Application Study

Application: Analysis of milk and creation of a model for the prediction of quality parameters in milk

Carried Out By:

Steinbeis Transfer Center for
Innovative Process Analytics (IPAT), Reutlingen

Steinbeis-Transferzentrum
Innovative Prozessanalytik
(IPAT)

NOVA
INDUSTRIAL ANALYTICS

+1 (518) 768-4479 info@novaitx.com

Analysis of milk and creation of a model for the prediction of quality parameters in milk

1. Objective and procedure

The aim of the study was to model the milk composition, i.e. the fat and protein concentration, using NIR spectroscopy. For this purpose, firstly, a PerkinElmer FT-NIR laboratory spectrometer (800-2500 nm) was used as a well-established reference to model Partial Least Square Regression (PLS-R) with cross-validation and, secondly, to evaluate the performance of a Phoenix inline spectrometer (NOVA Industrial Analytics), working within two Wave Length Ranges (WLR), namely, 1750-2150 nm (WLR I, in reflection modus) and 930-1132 nm (WLR II, in transmission modus), used to model the PLS-R with cross- and external validation. Table 1 shows the fat and protein concentration of milk for raw and pasteurized samples. The same two sets were used twice to prepare samples for external validated models. Thus, the total number of samples was 28 (raw/pasteurized-14/14). All milk samples were measured sequentially in three steps: first with FT-NIR, then with Phoenix (WLR I, then WLR II) spectrometer, using 0.2 mm thick cuvette at room temperature (20°C) in all cases. Before the measurement, up to 10 ml of the milk sample were injected through the cuvette with a syringe in order to eliminate residues from the previous sample and to achieve constant composition of the new sample. For each spectrometric measurement, the cuvette was refilled. In this way, 84 measurements were carried out, each with 5 spectra and a total of 420 spectra.

Table 1: Raw and pasteurized milk samples with different calibrated fat and protein concentrations.

Raw milk			
Nr.	Sample	Fat	Protein
1	RO-24-104	2.05	4.71
2	RO-22-094	2.74	4.38
3	RO-24-106	3.22	4.01
4	RO-22-098	3.58	3.6
5	RO-24-108	3.71	3.19
6	RO-22-096	4.15	3.37
7	RO-23-102	4.58	2.94

Pasteurized milk			
Nr.	Sample	Fat	Protein
1	PAM-22-053	0.1	3.5
2	REB-12-243	0.9	4.8
3	REB-12-241	1.5	3.4
4	REB-12-241	2.0	3.2
5	PAM-24-055	2.7	4.77
6	PAM-18-045	3.15	4.11
7	PAM-18-046	3.89	3.29

2. Creation of chemometric models with cross-validation and external validation

Figure 1 shows the spectra of raw and pasteurized milk. The colors of the spectra correspond to the protein concentration in raw milk, suggesting an inverse correlation – the higher the protein concentration regardless of fat content, the lower the overall absorption level. In contrast, the higher the fat content of pasteurized milk, the higher the dispersion or total absorption, regardless of the protein content. This could indicate that the pasteurization process completely changes the chemical or conformational structure of the fat-protein interaction, which in turn changes the degree of homogeneity and the size of the fat droplets. In pasteurized milk, the scattering effect is strongly shifted into the short-wave range. For these spectra, PLS-R models for fat and protein content were created with cross-validation (Figure 2).

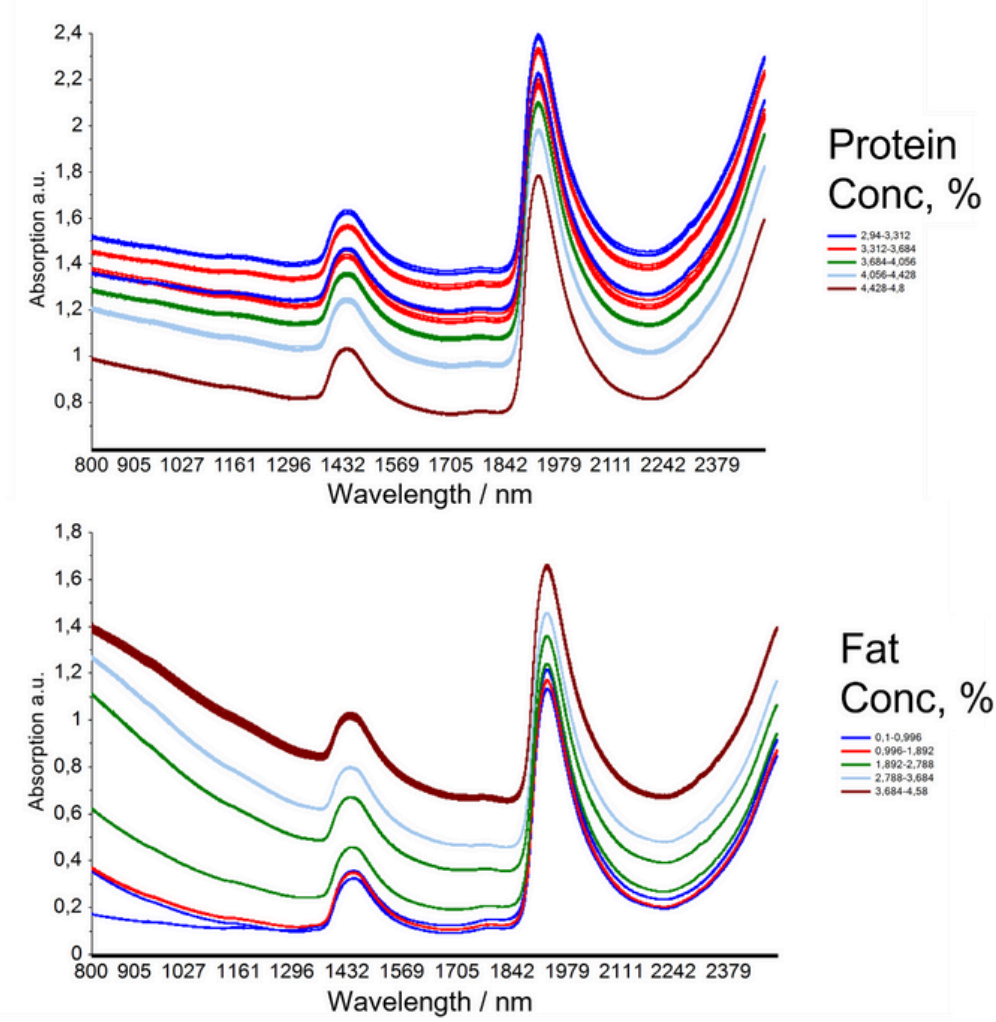

Figure 1: The NIR raw spectra (without pre-processing) of the calibrated raw and pasteurized milk samples (FT-NIR spectrometer, PerkinElmer).

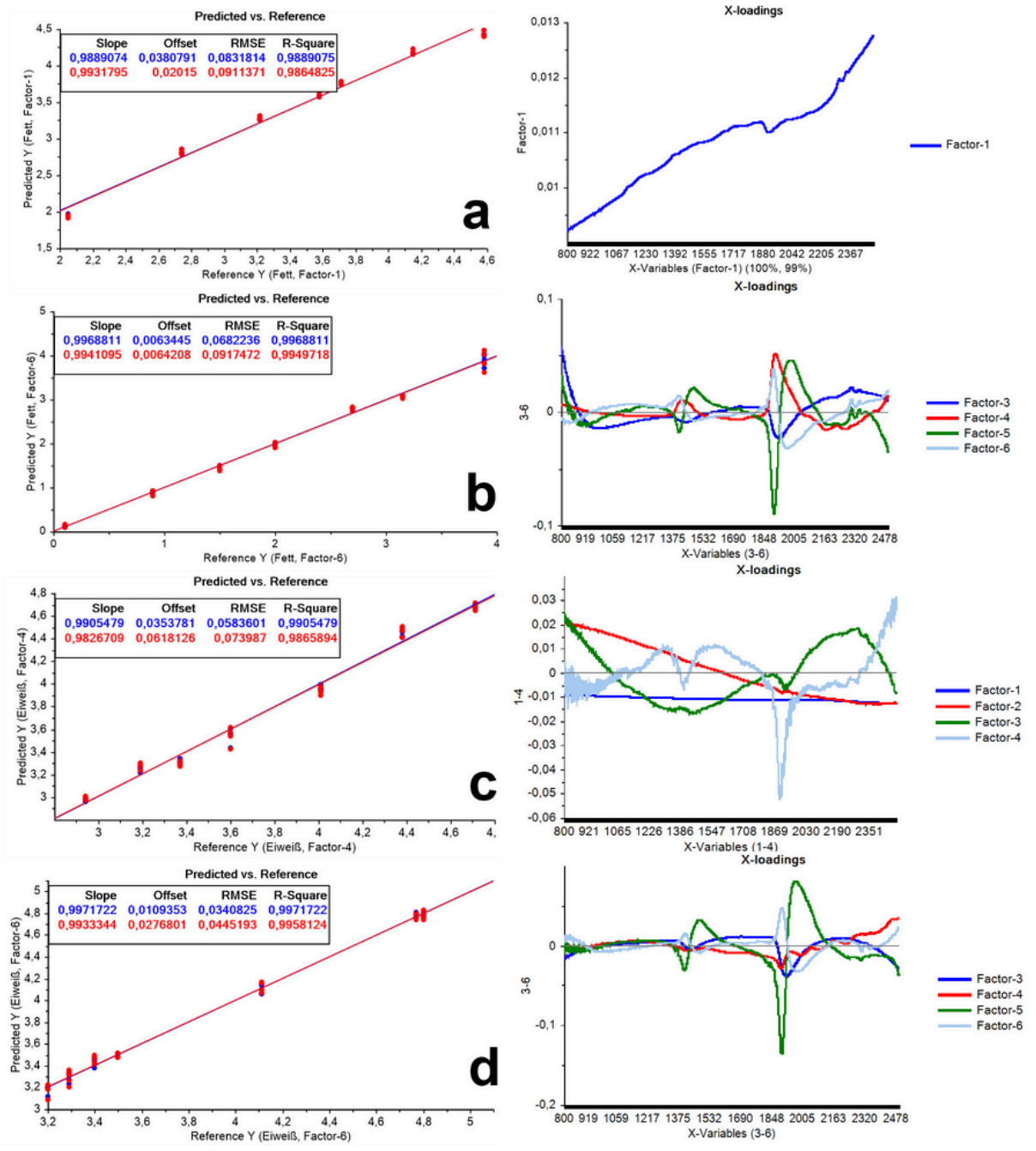

Figure 2: PLS models for the fat content of raw milk (a), pasteurized milk (c) and for the protein content of raw milk (b), pasteurized milk (d), acquired with PerkinElmer's FT-NIR spectrometer.

All models (Fig. 2) are of very good quality, with an error in the range of 1% relative and 0.04-0.09% absolute. However, with pasteurized milk, the prediction is more accurate. In this case, several factors must be taken into account in order to achieve the required accuracy of the model for pasteurized milk, while for the raw milk fat, as one factor, is sufficient. The first factor usually reflects the role of the dispersion scattering effect. The

other factors are responsible for the chemical and conformational composition of the dispersion shells of the fat droplets, which are also stabilized in the system by proteins. As mentioned above, pasteurized milk requires more factors, which can be explained by the increased complexity of the milk structure.

Figure 3 shows the spectra obtained in reflection mode with the Phoenix (WLR I) spectrometer. For reference and background, 4 mm thick Teflon was used. The spectra show the same results and dependencies as the spectra obtained for the corresponding wavelength range with PE FT-NIR spectrometer and are good reproducible.

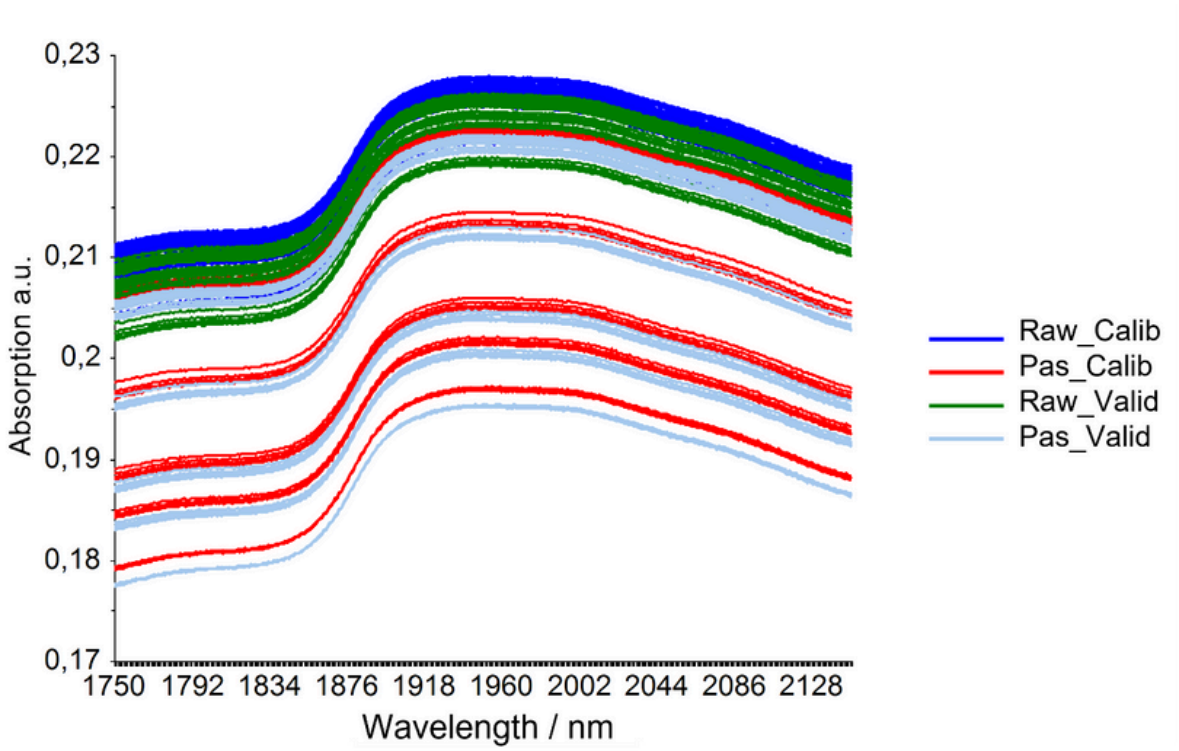

Figure 3: The NIR raw spectra (without pre-processing) of the calibrated raw (Raw) and pasteurized (Pas) milk samples (Phoenix, WLR I, reflection). Spectra are displayed for both calibration and validation.

Figure 4 shows the PLS-R models for raw milk. In cross-validation (**e** and **g**), unprocessed spectra give the best results, while in external validation (**f** and **h**), the spectra must be processed using the Standard Normal Variate (SNV) method. The effects associated with the dispersion are almost leveled out and the chemical information about the ingredients becomes the most relevant. Here, all models have good to very good quality, considering that the wavelength range is limited and the most information about scattering is missing.

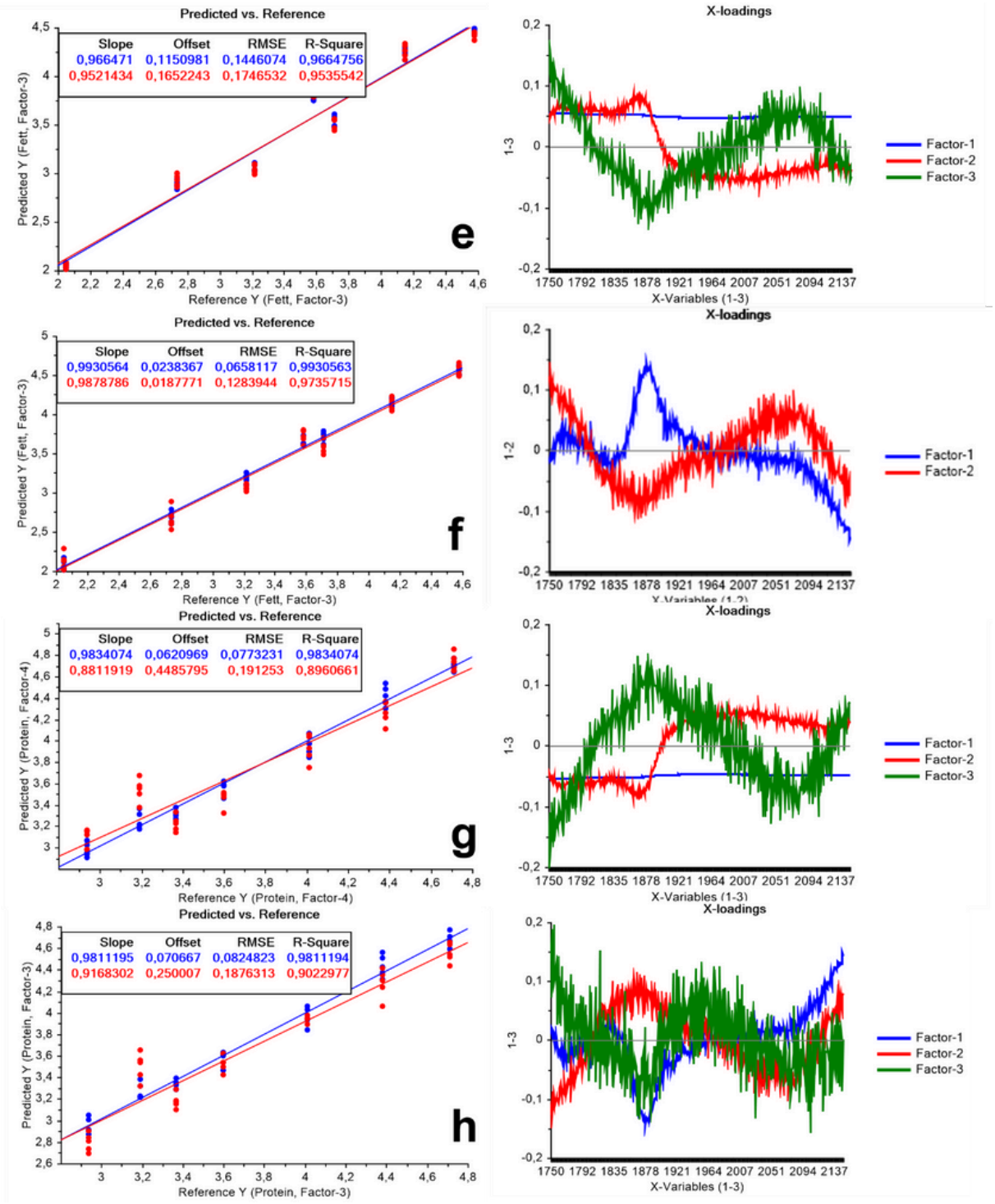

Figure 4: PLS models for the fat content (**e**, **f**) and protein content (**g**, **h**) of raw milk acquired with the Phoenix (WLR I) spectrometer. **e** and **g** – cross-validation, **f** and **h** – external validation.

To achieve the specified model accuracies, at least three factors must be considered. The minimal error is typical for determining fat content as compared to protein. External validation shows better results.

Figure 5 shows the models for fat and protein with cross-validation and external validation for pasteurized milk.

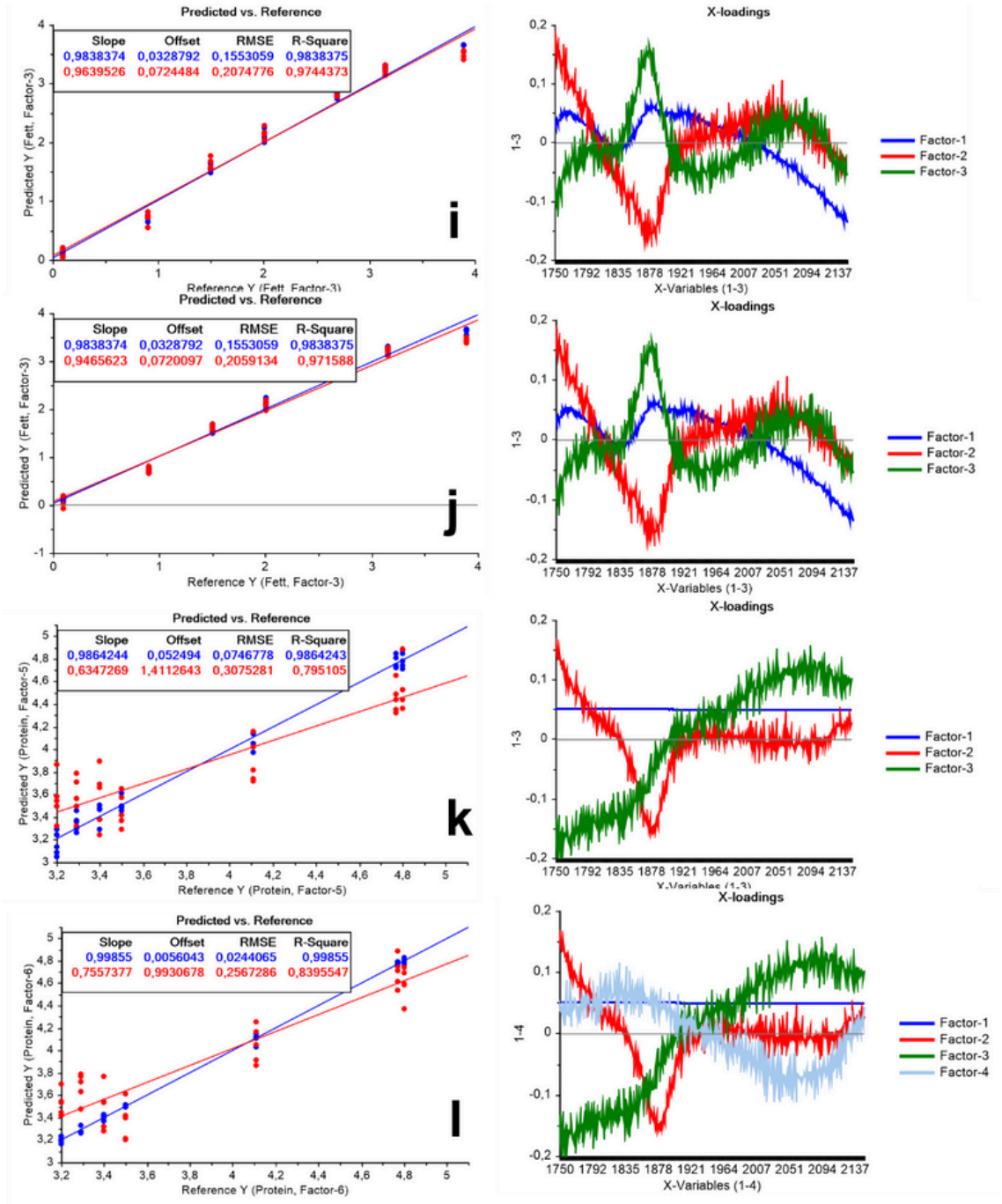

Figure 5: PLS models for fat content (**i**, **j**) and protein content (**k**, **l**) of pasteurized milk acquired with the the Phoenix (WLR I) spectrometer. **i** and **k** – cross-validation, **j** and **l** – external validation.

PLS models for fat have a high level of accuracy as compared to fresh milk, while for protein the error increases slightly being within acceptable limits. It should be noted, that for optimal results in determination of fat, the use of the SNV method is required, but not

for proteins, that confirms the results obtained with PE FT-NIR, according to which a complete structural reorganization of the scattering elements in the milk takes place.

Figure 6 shows the spectra obtained with the Phoenix (WLR II) spectrometer in the transmission mode. Air was used as a reference. The spectra show the same results and dependencies as the spectra obtained for the corresponding wavelength range with PE FT-NIR and are good reproducible. The finding that pasteurized milk scatters more strongly at short wavelengths is also confirmed. It should be noted, that for these wavelengths, only scattering information is characteristic, and chemical information is practically not present.

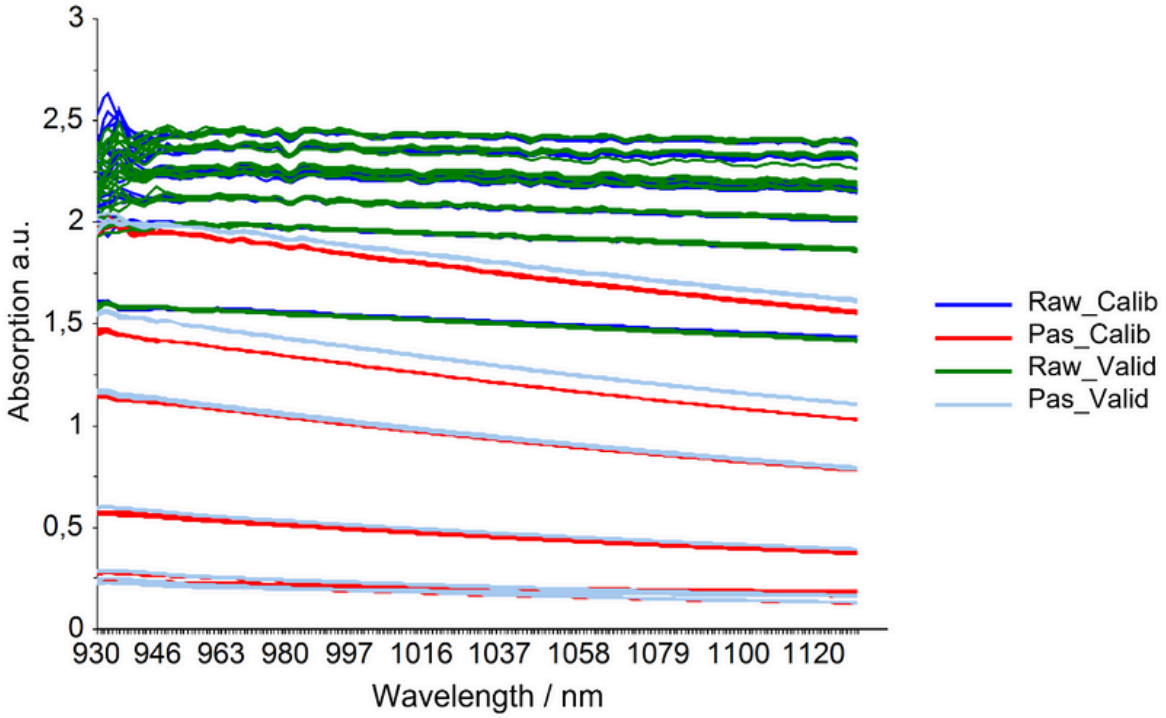

Figure 6: The raw NIR spectra (without pre-processing) of the calibrated raw (Raw) and pasteurized (Pas) milk samples (Phoenix WLR II, transmission). Spectra are displayed for both calibration and validation.

However, this information is sufficient to obtain high-quality PLS models for raw milk (Figure 7).

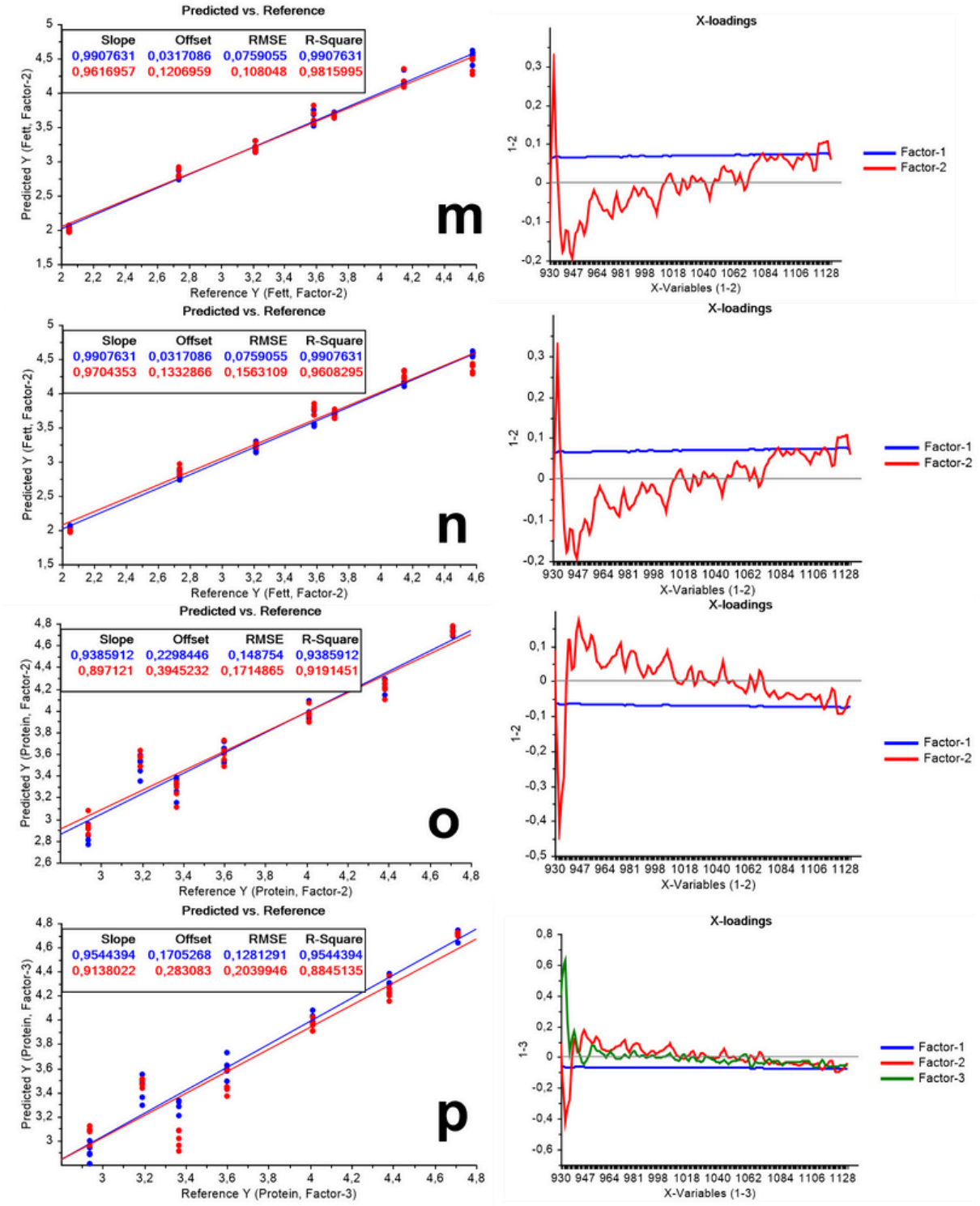

Figure 7: PLS models for the fat content (**m**, **n**) and protein content (**o**, **p**) of raw milk, acquired with Phoenix (WLR II) spectrometer. **m** and **o** – cross-validation, **n** and **p** – external validation.

Only protein (external validation) requires three factors, for the other three models two factors are sufficient. It should be noted, that all models use unprocessed spectra, which

also confirms that the latter contain almost only information on scattering, which is indirectly determined by the composition of milk.

Figure 8 shows models for pasteurized milk. Despite the good and acceptable quality of the models and the minimal number of factors, some information for this wavelength range is lost during pasteurization, which is consistent with previous conclusions. For example, for the external validation of a protein, it was not generally possible to obtain a model that had the required accuracy (cross-validation was only possible using SNV).

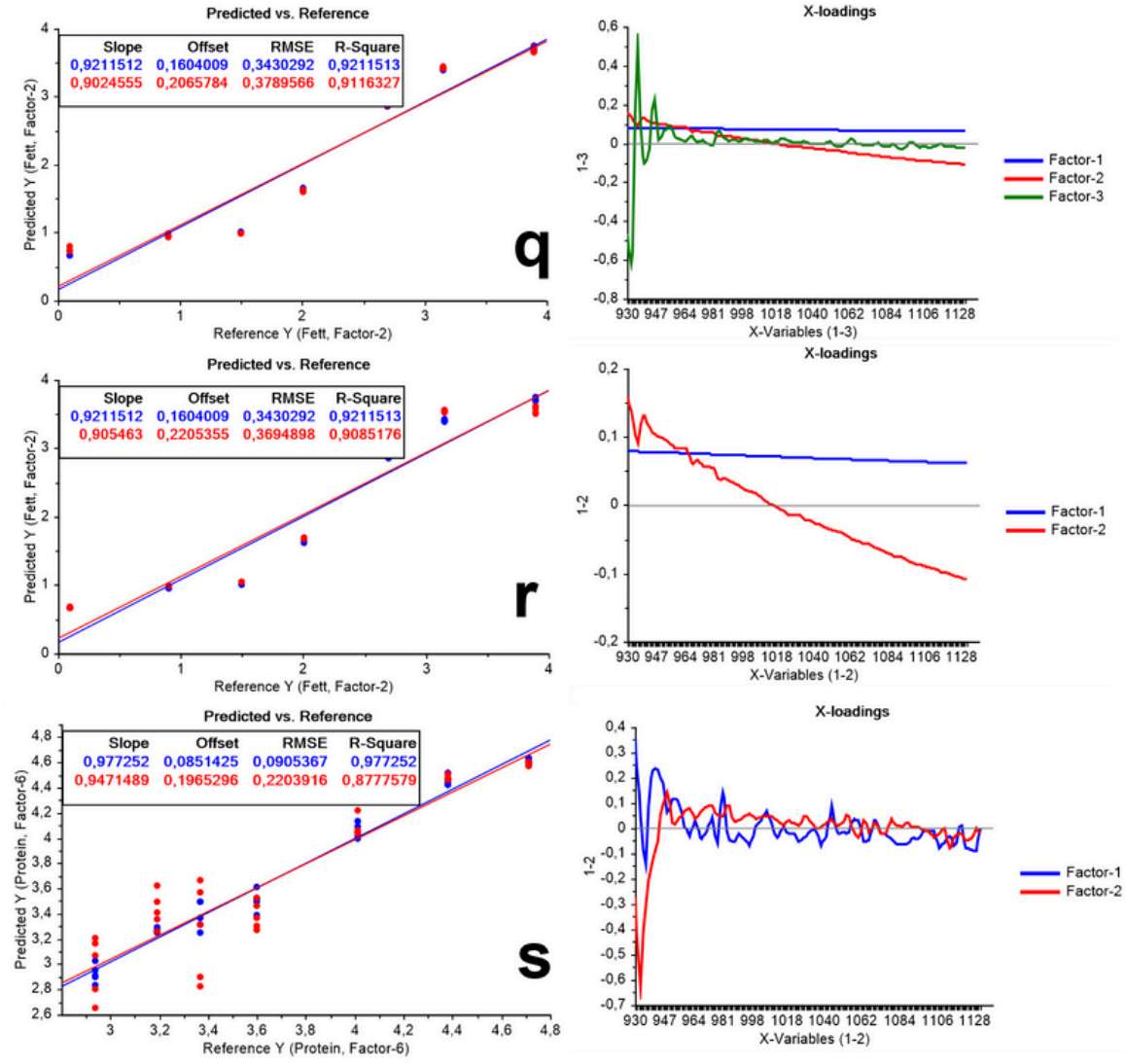

Figure 8: PLS models for the fat content (**q**, **r**) and protein content (**s**) of pasteurized milk acquired with the Phoenix (WLR II) spectrometer. **q** and **s** – cross-validation, **r** – external validation.

Table 2 shows the parameters of all the models obtained depending on the type and the wave length range of the spectrometer, the method of spectral preprocessing, the type of validation and the type of milk.

Table 2: Parameters of all PLS-R models obtained in the study.

Spectrometer	Milk type	Fat/Protein	Model number	Validation	Pre-processing	Number of factors	RMSE	R^2
PE FT-NIR	Raw	Fat	a	Cross	none	1	0.091	0.986
		Protein	c	Cross	none	4	0.074	0.987
	Pas	Fat	b	Cross	none	6	0.092	0.995
		Protein	d	Cross	SNV	6	0.045	0.996
Phoenix WLR I	Raw	Fat	e	Cross	none	3	0.175	0.954
			f	External	SNV	3	0.128	0.974
		Protein	g	Cross	none	4	0.191	0.896
			h	External	SNV	3	0.188	0.902
	Pas	Fat	i	Cross	SNV	3	0.207	0.974
			j	External	SNV	3	0.206	0.972
		Protein	k	Cross	none	5	0.308	0.795
			l	External	none	6	0.257	0.84
Phoenix WLR II	Raw	Fat	m	Cross	none	2	0.108	0.982
			n	External	none	2	0.156	0.961
		Protein	o	Cross	none	2	0.171	0.919
			p	External	none	3	0.204	0.885
	Pas	Fat	q	Cross	none	2	0.379	0.912
			r	External	none	2	0.369	0.909
		Protein	s	Cross	SNV	6	0.22	0.878
				External	n.a.	n.a.	n.a.	n.a.

Here, the quality of models is classified by color depending on the error and **R2**. Green means very good quality of the model, light green means good quality, yellow means acceptable quality, red means no acceptable model. The best possible reference quality of the fat and protein models was achieved for the PE FT-NIR, which corresponds to the widest wavelength range and measurement principle. As compared to reference chemometric models for fat and protein, very good results were achieved with the Phoenix spectrometer for both wave length ranges, covering both chemical and scattering information. In general, fat can be determined most accurately compared to protein by the Phoenix spectrometer (WLR I and II). With PE FT-NIR, both components are determined equally well. Raw milk is best suited for analysis with the selected process spectrometer.

Contact Nova Industrial Analytics
info@novaitx.com

North America: +1 (518) 768-4479
Europe: +49 7175 3844900