proficient · reliable · analytical · e

Atypical Spectra Screening: Applications for Monitoring Infrared Instruments and Model Predictions

Lukas Spieß, PhD

Data scientist, Innovation and Business Development Qlip

ICAR 2021

Qlip

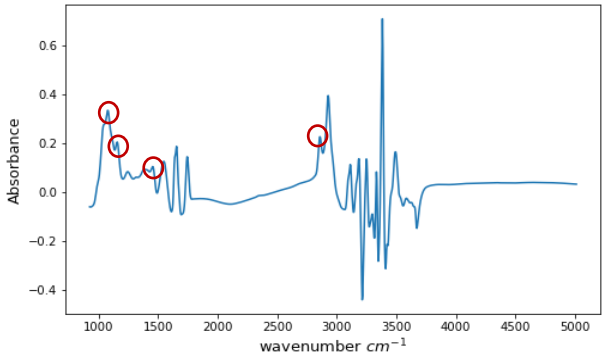
Quality assurance in dairy

Background

Milk sample

FT-IR spectrometer

FT-IR milk spectrum



- **F**ourier-**T**ransform mid **I**nfra**R**ed spectrometry (FT-IR) of liquid milk
- Composition milk profile predicted (i.e., calculated) from IR spectrum

Mathematical prediction model

4.3% fat
3.4% protein
4.5% lactose

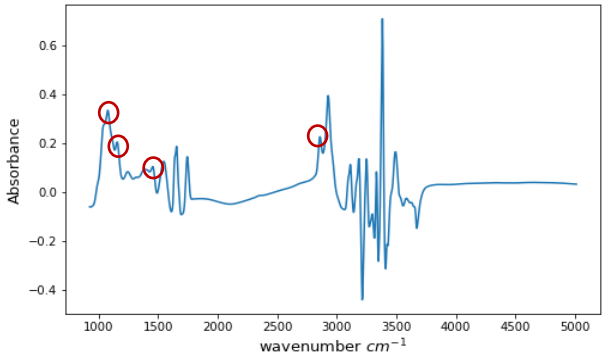
Qlip
Quality assurance in dairy

Background

Milk sample

IR spectrometer

FT-IR milk spectrum



- Quality of predictions dependent on:
  - Quality of the model (e.g., robustness)
  - Quality of the spectra (noise and artifacts)

Mathematical prediction model

4.3% fat
3.4% protein
4.5% lactose

Qlip
Quality assurance in dairy

Background

- Each spectrum results from **measurement process** performed under…
    - specific conditions
    - using a specific IR instrument

Quality assurance in dairy

Background

- Each spectrum results from **measurement process** performed under…
  - specific conditions
  - using a specific IR instrument

- Changes in measurement process (e.g., instabilities in the IR instrument) can lead to changes in the spectra

- **Problematic**:
  - …if this happens in spectral regions containing chemical information used by prediction models
  - …because the instrument is not working properly

Qlip
Quality assurance in dairy

Background

Slope / intercept correction

- ✓ Fat
- ✓ Protein
- ✓ Lactose

Quality assurance in dairy

Background

- Limited use-case for slope / bias corrections

Slope / intercept correction

	✓ Fat	✗ Pasture intake
	✓ Protein	✗ Methane emissions
	✓ Lactose	✗ NEFA, BHB_{Blood}
Present in control milk?	✓	✗

Qlip
Quality assurance in dairy

Background

- Limited use-case for slope / bias corrections

- **We need a general approach for monitoring the quality of our data (spectra, predictions)**

Slope / intercept correction

	Fat	Pasture intake
	Protein	Methane emissions
	Lactose	NEFA, BHB
Present in control milk?	✓	✗

Quality assurance in dairy

Background

- Limited use-case for slope / bias corrections

- **We need a general approach for monitoring the quality of our data (spectra, predictions)**

- Monitor spectra for systematic deviations due changes in the measurement process (i.e., instrument instabilities).

    - Use mathematical models originally developed to screen for non-specific adulteration in milk samples

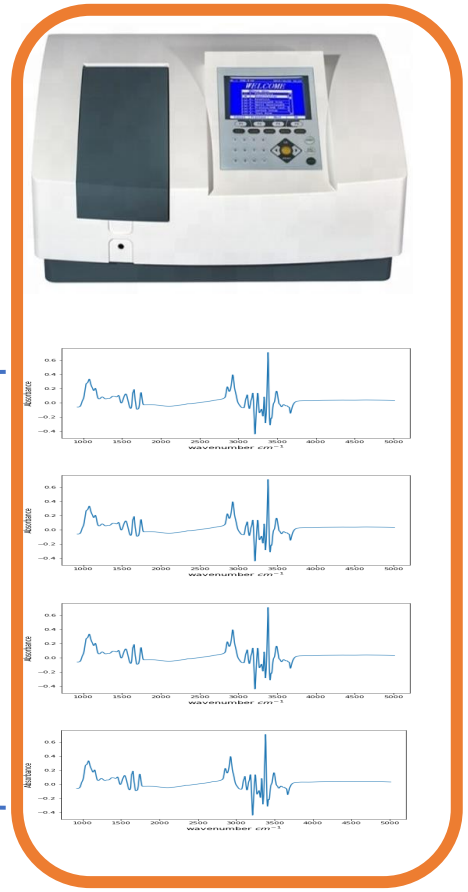
Slope / intercept correction

	✓ Fat	✗ Pasture intake
	✓ Protein	✗ Methane emissions
	✓ Lactose	✗ NEFA, BHB
Present in control milk?	✓	✗

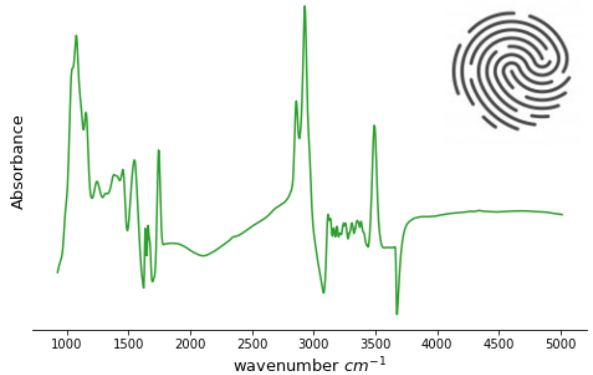
Qlip
Quality assurance in dairy

Methods & Results

Milk spectra with normal variation



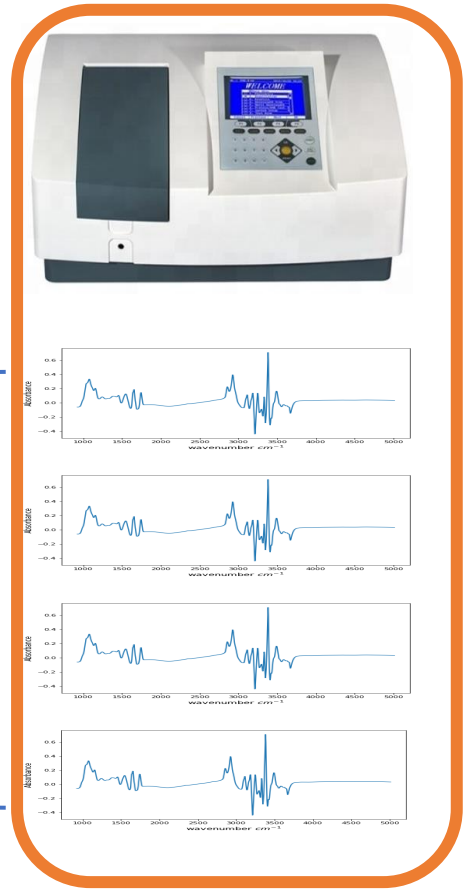
Instrument-specific milk fingerprint



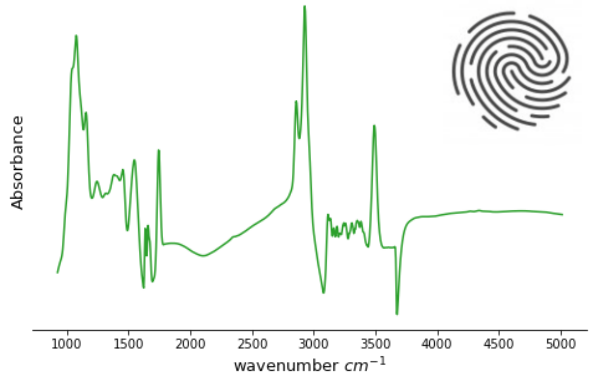
Quality assurance in dairy

Methods & Results

Milk spectra with normal variation

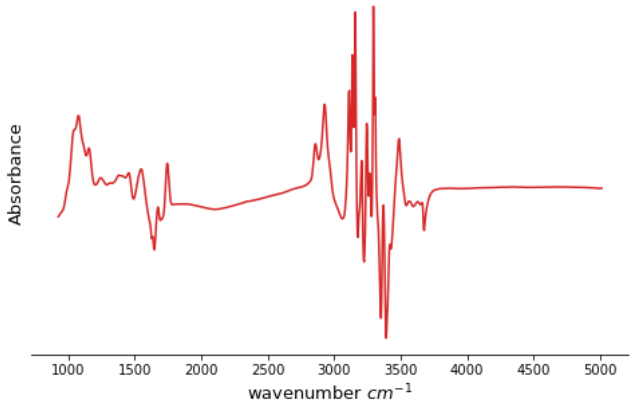


Instrument-specific milk fingerprint



vs.

New spectrum from same instrument



Compute anomaly scores for each spectrum from the same instrument

Qlip
Quality assurance in dairy

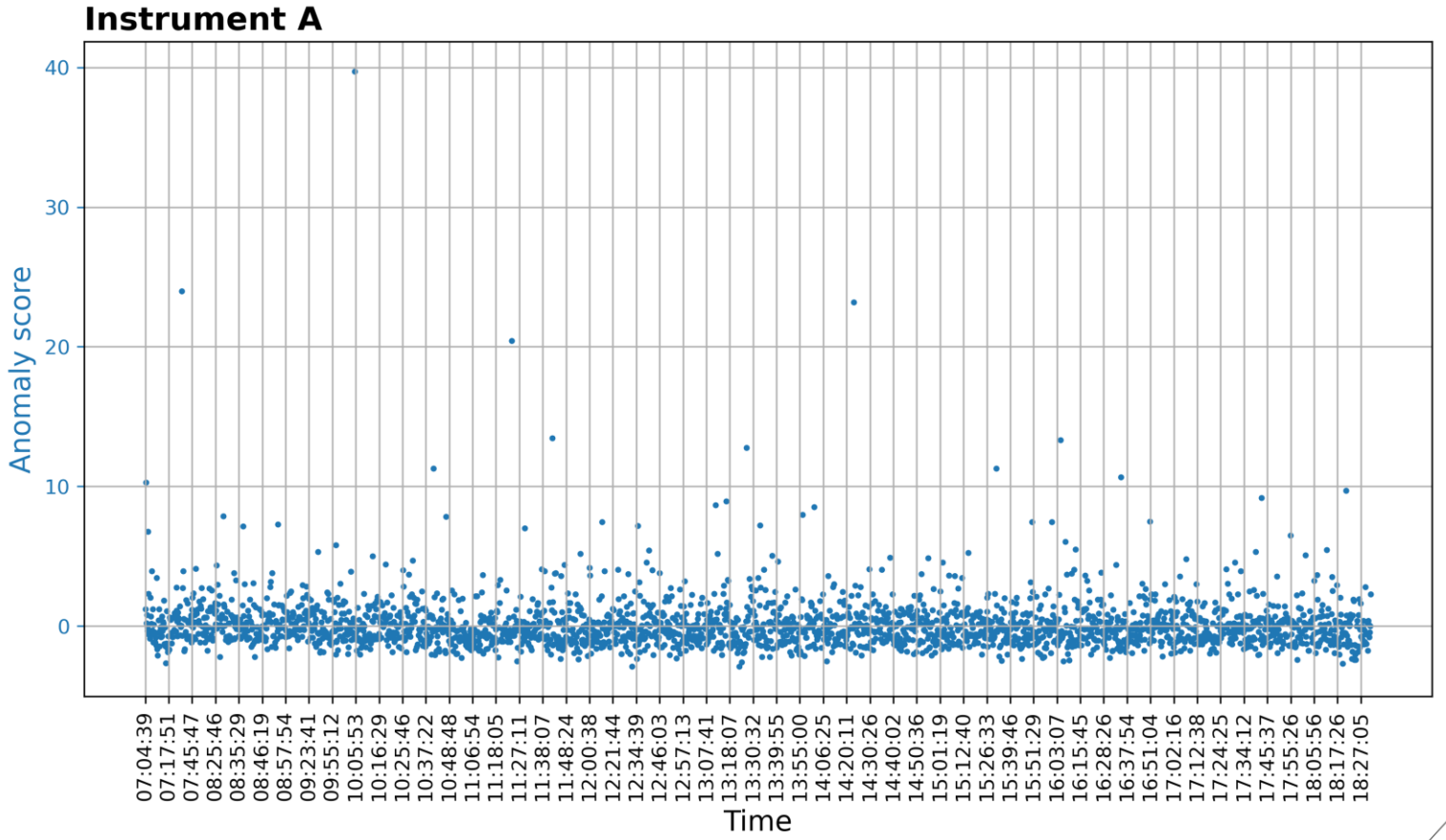
Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**

Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**

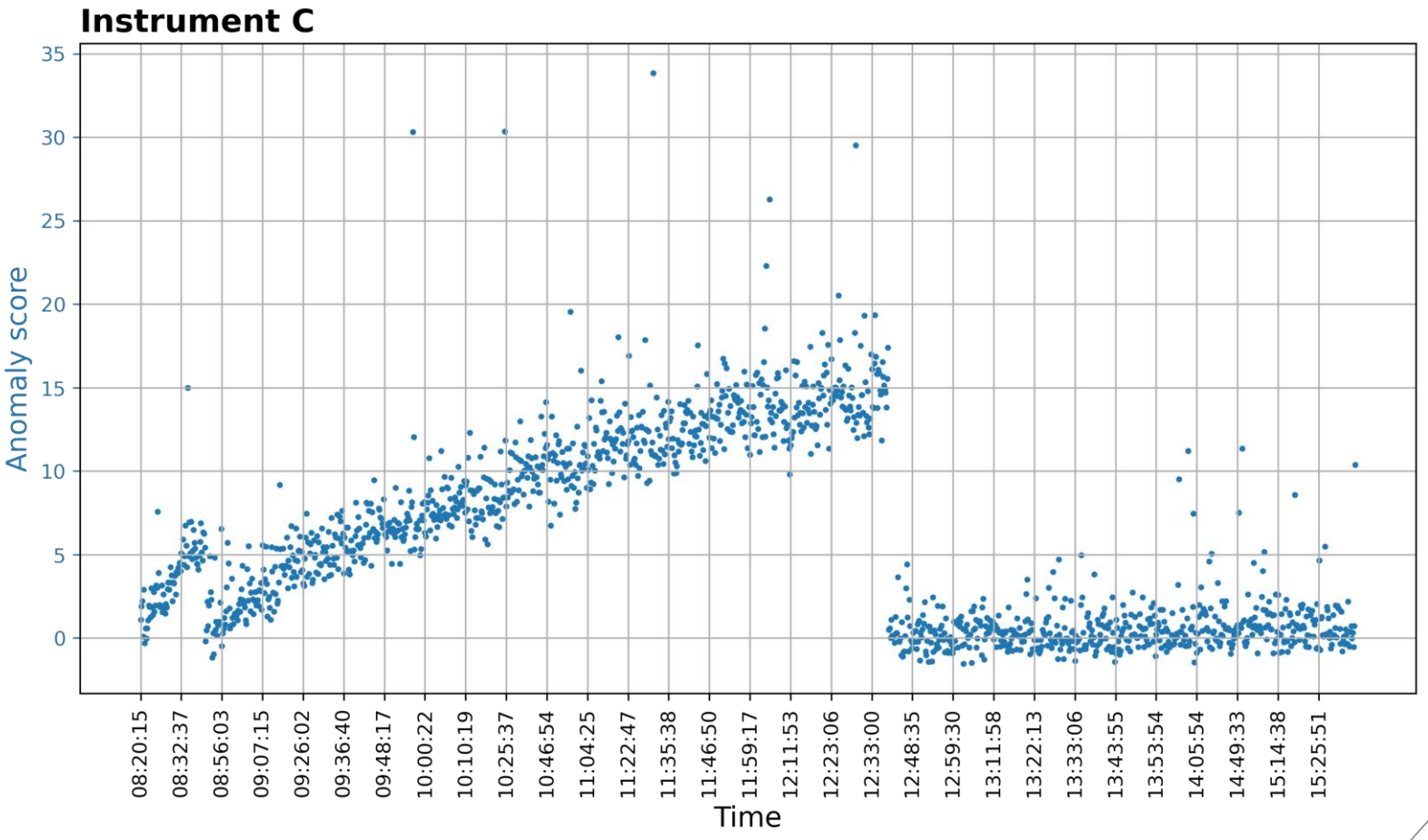


No pattern; mostly random noise with zero mean

Qlip
Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**

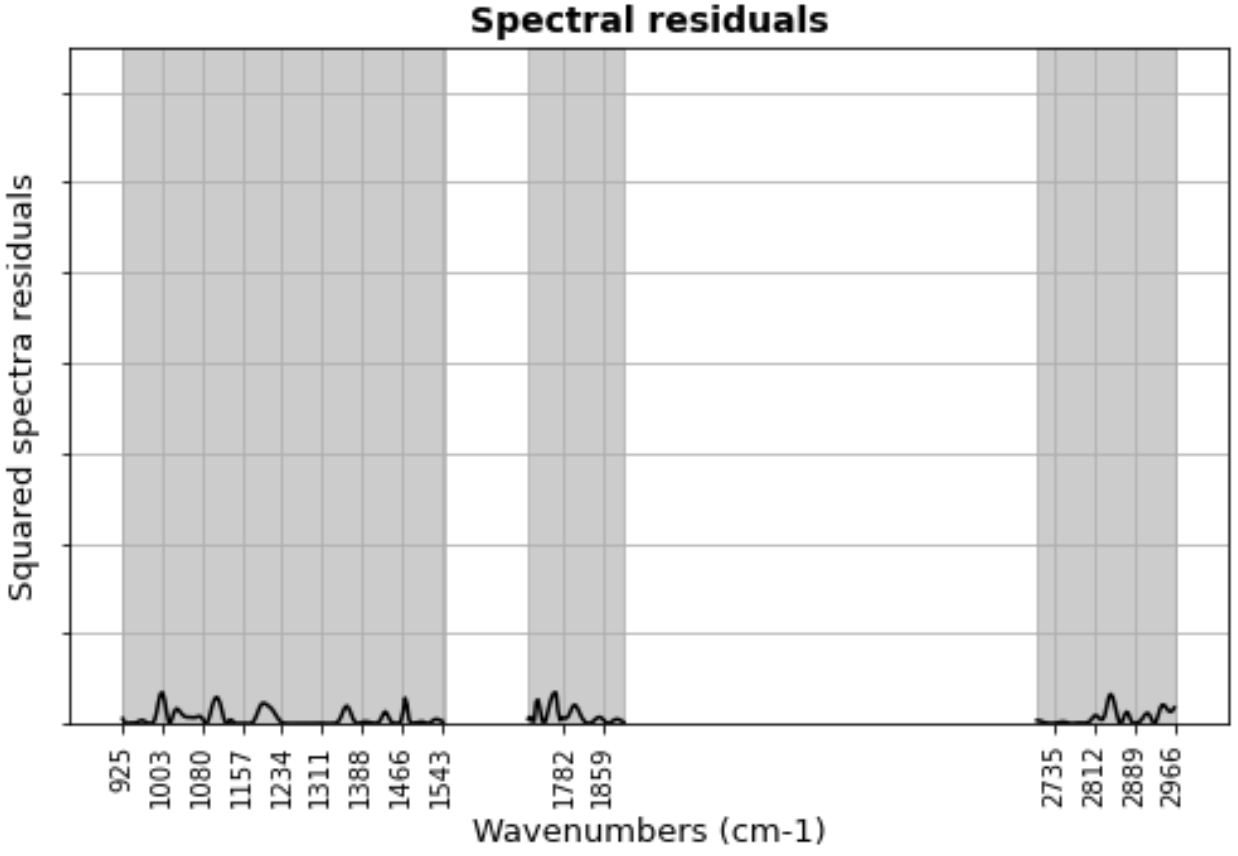


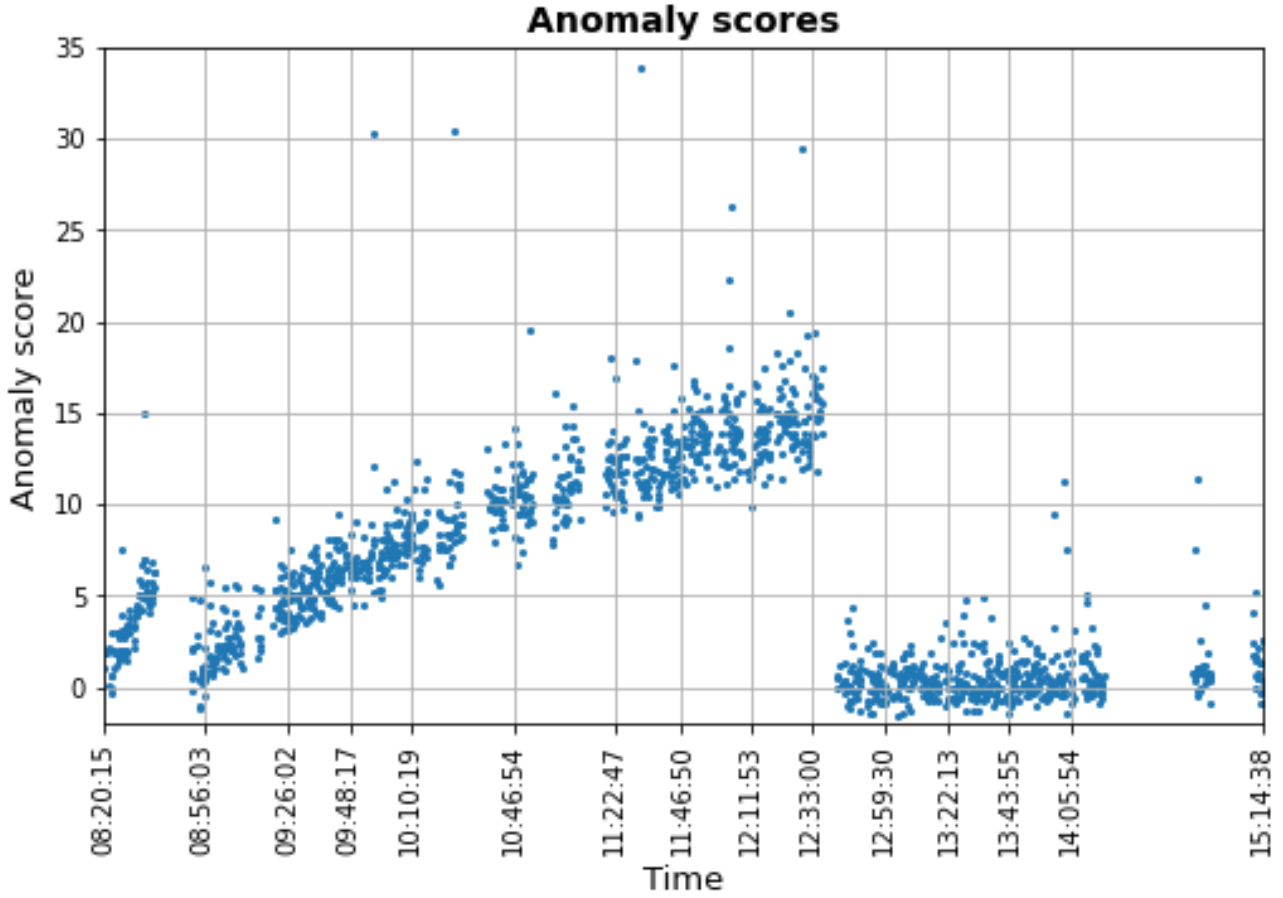
Gradual increase of anomaly scores

Qlip
Quality assurance in dairy

Methods & Results

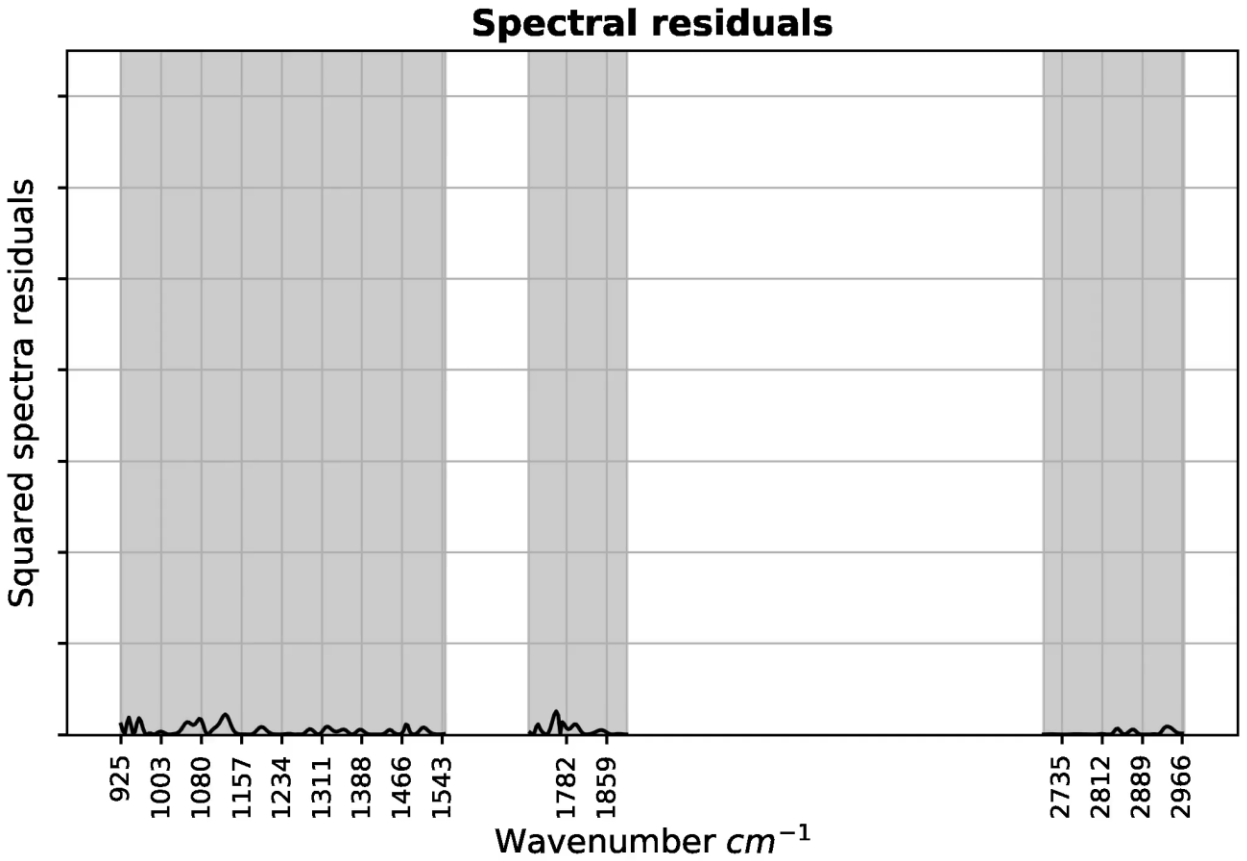
- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**

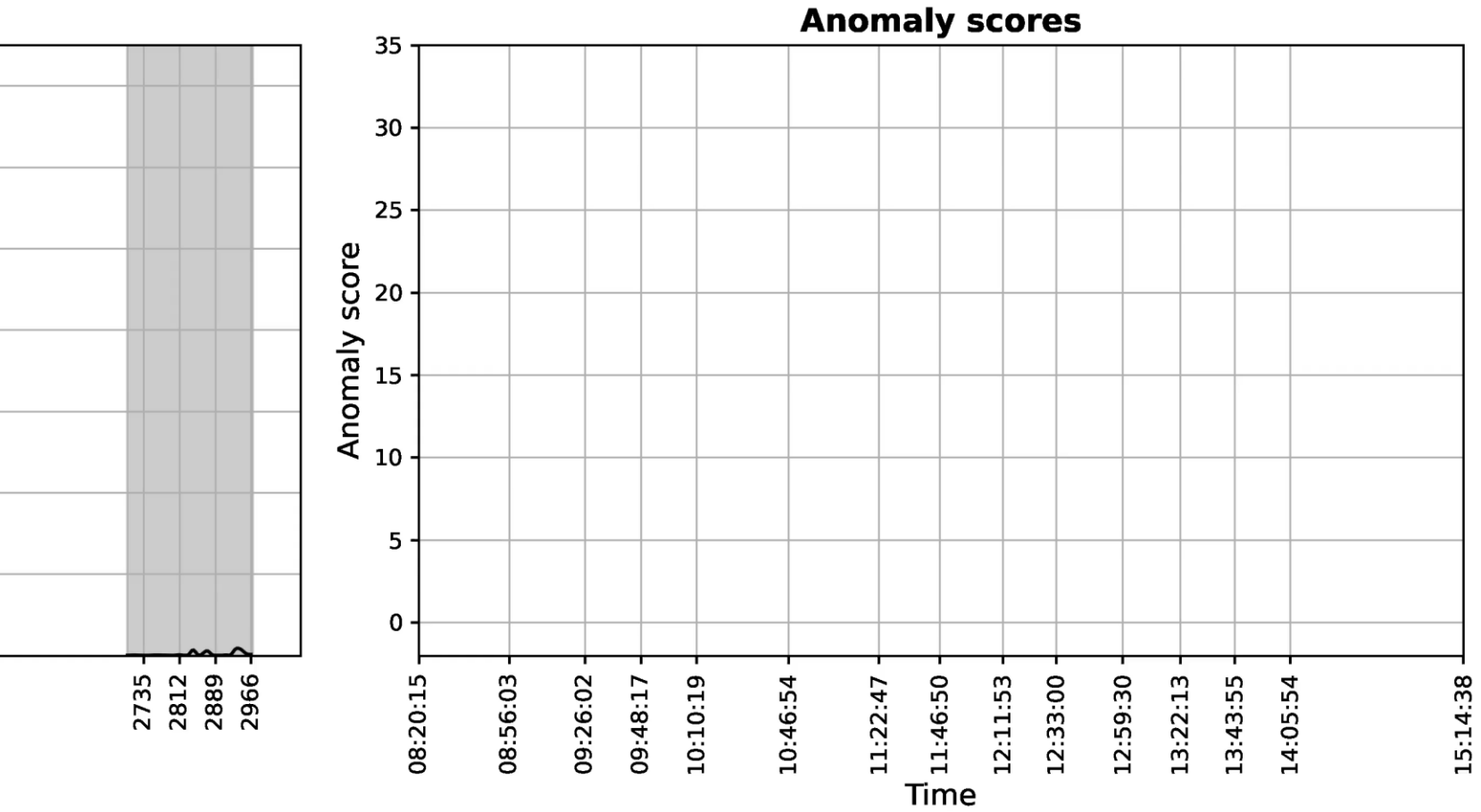




Methods & Results

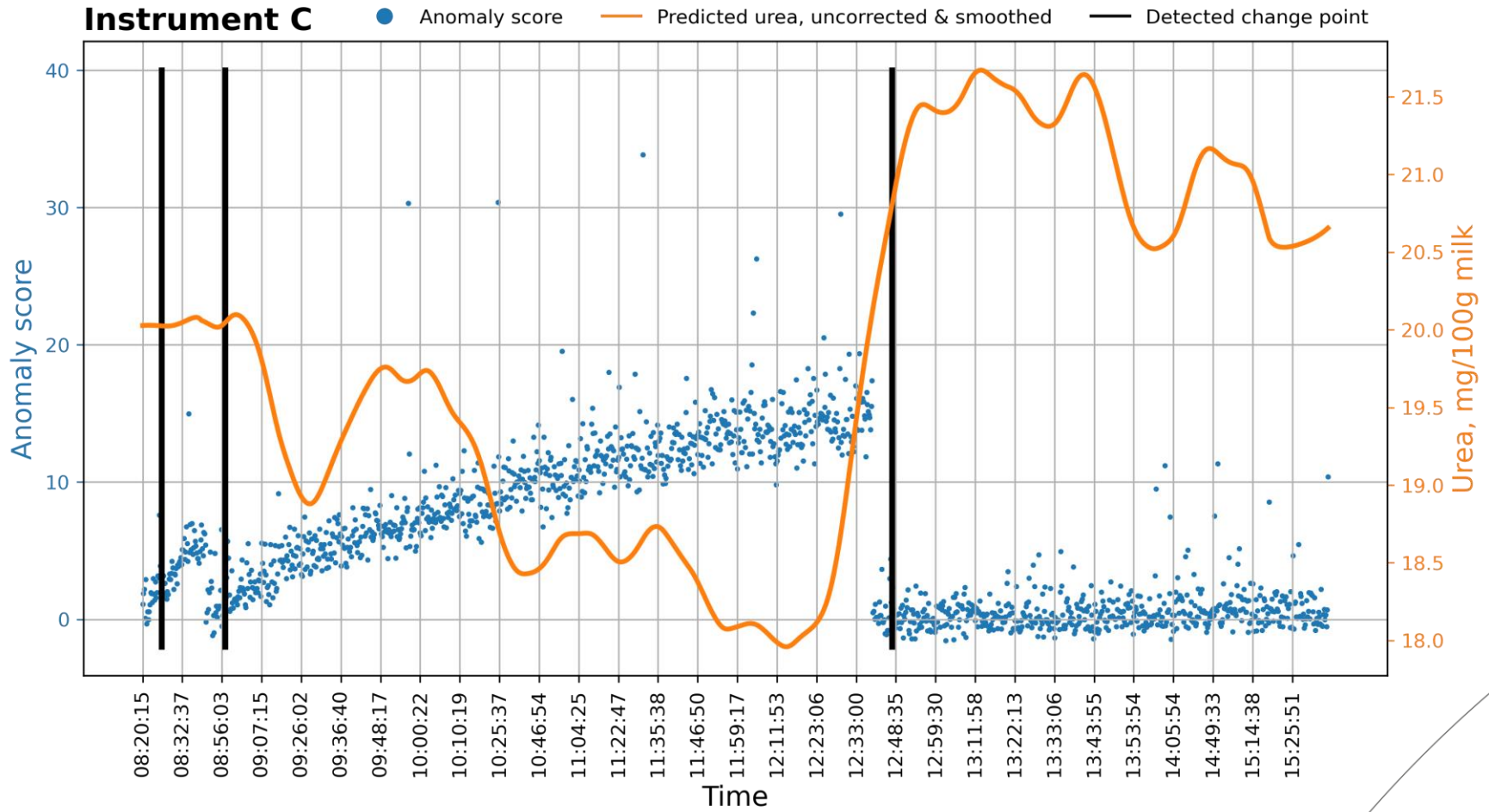
- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**





Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**

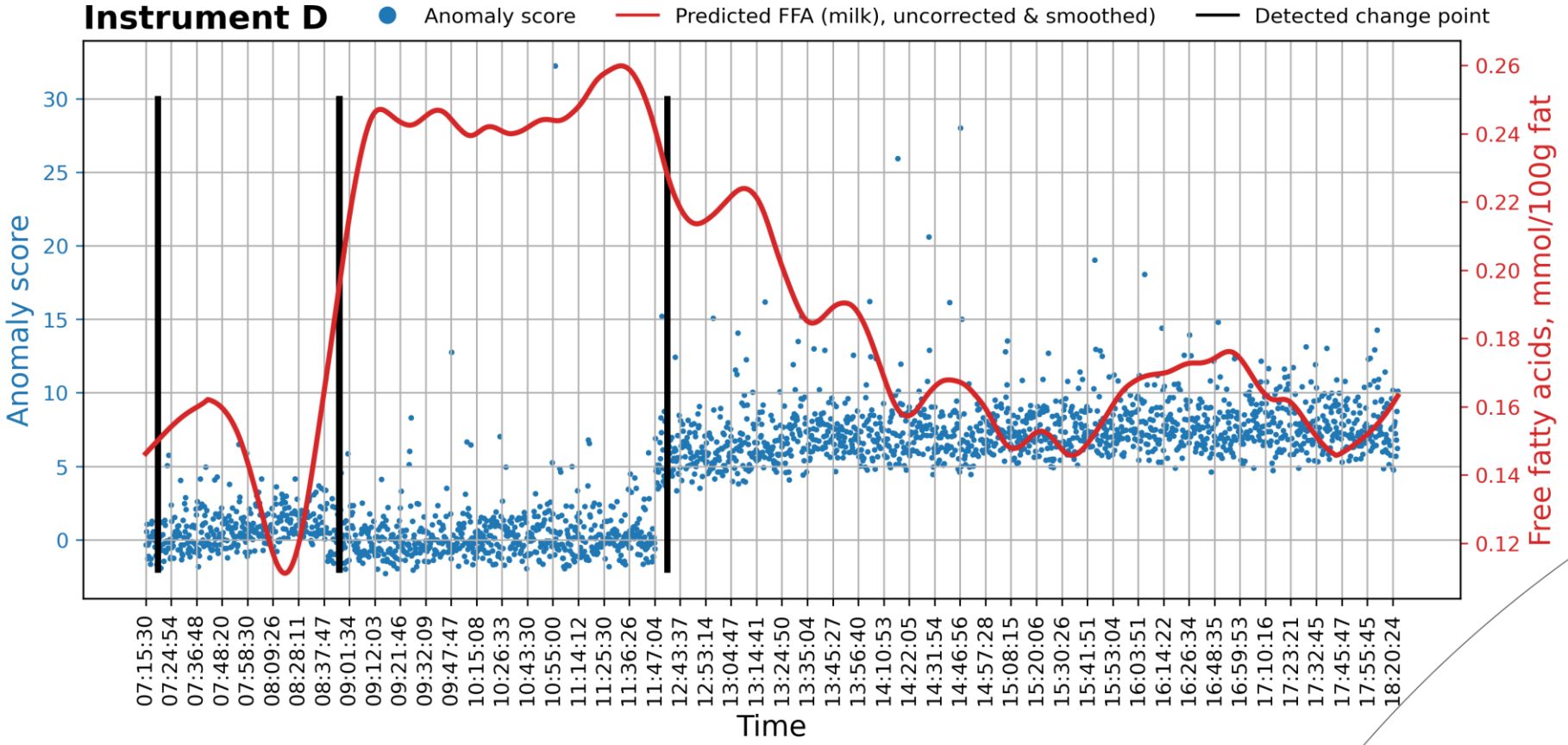


Change in anomaly scores coincides with bias in urea predictions

Qlip
Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- **Look for patterns over time**



Change in anomaly scores coincides with bias in free fatty acid predictions

Qlip
Quality assurance in dairy

Methods & Results

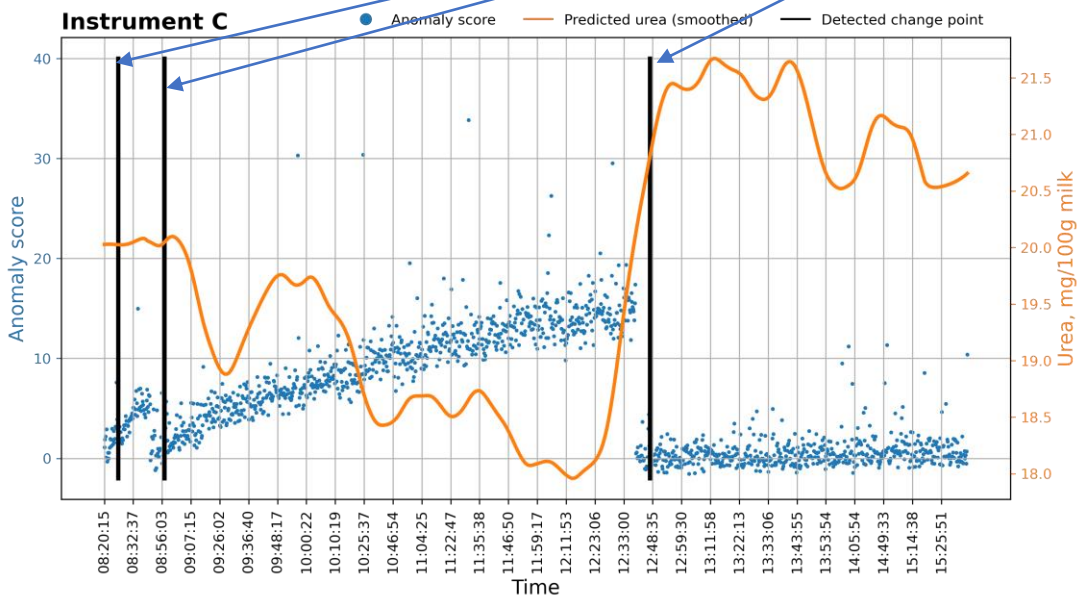
- Compute anomaly scores for each spectrum from the same instrument
- Look for patterns over time
- **Detect systematic deviations automatically and real-time**

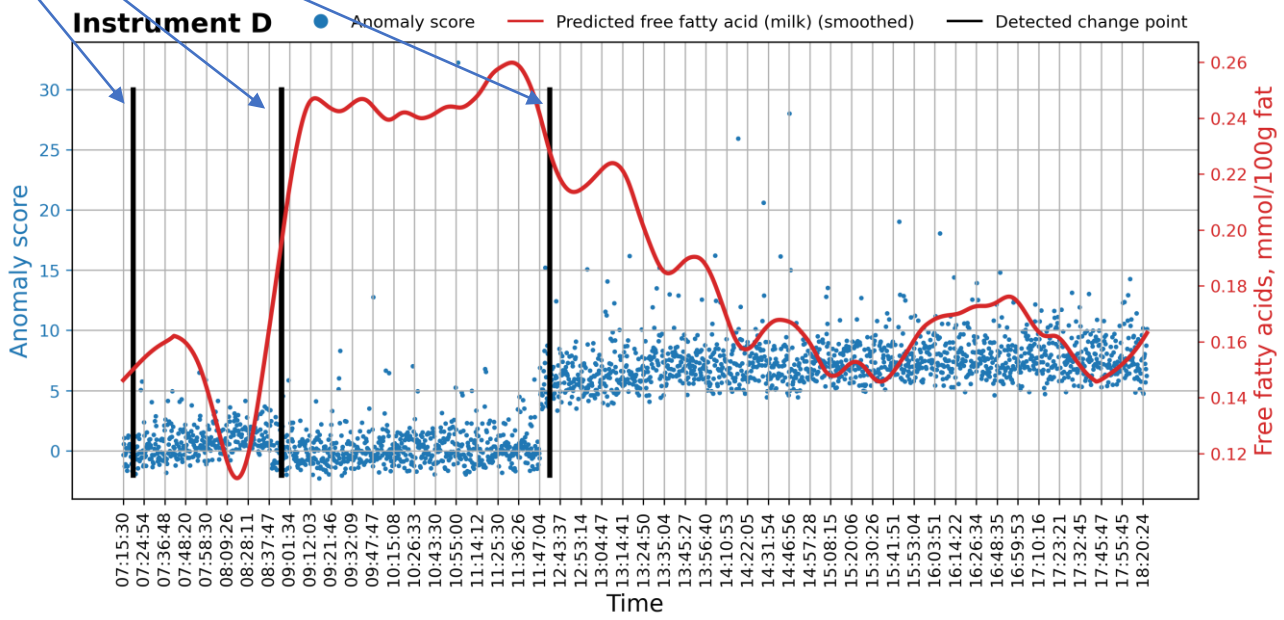
Qlip
Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- Look for patterns over time
- **Detect systematic deviations automatically and real-time**

Bayesian online change point algorithm





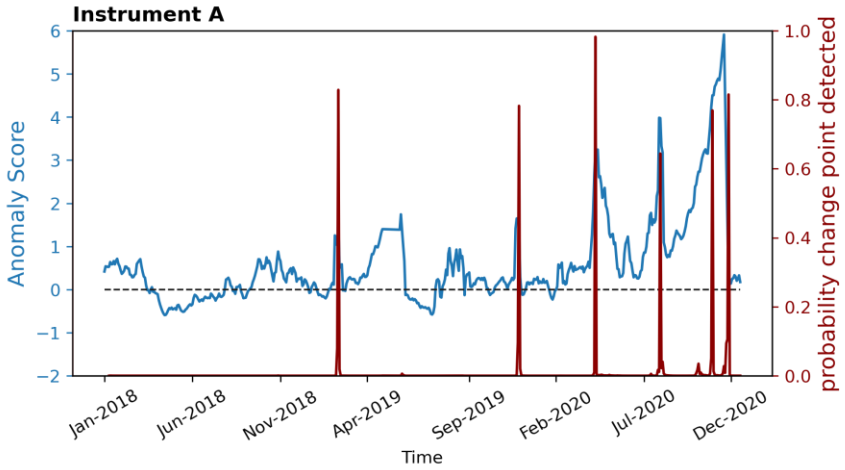
Methods & Results

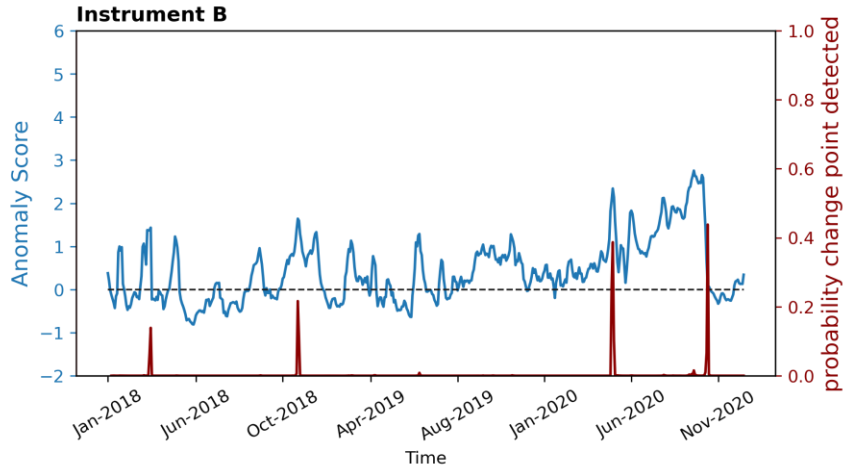
- Compute anomaly scores for each spectrum from the same instrument
- Look for patterns over time
- **On a scale of months to years**

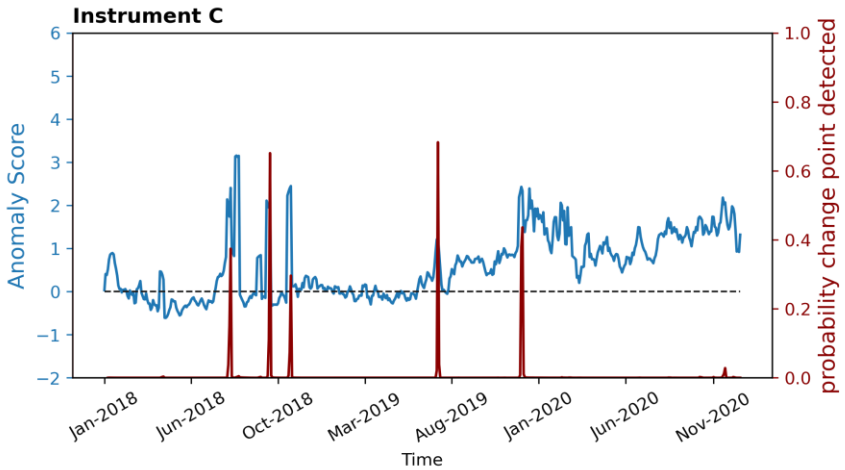
Quality assurance in dairy

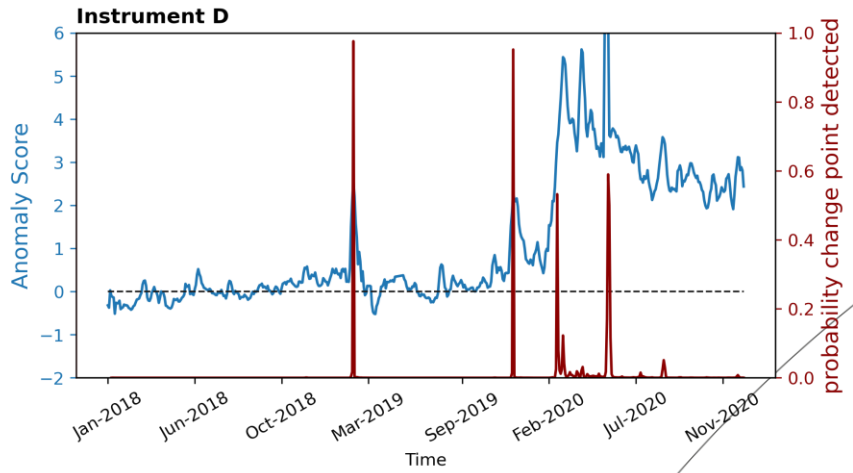
Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- Look for patterns over time
- **On a scale of months to years**









Steady drifts

&

Sudden jumps

Qlip

Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- Identify episodes with instabilities
- **Find patterns in spectral residuals**

Qlip
Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- Identify episodes with instabilities
- **Find patterns in spectral residuals**

Collected > 20.000 spectra from 38 episodes
with instabilities identified across four instruments
within a period of three years (> 5.9 million spectra)

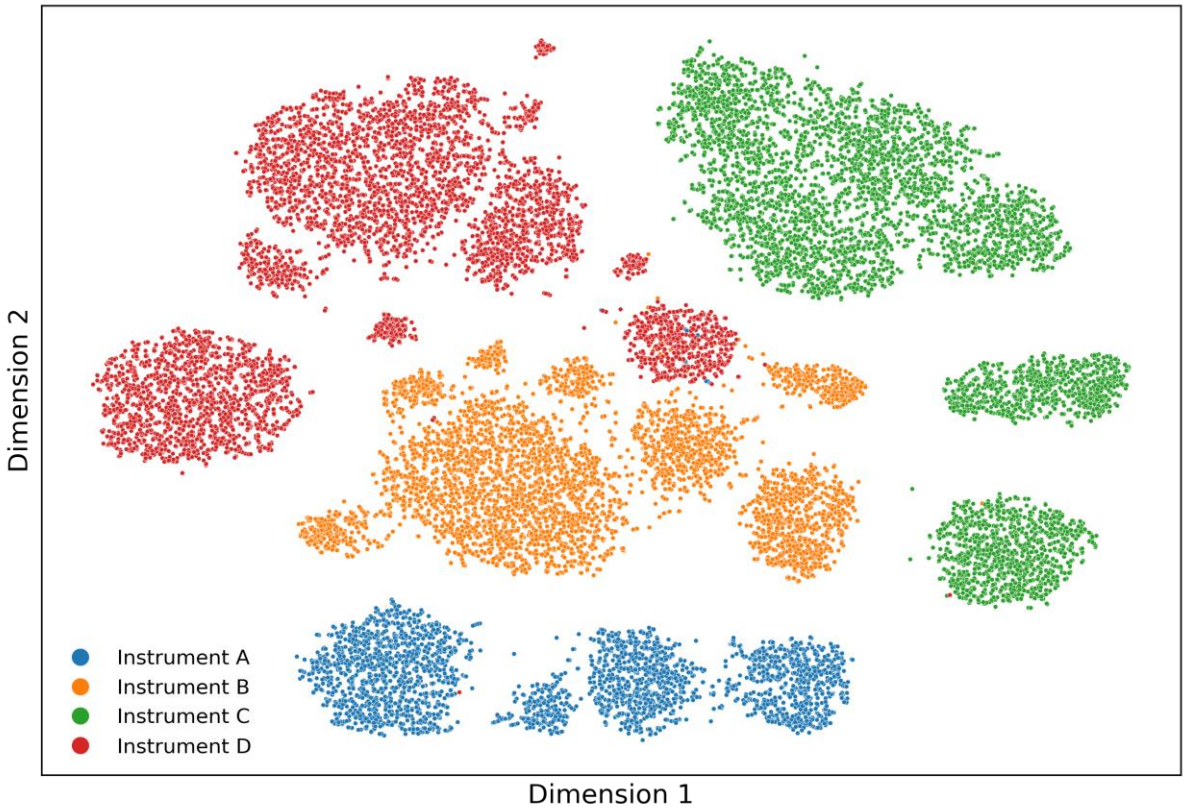
Qlip
Quality assurance in dairy

Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- Identify episodes with instabilities
- **Find patterns in spectral residuals**

Collected > 20.000 spectra from 38 episodes with instabilities identified across four instruments within a period of three years (> 5.9 million spectra)

PCA + *t*-distributed stochastic neighbor embedding of spectral residuals



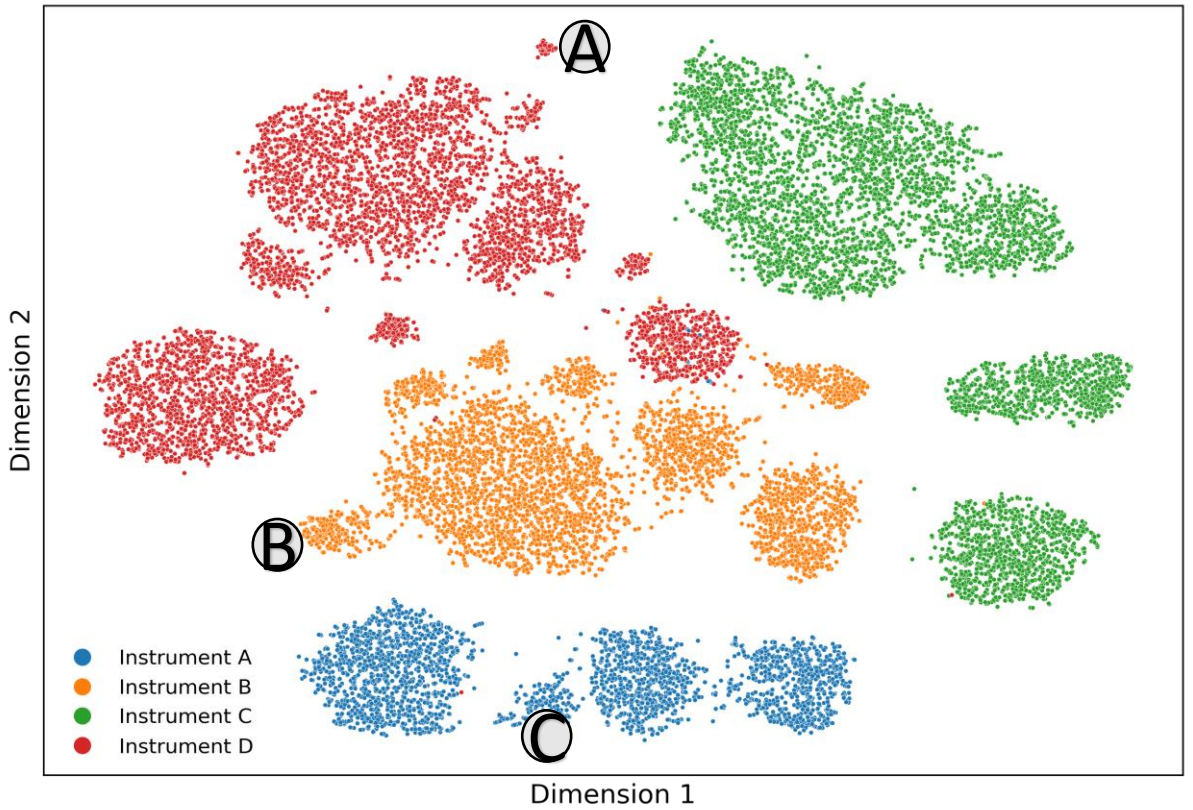
Methods & Results

- Compute anomaly scores for each spectrum from the same instrument
- Identify episodes with instabilities
- **Find patterns in spectral residuals**

Collected > 20.000 spectra from 38 episodes with instabilities identified across four instruments within a period of three years (> 5.9 million spectra)

PCA + *t*-distributed stochastic neighbor embedding of spectral residuals

Many instabilities uniquely occur for short period Ⓐ Ⓑ Ⓒ



Methods & Results

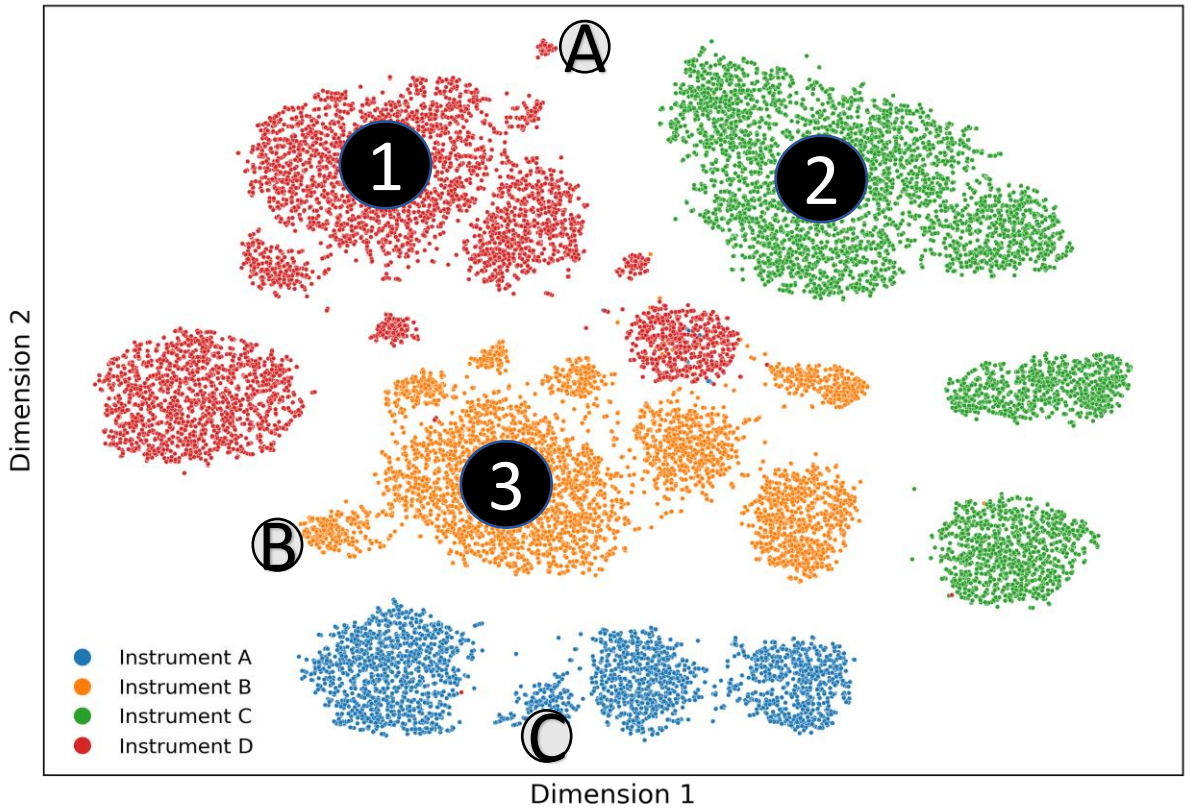
- Compute anomaly scores for each spectrum from the same instrument
- Identify episodes with instabilities
- **Find patterns in spectral residuals**

Collected > 20.000 spectra from 38 episodes with instabilities identified across four instruments within a period of three years (> 5.9 million spectra)

PCA + *t*-distributed stochastic neighbor embedding of spectral residuals

Many instabilities uniquely occur for short period Ⓐ Ⓑ Ⓒ

Some instabilities appear more often ❶ ❷ ❸



Summary

- Our approach can be used as tool for:
  - Monitoring the validity of our model predictions
  - Monitoring the status of our instruments
- Does not rely on expensive chemical reference analyses or control milk samples.
  - Instead, it is based on the spectra that are obtained by routine IR analysis of milk samples
- Can be developed with the tools that already exist

- Can be developed and used by every laboratory

Qlip
Quality assurance in dairy

Thank you for your attention!

Qlip
Quality assurance in dairy