WHC measured by NIR

Six *S. thermophilus* strains, A, B, C, D, E and F, which were chosen based on similar milk acidification speed and differences in WHC, were used to explore NIR to distinguish between the low *vs.* high WHC samples. Five *S. thermophilus* strains with similar acidification speed but differences in WHC (G, H, I, J, and K) were subsequently used for the validation of NIR analysis. The pre-processed spectra of strains A–K contained two large bands at 6980 and 5180 cm^{-1}, that were assigned to the 1st overtone and the combination band of water, respectively (Fig. 3). It is also seen that the pre-processed spectra of high and low WHC strains can be differentiated using these bands. These bands have also been previously found optimal for prediction of water content of yogurt samples dissolved in acetonitrile (Adam *et al.*, 2009).

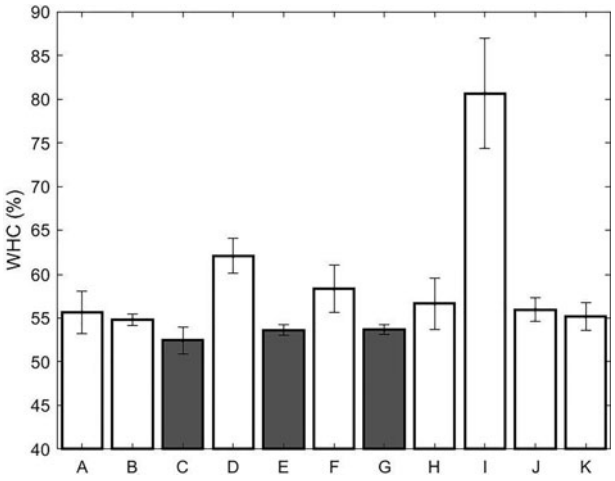

Fig. 2. Water-holding capacity (WHC) measured by centrifugation and expressed as the curd yield. Data are shown as mean and standard deviation of three replicates. The *Streptococcus thermophilus* strains with open bars are high WHC strains and the strains with closed bars are low WHC strains. With this separation there is a significant difference ($P < 0.05$) between the two groups.

An outlier detection was performed based on inspection of pre-processed spectra, Q residuals vs. Hotelling T^2 plot and the PCA score plot. The few outliers found were only removed if they were traced back to a human error. A clear distinction between high and low WHC strains on PC 1 and 2 of the score plot was found for SHT-SM and past-LFM samples (Fig. 4). No pattern differentiating samples from the three individual experiments with strain A–F (calibration samples) was found. Furthermore, the replicates of each sample were found to be precise (results not shown). Hence, the analysis was evaluated to be reproducible.

The separation of high and low WHC strains was highly influenced by the two bands assigned to water at 6980 and 5180 cm^{-1} on PC 1 and 2. The loadings of PC 2 at 6980 and 5180 cm^{-1} were inversely correlated (Fig. 5). This inverse correlation of the spectra of samples with high and low WHC was also seen in Figure 3. High absorbance at 6980 cm^{-1} was associated with high WHC strains and the spectra of samples with high absorbance at 5180 cm^{-1} was associated with low WHC strains. This was consistent with a previous study on the bands influencing WHC in chicken breasts (Bowker *et al.*, 2014). This can be explained with the theory of anharmonicity. NIR measures highly anharmonic bonds, such as O–H, N–H, C–H, etc. When a hydrogen atom of a water molecule forms a bond with a neighbouring molecule, e.g. protein or EPS, the vibration of O–H becomes more harmonic and can therefore not be detected by NIR and hence the absorbance decreases (Miller, 2001).

Validation of NIR for WHC measurement

As NIR was found to separate high and low WHC strains using PCA, we tested whether it was possible to quantitively predict the WHC. The calibration set was constructed of strains A, B, C, D, E, and F and the validation set was constructed of strains G, H, I, J, and K. SHT-SM and past-LFM samples were analysed with PLS-DA separately due to focusing on the WHC of the strains and not the difference in dry matter content of the milk types. The data were examined by PCA models for detection of

https://doi.org/10.1017/S0022029922000279 Published online by Cambridge University Press

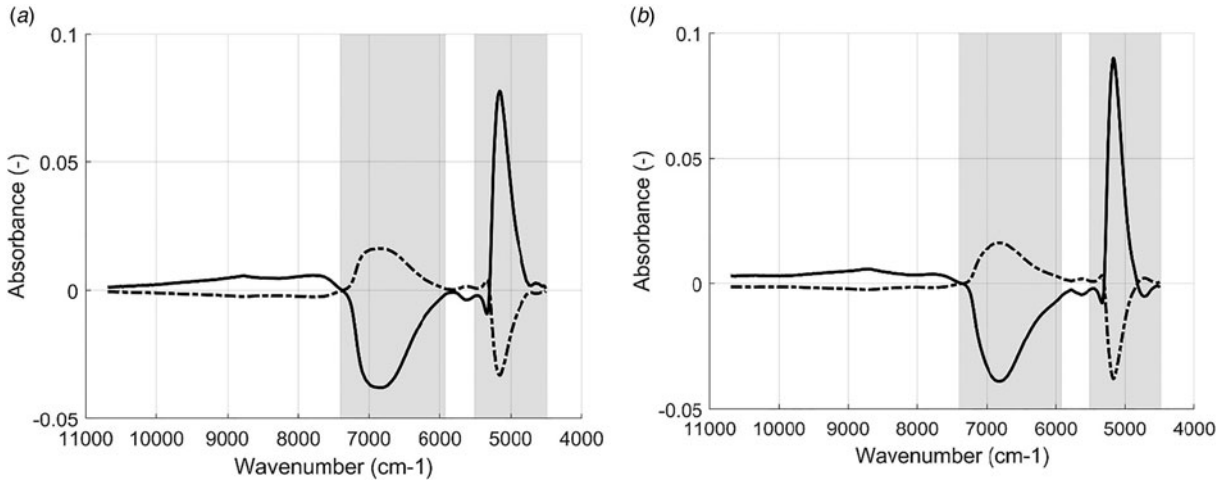

Fig. 3. The mean NIR spectra of (a) SHT-SM and (b) Past-LFM samples with high water-holding capacity fermented with *Streptococcus thermophilus* strains A, B, D, F, G, H, I, J and K (dashed line) and low water-holding capacity fermented with strains C, E and G (solid line). NIR spectra were obtained in the NIR region of 10 700-4500 cm^{-1} and pre-processed with standard normal variate and mean centring.

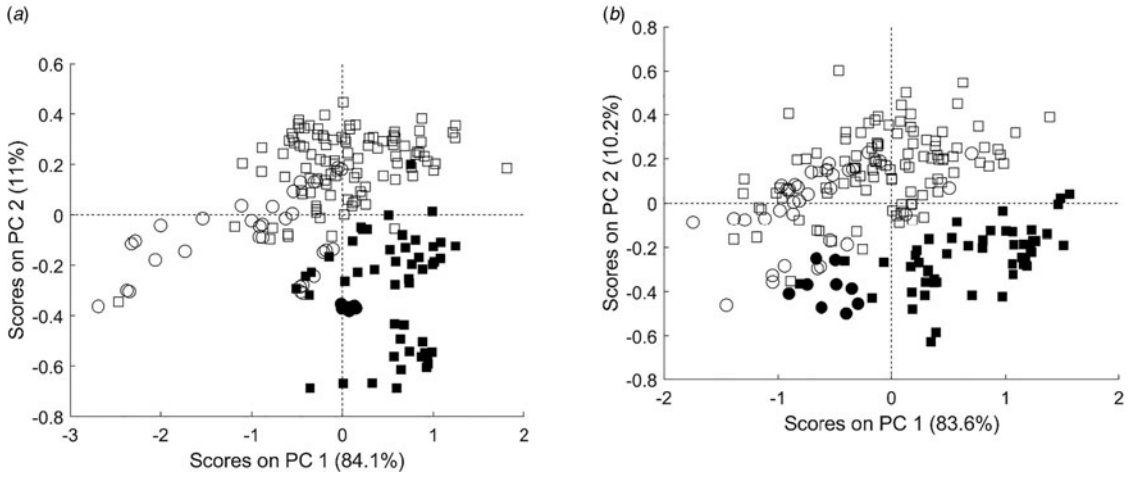

Fig. 4. PCA score plot of samples containing fermented (a) SHT-SM and (b) past-LFM samples inoculated with 11 different *Streptococcus thermophilus* strains (A–K). The samples were marked according to expected water holding capacity (WHC) and whether the sample is a calibration or validation sample; the open square represents calibration samples with high WHC, closed square represents calibration samples with low WHC, open circle represents validation samples with high WHC and closed circle represents validation samples with low WHC.

outliers as described in the previous section. The data were pre-processed with SNV and MC. Two PLS-DA models, one for each type of milk, were computed. The number of LVs was two for both the SHT-SM model and past-LFM model.

The PLS-DA models were able to predict the strains fermented in SHT-SM or past-LFM that exhibited high WHC (Fig. 6). The discrimination line shown in Figure 6 divided the samples classified as belonging to either Class 0 (low WHC strains) or Class 1 (high WHC strain). The same classification of strains into low *vs.* high WHC was obtained for SHT-SM (Fig. 6a) and past-LFM samples (Fig. 6b). The MCC of SHT-SM model were 0.941, 0.896 and 0.341 for the calibration model, the CV and the prediction, respectively, whereas the MCC of past-LFM samples were 1, 0.881 and 0.858 for the calibration model, the CV, and the prediction, respectively. The CV was chosen to be leave-one-strain-out and the remaining strains was used to predict the strain that was left out. As the MCC of calibration and CV for both SHT-SM and past-LFM were close to 1, the models showed ability to predict both the strains included in the calibration set and new strains. This was confirmed by MCC of prediction, which was lower but indicated that the calibration set was largely able to predict new strains. SHT-SM fermented with strain K was misclassified, which was also the reason for lower MCC of validation in SHT-SM compared to past-LFM. The ability to predict whether a given strain has a high or low WHC was affected by the band at 5180 cm^{-1} and 6980 cm^{-1} (Fig. 7). This was consistent with the cause of differentiation of high and low WHC strain explored by PCA.

In this study, the WHC measurement proved feasible for the two types of fermented cow milk with different dry matter content, SHT-SM and past-LFM, using 11 different strains. The study is a proof of concept and in order to implement the method, it would be beneficial to include more samples to the prediction model. It is expected that the method is feasible for other types of milk and strains if the variation of milk and

https://doi.org/10.1017/S0022029922000279 Published online by Cambridge University Press

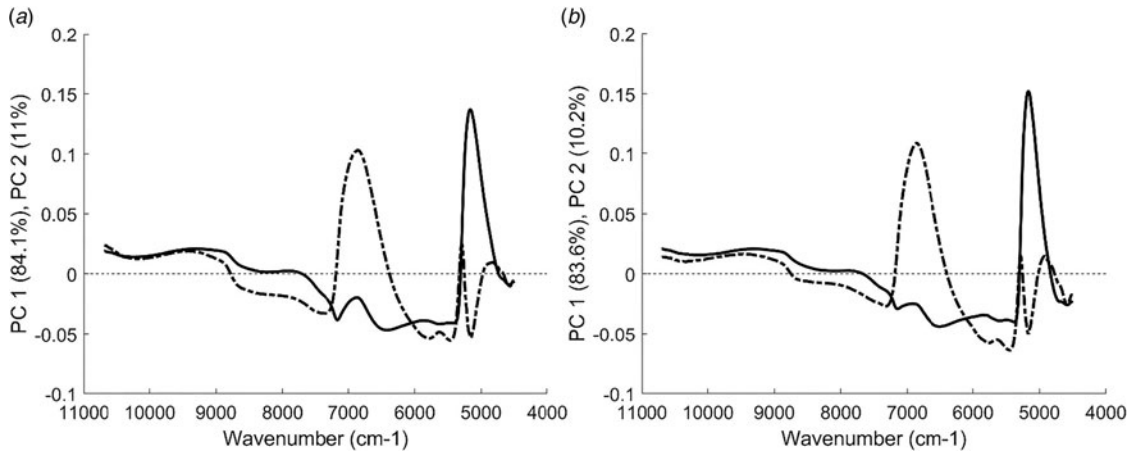

Fig. 5. The loading plot of PC 1 (solid line) and PC 2 (dashed line) of samples containing (a) SHT-SM and (b) past-LFM samples fermented with 11 different *Streptococcus thermophilus* strains (A–K).

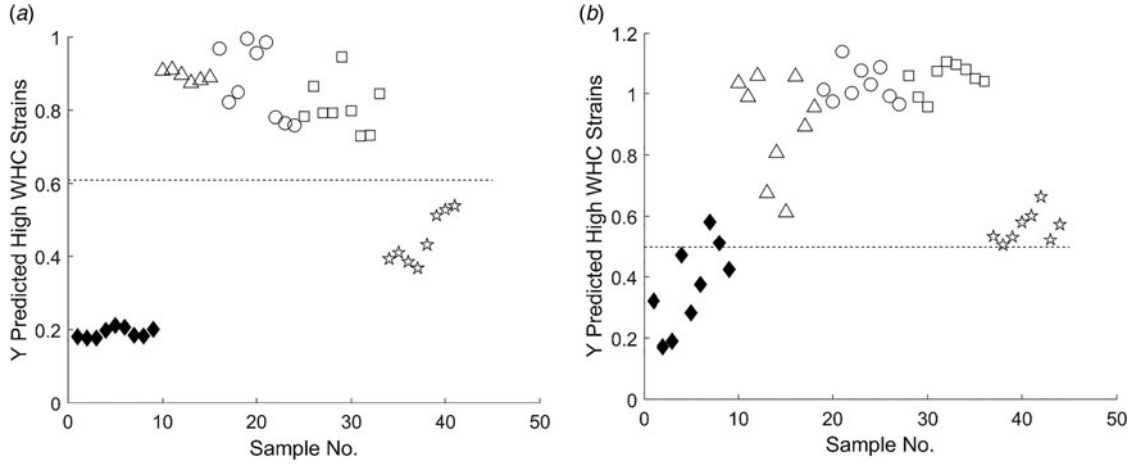

Fig. 6. PLS-DA for prediction of water-holding capacity (WHC) of (a) SHT-SM or (b) Past-LFM samples using *Streptococcus thermophilus* strains G–K. The dashed line is the discrimination line that separates the high WHC strains above the line (open symbols) and the low WHC strains below the line (closed symbol). The strains are shown on the plots as, G = closed diamond, H = open triangle, I = open circle, J = open square, and K = open star.

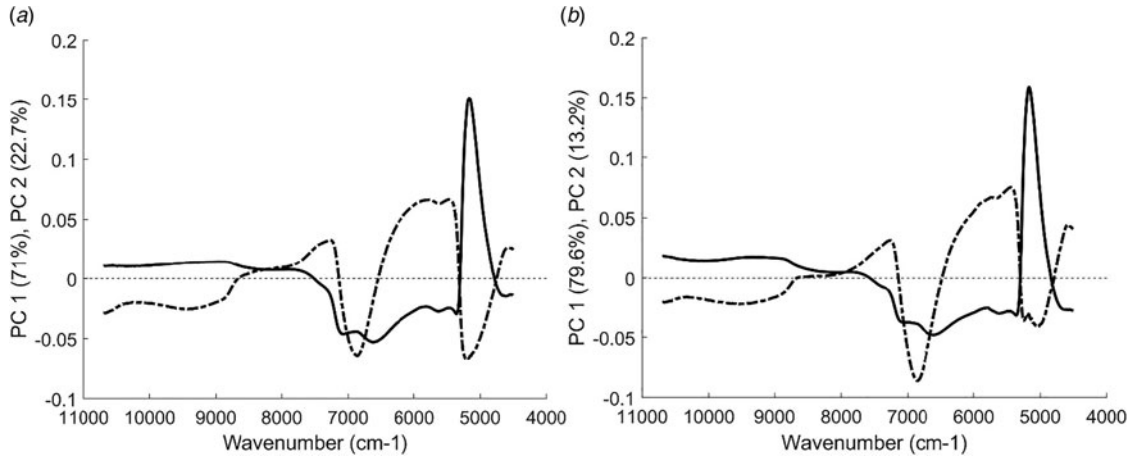

Fig. 7. The loading plot of LV 1 (solid) and LV 2 (dashed line) of samples containing fermented (a) SHT-SM or (b) Past-LFM. The dashed line indicates zero on the *y*-axis.

https://doi.org/10.1017/S0022029922000279 Published online by Cambridge University Press

similar strains are included as calibration samples and the classification of high and low WHC strains are based on the same method as applied in this study.

In conclusion, we were aware that a rapid non-destructive analysis would be beneficial for easy screening of high *vs.* low WHC strains for development of starter cultures in the dairy industry. NIR was shown to be a promising rapid method that can be used to screen for strains resulting in fermented milk with high WHC or low WHC. Similar patterns were obtained when classifying fermented milk samples into high- *vs.* low-WHC using centrifugation data obtained in 200 ml scale and NIR data obtained in 40 ml scale. Eleven *S. thermophilus* strains were used for successful prediction of WHC using PLS-DA, using two different milk types. Further data collection and calibration work is needed to implement the analysis to cover the variation of milk and bacterial strains.

Acknowledgments. We thank Kasper Borg Damkjær for guidance on using NIR as a method for WHC measurement and discussions on data analysis, Nanna Viereck and Søren Balling Engelsen for discussions on the theory of NIR and Elahe Ghanei Moghadam for help with performing the experiments.

References

Adam M, Dobiáš P, Bajerová P and Ventura K (2009) Comparison of various methods for determination of water in white yoghurts. *Food Chemistry* **115**, 1069–1073.

Amatayakul T, Sherkat F and Shah NP (2006) Syneresis in set yoghurt as affected by EPS starter cultures and level of solids. *International Journal of Dairy Science* **59**, 216–221.

Awad S, Hassan AN and Muthukumarappan K (2005) Application of exopolysaccharide-producing cultures in reduced-fat cheddar cheese: texture and melting properties. *Journal of Dairy Science* **88**, 4204–4213.

Bowker B, Hawkins S and Zhuang H (2014) Measurement of water-holding capacity in raw and freeze-dried broiler breast meat with visible and near-infrared spectroscopy. *Poultry Science* **93**, 1834–1841.

Bro R and Smilde AK (2014) Principal component analysis. *Analytical Methods* **6**, 2812–2831.

Broadbent JR, McMahon DJ, Oberg CJ and Welker DL (2001) Use of exopolysaccharide-producing cultures to improve the functionality of low fat cheese. *International Dairy Journal* **11**, 433–439.

Brøndum J, Munck L, Henckel P, Karlsson A, Tornberg E and Engelsen SB (2000) Prediction of water-holding capacity and composition of porcine meat by comparative spectroscopy. *Meat Science* **55**, 177–185.

Burgain J, Scher J, Francius G, Borges F, Corgneau M, Revol-Junelles AM, Cailliez-Grimal C and Gaiani C (2014) Lactic acid bacteria in dairy food: surface characterization and interactions with food matrix components. *Advances in Colloid and Interface Science* **213**, 21–35.

Chicco D (2017) Ten quick tips for machine learning in computational biology. *BioData mining* **10**, 35.

Costa NE, Hannon JA, Guinee TP, Auty MA, McSweeney PL and Beresford TP (2010) Effect of exopolysaccharide produced by isogenic strains of *Lactococcus lactis* on half-fat cheddar cheese. *Journal of Dairy Science* **93**, 3469–3486.

Dabour N, Kheadr E, Benhamou N, Fliss I and Lapointe G (2006) Improvement of texture of reduced-fat cheddar cheese by exopolysaccharide-producing lactococci. *Journal of Dairy Science* **89**, 95–110.

Derzelle S, Bolotin A, Mistou M-Y and Rul F (2005) Proteome analysis of *Streptococcus thermophilus* grown in milk reveals pyruvate formate-lyase as the major upregulated protein. *Applied and Environmental Microbiology* **71**, 8597–8605.

Eigenvector (2018) PLS Toolbox 8.6. Manson, WA. Retrieved 2019-03-04 from http://www.eigenvector.com.

El Jabri M, Sanchez MP, Trossat P, Laithier C, Wolf V, Grosperrin P, Beuvier E, Rolet-Répécaud O, Gavoye S, Gaüzère Y, Belysheva O, Notz E, Boichard D and Delacroix-Buchet A (2019) Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. *Journal of Dairy Science* **102**, 6943–6958.

Franzoi M, Ghetti M, Di Monte L and De Marchi M (2021) Investigation of weight loss in mozzarella cheese using NIR predicted chemical composition and multivariate analysis. *Journal of Food Composition and Analysis* **102**, 104002.

Gilbert A, Rioux LE, St-Gelais D and Turgeon SL (2020) Characterization of syneresis phenomena in stirred acid milk gel using low frequency nuclear magnetic resonance on hydrogen and image analyses. *Food Hydrocolloids*, **106**, 105907.

Hinrichs R, Götz J, Noll M, Wolfschoon A, Eibel H and Weisser H (2004) Characterisation of the water-holding capacity of fresh cheese samples by means of low resolution nuclear magnetic resonance. *Food Research International* **37**, 667–676.

Lee WJ and Lucey JA (2010) Formation and physical properties of yogurt. *Asian-Australasian Journal of Animal Science* **23**, 1127–1136.

Lee LC, Liong CY and Jemain AA (2018) Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *The Analyst* **143**, 3526–3539.

Low D, Ahlgren JA, Horne D, McMahon DJ, Oberg CJ and Broadbent JR (1998) Role of *Streptococcus thermophilus* MR-1C capsular exopolysaccharide in cheese moisture retention. *Applied Environmental Microbiology* **64**, 2147–2151.

Lucey JA and Singh H (1997) Formation and physical properties of acid milk gels: a review. *Food Research International* **30**, 529–542.

Metrohm (2013) Monograph: A guide to near-infrared spectroscopic analysis of industrial Manufacturing processes. *Monograph*. Retrieved 2020-03-04 from https://www.metrohm.com/da/documents/81085026.

Miller CE (2001) Chemical Principles of Near-Infrared Technology (2nd edition, Chapter 2). Near-Infrared technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Inc.

Nguyen ATB, Nigen M, Jimenez L, Ait-Abderahim H, Cunault C, Marchesseau S and Picart-Palmade L (2018) A multi-scale approach to identify the role of heat treatment, milk protein composition and starter culture on the gel formation and the texture defects of acid milk gel. *Food Hydrocolloids* **85**, 299–310.

Poulsen VK, Derkx P and Oregaard G (2019) High-throughput screening for texturing *Lactococcus* strains. *FEMS Microbiological Letters* **336**, 2.

Tarazanova M, Huppertz T, Kok J and Bachmann H (2018a) Altering textural properties of fermented milk by using surface-engineered *Lactococcus lactis*. *Microbial Biotechnology* **11**, 770–780.

Tarazanova M, Huppertz T, Kok J and Bachmann H (2018b) Influence of lactococcal surface properties on cell retention and distribution in cheese curd. *International Dairy Journal* **85**, 73–78.

Terzaghi BE and Sandine WE (1975) Improved medium for lactic streptococci and their bacteriophages. *Applied Microbiology* **29**, 807–813.

Zhu J and Ding W (2010) Quantified analysis of textural properties of goat yoghurt by near infrared spectroscopy. *Food Science* **19**, 170–173.

https://doi.org/10.1017/S0022029922000279 Published online by Cambridge University Press