Data digging for better calibration lines in NIRS

Maarten Scholtes-Timmerman
Trouw Nutrition Global NIR Team

Danish Chemometrics Society conference dsk.2020
November 5, 2020

trouw nutrition
a Nutreco company

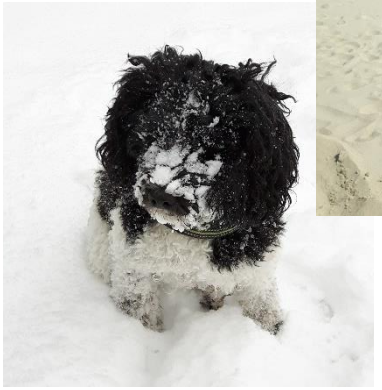
Who am I ?

Name: Maarten Scholtes-Timmerman

Function: NIR Application Specialist

Background

- Joined Trouw Nutrition in October, 2018
- MSc. in analytical chemistry, specialized in spectroscopy applications
- Strong focus on combining applications with (multivariate) data science



- Hobbies: digital photography, cooking, DIY
- Lives in Giessenburg (NLD)
- Married to Wanda, father to Tjibbe (11 years) and Veerle (10 years)

trouw nutrition
a Nutreco company

Trouw Nutrition

- Founded in 1931
- Member of the SHV Family Holding since 2015
- Feed premixes, minerals, additives, animal health products

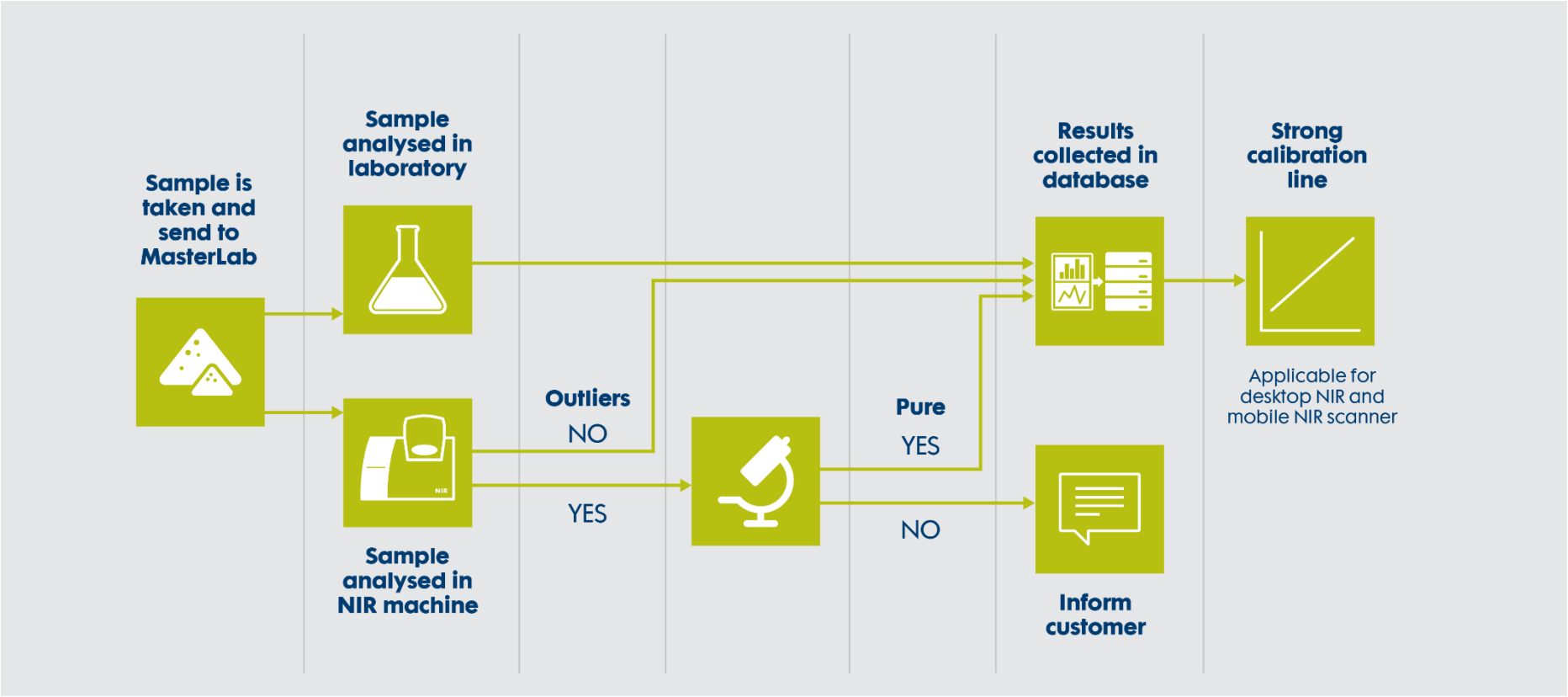
Global NIR Team

- Part of MasterLab
- Vital for Trouw Nutrition's Quality Control
- Manages network with >400 NIR instruments connected worldwide
- Calibration lines for 100's of products (raw and finished)

trouw nutrition
a Nutreco company

Trouw Nutrition Global NIR Team – modeling

- Multivariate models (calibrations) based on (modified) PLS
- Models predict nutritional values of animal feed ingredients or mixes, using lab references
- Frequent updates to incorporate new products, new harvest, new variants



trouw nutrition
a Nutreco company

Scenario

trouw nutrition
a Nutreco company

Problem situation

- Customer producing compound feed
- 7 FOSS DS2500 instruments on different plants

- The customer reports an issue
  - Differences found in Protein values of a product (Sunflowermeal), measured on two of their machines
    - Instrument “A”, performs as expected
    - Instrument “B”, gives ~3.5% absolute protein too low (compared to A)
  - Only in this product: other products are fine

What would you do?

trouw nutrition
a Nutreco company

What would you do?

Blame it on the customer

Blame it on sampling

Blame it on the model

Ignore the problem and hope it subsides

Add customer data to model to add robustness

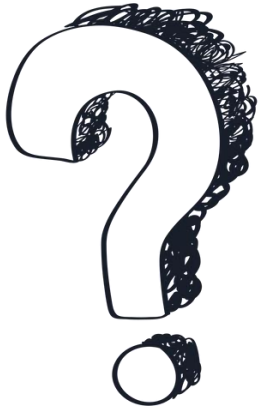
Do data digging to find the root cause

trouw nutrition
a Nutreco company

Prior findings

- The customers uses the same calibration line on all machines
  (no version conflict, all instrument updates are run)
- Sampling effect is ruled out by measuring the *exact same samples* on A and B

So, what is happening here?



trouw nutrition
a Nutreco company

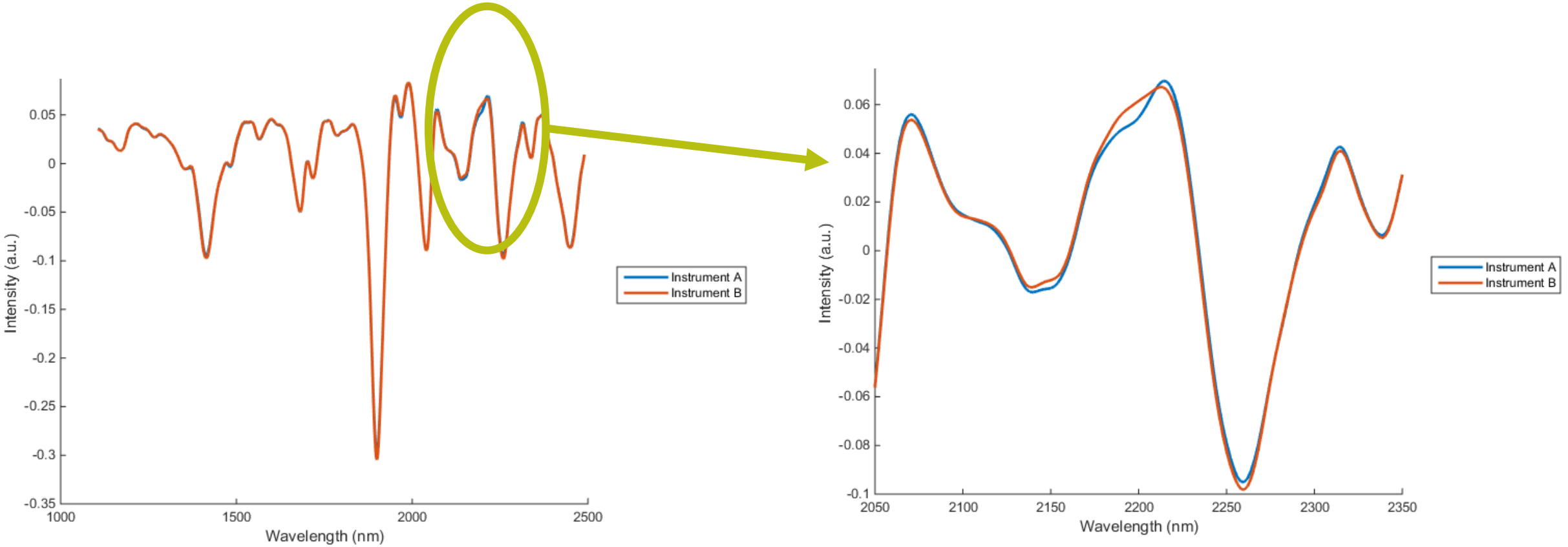
Data digging

trouw nutrition
a Nutreco company

Spectral analysis

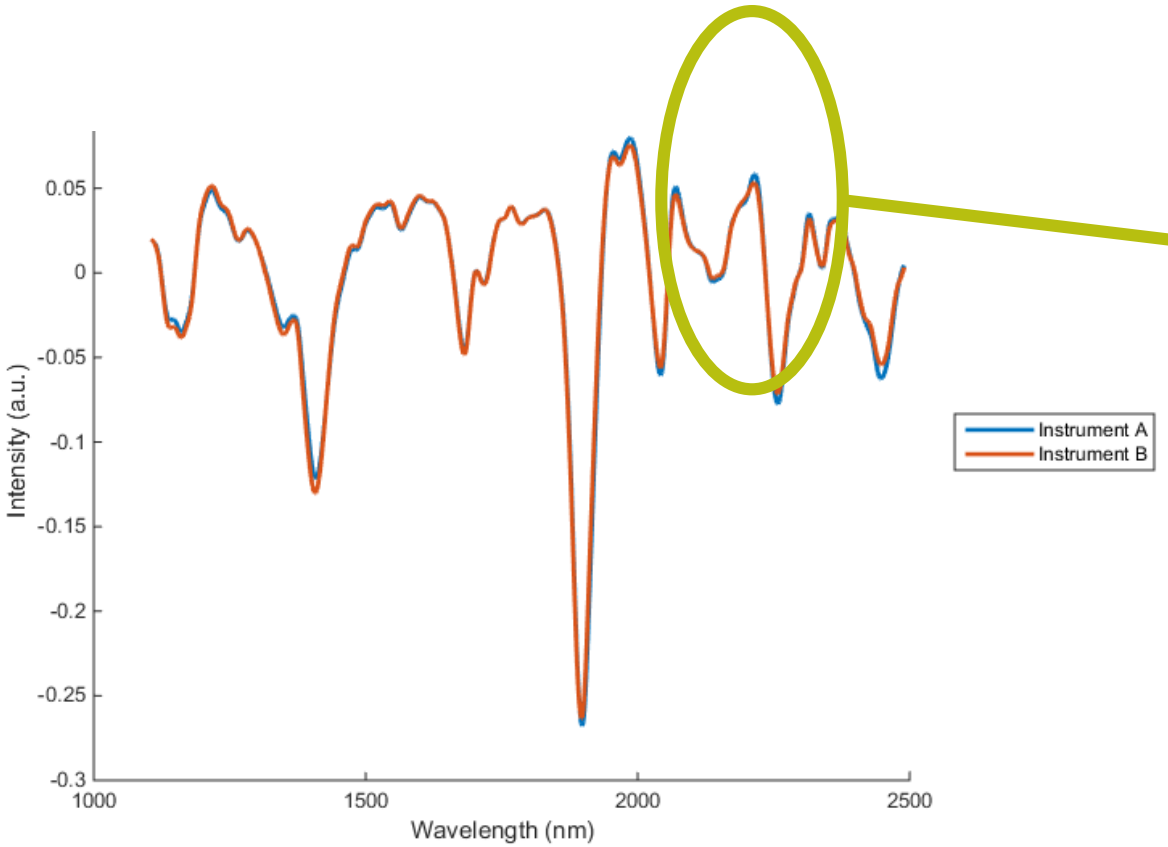
- View the differences in NIR spectra, of the same sample on instruments “A” and “B”
  - Affected product: Sunflowermeal
  - Unaffected products
    - Soyabeanmeal
    - Rapeseedmeal
    - Wheat

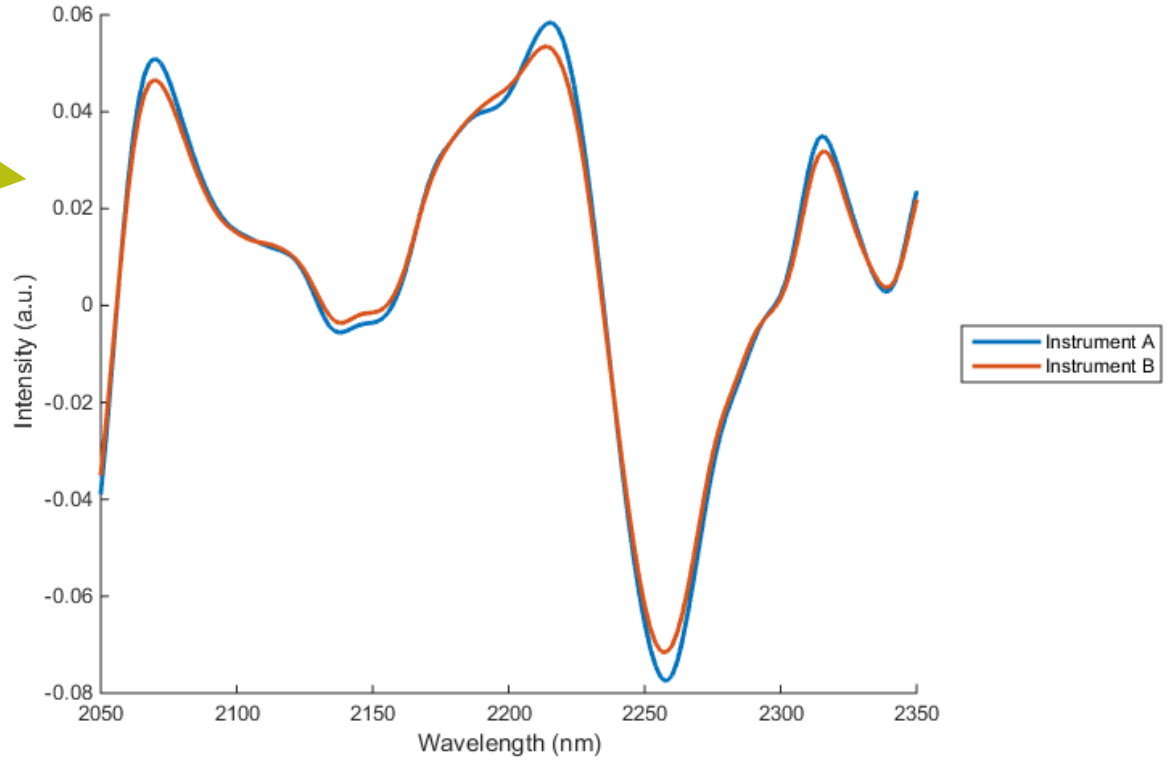
Spectral analysis: Sunflowermeal



trouw nutrition
a Nutreco company

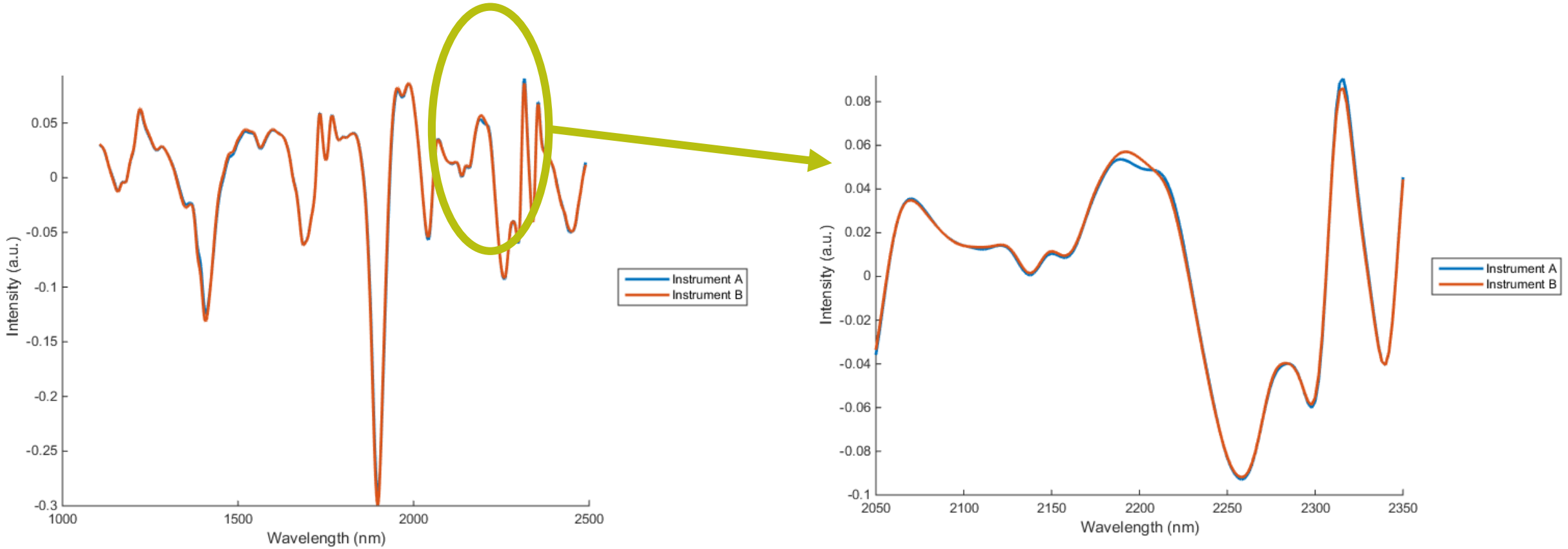
Spectral analysis: Soyabeanmeal





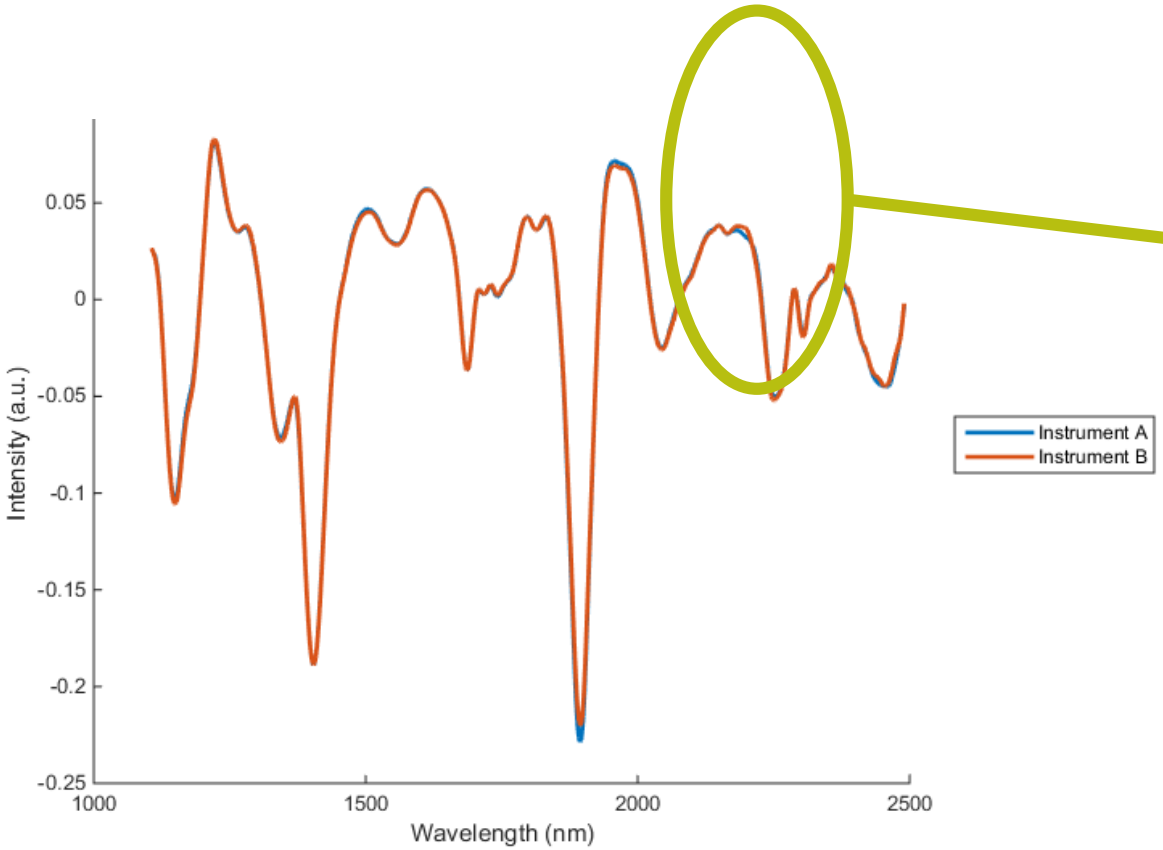
trouw nutrition
a Nutreco company

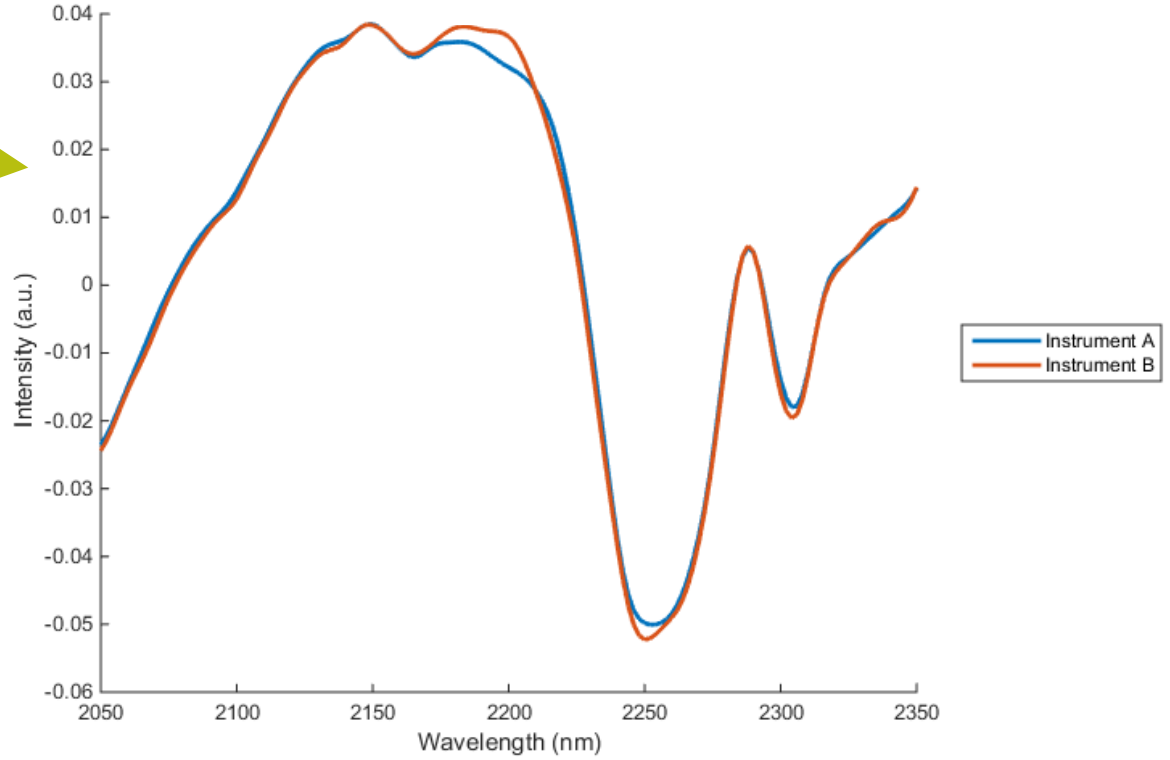
Spectral analysis: Rapeseedmeal



trouw nutrition
a Nutreco company

Spectral analysis: Wheat





trouw nutrition
a Nutreco company

Spectral analysis

Spectral analysis revealed

- A consistent spectral artefact is present
- Artefact is seen in all products
- Some systematic effect

… but why are only PLS predictions in 1 product affected?

→→→ understanding of how PLS models create a prediction

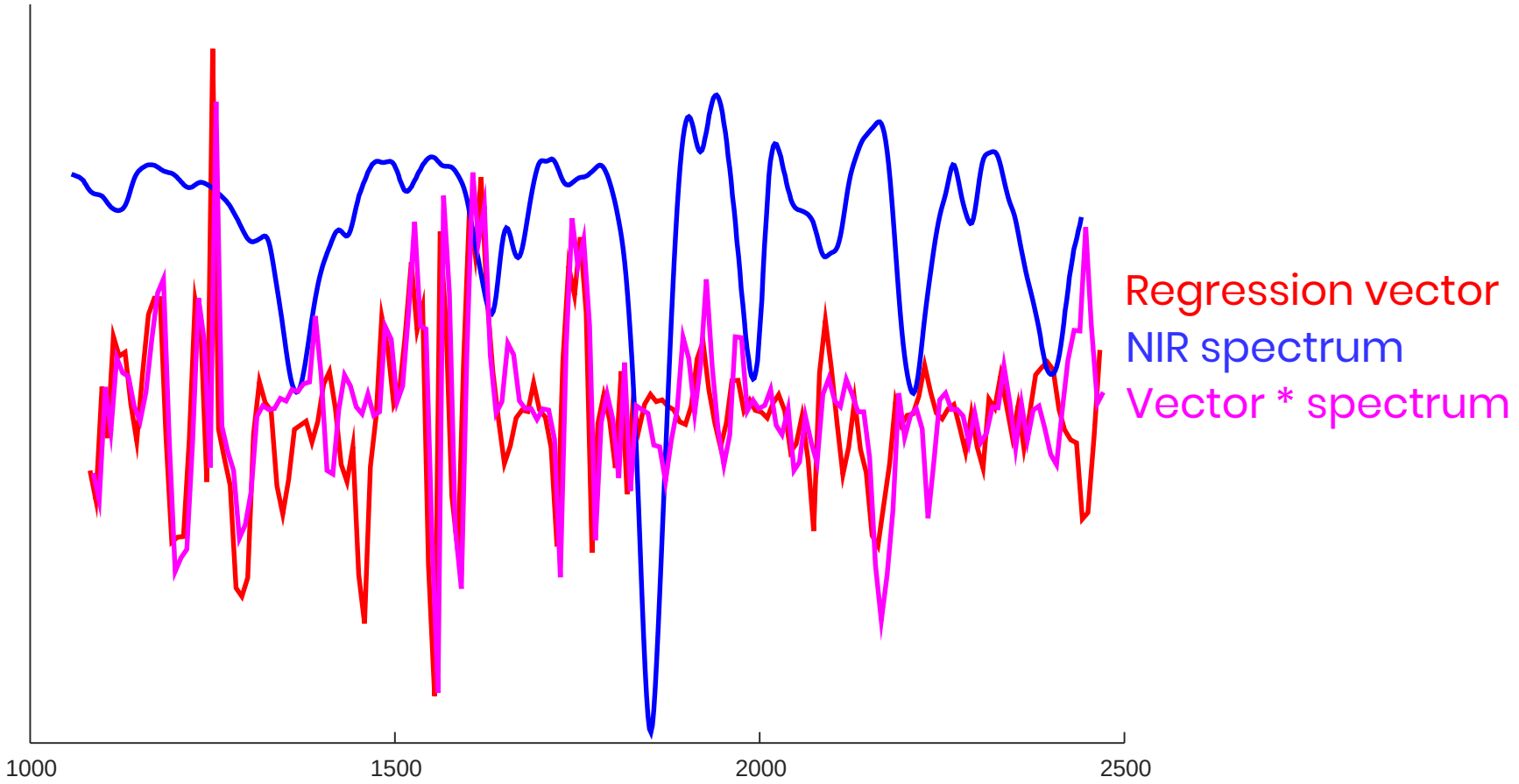
How does the PLS model calculate a prediction?

$$\widehat{Y} = \beta_0 + \sum_{i=1}^{n} \beta_i \times S_i$$

- $\widehat{Y}$ predicted value (here, % protein in the sample)
- β_0 model offset value
- β_i regression vector factor at channel i
- S_i spectral value at channel i
- n number of data points

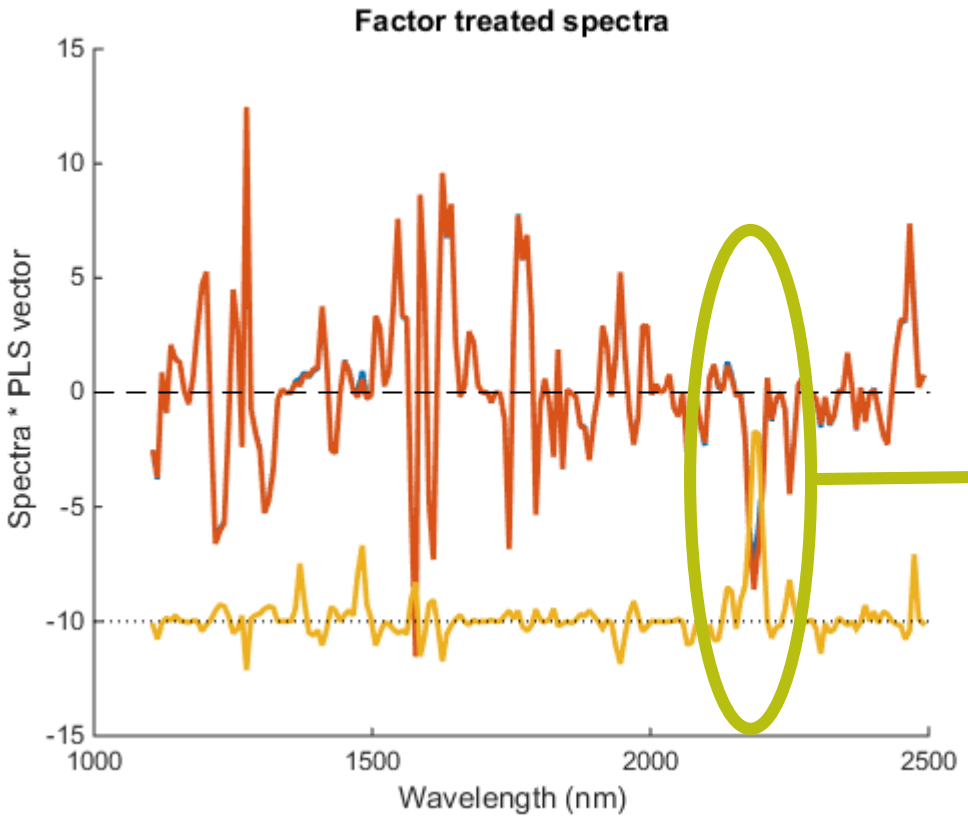
trouw nutrition
a Nutreco company

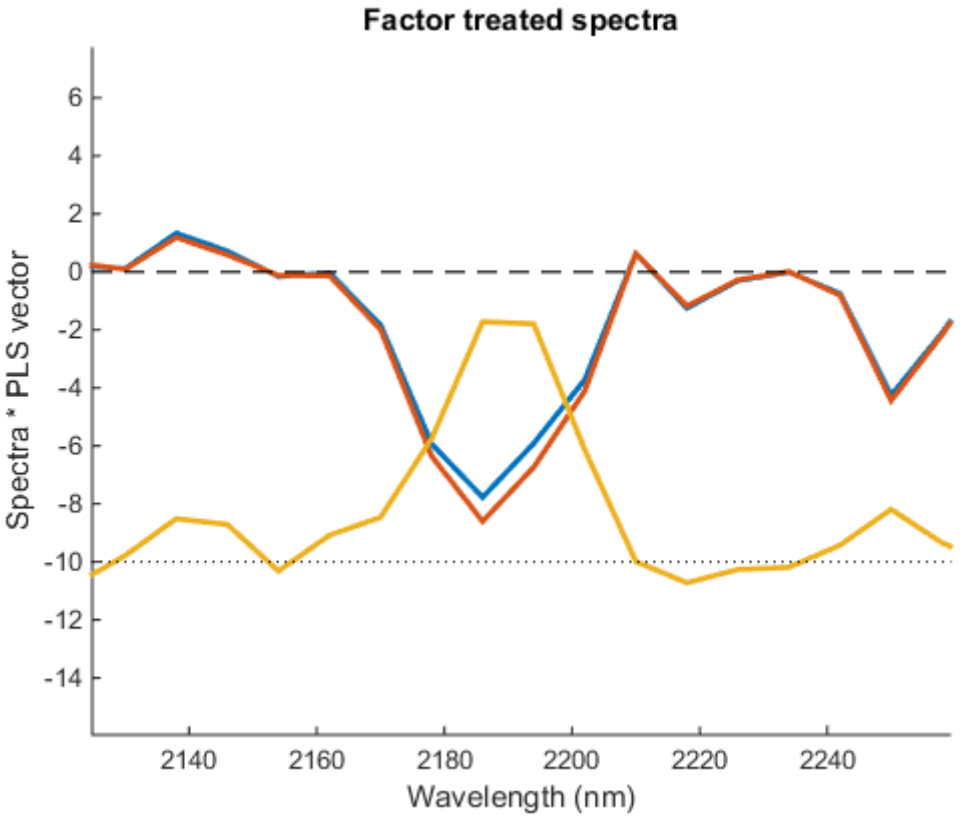
How does the PLS model calculate a prediction?



trouw nutrition
a Nutreco company

How does the PLS model calculate a prediction?

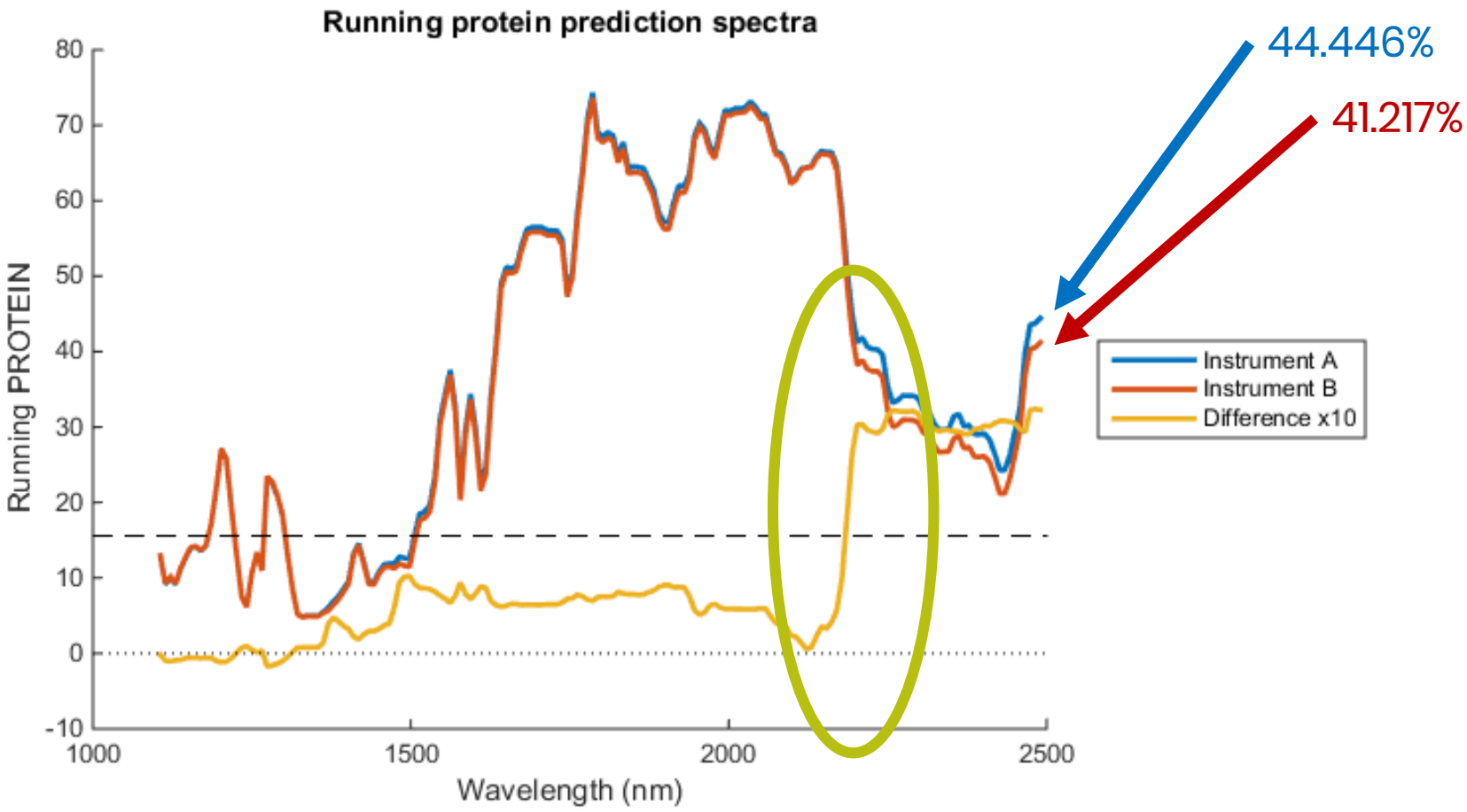




Sum from L to R: “Running protein”

trouw nutrition
a Nutreco company

How does the PLS model calculate a prediction?



44.446%

41.217%

PROTEIN [%]
44,4
41,2

trouw nutrition
a Nutreco company

Conclusion

- Only 5 channels are “responsible” for this large deviation in NIR prediction
- **This is only for this product as the regression vector of other tested products is insignificant in these channels**

- The NIR absorption in this region is different between instruments

- The customer pointed at instrument B to deviate, based on results, …
  but actually instrument A was deviating!
  Overall results lower than expected

- … what would have happened if we blindly added spectra of B to the model?

trouw nutrition
a Nutreco company

Take home message

- When using (m)PLS, keep in mind what the model does
- Always check *why* a user expects a result.
  In the shown example, the user's expectations made us chase the 'good' instrument first!
- Do not blindly update calibration lines to avoid adding 'wrong' data

Data digging can be laborious but worth your time!

trouw nutrition
a Nutreco company

Thank you for listening

trouw nutrition
a Nutreco company