Atypical spectra screening: applications for monitoring infrared instruments and model predictions

L. Spieß

Qlip B.V., P.O. Box 119, 7200 AC Zutphen, The Netherlands
Corresponding Author: spiess@qlip.nl

Abstract

Fourier-Transformmid-InfraRed (FT-IR) spectrometry is a recognized and widely used method to determine the compositional quality of raw milk and other liquid milk products. In recent years, mathematical models were developed that are capable of detecting atypical milk samples. Here we show how atypical spectra screening can also be used to monitor variations in the spectra that are related to the instrument performance rather than to the characteristics of the milk. By studying the temporal dynamics of spectra anomaly scores, it is possible to detect measurement instabilities caused by changes in the measurement context (i.e., instabilities of FT-IR instruments). We also show how algorithms can be developed to detect episodes of measurement instabilities automatically and in real-time.Such information can provide insights in the maintenance status of infrared instruments and changes in the measurement context. This, in turn, is important to ensure a consistent quality of infrared measurements and chemical profiling of milk samples.

Key words: Milk, Infrared spectrometry, atypical spectra screening.

Introduction

Fourier-transform mid infrared spectrometry (FT-IR) is a recognized and widely used method to rapidly determine the compositional quality of raw milk and other liquid milk products. This is done by generating spectra based on the absorption of light at different frequencies that result from different chemical bonds present in the milk. On the basis of these spectra, mathematical prediction models can predict (i.e., calculate) the concentration of fat, protein, lactose, and many other parameters. For proper predictions in routine operation, the accuracy has to be continuously monitored. If needed, slope/intercept settings can be adjusted based on analysis of calibration sample sets with chemical reference values.

In the recent years, FT-IR spectrometry has gained increasing nterest for its potential to predic tindividual fatty acids or groups of fatty acids (Fleming *et al.*, 2017), minerals (Stocco, Cipolat-Gotet, Bonfatti, Schiavon, Bittante and Cecchinato, 2016), green house gas emissions (Vanlierde*et al.*, 2016), energy status of the cow (McParland *et al.*, 2012), pasture (Coppa *et al.*, 2020)dry matter intake (Dórea, Rosa, Weld and Armentano, 2018), adulteration of milk (Hansen and Holroyd, 2019), and others. These parameters are of interest not only to individual farmers for monitoring and improving herd management practices, but also to feed advisors, veterinarians, dairy processors and researchers. Due to the economic impact, quality control procedures need to secure that the predictions are valid. However, with more and more parameters that need to be monitored, the costs associated with chemical reference analyses of control milk samples become increasingly uneconomic. Moreover, for parameters such as pasture

intake or gas emissions, chemical reference analyses that can be applied to the milk do not exist. As a consequence, a different and more general approach to monitor the validity of such predictions is required.

Since FT-IR predictions are eventually based on the information present in the spectra, the spectra need to be of good quality. Importantly, acquiring spectra from milk samples is, in fact, a measurement process that takes place under certain conditions (e.g., ambient temperature, physical sample characteristics) and by using a particular FT-IR instrument that is subject to wearing and measurement instabilities. Instabilities at the level of individual FT-IR instruments can manifest gradually (e.g., built up of a film inside the cuvette) or abrupt (e.g., suboptimal functioning of the homogenizer) and can occur over long (e.g., dissolving or wear of the cuvette) or short time scales (e.g., temporary issues with a moving mirror inside the cell). In principle, all these factors can lead to systematic changes in the obtained spectra. If this occurs at spectral bands containing information used by prediction models it can lead to erroneous or invalid predictions.

Monitoring routinely analyzed FT-IR spectra for systematic temporal deviations can therefore give valuable information about the validity of the predictions calculated from these spectra and, potentially, the functioning of the instrument. In the present paper we show how untargeted methods originally used to screen spectra for adulterated milk (e.g., Hansen and Holroyd, 2019) can be slightly modified to screen for systematic temporal deviations in routinely analyzed milk spectra. In short, this approach is based on establishing an FT-IR milk fingerprint that is unique to each FT-IR instrument and by tracking, over time, how individual milk spectra deviate from this fingerprint. Because the focus is on systematic deviations at the level of spectra rather than predictions calculated from them, the approach is more general compared to the use of pilot milk samples. Moreover, no cost and labor-expensive chemical reference analyses and pilot milk samples are required.

Materials and methods

Data

The data set contained 345473 spectra that correspond to Dutch bovine herd bulk milk samples randomly collected between January 2018 and November 2020. All milk samples were routinely analyzed for milk payment. For acquisition of the spectra, milk samples were randomly assigned to one of four FT-IR instruments (Milkoscan FT+, FOSS Analytical A/S, Hillerød, Denmark) where they all underwent the same pre-treatment before the scan took place. FT-IR spectra were obtained in the mid infrared region with wavelengths between 10.8 µm (926 cm^{-1}) and 1.995 µm (5012 cm^{-1}). AllFT-IR instruments were standardized monthly using the FOSS equalizer application in accordance with the manufacturer's instructions (Winning, 2015).

Untargeted spectra screening

For the development of a mathematical model to identify non-specific deviations in milk spectra, we followed a conceptually similar approach as described by (Hansen and Holroyd, 2019) and employed by manufacturers of FT-IR instruments (FOSS, 2014). For each of the four FT-IR instruments in the Qlip payment testing laboratory a separate model was developed on the basis of ca. 85000 spectra. Preprocessing of the spectra consisted of selecting wavenumbers between 925 cm^{-1} and 1550 cm^{-1}, 1710 cm^{-1} and 1900 cm^{-1}, and between 2700 cm^{-1} and 2971 cm^{-1}. Moreover, the spectra from each instrument were standardized to have, per wavelength, zero-mean absorption with unit-variance. We then used a principal component analysis (PCA) with ten components to project the spectra to the latent space. After transforming the spectra to the latent space, we computed the covariance matrix and calculated per spectrum

the Mahalanobis distance. In the next step, we used the PCA to perform an inverse transformation on the spectra in the latent space in order to obtain the reconstructed spectra in the original space. By calculating, across all wavelengths, the root-mean-square error between the original and reconstructed spectra, we obtained the spectral residuals. In the last step, the Mahalanobis distances and the spectral residuals were each standardized to have zero mean with unit variance before they were summed to a single score per spectrum: the spectrum anomaly score. The higher the spectrum anomaly score, the more a spectrum deviates from all the spectra that were used to develop the screening model.

Identification and analysis of systematic deviations in IR spectra

The use of untargeted spectra screening for identifying instrument instabilities is grounded on two critical assumptions. First, instrument instabilities need to be events that have a duration such that a sufficient number of spectra is influenced. Instabilities affecting individual spectra entirely incidental and temporally unpredictable will not be detected. Second, the chemical composition of the milk samples has to be independent of the time at which the spectra are obtained. That is, milk samples from different groups (e.g. cow milk and goat milk) should not be analyzed in discrete groups. Only under these conditions it is possible to identify and relate systematic deviations in spectra to instabilities during their acquisition. The detection of systematic deviations in IR spectra therefore comes down to detecting deviations in the time course of anomaly scores at the level of individual instruments over a pre-defined time window. As can be seen in Figure 1, such systematic changes in the anomaly scores can present themselves as drifts (upper panel) or sudden jumps (lower panel) and can even impact routinely predicted milk compositional parameters such as urea and free fatty acids in the milk obtained from commercial prediction models installed on the respective instruments.

By visually inspecting for each instrument the anomaly scores on a daily basis over a period of three years, we identified 48 events with comparable instabilities. Of these, seven events were related to instrument A, sixteen to instrument B, eleven to instrument C, and fourteen to instrument D. For an actual implementation of such a tool in routine FT-IR analysis of milk samples, it is desirable to detect such incidences in real-time and as early as possible. To do so, we fed the anomaly scores of one day, one after another, to a Bayesian online change point detection algorithm (Adams and MacKay, 2017). The algorithm uses Bayesian inference to compute a distribution over the next unseen anomaly score in a time series, given only the anomaly scores it has seen before. It then computes for each individual anomaly score the probability that it reflects a change point on the basis of the past k anomaly scores. We found that $k = 24$ successive anomaly scores were sufficient to reveal most of the previously identified instrument instabilities. In other words, systematic changes in the anomaly scores could be detected with a latency of 24 measured spectra. The resolution at which a change point can be detected is at the level of an individual spectrum (see Figure 1 for examples).

So far, we focused on analyzing systematic deviations in anomaly scores on a spectrum-by-spectrum basis. However, instrument instabilities or drifts can also manifest over longer time scales. Figure 2 illustrates for each of the four instruments the change in anomaly scores over a period of three years. We also applied the Bayesian online change point algorithm to these time series. This time, however, we computed the average anomaly score over a time window of two consecutive days and computed the probability of each possible change point using the last three two-day-averages. This means that systematic changes in the anomaly scores could be detected with a latency of six days and a resolution of two days. As can be seen in Figure 2, most of the major peaks were readily identified by the algorithms.

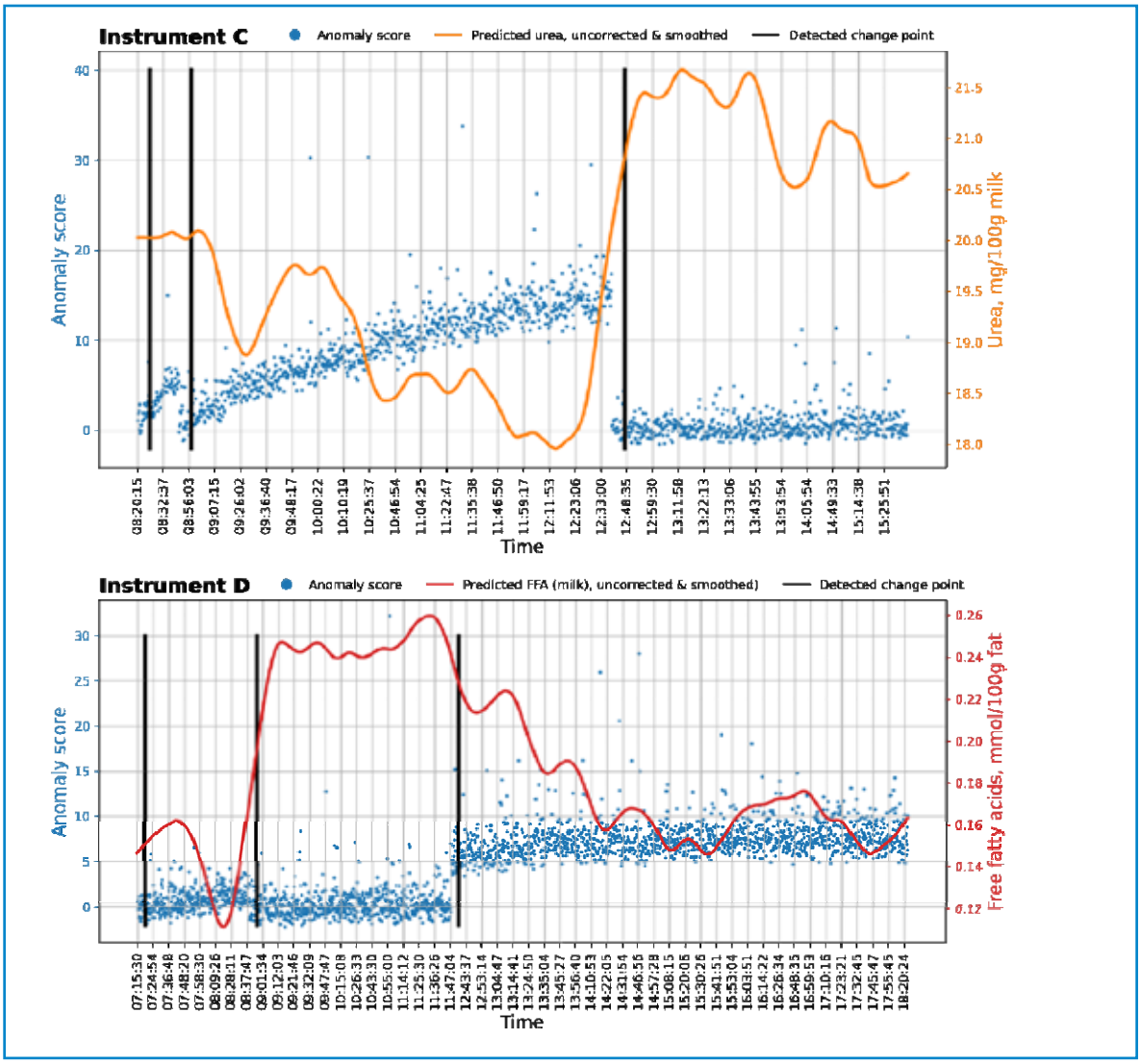

Figure 1. Examples of measurement instabilities that manifest as gradual (upper panel) or sudden (lower panel) temporal changes in the spectra anomaly scores over a day. Each blue dot corresponds to the spectrum anomaly score of an individual milk sample measured on that instrument. The orange and red lines depict LOWESS curves fitted to the uncorrected predictions for urea and free fatty acids in the milk, respectively. The vertical black bars reflect the moment in time where the Bayesian change point algorithm detected an unpredictable change in the anomaly scores. Note, how the change in the anomaly scores coincides with systematic biases in the predictions of urea and free fatty acids.

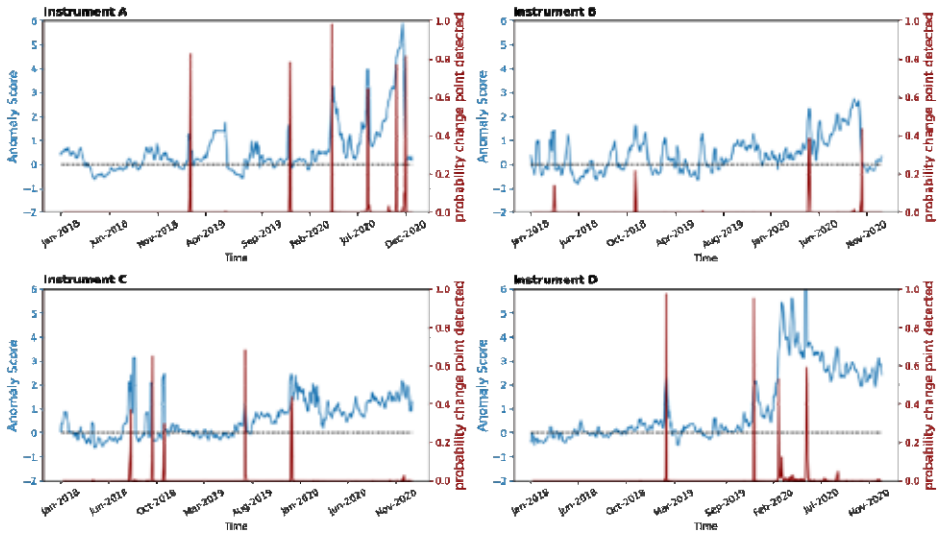

Figure 2. Long-term average of anomaly scores for each of four instruments from 2018 to 2020. The blue lines reflect the average anomaly score. The red spikes reflect the moment in time where the Bayesian change point algorithm identified an unpredicted change in the average anomaly score. The height of the spike corresponds to the probability or confidence that a change point has been detected.

Interestingly, the anomaly scores for instrument D appear to change drastically from early 2020 onwards. Around the same time, our colleagues began to notice increased fluctuations in the predictions of urea and free fatty acids on our pilot milk samples. These problems persisted until end of March 2020, when the cuvette was eventually replaced. Aside from abrupt and temporary fluctuations in the anomaly scores, all four instruments reveal a trend towards higher anomaly scores as time passes by. This can be a consequence of dissolving or wear of the cuvette or the built-up of a film inside of it.

For the majority of the instabilities that we discovered, it was not possible to relate it to documented maintenance work or other disturbances reported around that time. This is not surprising given that none of these events was detected or issued by the quality control sensors present in FT-IR instrument. Nonetheless, we can use the spectra to get a basic understanding about how systematic the effects of such instabilities are and whether certain types of instabilities reappear. To do so, we created a dataset with all milk spectra that were measured during the 48 episodes with instabilities. This amounts to ca. 20000 individual milk spectra. We then used a combination of PCA and *t*-distributed stochastic neighbor embedding to visualize possible patterns in the residuals. As can be seen in Figure 3, on a macro level, the four different instruments are well separated from each other. This emphasizes the need for the development of separate models per instrument. At the micro level, it can be seen that the number of distinct clusters (n = 23) is about half the number of identified incidences with instrument instabilities (n = 48). On the one hand, this can be an indication that we identified 23 distinct and unique categories of instrument instabilities. On the other hand, some of the clusters, particularly the small ones, only contain spectra from a single day. Because seasonal variations are not explicitly modelled in the untargeted screening model, they cannot be removed from the spectral residuals. This means that the same type of instrument instability that occurs during one day in the winter and one day in the summer can lead to distinct clusters. Nevertheless, some clusters contain spectra obtained during instrument instabilities that occurred on more than a single day. This suggests that certain instrument instabilities reappear.

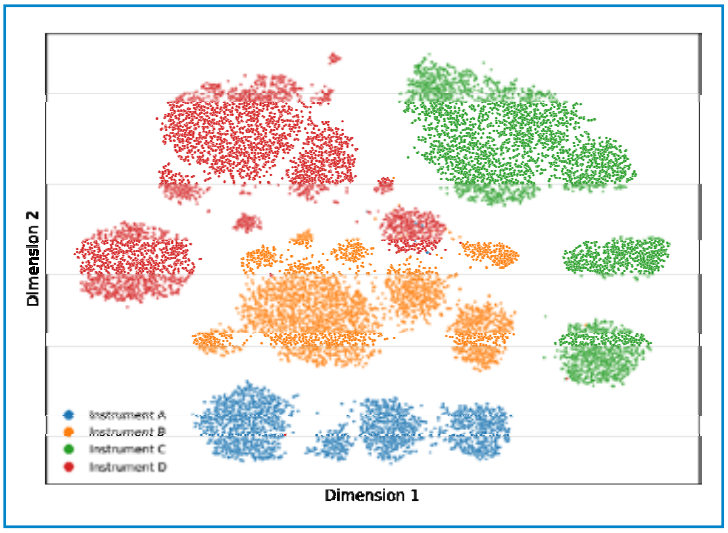

Figure 3. Clustering of milk spectra collected during episodes of instrument instabilities (n = 48). Each dot reflects an individual milk spectrum positioned on the two dimensions of a t-distributed stochastic neighbor embedding. The colors correspond to the different instruments.

Discussion

In the present paper, we demonstrated how untargeted models, originally developed to screen for adulterated liquid milk samples, can also be used to screen IRspectra for instrument instabilities. Our approach is based on the development of an FT-IR milk fingerprint that is unique to each instrument. When combined with change point algorithms, we have shown how instrument instabilities can be identified automatically and in real-time without requiring pilot milk samples. The spectral regions that our finger prints are based on contain chemically relevant information used by various prediction models. Instrument-related artifacts in these spectral regions can and occasionally did lead to biased predictions. Moreover, in the case of frequent short-term instabilities (in the range of minutes to hours) or strong drifts in the long-term (in range of months to years), additional maintenance measures could be considered.

In our admittedly small investigation, we have shown that instrument instabilities can be strong enough to bias predictions from commercial models of urea and free fatty acids in the milk. It should be noted, however, that instrument instabilities do not always have to affect model predictions and are certainly not the only factor affecting these predictions. Moreover, predictions of the chemical components such as fat, protein, and lactose are probably less susceptible to spectral deviations than models predicting parameters that are not directly present in the milk itself (e.g., fatty acids in the blood, pasture intake, or gas emissions). This is particularly problematic when these parameters are weakly correlated with the milk spectrum. In this case, prediction models are often more susceptible to spectral noise in general. It should also be noted that we could not relate the identified instrument instabilities to a particular cause. This is not surprising given that these instabilities were not detected or reported by the quality control sensors implemented in the instrument. Interestingly, the majority of temporary instabilities disappeared without human intervention. While individual episodes of instability may not always signal a need for maintenance, a change in the frequency

of such episodes, however, deserves attention. Finally, it should be kept in mind that systematic changes in the spectra can be due to any change in the measurement conditions. These do not always have to be restricted to the FT-IR instrument but can also be due to changes in ambient temperature, humidity, sample pre-treatment, and even the physical properties of the milk sample itself.

When it comes to a practical application of our approach, the specific implementation may differ depending on the use case. When used as a quality control tool for monitoring model predictions, the change points in combination with the anomaly scores can be used to indicate whether an individual spectrum falls within an episode of instability or not. If it does fall in an episode of instability, one can compute the mean of the predictions before, during, and after the episode and compare them for statistical differences. When used as an additional quality control tool for instrument functioning, the detection of a change point in the anomaly scores can be used as a signal to have the respective instrument checked by the operator and, if needed, flagged for maintenance or repair.

List of references

Adams, R.P.and MacKay, D.J.C., 2007. Bayesian online change point detection. arXiv preprint arXiv: 0710.3742

Coppa, M., Martin, B., Hulin, S., Guillemin, J., Gauzentes, J.V., Pecou, A. and Andueza, D., 2020. Predictions of indicators of cow diet composition and authentication of feedings pecifications of protected designation of origin cheese using mid-infrared spectroscopy on milk. J. Dairy Sci., 104, 112-125. doi.: https://doi.org/10.3168/jds.2020-18468

Dórea, J.R.R., Rosa, G.J.M., Weld, K.A.and Armentano, L.E., 2018. Mining data from milkinfrared spectroscopy to improve feed intake predictions in lactating dairy cows. J. Dairy Sci.,101 (7),5878-5889. doi.: https://doi.org/10.3168/jds.2017-13997

Fleming, A., Schenkel, F.S., Chen, J., Malchiodi, F., Bonfatti, V., Ali., R.A., Mallard, B., Corredig, M. andMiglior, F., 2017. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. J. Dairy Sci., 100 (6), 5073-5081. doi.:https://doi.org/10.3168/jds.2016-12102

FOSS, 2014. Abnormal spectrum screening (ASM). Retrieved from:https://www.dairyfoods.com/ext/resources/White_Papers/WhitepaperDairyASM-US.pdf

Hansen, P.W. and Holroyd, S.E., 2019. Development and application of Fourier transform infrared spectroscopy for detection of milk adulteration in practice. International Journal of Dairy Technology, 72 (3), 321-331. doi.: 10.1111/1471-0307.12592

McParland. S., Banos, G., McCarthy, C., Lewis, E., Coffey, M.P., O'Neil, B., O'Donovan, M., Wall, E. and Berry, D.P., 2012. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. J. Dairy Sci., 95 (12), 7225-7235. doi.: 10.3168/jds.2012-5406

Stocco, G., Cipolat-Gotet, C., Bonfattie, V., Schiavon, S., Bittante, G. and Cecchinato, A., 2016. Variations in major mineral contents of Mediterranean buffalo milk and application of Fourier-transform infrared spectroscopy for their prediction. J. Dairy Sci., 99, 8680-8686. doi.: https://doi.org/10.3168/jds.2016-11303

Vanlierde, A., Vanrobays, M.-L., Gengler, N., Dardenne, P., Froidmont, E., Soyeurt, H., McParland, S., Lewis, E., Deighton, M.H., Mathot, M. and Dehareng, F., 2016. Milk mid-infrared spectra enable prediction of lactation-state-dependent methane emissions of dairy cattle within routine population-scale milk recording schemes. Animal Production Science, 56, 258-264. doi.: http://dx.doi.org/10.1071/AN15590

Winning, H. 2015. Standardization of FT-IR instruments. Foss. Retrieved from: https://www.google.com/url?sa=tandrct=jandq=andesrc=sandsource=webandcd=andcad=rjaanduact=8andved=2ahUKEwji5IXX0L7wAhVJyaQKHXTeBI8QFjAAegQIAxADandurl=

https://www.fossanalytics.com/-/media/files/documents/papers/dairy-segment/standardization-of-ft-ir-instruments_gb.pdfandusg=AOvVaw2hu8dwViWhw6fAwVe47QO6