CLASSIFICATION AND ANALYSIS OF CHINESE AND GERMAN FLOUR SAMPLES: HANDHELD NEAR-INFRARED SPECTROSCOPY IN COMBINATION WITH CHEMOMETRIC DATA EVALUATION

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

in der Fakultät für Chemie

der Universität Duisburg-Essen

vorgelegt von

Xinyu Chen

aus Zhenjiang (China)

Oktober 2022

DOI: 10.17185/duepublico/78271
URN: urn:nbn:de:hbz:465-20230420-094827-2

1. Gutachter: Prof. Heinz W. Siesler
2. Gutachter: Prof. Eckhart Hasselbrink

Tag der mündlichen Prüfung: 21.02.2023

Statement

The work described in this PhD thesis was carried out in the laboratories of the Department of Physical Chemistry, University of Duisburg-Essen (Campus Essen), Germany, and in the laboratories of the School of Food and Bioengineering, Jiangsu University, China, between July 2019 and July 2022. Except the determination of the reference values of the flour calibration samples, this work is entirely that of the author.

Acknowledgments

I would especially like to thank Prof. Heinz W. Siesler (Department of Physical Chemistry, University of Duisburg-Essen) for his supervision and unequivocal support of my experimental work and the interpretation of the data and formulation of the final text in numerous discussions.

I thank Prof. Eckart Hasselbrink for taking over the task as second reviewer.

For experimental, instrumental and theoretical support I would like to thank Dr. Wei Hansen (Rotterdam, The Netherlands), Prof. Yan Hui (School of Biotechnology, Jiangsu University of Science and Technology, China), Prof. Xiong Zhixin (School of Light Industry and Food, Nanjing Forestry University, China), Dipl-Chem. Frank Pfeifer (Department of Physical Chemistry, University of Duisburg-Essen, Germany), Dr. Marina de Gea Neves (Department of Physical Chemistry, University of Duisburg-Essen, Germany), M. Eng Tian Jing (Qinhuangdao, Hebei Province, China), Dr. Chen Tong (College of Biological and Chemical Engineering, Guangxi University of Science and Technology, China) and IAS Company (Intelligent Analysis Service, Wuxi, China) for their spectrometer assistance and financial support.

除此之外，还需要感谢我的父母，我的外公外婆，以及我的祖国，因为有了你们的无私支持，我才能够顺利的完成二十三载的求学之路。

我不知道自己何德何能，在你们还没有踏上欧洲这片土地的时候，我已经可以如愿的为自己的将来做出一次勇敢的探索。谢谢你们！

Abstract

Classification and Analysis of Chinese and German Flour Samples: Handheld Near-Infrared Spectroscopy in Combination with Chemometric Data Evaluation

1. Several major reference values of wheat flour were modeled using the near-infrared (NIR) spectra of flour samples and the multivariate partial least squares (PLS) method. The calibration models for crude protein, moisture and wet gluten of flour were developed, and the root mean square errors of prediction (RMSEPs) were 0.3784% (w/w), 0.2624% (w/w) and 1.4653 % (w/w), respectively. The correlation coefficients squared for prediction (R_p^2) were 0.7922 for the best crude protein model, 0.6096 for the best moisture model, and 0.8346 for the best wet gluten model. The results showed that the analysis of the main parameters of wheat flour by benchtop and handheld NIR spectroscopy is feasible and good prediction models can be obtained.

   Furthermore, the differences in origin of the investigated flour samples (German flour and Chinese flour) can be discriminated by the analysis of their NIR spectra. This discrimination is based on the fact, that the NIR region contains overtone and combination absorption bands of CH, NH, and C=O functionalities which reveal the chemical differences of wheat flour samples of different geographical origin. By using principal component analysis (PCA) German and Chinese flour samples could be well classified by their NIR spectra. However, by using the partial least squares discriminant analysis (PLS-DA) method, the assignment of flour samples to a specific geographical origin can be significantly and effectively improved.
2. In this PhD project, five NIR spectrometers (two benchtop and three handheld systems) are used to measure the NIR spectra of a total of 50 German and 163 Chinese flour samples. The signal-to-noise ratio, spectral resolution, and accuracy of absorbance values of handheld spectrometers are generally lower than those of benchtop spectrometers.

   According to the near-infrared spectral characteristics of the samples, a PCA model was established, in which IAS 3100 and MicroNIR (VIAVI) achieved 100% correct

classification of German and Chinese flour. Furthermore, PLS calibration models for protein, moisture and wet gluten were also established for samples from both countries. Experimental results demonstrate that accurate calibration can be achieved using a benchtop spectrometer with better instrument performance and a handheld spectrometer with more flexible measurement operations. However, calibration models developed using spectra measured with a benchtop spectrometer outperformed those acquired with a handheld spectrometer.

3. In order to demonstrate the transfer of calibration models based on the NIR spectra of different instruments, model transfer methods are discussed in this thesis. For this purpose one spectrometer was defined as master and the others were defined as target instruments and the effects of three spectral standardization methods, direct standardization (DS), piecewise direct standardization (PDS) and simple linear regression direct standardization (SLRDS) algorithms, regarding the sharing of calibration models across instruments were investigated. Applying the three methods, the variability of the spectral data between instruments was significantly reduced and the prediction accuracy of the calibration models for wheat flour parameters was improved.

Keywords: Determination of wheat flour parameters in Chinese and German flour samples; handheld NIR spectroscopy; application of qualitative and quantitative chemometric data evaluation methods; spectral standardization and model transfer.

Contents

1 Introduction

1.1 Background and Significance

Wheat is a cereal plant, which is widely grown all over the world and is one of the three most important cereal crops produced and consumed worldwide. It is an essential food crop for human survival, reproduction, and daily living, with more than half of the world's population eating it [1]. For example, China was one of the first countries in the world to cultivate wheat, and its wheat production in 2018 totaled 256.7 billion kg, making it a vital component of agricultural production. Wheat accounts for a substantial component of the Chinese grain system, with the main production area in the north and a large proportion of all grain crops cultivated. Wheat is in high demand in China, which outnumbers all other countries in terms of production and planted area [2]. As a result, maintaining a balance between wheat supply and demand, as well as a fair structure of wheat quality for the international market and food production is critical for all countries [1]. Wheat can be used to make flour, bread, cookies, noodles, and buns, as well as fermented items like beer, alcohol, and spirits (e.g. vodka or biofuel).

The focus of this study is the rapid detection of several common quality indicators of wheat flour using handheld near-infrared (NIR) spectrometers. The application of handheld NIR spectroscopy equipment has been launched for the investigation of wheat flour in order to explore the enhancement of quality consistency of wheat flour, because of its non-destructive, rapid, and convenient detection approach in comparison to wet chemical analysis. It is hoped, that the results of this study will contribute to the quality assurance of wheat flour in processing, transportation, and sale. Furthermore, it is hoped to have an impact on quality control of raw materials in general and the advancement of handheld NIR spectroscopy instrument technology.

1.2 Flour

Flour is the powder made from peeling and grinding wheat. Water, protein, carbohydrate, fat, minerals, vitamins, and enzymes make up the majority of its composition [3]. Flour quality mainly refers to the edible quality, processing quality, nutritional quality and storage quality of flour, and all these qualities are closely related to the basic components of flour, such as moisture content, ash content, protein content and gluten content [4]. Due to the differences in wheat varieties and milling processes, the proportions of each component in various brands of flour produced in different countries for different purposes are also different. The main parameters, that reflect the nutritional and processing quality of flour are moisture, protein, ash, starch, wet gluten, dry gluten, and gluten index [3, 5]. The moisture level of flour has an impact on its water absorption and, as a result, the freshness of flour products. The water absorption of flour and the tensile strength of dough, as well as the edible flavor of flour products, are affected by the crude protein level of flour. The ash component of flour shows its mineral content and is the most important component influencing the sensory evaluation of flour products. The amount of starch in flour has an impact on its shine, viscosity, elasticity, and soft palatability. The protein concentration of wet gluten is closely connected to the protein content of dry gluten; the gluten index shows the quality of protein [6]. As a result, the physical and chemical indexes of flour can be used to infer that flour produced by flour mills meets the highest possible edible and processing quality parameters for residents.

Protein, carbohydrates, vitamins and minerals as well as trace elements such as vitamin A and vitamin C, are all abundant in milled wheat flour [7]. As a result, flour is often used as a carrier for iron, zinc, and fortification of citizens' micronutrient intake. High-gluten flour, in particular, is high in protein and carbohydrates and has been shown to balance potassium and salt, reduce edema, promote immunity, lower blood pressure, control fat metabolism, supply dietary fiber, conserve protein, detoxify, and improve intestinal function [8, 9]. In terms of the distinctions between German and Chinese flours, in addition to the basic variables that lead to major disparities, such as the variety of wheat farmed, the region, and the environment, the

two countries have separate flour classifications. In Germany, flour is classified into species depending on mineral concentration, whereas in China, it is defined depending on its protein content.

1.2.1 German Flour Characteristics

Wheat (including the subspecies spelt, two-grain wheat and single-grain wheat), rye, oats, barley, millet, corn, and rice are used to make flour in Germany. Soft wheat, Durum wheat, Spelt, and Rye flours (bread grains) are appropriate for making "European-style bread," which is baked in a steam oven to create water vapor and achieve a crispy crust and soft interior [10].

Mineral content (or how many milligrams of minerals are included in 100 grams of flour), is used to classify the different kinds of flour in Germany. Flour TYPE 405, for example, contains 405 mg of minerals per 100 g of flour. As a result, the higher the number, the higher the mineral content of the flour and the darker it is. On the contrary, if the mineral content is relatively low, the color is brighter. The Tab. 1.1 depicts the main classification types of flour in Germany, and the DIN standard 10355 has been in use in Germany since 1992 to represent milled products made from common wheat, rye, and spelt [11,12]. To determine the mineral content of flour, a small amount of flour is burnt in a muffle furnace at 900 ℃ to ascertain the type. The remaining (non-combustible) components roughly correspond to the mineral content of the flour. They have also been called "flour ash" in the past [12]. Whole grain foods (flour, semolina) are classified according to DIN 10355 without specifying their content, and thus no type of number is provided. Whole grain products always contain young seedlings. Baking flour differs from whole wheat flour in that it no longer contains any seedlings [12] (Tab. 1.1).

Tab. 1.1 Flour Types according to DIN 10355 [11]

Designation	Types	Baking characteristics	Minimum mineral content (% (w/w) in dry matter)	Maximum mineral content (% (w/w) in dry matter)
		Wheat		
Wheat flour WM	405	preferred household flour, good baking properties	——	0.50
Wheat flour WM	550	Powerful for fine-pored doughs and can be used as a multi-purpose flour	0.51	0.63
Wheat flour WM	812	for light mixed breads	0.64	0.90
Wheat flour WM	1050	for mixed breads or baked goods in the household	0.91	1.20
Wheat flour WM	1600	for dark mixed breads	1.21	1.80
Durum wheat flour DM	1600	Durum wheat flour	1.55	1.85
Wheat baked meal WBS	1700	without a seedlings	——	2.10
		Spelt (DinkelMehl)		
Spelt flour DM	630	In terms of baking technology, it can be used in a similar way to wheat flour type 550	——	0.70
Spelt flour DM	812	——	0.71	0.90
Spelt flour DM	1050	——	0.91	1.20
		Rye (RoggenMehl)		
Rye flour RM	815	only rarely used, mostly in southern Germany, for light rye breads	——	0.90
Rye flour RM	997	for mixed breads, distributed differently from region to region	0.91	1.10
Rye flour RM	1150	for mixed breads, distributed differently from region to region	1.11	1.30
Rye flour RM	1370	typical "bakery flour" for Rye and mixed rye breads	1.31	1.60
Rye flour RM	1740	typical "bakery flour" for Rye and mixed rye breads	1.61	1.80
Baked rye meal RBS	1800	without seedlings	——	2.20

1.2.1.1 Types of Flour with low Mineral Content

TYPES 405/480/400 are the most popular flour species with a protein concentration of 9.0-9.82% (w/w) and good viscosity for doughs and batters, as well as the basic German cake baking flour, for muffins, cakes, cookies, pies, toast, doughnuts, and other baked goods [13].

TYPE 550 is a multi-purpose flour that contains 11.0-11.5% (w/w) protein and is a medium gluten flour with a higher viscosity. It is mostly used for home baking and it is an excellent bread baking flour. It can also be used to make crusty white bread (similar in texture to French baguette). Bleached or unbleached all-purpose flour is a blend of soft and durum wheat (bleached flour is whiter than unbleached flour with less vitamin E) and has gluten composition of 9.0-11.0% (w/w). Multipurpose flour is mostly used in home baking since it is the most adaptable flour available; it may be used to bake bread as if it were a special bread flour, but not for cakes or pastry [13].

1.2.1.2 Types of Flour with high Mineral Content

TYPE 812 is a firm, frosted flour for mixed bread (Mischbrot) with increased softness. It is ideal for baking bread with a consistent form. The flour is derived from durum wheat and has 11.0-13.0% (w/w) gluten. When rubbed between the fingers, it has a light yellow color and a slightly gritty texture. This is the best flour on the market. Its high gluten content improves the bread's structure and makes it easier to keep its shape according to bread-making requirements [14].

TYPE 1050 is a durum wheat flour with a high gluten content that is used to make whole wheat bread (Vollkornbrot). It can be combined with other flours to increase the gluten content of the flour, with the proper proportions blended according to the needs. High gluten flour is created from white durum wheat flour that has at least 13.0-14.5% (w/w) gluten and can be mixed with flour from other grains to add structure. It can be used to make bread and pizza and the resultant bread has a high degree of elasticity and [13]. High gluten flours and gluten-free flours can be used to modify the gluten content of other flours by admixture.

TYPE 1060 is a whole-wheat flour with a brown hue comparable to rye, which is more nutritious and fiber-rich than white flour and can be used to make bran or black bread. The core kernel of whole wheat, which includes germ and bran, yields TYPE 1600 whole wheat flour, which has a brown hue similar to rye. It has more minerals, fiber, and fat than white flour and is more appetizing [14]. When baking bread, whole wheat flour is used because it has a high texture and a nutty flavor, and because whole wheat flour bread is heavier and more solid, it is frequently blended with white flour.

Rye Flour (RoggenMehl TYPE 1150) is the same as medium rye flour, which has more vitamin B and E than wheat and flour. Rye flour is the most often used flour in bread production, and it has a somewhat sour flavor. Rye flour bread has a longer shelf life and a better taste. It is, however, frequently combined with other flours due to its low gluten content [15]. There are two subcategories in the classification of rye flour. In southern Germany, TYPE 815 is primarily converted rye flour, and it is less commonly used than light baking [16]. Wheat and rye of TYPE 997 are grown in different parts of Germany [17].

Spelt white flour (Dinkel Mehl TYPE 630), also known as white flour in Germany, is a finely ground flour, which is frequently substituted for flour TYPE 405 [18]. This is an excellent flour for making bread, but it is usually combined with other flours due to its high gluten content. It can be quickly kneaded into gluten, the finished product is slightly dry, and the dough has a relatively low moisture content. TYPE 812 and TYPE 1050 spelt flours are also equal to all-purpose wheat flour and have the same impact as TYPE 630.

1.2.2 Characteristics of Chinese Flour

China's wheat cultivation stretches across the country; the main regional divisions are southwest wheat region, middle and lower regions are the Yangtze River wheat region, the Yellow Huaihai wheat region and northwest wheat region (including northeast spring wheat, northern spring wheat, northwest spring wheat, Xinjiang winter and spring wheat, Qinghai-Tibet spring and winter wheat, northern winter wheat, Yellow Huaihai winter wheat, middle and lower areas of the Yangtze River winter wheat, southwest winter wheat, southern

China winter wheat, a total of 10 sub-regions). Yellow Huaihai wheat region includes Henan, Shandong, Hebei, northern Jiangsu, northern Anhui, Shaanxi Provinces. The middle and lower reaches of the Yangtze River wheat region including Sichuan, Hubei, southern Henan and Anhui, Jiangsu's riverine areas. The southwest wheat region includes Yunnan, Guizhou, Sichuan and Chongqing Provinces [22].

In China the protein level is used to classify flour. It can be divided into several groups based on the classification criteria used. According to the protein content of flour, flour can be classified into three categories: high gluten flour with 10.5%-13.5% (w/w) protein content, used for bread making), medium gluten flour with 8.0%-10.5% (w/w) protein content, used for noodles and snacks), and low gluten flour with 6.5%-8.5% (w/w) protein content, used for snacks and dishes [19]. Based on protein and mineral content, flour can be classified as prime, first, or second grade, with prime flour containing 7.2% (w/w) and 0.32% (w/w) protein and ash, respectively, first grade flour containing 12.7% (w/w) and 0.43% (w/w) protein and ash, respectively, and second grade flour containing 13.5% (w/w) and 0.54% (w/w) protein and ash, respectively [20]. According to processing precision and use, flour may be split into two categories: grade flour and speciality flour, and grade flour can be divided into three categories: special flour, standard flour, and common flour [21]. Speciality flours, such as bread flour and cake flour, are manufactured by milling particular types of wheat or by admixing edible whitening agents, edible bulking agents, edible flavors, and other substances on the basis of the flour grade for the intended use [21, 22].

1.3 Current Status of NIR Spectroscopy Research

Compared with the current standard measurement methods, NIR spectroscopy has the characteristics of high efficiency, speed and convenience. Thus, in the food field NIR spectroscopy is gradually replacing the traditional methods for the determination of physical and chemical indicators of food as a “rising star” in the history of rapid food quality testing, which has greatly contributed to the rapid and vigorous development of the food industry and is already developing into a very mature technology [23]. In 1987 Williams and Norris

proposed the NIR technique for the analysis of wheat breeding, cereals, fruits and oilseeds with reference to data analysis, instrumentation, commercial applications and factors affecting spectral analysis [24]. NIR spectroscopy is currently being used to determine the quantity of significant ingredients in cereals, fruits, seafood, meat products, tobacco, tea, and pasta products [25-32]. The development of research on NIR approaches in cereals and flour will be discussed in depth in the following sections 1.3.1 and 1.3.2.

1.3.1 Current Status of Research on Cereals

NIR spectroscopy was first used to determine cereal quality indicators, and it was a huge success. For example, NIR spectroscopy detection of wheat began relatively late and was hampered by the slow development of spectroscopic technology. In most countries around the world, wheat quality assessment is based on subjective guesses, and the judging method has no scientific basis. Chemical procedures for evaluating wheat are also extensively used in scientific research institutions and food safety testing, which are time-consuming, expensive, and necessitate professional instruments and testing staff. The chemical method usually uses the Kjeldahl method to detect the protein content of the sample [33] and obtains a high testing accuracy, but the long testing time required (about 2 days for proteins) and the large number of testing reagents make the cost of testing higher; furthermore, the chemical method can only do sampling and cannot test in real time. The drying method is commonly used to detect water content in samples [34], but the procedure is time-consuming and labor-intensive, and the results can alter depending on the sample's instability.

Yu [35] used NIR spectroscopy and Fourier-transform mid-infrared (FT-MIR) spectroscopy for the determination of crude protein and water in wheat using different spectral pretreatment methods and selection of characteristic wavelengths, and the results showed that the best NIR model for crude protein was developed in the 1400-2500 nm spectral range using 1st derivative and standard normal variate (SNV) pretreatment methods, showing excellent predictive performance (R^2=0.97); the best mid-infrared spectral model was developed using the 1750-1100 cm^{-1} region of SNV-corrected spectra, which yielded a good predictive model

(R^2=0.90). Regarding moisture determination, the best model obtained by NIR spectroscopy (1100-2500 nm based on the original spectrum) showed good predictive performance (R^2= 0.86), while the best model generated by the MIR technique using the full wavenumber range and a 1st derivative-SNV pretreatment spectrum gave only approximate quantitative predictions (R^2= 0.72).

The possibility of quantitative prediction of crude protein content of wheat seeds based on selected short-wave NIR spectral variables was demonstrated by Lu [36]. 52 samples of wheat seeds were collected and their spectra were obtained in the 900-1700 nm wavelength range, which were then used to optimize the preprocessing method and to preferentially select protein-sensitive variables for wheat seeds to establish a PLS calibration model for protein prediction The results show that the combination of multiplicative scatter correction (MSC) and wavelet transform (WT) is the optimal preprocessing method for predicting crude protein content of wheat seeds; using the statistical results of 200 competitive adaptive reweighted sampling (CARS) variable preferences, 12 variables were selected as wheat seed protein sensitive variables, and the PLS was established by combining preprocessing optimization and variable preferences. R_p^2 and $RMSEP$ were 0.96 and 0.369 % (w/w), respectively.

Sun [37] developed a PLS prediction model for the determination of wheat protein content using NIR transmission spectroscopy, and the results showed that the model was able to predict wheat protein content more accurately, with prediction correlation coefficient, prediction mean square error and mean relative error of 0.98, 0.113% (w/w) and 1.973% (w/w), respectively. Barton [38] studied a total of 2203 wheat samples from various regions to establish a complete database and used it as a "global" calibration. Zheng [39] used short-wave near-infrared detection to obtain wheat proteins, and the calibration model was developed by PLS to obtain the best detection wavelength for wheat protein composition.

Since the 1960s, near-infrared spectroscopic (NIRS) techniques have been used for the quantitative analysis of some components in cereals [40], and the measurement range has been gradually widened since then. In 2006 Zhu [41] used a filter-based NIRS instrument for the determination of 7240 samples of 17 crops, including rice, wheat, corn, soybean, sorghum,

grain, barley, and edible beans. The results showed that the filter NIRS could obtain satisfactory results for the quantitative analysis of different crops and different nutritional quality traits. For example, Sohn [42] applied this technique to correctly determine the straight-chain starch content and protein content of rice during cooking and to predict its structure. Sinelli [43] applied NIR spectroscopy to study the optimum cooking time and recommended cooking time of rice and obtained good results. In addition, NIR spectroscopy can be used to detect the degree of adulteration, storage time and pest infestation of grains.

Villareal [44] collected 250 rice samples and tested them for amylose content (AC). AC selection in breeding programs was found to be sufficient using the NIR transmission spectrum of unground brown rice or milled rice. Milled rice calibration is easier to transfer to other units than brown rice calibration.

1.3.2 Current Status of Research on Flour

In the evaluation of flour quality, the solvent retention capacity (SRC) [45] and ultrasonic method have been developed rapidly in recent years. By establishing a correlation between the retention capacity of flour in flour solutions, namely water, sucrose, sodium bicarbonate, and lactic acid, and the physical and chemical indexes of flour or the traits of flour products, the SRC method can predict the physical and chemical indexes of flour or the traits of flour products, but the correlation between the two is poor. The method, however, cannot achieve accurate prediction of the physical and chemical indexes of flour and can only make a rough judgment of the nature of the finished flour products due to the poor correlation between the two. Thus, the method does not provide good guidance for the specific production process [45]; the ultrasonic technique can only provide a qualitative assessment of the flour category, not an exact forecast of the physical and chemical indexes of flour; therefore it is ineffective for guiding the manufacturing process of flour and flour-processing businesses [46].

NIR spectroscopy has been effectively employed in grain storage companies to detect several physical and chemical markers of wheat, the raw material for flour. NIR spectroscopy is now frequently also used in flour mills and flour processing plants [48]. For single indicators such

as flour moisture content, protein content, ash content, gluten index, wet gluten coefficient, and sedimentation value, NIR spectroscopic calibration models have been established [47]. These models can be used to quickly determine physical and chemical parameters of flour, providing a point of reference for evaluating flour's nutritional and processing quality. In addition, in 2003, Mendoza [49] attempted to use NIR spectroscopy to detect insects in flour, but the results showed that the accuracy of detection could not reach the level of 75 insects/50 g of flour as specified by the U.S. Food and Drug Administration (FDA) but could only distinguish between two types of flour with insect counts greater than and less than 130 insects / 50 g of flour.

Hrušková & Faměra [50] detected the moisture and protein content and sedimentation values in various commercial wheat samples by NIR spectroscopy, used modified partial least square (MPLS) and PLS methods to calculate the spectral properties of wheat, and successfully predicted the accuracy of the parameters of various commercial wheat samples almost the same. Statistically significant correlations ($P < 0.01$ for probability) between predicted and measured values of protein content and Zeleny sedimentation were determined in a variety of commercial flour with cross and independent validation.

Cocchi [51] tried to develop the wavelet interface to linear modelling analysis (WILMA) algorithm and used this algorithm to quantify the extent of adulteration of durum wheat flour with ordinary bread wheat flour by analysis of NIR spectra.

Jiang [52] developed a new portable NIR spectroscopy software. First he corrected the spectra by the standard normal variables (SNV) method, then used variable combined population analysis (VCPA) to optimize the wavelength variables of the SNV-corrected spectra, determined the characteristic wavelength variables that are highly correlated with fatty acid values, and finally used an extreme learning machine (ELM) to construct a detection model for fatty acid values.

Wesley [53] obtained samples directly from wheat breeders and compared the results of the curve fitting method with the more commonly used partial least squares curve fitting of the component protein spectra. The content of gliadin and glutenin was measured using

size-exclusion high performance liquid chromatography (SE-HPLC) and used to develop partial least squares calibration, with the results compared to curve fitting methods.

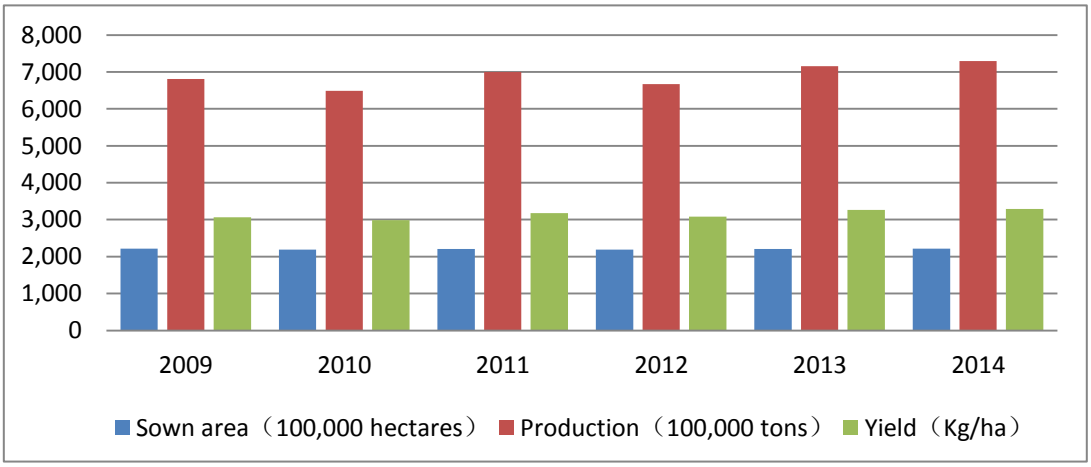
Yan [54] achieved a rapid detection of flour moisture by combining chemometric methods and NIR-based spectral detection technology, developing efficient mathematical models (PCA, PLS, and MPLS) and improving the model database.

2 Theory

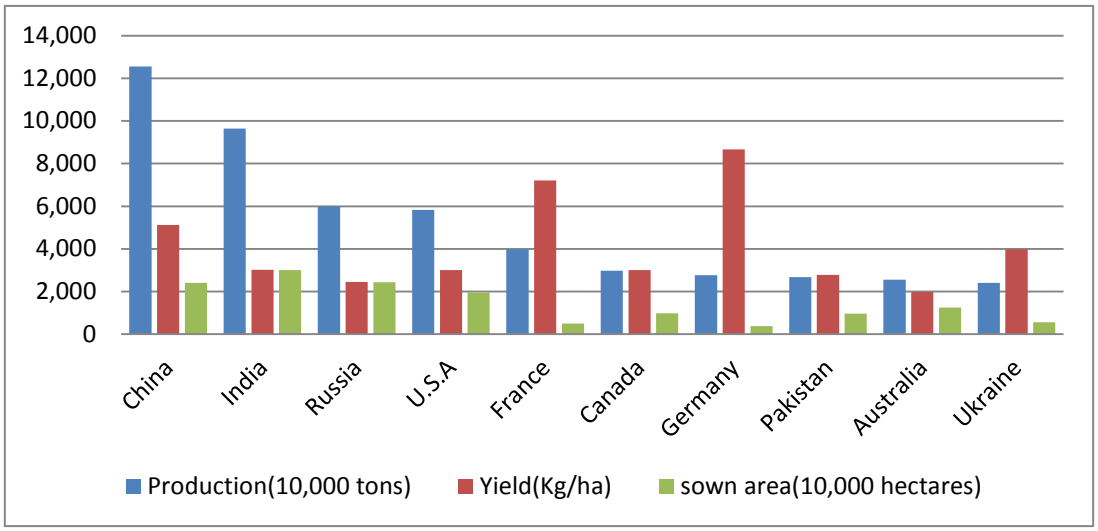
2.1 Overview of Flour

Wheat flour, as an easily digestible natural product, is rich in beneficial substances such as carbohydrates, proteins, fibers, and minerals, and it plays an important role in the daily dietary structure, being used in the preparation of a wide range of baked goods and pasta products [3]. According to the Food and Agriculture Organization of the United Nations (FAO), the world wheat sown area amounts to 221.62 million hectares in 2014, with a total production of 728.97 million tons and an average yield of 3289 kg/ha (Tab. 2.1). Some countries in Asia, Europe, and the Americas produce the majority of the world's wheat (Tab. 2.2). These three continents and their regions account for 89.82 % of total sown area and 92.89 % of total wheat production in the world. For example, China is the world's second largest wheat sown area but the world's first in wheat production [55].

Tab. 2.1 Global wheat sown area and production (2009-2014) [55]



Tab. 2.2 Wheat production in the top 10 (producing) wheat-growing countries worldwide in 2014 [55]



Because wheat grows in different geographical areas, the surface shape and structure of the wheat grain can vary. Wheat flour has a wide range of physical and chemical properties, including protein, wet gluten, moisture, ash, and sediment content, all of which have a direct impact on its application range.

2.1.1 Flour Composition and Quality Parameters

2.1.1.1 Physical Parameters of Flour

Color: The mill grinds wheat to make finely ground flour from its endosperm. The color of the flour varies due to the different skin color and grain quality of wheat [3]. White wheat flour is generally whiter than red wheat flour, and durum wheat flour is darker than soft wheat flour. This is due to the fact that it is impossible to avoid incorporating bran into the flour during the manufacturing process; the skin color of white wheat is less noticeable in the flour, whereas red bran mixed into the flour causes the flour to be maroon in color [56]. The endosperm of durum wheat is slightly creamy yellow, and the endosperm of silty wheat is white; the raw material contains too much lime soil or there are more mustard seeds. Without thorough cleaning and selection, the color of the flour is bluish gray or has extremely fine

black spots; if the rolling distance of the grinding roller is too tight, this will cause the grinding roller to be heated and the powder particles will be dark gray [57].

Coarseness and **fineness**: These physical properties are determined by the sieve specification. The particle size requirements are not consistent due to differences in flour quality and usage. The thickness of the flour reflects the precision with which wheat flour is processed. Wheat flour particles can be ground into flour when they must have a specific standard particle size. Flour is made up of three main components: endosperm pieces with a particle size of less than 15 μm, large starch granules with a particle size of 15-40 μm, and protein fragments with a particle size of less than 15 μm [53].

Water absorption: It refers to the amount of water added to the flour when it is made into dough. Because the quality of flour varies, so does the water content, and so does the water absorption. The water absorption of flour is closely related to its protein content; high protein content equals large water absorption; low protein content equals low water absorption. Water absorption is also linked to the amount of starch in flour. The water absorption rate of damaged starch is approximately 5 times that of intact starch (the water absorption rate of intact and damaged starch is 0.44 % (w/w) and 2.0 % (w/w), respectively); thus, the more damaged starch is in the flour, the higher the water absorption rate of the flour [58]. However, the water absorbed by damaged starch will seep from inside during dough fermentation, affecting the quality of flour products. As a result, in some countries, the maximum percentage of damaged starch is also listed as a flour quality indicator [59].

Sedimentation value: It is a comprehensive indicator of wheat quality and is measured by the sedimentation test. The sedimentation test was first proposed by L. Zeleny in 1947, and the basic method is that a certain amount of wheat flour absorbs water and swells when exposed to a weak acid medium, forming flocs that settle slowly. The sedimentation volume within a given time period is referred to as the sedimentation value, which is expressed in milliliters (mL) [60]. The settling speed and volume reflect the gluten content and quality; the higher the

measured value, the greater the gluten strength, indicating that the higher the gluten strength, the better the baking quality of wheat flour. Furthermore, because its determination method is relatively simple, it has been highly valued and widely used by breeders and grain chemistry researchers both at home and abroad, and it has been standardized [61, 62].

Numerous studies have found that the sedimentation value very significantly correlated with the quality of wheat flour food processing (baking, steaming, cooking, and so on), and that the sedimentation value has a high genetic capacity and is effective in breeding early generation selection [63-65]. Scholars in some European countries believe that combining sedimentation value and protein content is the best way to assess wheat quality, replacing gluten quantity analysis [66-68]. Seed hardness, protein content, and sedimentation value, according to Luo et al., are measurements that respond well to selection for quality when seed quantity is limited (e.g., in early breeding generations) [69]. In Germany, the sedimentation value has been identified as a crude measure of processing quality [70].

2.1.1.2 Chemical Composition Parameters of Flour

The chemical composition of wheat flour greatly influences its quality. Wheat flour's main components are protein (about 10-12 %) and starch (about 70-75 %), with polysaccharides (about 2-3 %) and lipids as minor components (about 2 %) [71]. In Tab. 2.3 the chemical composition ranges of wheat and flour are summarized.

Moisture: Moisture content of wheat flour ranges between 13.5% (w/w) and 14.0 % (w/w). When the moisture level exceeds 14.0% (w/w), storage under high temperature and high humidity conditions increases cellulose loss; when the moisture level exceeds 15.0% (w/w), mildew can reproduce; and when the moisture level reaches 17.0% (w/w), not only molds, but other bacteria can reproduce [72].

Tab. 2.3 Chemical composition of wheat flour (weight %)

Name \ Ingredients	Moisture	Starch and Sugar	Protein	Fat	Cellulose	Minerals
Wheat	13.84	68.74	9.42	1.47	4.43	2.07
High-precision flour	13.68	75.65	9.12	0.90	0.06	0.59
Ordinary flour	13.48	72.57	10.37	1.70	0.79	1.09
Bran	11.00	56.00	13.00	4.20	10.50	5.30

As the moisture content rises, so does the activity of various enzymes, resulting in the breakdown of nutrients and the generation of heat, as well as the proliferation of microorganisms and insects, which eventually leads to flour spoilage. Furthermore, for every 1% (w/w) increase (or decrease) in moisture in wheat flour, the dry matter content decreases (or increases) by about 0.86 % (w/w) [73]. Any consumer or manufacturer will suffer losses if the moisture content is too high or too low. As a result, the moisture content of wheat flour must be strictly controlled during the flour-making process.

Starch: Starch is a polysaccharide that is one of the main components of wheat flour. In the human body, starch is hydrolyzed to produce glucose, which is then absorbed and utilized. Starch is used by the human body at a rate of more than 90%. Starch provides approximately 70% of the energy required for various physiological activities and labor, making it one of the most cost-effective and important nutrients for human beings [74].

Gluten: Gluten is a substance that has extension and elasticity when wheat flour is mixed with water and formed into dough. Then, water is used to rinse off the starch, bran, and water-soluble substances in it, and it is finally rendered insoluble in water [75]. The chemical composition of gluten is represented in Tab. 2.4:

Tab. 2.4 Composition of gluten (weight %)

Chemical Composition	Protein			Sugar		Fat	Ash
	Gliadin	Glutenin	Albumin Globulin	Soluble Sugar	Starch		
	43.10	39.10	4.40	2.13	6.45	2.80	2.10

Gluten is important in food quality, because the volume and shape of bread and buns (the ratio

of height to bottom area diameter) are directly related to the quantity and quality of gluten. Wheat flour with a high gluten capacity absorbs more water, has a high elasticity, a high gas retention capacity, a large dough height after fermentation, a small bottom area diameter, a loose and uniform internal structure, and is suitable for baking bread and other foods [76]. If the gluten capacity of wheat flour is poor, with less dough absorption, weak elasticity, poor gas retention capacity, large dough dispersion area after fermentation, and a small height, this flour is suitable for baking cookies or pastries [77]. The amount of gluten content is determined by the quality of the wheat variety. In general, durum wheat has a high protein content as well as a high quality [76]. Wheat storage conditions also have an impact on gluten quality. Gluten quantity and quality will be reduced in sprouted, heated, frostbitten, insect-infested, or moldy wheat. In conclusion, gluten content is related to wheat variety and quality; thus, wheat gluten content serves as the foundation for matching processing and ensuring process quality.

Fat: Fat content of wheat ranges from 0.7 (w/w) to 1.9 % (w/w), with the majority of it being unsaturated fatty acids found primarily in the embryo and paste flour layers (Aleurone layer). The fat content of wheat flour varies after processing due to the different precision of wheat flour; in general, high-precision wheat flour contains less fat, while low-precision wheat flour contains slightly more fat [78]. If the fat content of wheat flour is low, it is usually necessary to increase the fat content when making food to increase the nutrient content and improve and enhance the taste. However, from the standpoint of storage and transportation, wheat flour, such as high fat content, will cause acidification in certain hot and humid conditions, resulting in wheat flour deterioration.

Cellulose: The cellulose in wheat flour is derived from finely ground wheat hulls and the paste flour layer scraped off the hulls during the flour-making process. When the processing precision of wheat flour is high, the flour yield is low and the crude fiber content is high when the processing precision is low, the flour yield is high and the crude fiber content is

lower [79]. Tab. 2.5 shows the cellulose content for various precisions of wheat flour.

Tab. 2.5 Cellulose content for various precisions of wheat flour [79]

Precision of wheat flour (flour extraction rate)%	75	85	90	95	100
Crude fiber content % (w/w)	minimum	0.55	1.00	1.50	1.95

Although fiber is a sugar that cannot be digested or absorbed by the body, it can promote gastrointestinal peristalsis, stimulate the digestive glands to secrete digestive juices, aid in the digestion of other nutrients, and play an important role in the prevention of colon cancer, among other things [80].

Minerals: Phosphorus, potassium, magnesium, calcium, sodium, iron, copper, and other elements are found in wheat flour [81]. Wheat flour contains various elements in the form of inorganic salts. Because of the role of phytic acid and cellulose in whole wheat flour and high flour yield wheat flour, calcium is recovered by phytic acid, and the effective value of iron is also reduced, whereas high-precision wheat flour contains significantly less iron and calcium; thus the lack of these two minerals in wheat flour is an important addition to the fortification of wheat flour [82].

Vitamins: Vitamins are a class of compounds that the body requires to maintain normal physiological functions [83]. Wheat used to have more B vitamins, but the majority of these vitamins are now concentrated in the germ, paste flour layer, and skin layer. After processing, the vitamins in wheat flour are greatly reduced, and their content ranges from 10%-30% (w/w) of the wheat grain [84]. The vitamin content of high-precision wheat flour is even lower. In China, the standard flour yield of wheat is about 80%-82% (w/w), with a retention rate of about 95% for vitamin B1 and E, 50% for vitamin B2, and 35% for nicotinic acid. Specially made flour has a flour yield of about 70%, with a retention rate of about 10% for vitamin B1, 60% for vitamin E, 35% for vitamin B2, and 20% for nicotinic acid [85]. There is also a significant loss of vitamins during the cooking process due to various methods.

2.1.2 Composition Quality Analysis of Flour

In general, the variety of properties of the final products implies, that analytical techniques for determining the parameters, which represent the property profile of a specific type of flour must be available. Given the variability of grain components, producing flour with correct and consistent quality parameters is a difficult task, which necessitates continuous monitoring of flour parameters during the manufacturing process.

2.1.2.1 Determination of Moisture Content of Flour

The weight measurement method is the most commonly used method for determining moisture content. It is divided into atmospheric drying method, decompression drying method, high-temperature drying method, infrared drying method, microwave drying method, and distillation method [86]. Because atmospheric pressure drying, reduced pressure drying, and high temperature timing drying lose some volatile substances while removing moisture, the measured moisture content is inaccurate and time-consuming [87]. When using infrared technology to dry flour, decomposition and Merad reactions will occur inside the flour, causing some moisture to affect the experimental results [88]. At the same time, the physical state of the material influences measurement accuracy. The microwave method is sensitive, fast, safe, does not damage the material, has moderate price, and can be applied as on-line measurement [89]. The measurement signal can be used for online digitalization, and visualization. The sampling device used in on-line measurement requires only low loading, the measured result is the overall moisture and superior to the surface measurement technology. But the lower detection limit is not low enough and the measured value is related to the composition of the material; different varieties need to be calibrated separately. The distillation method consists of heating the sample together with an organic solvent, so that the moisture escapes by vaporization and volatilization, and the moisture collected by the condensing device determines the moisture content of the sample. This method easily causes the sample to blacken and scorch dehydration carbonization thus leading to higher

measurement results due to the high boiling point of organic solvent [90]. The determination of sample moisture content can also be accomplished using chemical methods, such as the Karl Fischer method, which uses a chemical reaction between Karl Fischer reagent and water to determine the moisture in the sample by titration, which is too expensive and requires grinding, which loses some of the moisture, resulting in inaccurate measurement results that frequently require calibration [91]. Furthermore, resistive, capacitive, NMR, and acoustic methods can be used to quickly determine the moisture content of flour. The resistive and capacitive methods are unsuitable for the moisture content range to be measured as well as the measurement environment. Although NMR is fast, accurate, and has a wide application range, it is influenced by uncontrollable factors such as material flow, stack density, and temperature [92]. The acoustic method is highly accurate, reproducible, and can be used for on-line detection, but it is expensive [93]. The moisture measurement method used for the parameter values covered in this PhD thesis is: ICC 110/1 (mod., Brabender MT):1976 [201] (see Section 3.1.2.3).

2.1.2.2 Determination of Protein Content of Flour

Protein is an important component of all cells and tissues in the human body, is involved in all components of the organism, and serves as the material foundation and primary bearer of life activities [94]. The use of physical properties such as refractive index method, ultraviolet (UV) absorption, infrared spectroscopy, spectrophotometry are the most common protein determination methods. Another method is to use its chemical properties, such as Kjeldahl nitrogen method, Biuret reagent method, hydrogen peroxide method, dye binding method, Folin-Phenol reagent method (Lowry method) and Dumas combustion method. The Kjeldahl method has the best precision and accuracy, good reproducibility, and is suitable for determining the protein content of various materials, but its determination process is long and has low sensitivity, and the complex operation cannot be applied for on-line monitoring [95]. The Biuret reagent method is fast and easy to measure, but its accuracy is poor [96]. The alkali-treatment method and colorimetric method have high detection limits for materials with

high protein content [97]. The hydrogen peroxide method is faster and the results are not significantly different from the national standard method, but its precision is slightly poor [98]. The Folin-Phenol reagent method is sensitive and suitable for determining low protein content of materials, but it takes a long time and contains more interfering substances [99]. Although the UV absorption method is simple, sensitive, fast, and does not consume the sample, its accuracy is poor and there are more interfering substances [100]. The dye binding method is simple and easy to use, but there are large deviations when used to determine different proteins [101]. The protein measurement method used for the parameter values covered in this Ph.D thesis is: ICC 167:2000 [197] (See Section 3.1.2.1).

2.1.2.3 Determination of Ash Content of Flour

Ash is the residue of high-temperature burning of flour, and it is composed of metal oxides. The ash value reflects the amount of flour containing the skin layer, flour purity (whiteness) or bran, wheat germ, endosperm separation of the thoroughness, and thus the quality of flour [102]. The flour ash index has a special significance for flour production, and flour mills measure the accuracy level of flour with the help of ash determination. The dry ashing method [103], electrical conductivity method [104], mid-infrared spectroscopy technique [105], and X-ray scattering technique [106] are the most commonly used methods for determining ash content. The ash measurement method used for the parameter values covered in this article is: ICC 104/1:1990 [203] (See Section 3.1.2.4).

2.1.2.4 Determination of Wet Gluten Index and Dry Gluten Index of Flour

Gluten has a high nutritional value as a high quality vegetable protein source [107]. Gluten content is an important quality indicator and a factor influencing the quality of processed foods. Because gluten content varies depending on the physical state of the flour, both the dry gluten content and the wet gluten content are commonly used to indicate both. There are two

methods for determining the above two values: hand-washing and machine-washing, both of which work on the same principle but in different ways. The hand-washing method is complex [108], time-consuming, and labor-intensive, whereas the machine-washing method can speed up the measurement to a certain degree, but cannot be monitored on-line [109]. The gluten index is used to evaluate gluten quality, and it is based on determining the wet gluten content, but the measurement process is too long and heavily influenced by external factors. Wieser [110] used a microscale combined extraction-HPLC procedure to determine the amounts of the various gluten protein types in wheat flour. Czaja [111] made use of a method for quantifying gluten in wheat flour based on PLS treatment of FT-Raman data. The measurement method for wet gluten and gluten index used for the parameter values covered in this thesis is: ICC 155:1994 [199] (See Section 3.1.2.2).

2.2 Overview of NIR Spectroscopy

2.2.1 The Development Process of NIR Spectroscopy

Astronomer William Herschel discovered electromagnetic waves in the wavelength range of 800-2500 nm [112]. Scientists were able to analyze some of the material information contained in the NIR spectrum by the early nineteenth century, but the application of NIR spectroscopy was limited due to its low intensity, complex spectral bands, and overlapping characteristics, and the low level of theory and technology at that time could not fully extract the information contained in the spectra. Scientists discovered that the particle size of the sample has a significant impact on NIR spectra. Thus, methods to homogenize the test sample before acquiring the spectrum can greatly improve the technique's accuracy and applicability. With the development of applied mathematics, various mathematical algorithms and spectral data processing methods have been developed to further eliminate the testing errors caused by spectral overlap and broaden the application of the technique from the NIR diffuse reflectance region (1100-2500 nm) to the NIR transmission region (700-1100 nm) where the absorption signal is weaker but the penetration ability is stronger [113,114]. The advancement of NIR

spectroscopy has resulted also in the further instrumental development of NIR spectrometers. Kaye was the first to develop a transmission NIR spectrometer in the mid-1950s, and since then, several manufacturers have begun to develop NIR spectrometers [115]. By the 1980s, scientists have gradually started to apply multivariate calibration methods as well as modern data processing techniques for NIR spectroscopy, which resulted in the evolution of a valuable analytical technique. Various instrument manufacturers competed to develop special NIR spectrometers at the same time, and high-resolution Fourier transform NIR spectrometers emerged [115]. The widespread interest of scientists and instrument manufacturers in the further development of NIR spectroscopy contributed to the successful convening of the First International Conference on NIR Spectroscopy in Norway in 1987, and NIR spectroscopy has become increasingly popular since then [116].

2.2.2 Basic Theory of IR/NIR Spectroscopy Technology

The measurement method used in the present work is based on vibrational spectroscopic investigations in the infrared (near-infrared) range. IR spectroscopic applications make use of the fact that organic molecules can be excited to vibrations and rotations by absorption of electromagnetic radiation [117-121]. Only a small spectral range of wavelengths from 500 to 0.8 µm (corresponding to 20-12500 cm^{-1}) is assigned to IR spectroscopy, which is further subdivided into NIR, MIR, and FIR ranges (Fig. 2.1).

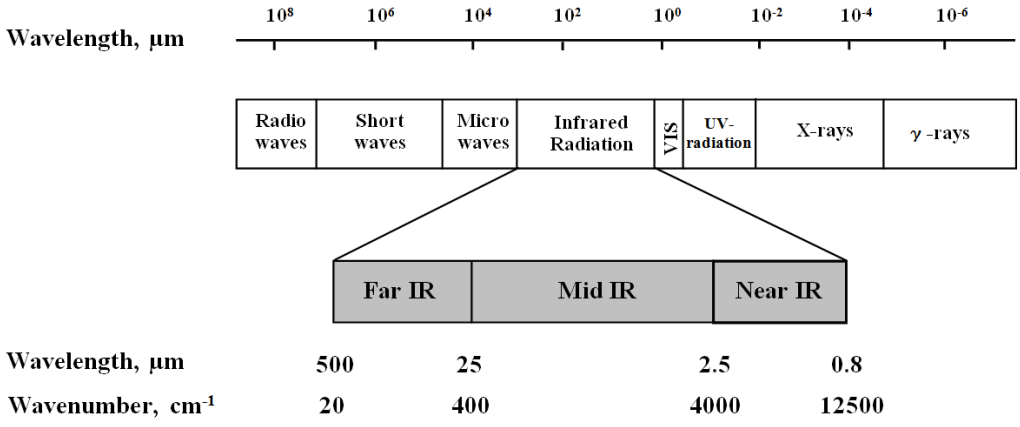


Fig. 2.1 The electromagnetic spectrum and its subdivision into different wavelength ranges

Electromagnetic radiation consists of sinusoidally alternating fields that travel at the speed of light in vacuum. If there is a molecule in the alternating field, it can absorb energy, causing the radiation energy to be attenuated. Absorption of IR radiation leads to excitation of vibrations. A molecule of N atoms has 3N degrees of freedom of motion; three of which are translational and three rotational. From this fact, the number of vibrational degrees of freedom is 3N-5 for linear molecules and 3N-6 for non-linear molecules. When a change in dipole moment occurs during a vibration, a molecule absorbs IR radiation for vibrational excitation. Harmonics (overtones) occur when the frequency of the fundamental is doubled (1st harmonic) or tripled (2nd harmonic). Combination oscillations are caused by the addition of two fundamental frequencies of the same or very close neighboring functionalities. Depending on whether bond distances or bond angles change, oscillations are referred to as stretching or deformation vibrations.

The harmonic oscillator model, based on a linear, diatomic molecule, is used to explain the oscillation process [117-121]. The harmonic oscillator is made up of two mass points, m_1 and m_2, connected by a spring. At the equilibrium nuclear distance r_0, the spring represents the bonding force, which opposes the deflection of the atoms involved in the bond. Hooke's law describes the deflection of the molecular model from its equilibrium position by the distance Δr:

$$F = -k \cdot \Delta r \qquad \text{Eq.1}$$

F: Repulsive force [N]

k: Force constant [10 ²N m^{-1}]

Δr: Deflection [m]

For the oscillation process of a diatomic molecule, the oscillation frequency ν is as follows:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad \text{Eq.2}$$

where k is the force constant and a measure of the strength of the chemical bond between the vibrating atoms and μ is the reduced mass (see below). Thus, the smaller the masses of the atomic oscillators and the stronger the bonds between them, the higher the frequency of oscillation will become. A parabola and the following equation describe the change in potential energy:

$$E_{pot} = \frac{1}{2} k \Delta r^2 = 2\pi^2 \mu \nu^2 \Delta r^2 \qquad \text{Eq.3}$$

E_{pot}: Potential energy [J]

ν : Vibration frequency [s^{-1}]

$\mu = \frac{m_1 \cdot m_2}{m_1 + m_2}$: Reduced mass [kg]

A quantum mechanical treatment shows, that the vibrational energy may only have certain discrete values called energy levels. For the harmonic oscillator these energy levels are given by:

$$E = h\nu(\upsilon + \frac{1}{2}) = \frac{h}{2\pi}\sqrt{\frac{k}{\mu}}\left(\upsilon + \frac{1}{2}\right) \qquad \text{Eq.4}$$

where the oscillation quantum number can take the form $\upsilon = 0,1,2,\ldots$ .The vibration energy is not zero at the lowest vibration level, but has the value $E_0 = \frac{1}{2} h\nu$. Equation 4 then yields the equidistant oscillation terms:

$$\frac{E}{h \cdot c} = \tilde{\nu}\left(\upsilon + \frac{1}{2}\right) \quad \text{with } \Delta\upsilon = \pm 1 \qquad \text{Eq.5}$$

The indication of the wave number $\tilde{\nu}$ is used in IR spectroscopy because the oscillation frequency ν of the absorbed radiation is directly proportional to the energy E. The following relationship holds true:

$$\tilde{\nu} = \frac{\nu}{c} = \frac{1}{\lambda} \qquad \text{Eq.6}$$

$$\Delta E = E_{\upsilon+1} - E_{\upsilon} = h \cdot \nu = \frac{h \cdot c}{\lambda} = h \cdot c \cdot \tilde{\nu} \qquad \text{Eq.7}$$

$E_{\upsilon+1}$: Energy after the absorption of a photon [J]

E_{υ}: Energy before a photon is absorbed [J]

c: Speed of light 2.99793 $\cdot 10^{10}$ [cm s^{-1}]

h: Planck's constant 6.62618 $\cdot 10^{-34}$ [J s]

ν: Vibration frequency [s^{-1}]

$\tilde{\nu}$: Wavenumber [cm^{-1}]

λ: Wavelength [cm]

Only at small deflections (Δr) do real molecules behave like harmonic oscillators. The anharmonicity increases as the amplitude of the oscillation increases. Excessive stretching of the molecular bond can cause the molecule to dissociate. If the two atoms are too close together, they repel each other. As a result, rather than a parabolic function, the Morse function with the following equation better describes the dependence of a diatomic molecule's potential energy on bond distance:

$$E = h\nu(\upsilon + \frac{1}{2}) - \chi_D h\nu(\upsilon + \frac{1}{2})^2 \qquad \text{Eq.8}$$

χ_D: Anharmonicity constant

The equation implies that the oscillation energy distances are not equidistant, but rather move closer together as the oscillation quantum number υ increases. The harmonic oscillator's strict selection rule $\Delta\upsilon = \pm 1$ no longer applies to the anharmonic oscillator. Transitions with $\Delta\upsilon = \pm 2$ (1^{st} harmonic) and $\Delta\upsilon = \pm 3$ (2^{nd} harmonic) are now permitted, but the intensity of the associated absorption bands is reduced by a factor of 10-100 each time (Fig. 2.2). The greater the term difference $\Delta\nu$, the larger the required excitation energy and the shorter the absorption band wavelength [117-121].

The main applications of IR spectroscopy are structure elucidation of unknown substances via characteristic group frequencies and quantitative analysis by determination of band intensities [122-125].

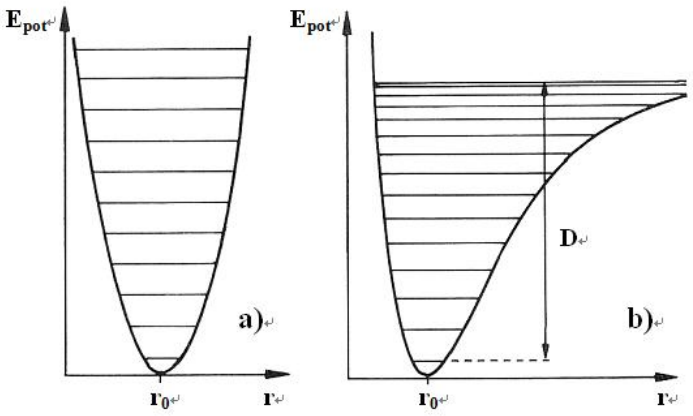


Fig. 2.2 Potential energy representation of a harmonic (a) and an anharmonic oscillator (b) as a function of the atomic distance r (D = dissociation energy) [123]

2.2.3 Principles of NIR Spectroscopy Technology

The near infrared wavelength range is 800 to 2500 nm (12500-4000 cm^{-1}). Overtones and combination vibrations of OH, NH, and CH functionalities with large mechanical anharmonicity (large mass difference of the vibrating atoms) can be observed in the NIR spectral range [116, 127-130]. The fundamental vibrations of the aforementioned groups can be found in the mid-infrared range, with wavenumbers ranging from 2500 to 4000 cm^{-1}.

With increasing order, the intensities of the harmonics of the corresponding fundamental vibrations decrease very quickly. Combination oscillations occur at wavenumber sums of at least two fundamental oscillations and result from the simultaneous excitation of multiple oscillations [131-134]. In the following Tab. 2.6 the near-infrared absorption regions and their assignment to vibrational motions of selected chemical functionalities are summarized.

Tab. 2.6 Assignment of some absorption bands in the NIR wave number range for CH, OH and NH functionalities [134]

Wavenumber [cm^{-1}]	Wavelength [nm]	Vibration type
4200-4400	2439-2273	C-H combination vibrations
4500-5200	2222-1923	N-H combination vibrations
4900-5400	2041-1852	O-H combination vibrations
5500-6300	1818-1587	C-H 1st overtone
6500-7000	1538-1429	N-H 1st overtone
6400-7500	1563-1333	O-H 1st overtone
6900-7700	1449-1299	C-H combination vibrations
7800-9100	1282-1099	C-H 2nd overtone

Larger sample thicknesses can compensate for the much lower band intensities of the near-infrared region compared to the mid-infrared fundamental vibration range. This, however, places much lower demands on the specimens' sample preparation. Because many organic compounds absorb in the near infrared, measurements in this range provide good conditions for practical spectroscopy, allowing for the universal use of this spectroscopic measurement method [131, 132].

The absorbance (A) $log\ 1/T$ (for transmission measurements) or $log\ 1/R$ (for diffuse reflection measurements) versus wavelength or wavenumber is plotted for the representation of NIR spectra. These intensity parameters are proportional to the concentration of the sample under consideration. Various recording techniques (transmission, transflection, and diffuse reflection (Fig. 2.3)) can be used depending on the type and physical state of the samples [131]. For transmission measurements the Bouguer-Beer-Lambert law is used:

$$A = \varepsilon \cdot c \cdot d \qquad \text{Eq.9}$$

The wavelength-dependent molar extinction coefficient ε (in L • mol^{-1} • cm^{-1}) is a constant property of each chemical compound in this equation, d is the layer thickness (in cm), and c is the concentration (in mol • L^{-1}). As transmittance T and reflectance R, a relationship is established between the radiation intensity I attenuated by the sample and the measured reference intensity I_0:

$$A = \log \frac{I_0}{I} = \log \frac{1}{T} \qquad \text{Eq.10}$$

For diffuse reflection measurements, the Lambert-Beer law is modified by the scattering coefficient s (in cm^{-1}) to account for scattering between particles of different sizes:

$$\log \frac{1}{R} \approx \frac{\varepsilon \cdot c}{s} \qquad \text{Eq.11}$$

When compared to measurements in the mid-infrared, diffuse reflection measurements in the near infrared do not require additional dilution with an inert matrix such as KCl, which is regarded as a significant advantage of this method [131,135]. The three methods of near-infrared spectroscopy measurement are shown in the Fig. 2.3 below.

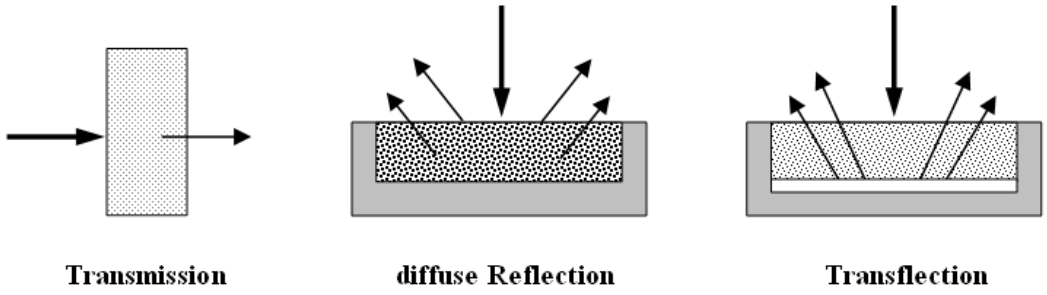


Fig. 2.3 Three methods of near infrared spectroscopy measurement techniques

For NIR spectroscopy, window materials such as glass and quartz are used, which are much less sensitive to water than the window materials of the MIR measurement range (NaCl, KBr and KRS5). Apart from its ease of use, short analysis times, and high signal-to-noise ratio (>10000), NIR spectroscopy has primarily grown in popularity for chemical process control due to the use of quartz light fibers [129, 131, 132, 135]. However, due to the comparatively low selectivity of NIR spectroscopy, chemometric methods are usually required for the qualitative and quantitative evaluation of NIR spectra with respect to the chemical and physical properties of the investigated substances [131, 136-138].

2.2.4 Reflection Spectroscopy

Reflection spectroscopy is a research technique that measures the intensity of reflected electromagnetic radiation from a sample surface that has been illuminated [139-143]. There are two types of limiting cases: regular or directional reflection and diffuse reflection. The geometrical optics laws govern directional reflection at an optically smooth, reflecting surface (metal, glass surface). Because of the behavior of light waves at interfaces, the angle of incidence and angle of reflection coincide (law of reflection) [141]. Snellius' law of refraction (Fig. 2.4) applies to the direction of the incident and refracted light beams:

$$n_1 \cdot \sin\alpha = n_2 \cdot \sin\beta \quad \text{with} \quad n_2 > n_1 \qquad \text{Eq. 12}$$

When radiation from an optically denser material strikes an optically thinner material, total reflection is observed above a certain limiting angle α_g, indicating that the entire energy of the incident radiation is found in the reflected radiation. This phenomenon is used in optical fiber technology to transport radiation over long distances or in attenuated total reflection (ATR) spectroscopy for analytical investigations thereby avoiding time-consuming sample preparation.

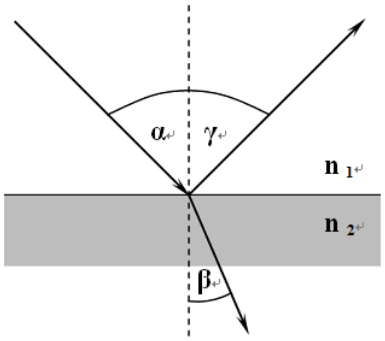


Fig. 2.4 Reflection and refraction of radiation at an interface [141]

Regular reflection is a process that refers to macroscopic, plane phase boundaries. Diffuse reflection, on the other hand, is obtained when there are centers on a rough surface that scatter the incident radiation. Cracks and stipples, as well as particles, can act as scattering centers.

Fig. 2.5 depicts a schematic representation of diffuse reflection of radiation by a powder

substance. Absorption, scattering, interference, diffraction, and refraction are all phenomena, that can occur in this process [139-141].

Because neither the classical Rayleigh scattering (particles much smaller than the wavelength of the radiation) nor the Mie theory (no interaction between the irradiated particles) apply to the powdered samples studied by NIR spectroscopy in this work, simplified laws are used to evaluate the NIR spectra measured in diffuse reflection [135, 139].

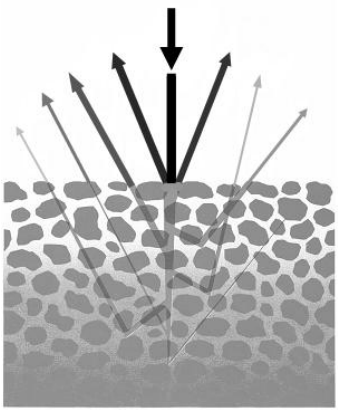


Fig. 2.5 Schematic representation of diffuse reflection in a powder substance

The function $F\ (R_\infty)$ proposed by Kubelka and Munk in the early 1930s [142] has proven useful for quantitative evaluation of the corresponding spectra for diffuse reflection measurements in the MIR:

$$F(R_\infty) \equiv \frac{(1-R_\infty)^2}{2R_\infty} = \frac{K}{s} = \frac{\varepsilon \cdot c}{s} \qquad \text{Eq.13}$$

R_∞ denotes the sample's diffuse reflectivity at large - theoretically infinite - layer thickness, which in practice is around 2 mm; K denotes the absorption coefficient, and ε and c denote the sample's absorptivity (extinction coefficient) and concentration, respectively. The scattering coefficient s is affected by the particle size distribution of the sample [139, 143].

Because of the high absorptivity of organic compounds in the MIR range, the sample must usually be diluted with a non-absorbing material (e.g. KCl) to obtain suitable intensities in diffuse reflection measurements. This is not required for NIR measurements, and diffusely

scattering substances (powders, rough surfaces, etc.) can be investigated without sample preparation [135, 140, 143].

For NIR measurements in diffuse reflection, the log 1/R function is used for quantitative analysis analog to absorbance A, which provides better proportionality to concentration than the Kubelka-Munk function [144, 145]:

$$A = \log \frac{1}{R} \approx \frac{\varepsilon \cdot c}{s} \qquad \text{Eq.14}$$

In practice, the reflected radiation is measured using an integrating sphere with a gold-plated inner layer (which also serves as a reference) or an optical accessory (e.g. light-fiber bundle), that collects the diffusely reflected radiation as efficiently as possible [135, 143, 146]. In most cases, however, a scatter correction is required as data pretreatment for the quantitative evaluation of NIR spectra measured in diffuse reflection [137, 138].

2.2.5 Interaction between Radiation and Matter in NIR Spectroscopy (Basic Principles of Qualitative and Quantitative Analysis)

Because the spectrum contains a wealth of information about the substance, NIR enables qualitative and quantitative analysis [147]. The characteristics of the spectral pattern are closely related to the composition and content of the substance itself. If a substance's structure and composition vary, as do the spectral patterns collected, and vice versa. Organic substances, that make up a multiplicity of materials contain different functional groups (such as O-H, N-H, C-H/CH_2/CH_3, C=O), which have specific vibrational frequencies in the near-infrared [147]. When a functional group in a substance is exposed to NIR radiation, only the incident wave with the intrinsic frequency of the specific molecular vibration of this functional is absorbed, causing the group to resonate, and the NIR spectrum of the substance can be obtained by measuring its absorption at this wavelength of the NIR spectrum. Therefore, the measured spectrum reflects information about the composition of the measured substance under investigation, laying the theoretical groundwork for qualitative and

quantitative NIR spectral analysis. NIR spectra contain a multiplicity of absorption bands related to hydrogen-containing groups (C-H, O-H, N-H) and because the main components of food are organic substances composed of these hydrogen-containing groups, NIR spectroscopy can be used for qualitative and quantitative analysis of food.

2.2.5.1 Qualitative Analysis of NIR Spectral

The purpose of qualitative NIR analysis is to determine the identity of an unknown sample by comparing the absorption intensity, frequency, and peak shape of the absorption peaks in the unknown sample's spectrum and the spectrum of a known reference sample set [148]. Because the reliability of direct comparison by human eyes is very low due to the complexity and variability of NIR spectra, it relies on computers and mathematical algorithms to separate and extract the information in the spectra and then compare and identify them. Pattern recognition is one of the most widely used methods for distinguishing and identifying spectral patterns using mathematical algorithms such as PCA, discriminant analysis (DA) and soft independent modeling of class analogies (SIMCA).

The practical application of NIR qualitative discriminant analysis can be subdivided into the following steps:

(1) Acquisition of the spectra of a standard sample set;

(2) Calibration and preprocessing of specta;

(3) Extraction of spectral features;

(4) Development of qualitative discriminant analysis models;

(5) Realization of qualitative discrimination of unknown samples.

2.2.5.2 Quantitative Analysis of NIR Spectral

The quantitative analysis of NIR spectra is based on the principle that the absorption spectra of samples with different content of each constituent are different. As a result of the selective

absorption of different frequencies of NIR radiation by each component of a sample, the content of each component can be determined by analyzing the optical density of the transmitted or reflected radiation through a detector [149]. Sample set's spectra are collected, and the sample set's chemical and physical parameters are determined by independent reference methods. Using chemometric methods, a correlation between the spectra and these parameters is established in a mathematical model and the content of each component of an unknown sample can be predicted by subjecting the corresponding spectrum to the developed calibration model.

2.2.6 Technical Characteristics of NIR Spectroscopy

When compared to traditional methods for determining physical and chemical properties of various substances, methods based on NIR spectroscopy have the following advantages [150, 151]:

(1) A simple testing procedure that eliminates the need for time-consuming sample preparation and chemical reaction steps.

(2) Fast testing speed, which can be completed in many cases within a few minutes.

(3) High testing efficiency; a single person can test multiple chemical indicators simultaneously.

(4) A clean testing process, which does not necessitate the use of a large number of chemical reagents, as well as low testing costs.

(5) Good accuracy and repeatability.

(6) Applicable for solid and liquid samples.

(7) The technique is non-destructive and can be applied in the chemical, pharmaceutical, food, environmental and material analysis.

(8) NIR light has good transmission properties for optical (especially quartz) fibers, making it useful for on-line analysis, remote monitoring of substance composition and in-vivo analysis.

However, NIR spectroscopy also has some disadvantages: the sensitivity is low due to the weak absorption of substances in the NIR region, and substance content of interest should be

greater than 0.1%, which means it cannot be used to determine trace substances; the calibration modeling work necessitates specialists and model maintenance requires ongoing work effort.

2.3 Instrumentation of NIR Spectroscopy

Prior to the 1850s, NIR spectroscopy was a forgotten spectral region due to its severe spectral absorption band overlap and poor selectivity; however, in the late 1950s, Karl Norris was the first to use short-wave NIR transmission to determine the moisture content of grains and rapeseed, kicking off the development of NIR spectroscopic instrumentation and its applications [152]. The group led by Karl Norris conducted a large number of spectroscopic method demonstrations in the 1960s, including comparisons of transmission, reflection, and transflection methods in the visible and near-infrared wavelength region. The availability of reflectance spectra of plant leaves and grains was the most significant achievement in this phase of work and the successful agricultural applications paved the way for further development of new NIR spectroscopic applications [153].

Nowadays, NIR spectroscopic instruments are gradually progressing towards process integration and miniaturization, which further advances the technology and also satisfies market demand. At present, miniaturization for in-the-field and on-site testing are an extremely important instrumentation trend and analytical application areas. Therefore, NIR spectrometers are widely used in environmental monitoring, food testing, biomedical research and other fields as an essential testing instrument in modern society.

2.3.1 Benchtop NIR Spectrometers

With the rise of computers, which made instrument control more precise and accurate, and the rise of chemometric techniques, which made data analysis relatively simple and could be used to process quantitative or qualitative information of complex systems, NIR spectroscopy became widely studied and applied in the 1970s [152]. Norris developed the world's first NIR

scanning spectrometer allowing data transfer to a microcomputer [154]. It was on this scanning spectrometer that the benefits of multiple linear regression analysis in extracting spectral information related to components were demonstrated, and this instrument became the prototype for the further development of NIR spectroscopy.

The American Society for Testing and Materials (ASTM) formed the Working Group on Near Infrared Spectroscopy in 1984, at Hirschfeld's initiative, to study the issue of standard methods for NIR spectroscopy [148]. Simultaneously, some well-known instrument manufacturers began to develop new NIR instruments, and the NIR spectrometer market began to blossom. The main technical lines are divided into two categories: NIR spectrometers developed from grating spectrometer principles commonly used in the visible region [155], and NIR spectrometers developed from optical interference Fourier-transform principles commonly used in the mid-infrared region [156]. The rise in demand for industrial analysis has accelerated the development of NIR instrumentation, including transmission-based systems for the analysis of liquids by immersion-probe systems, and diffuse reflection-based instruments for solid-state analysis. Some companies, primarily involved in the production of benchtop NIR instruments for laboratory use, have also begun the development of NIR spectrometers for on-line applications.

Compared to other analytical instruments, a multiplicity of NIR spectrometer types are in practical use. According to the spectroscopic principle, benchtop NIR spectrometers used in laboratories can be subdivided into filter type, grating dispersive type, Fourier-transform type and acousto-optic tuneable filter type. Fourier-transform type spectrometers have the largest market share [157]. Generally, NIR spectrometers are made up of a light source, a spectrometer system, a sample chamber, a detector, and a control and data processing system.

2.3.2 Handheld NIR Spectrometers

Since the late 20th century, the development of spectrometer instruments has gradually begun to develop in the direction of miniaturization and portability. Research and development of miniaturized spectroscopy instrumention has become the focus of attention of science and

industry departments in various countries [158].

Traditional spectroscopy instruments are not only large in size, but also frequently require pretreatment of specific samples before entering the analysis step, as well as strict working environment requirements, necessitating the use of professional analysts to complete the analysis. In comparison, miniaturized spectrometers have absolute advantages in terms of size, portability, speed, convenience of use and mobile power, which are the basic requirements for small spectrometers that do not take up much space [158, 159]. In addition, a handheld spectrometer has the characteristics of fast detection speed and suitable for online monitoring. Because of their small size, fast analysis speed, simple operation, lack of sample pretreatment, and low price, some portable or handheld NIR spectrometers are gradually making their way into various fields of production life [158].

The development of miniaturized NIR spectrometers is progressing and there are numerous miniaturized NIR spectrometers based on various spectroscopic principles on the market. At the moment, the most common types are grating micro spectrometers, microelectro-mechanical systems (MEMS)-based Fourier-Transform Near-infrared (FT-NIR), linear variable filter (LVF) type, Hadamard Transform type, Fabry-Perot tuneable filter type, array light-emitting diode (LED) type and acousto-optic tuneable filter (AOTF) type [160-162].

The introduction of fiber optic probes has accelerated the use of small NIR spectrometers, and the convenient optical path design is especially suited to various online inspection needs. Manufacturers of handheld near-infrared spectrometers include Si-Ware Systems (Cairo, Egypt), VIAVI Solutions Inc. (San Jose, California, USA), Spectral Engines Oy (Helsinki, Finland), Hamamatsu Photonics (Hamamatsu City, Japan), OtO Photonics (Hsinchu, Taiwan Province, China), Insion GmbH (Obersulm, Germany) and TrinamiX GmbH (Ludwigshafen, Germany), to name just a few [163].

Rajendran et al. [164] Minotto et al. [165] investigated portable near-infrared devices based on LEDs. Zeltex LLC (Hagerstown, Maryland, USA), manufactures the LED-based Zx-50 handheld grain analyser, which can be used to analyse wheat composition [166].

With the rapid development of computer science, the control system of small NIR spectrometer has been expanded from single PC control to microcontroller, Advanced Reduced Instruction Set Computer (RISC) Machines (ARM) embedded microcontroller, Raspberry Pi (RPi), smart phone and Industrial Personal Computer (IPC), making the NIR spectrometer smaller and lighter while providing convenience for users [167].

In terms of software operating systems, small NIR spectrometers have spread from the WINDOWS platform to the MacOS, Linux, Android, and Apple platforms, among others [167]. The richness of the possible software platform encourages the use of small NIR spectrometers, and the sharing of spectral data is more convenient, safe, and fast.

2.4 Data Pretreatment

Apart from baseline correction and 1st or 2nd derivative, scatter correction is frequently used for data pretreatment of spectra measured of solid samples in diffuse reflectance, because they exhibit scatter effects leading to different slope, band intensity, and baseline offset as a consequence of sample morphology (particle size, distribution, packing, or layer density) [168, 169, 170-172]. The use of mathematical methods results in an emphasis on important spectral properties and a reduction of difficult-to-model physical influences in the spectral data, resulting in a better correlation between the spectral data and the investigated properties. In this work, the scattering correction methods MSC, SNV, and Extended Multiplicative Scatter Correction (EMSC) are used and explained with reference to the following literature [170-179].

MSC: Geladi et al. [173] invented MSC correction in the 1980s. A mean spectrum $\bar{x}$ is first calculated from the available spectral data in this method. Individual spectra are then fitted to the calculated mean spectrum by multiplying each data point by a coefficient b_i and adding a constant a_i:

$$x_i = a_i + b_i\bar{x} \qquad \text{Eq.15}$$

The original spectra can be converted into the corrected spectra ($x_{i\,MSC}$) in three different ways: either by subtracting the additive influence (offset a_i) according to Eq.16, by taking into account the multiplicative influence (slope b_i) according to Eq.17, or by full MSC according to Eq.18.

$$x_{i,MSC^a} = x_i - a_i \qquad \text{Eq.16}$$

$$x_{i,MSC^b} = \frac{x_i}{b_i} \qquad \text{Eq.17}$$

$$x_{i,MSC^{a,b}} = \frac{(x_i - a_i)}{b_i} \qquad \text{Eq.18}$$

SNV: SNV scatter correction, which is similar to MSC, has been used to eliminate scatter effects and particle size differences in spectral data since the late 1980s, following the publication of Barnes et al [174].

$$x_{i,j,SNV} = \frac{(x_{i,j} - \bar{x}_i)}{\sqrt{\frac{\sum_{j=1}^{m}(x_{i,j} - \bar{x}_i)^2}{(m-1)}}} \qquad \text{Eq.19}$$

where $x_{i,j}$ is a data point in the spectrum x_i, $\bar{x}_i$ is the spectrum's mean value, and m is the number of data points on the spectrum.

The SNV calculation standardizes each spectrum by taking the mean and standard deviation of each individual spectrum into account. Each spectrum is pretreated independently of the others by first centering it on its mean value and dividing it by its standard deviation. The scatter correction algorithms SNV, MSC, and EMSC are applied to spectra in log 1/R, log 1/T, and Kubelka-Munk formats [171, 175, 180].

EMSC: Martens et al. [176-179] developed an evaluation algorithm known as EMSC. The EMSC method is unique in that the chemical effects (light absorption) and physical effects (light scattering) in the spectra of powders and turbid solutions can be separated, and the scattering dependence on wavelength is taken into account in addition to compensating for the multiplicative and additive effects. The previous MSC method is extended to include the wavelength-dependent terms $d_i \cdot \lambda$ and $e_i \cdot \lambda^2$ for this purpose.

The additive and wavelength-dependent effects, as well as the term b_i multiplicative effects (e.g. effective optical path length) in the measured spectrum x_i, are taken into account by the model parameters a_i (baseline offset), d_i (linear), and e_i (quadratic coefficients). Then, using Eq. 20, an EMSC-corrected spectrum ($x_{i,EMSC}$) is obtained:

$$x_{i,EMSC} = \frac{x_i - a_i - d_i\lambda - e_i\lambda^2}{b_i} \qquad \text{Eq.20}$$

All three scatter correction procedures (MSC, SNV, and EMSC) are implemented in the Unscrambler® v.9.7 software [181] and were used to create chemometric evaluation procedures in this work.

2.5 Chemometric Evaluation Methods

Chemometrics is a branch of chemistry that uses mathematical and statistical methods to design optimal measurement procedures and experiments and to gain the most relevant information from data analysis [182].

Due to the overlap of absorption bands of hydrogen-containing CH, OH, and NH groups that occurs frequently in the NIR spectral region, univariate calibration is usually insufficient for quantitative analysis of multicomponent systems in NIR spectroscopy. Instead, multivariate evaluation methods are used, with absorbance values from larger wavelength ranges typically used to establish a calibration. The individual chemometric evaluation methods are described in greater detail below.

2.5.1 Linear and Multiple Linear Regression (MLR)

Calibration establishes a relationship between experimental system variables x (e.g., absorbance) and one or more target variables y (e.g., concentration) for the quantification of a substance using chemometric methods. MLR, PCR, and PLS are suitable computational algorithms for establishing a quantitative calibration [168, 169, 180, 182-187].

The system properties of unknown samples can then be determined relatively quickly using a robust calibration model that has been developed. Generally, the Bouguer-Beer-Lambert law, which shows the linear relationship between absorbance (A) and concentration and sample thickness, is used to perform the quantitative evaluation:

$$A = \varepsilon \cdot c \cdot d \qquad \text{Eq.21}$$

In this equation, d denotes the thickness of the sample layer, ε the molar extinction coefficient, and c the concentration of the substance to be determined. In the linear calibration, the term $(\varepsilon \cdot d)^{-1}$ is expressed by the parameter b and then determined. The linear model equation of the calibration function or regression coefficient b, respectively, is:

$$y = b \cdot x \qquad \text{Eq.22}$$

Here, x represents the spectral data and y the reference data of the sample under consideration. Eq.22 is valid for univariate calibration where the component to be determined has a disturbance-free absorption band. The regression coefficient b calculated from the linear relationship of the system variable x and the target variables y then allows the target variables y (e.g. concentration) of unknown samples to be determined from their experimentally determined absorbance values. Due to the lack of isolated absorption bands, quantitative evaluation of multicomponent systems is typically not possible in NIR spectroscopy.

MLR is based on the linear regression with the Bouguer-Beer-Lambert law, whereby multiple wavelengths are selected for calibration. Unlike univariate calibration, multiple linear regression establishes the relationship between the dependent y-variable and multiple independent uncorrelated x-variables by calculating regression coefficients during calibration. The linear model equation system of Eq.23 is mathematically equivalent to univariate regression [168, 180, 183-185]:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m + e \qquad \text{Eq.23}$$

It consists of the parameter y to be determined, the absolute constant b_0, the model parameters b_m of the m^{th} wavelength, the absorbance values x_m of the m^{th} independent x −variables (wavelengths), and the error term e.

2.5.2 Factorial Methods

PCA, PCR, PLS-DA and PLS Regression are the most commonly used multivariate factorial methods [169, 186, 187, 205]. These methods are used to reduce the information in the calibration spectra's entire data matrix to a few key factors. As a result, the substance spectra are replaced by newly calculated spectra while no important information is lost.

2.5.2.1 Principal Component Analysis (PCA)

Pearson's works contain the first description and application of this method [188]. In the context of NIR spectroscopy PCA decomposes the original mean-centered data matrix X with n spectra and m data points into a new score matrix T and a loadings matrix P (see Fig. 2.6 and Eq.24) [168, 169, 180, 183, 184, 186]:

$$X = TP^T + E \qquad \text{Eq.24}$$

The E matrix represents the differences between the original X matrix and the data matrix generated by the loadings and score matrix. The transposed loadings matrix P, which consists of p rows (number of principal components) and m columns (data points in the spectrum), describes the relationship between the X −data and the corresponding principal components. The score matrix T contains information about the object's location in relation to the principal components, which is visualized in the score plot by object-projecting the data matrix onto the new principal component space. T is a matrix with n rows (number of calibration spectra) and p columns (number of factors). The residual matrix E contains the information that was not taken into account in the score or loadings matrix. It is made up of n rows and m columns.

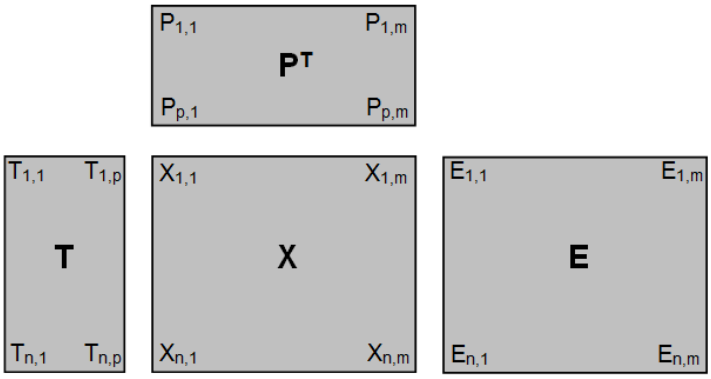


Fig. 2.6 Schematic matrix representation of the PCA algorithm

The uncorrelated variables are generated from the correlated variables by the coordinate transformation of the original data matrix X, with the first main axis PC1 having the highest variances. Following the criterion of maximum variance, the next PCs are determined by axis rotation of the new coordinate system and the orthogonality principle. In this case, the residual matrix E corresponds to the portion of the $X-$matrix that was not modeled by the $p-$principal components and should ideally consist only of measurement errors (noise) [189-191].

The relevant first principal components, which account for the majority of the total variance, are distinguished from the less relevant principal components by a maximum in the curve in the Unscramber® v.9.7 software's Explained Variance Plot [181]. PCA is widely used for qualitative product differentiation (e.g. in incoming raw materials).

2.5.2.2 Principal Component Regression (PCR)

The PCR is a procedure that connects PCA with multiple linear regression to compute a regression model for a target variable y based on the X-data [168, 169, 180, 183-185, 190]. As with PCA, a mean-centered data set X is first decomposed into the score matrix T and the loadings matrix P. After that, the regression model is used to calculate the regression model:

$$X = TP^T + E \qquad \text{Eq.25}$$

The calculated mean spectrum is subtracted from each calibration spectrum when centering. By transforming Eq.26, the score values T are given as projections of the X-data onto the loadings P, with all elements of the residual matrix E having the value zero:

$$T = XP \qquad \text{Eq.26}$$

The most important main components are used to develop a calibration model. The loadings Q of the Y matrix are obtained through multiple regression of scores and target variables:

$$Y = TQ^T + F \qquad \text{Eq.27}$$

The regression coefficient B, which describes the correlation between Y and X data, is contained in the loading matrix Q:

$$Y = XB + F \qquad \text{Eq.28}$$

and corresponds the product of the loadings matrices P and Q:

$$B = PQ \qquad \text{Eq.29}$$

The number of principal components required to describe the data set is critical for predicting unknown samples. If a principal component number is chosen that is too large, it can result in the "overfitting" effect, in which interfering noise components are included in the model. If the number of principal components is too low, the available spectral information is not fully captured, indicating "underfitting" [180, 183, 184].

2.5.2.3 Partial Least Squares Regression (PLS)

H. Wold created it in the mid-1970s, S. Wold and H. Martens refined it in the 1980s into the simpler PLS representation with X and Y data sets [168]. PLS regression, which is also used in this work, is the most commonly used method for quantitative evaluation in NIR spectroscopy. In this method, the scores (T) are calculated using both the spectral data (X) and the target variables (Y) [169, 186, 187, 192]. As a result, the first PLS principal components

have the most variance in the spectral data (X) and the greatest correlation to the target variables (Y). The PCR/PCA algorithms, on the other hand, account for variance maximization solely by using the X-matrix data to calculate the score matrix T [191].

The X and Y matrices are decomposed into scores and loadings in the same way as the PCR calculation (Fig. 2.7) and the Eqs.33, 34:

$$X = T \cdot P^T + E \qquad \text{Eq.33}$$

$$Y = T \cdot Q^T + F \qquad \text{Eq.34}$$

The score matrix is T, the loadings matrices are P and Q, and the residual matrices of the two data sets are E and F. The residual matrices describe the unexplained variance or calibration error between the measured and reconstructed spectra [193].

In contrast to the PCR, an additional loadings matrix is required for the calculation of the B-regression coefficients according to Eq.35:

$$B = W(P^T W)^{-1} Q^T \qquad \text{Eq.35}$$

In PLS analysis, the W-loadingsweight matrix describes the relationship between the X- and Y-data. When predicting unknown samples from spectral data, the loadingsweights are used to calculate the corresponding scores. The W and T matrices are orthogonal to one another. For both B-coefficient calculation and sample prediction, the P-and Q-loadings matrices, as well as the W-loadingsweight matrix, are used:

$$Y = (W(P^T W)^{-1} Q^T) X \qquad \text{Eq.36}$$

The Nonlinear-Iterative Partial Least Squares (NIPALS) algorithm is used to perform the PLS calculation iteratively [168, 180, 184].

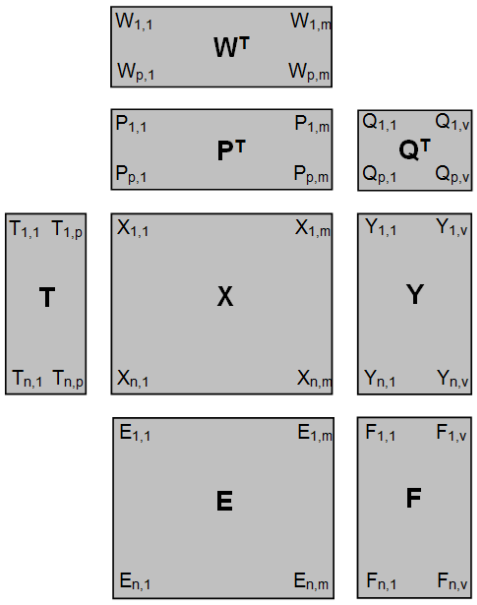


Fig. 2.7 Scheme of the X and Y matrix decomposition in the PLS regression

The optimal number of principal components is critical for robust calibrations. For calibration of the same data set, it was discovered that the PLS method requires fewer principal components than the PCR method [184]. According to Esbensen [180], the PLS method produces better results in terms of data reduction than the MLR and PCR methods.

When calibrating multiple Y-variables at the same time, the PLS2 algorithm is used, which determines the principal components for all variables at the same time. Due to the better prediction results, however, the PLS1 regression is preferred for calibration in quantitative NIR spectroscopy. Because the PLS1 regression determines only one target variable (i.e. one vector y), the Q-loadings matrix and the residual matrix F are also reduced to one column, i.e. to vectors [168, 183].

2.5.2.4 Partial Least Squares Discriminant Analysis (PLS-DA)

PLS-DA is a statistical method related to principal components, which reduces the dimensionality of the data to establish a regression model and performs discriminant analysis on the results. It is similar to PCA, except that PCA is unsupervised but PLS-DA is supervised [205].

When the differences between sample groups are large and the differences within each sample group are small, unsupervised analysis methods can be very good at distinguishing the different groups. Conversely, if the differences between groups are small, and the differences within each group are large, then the unsupervised analysis method will be very poor. Under this condition, the group that has a larger sample size will dominate the evaluation criteria of the model. Using the discriminant analysis method (PLS-DA) can solve these problems very well [229].

True samples are attributed to Class 1 and false samples are attributed to Class 0. If the real sample is recognized as a true sample, it is recorded as TP (True Positive); if the real sample is recognized as a false sample, it is recorded as FN (False Negative); if a fake sample is identified as a false sample it is recorded as TN (True Negative); if a fake sample is identified as a true sample it is recorded as FP (False Positive). Based on the above, three important parameters can be calculated according to the following equations [230]:

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \qquad \text{Eq.30}$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \qquad \text{Eq.31}$$

$$Accuracy = \frac{TN+TP}{TP+TN+FP+FN} \times 100\% \ \ or \ \ {}^{(Sens.+Spec.)}/_{2} \qquad \text{Eq.32}$$

2.6 Model Transfer Methods

In practical applications of NIR spectroscopy, frequently a calibration model built with the spectra measured on one specific (master) instrument, is not applicable for the data measured on another (target) instrument. The main causes of this model mismatch are the changes in hardware between a master and a target instrument or instrument aging in the case of the same type of instrument. The current main research methods for model transfer include S/B, DS, PDS and SLRDS algorithms.

2.6.1 Slope/Bias (S/B) Algorithm

To standardize the prediction results, the slope/bias correction method is used. The basic idea is to establish a linear relationship between the master and target spectrometers' spectral predictions. Let B be the model's regression coefficient, S_m be the standardized sample's master spectrum, and S_s be the standardized sample's target spectrum [194].

The predicted values of the master spectra are:

$$Y_{m,j} = S_{m,i} \times B \quad \text{Eq.37}$$

The predicted values for the target spectra are:

$$Y_{s,j} = S_{s,i} \times B \quad \text{Eq.38}$$

A one-dimensional linear fit is performed to obtain its least-squares solution as:

$$slope = \frac{\sum(Y_{s,j}-\bar{Y}_s)(Y_{m,j}-\bar{Y}_m)}{\sum(Y_{s,j}-\bar{Y}_s)^2} \quad \text{Eq.39}$$

$$bias = \bar{Y}_m - slope \times \bar{Y}_s \quad \text{Eq.40}$$

Then the predicted value of the concentration of the unknown sample after correction of the target spectrum is:

$$Y_{un} = slope \times (S_{un} \times B) + bias \quad \text{Eq.41}$$

S_{un}: Spectrum of the unknown sample measured with the target spectrometer

Y_{un}: Predicted results after correction

2.6.2 Direct Standardization (DS) Algorithm

The DS algorithm mainly establishes a spectral normalized transition matrix F through the mathematical relationship between the master spectrometer and target spectrometer standard sample set spectra, and uses F to correct the spectrum on the target. The specific steps are as follows:

① Select the standard set from the calibration set samples of the Master and Target, and then

establish the mathematical relationship between the standard set spectral matrices X_m and X_t collected on the Master and Target, and establish the transfer matrix F [195]:

$$X_m = X_t F \quad \text{Eq.42}$$

$$F = X_t^+ X_m \quad \text{Eq.43}$$

X_m: Centered spectral matrix of the standard set measured on the master spectrometer

X_t: Centered spectral matrix of the standard set measured on the target spectrometer

F: Transfer matrix

X_t^+: Generalized inverse matrix of X_t

② With the help of the transfer matrix F, the spectrum measured on the target spectrometer is transformed, thereby reducing the difference in the spectral data of the same sample measured between different spectrometers, so as to achieve the purpose of transferring the model.

$$X_{std} = X_{unknow} F \quad \text{Eq.44}$$

X_{unknow}: Unknown sample spectral matrix measured on the target spectrometer

X_{std}: Spectral matrix after normalizing X_{unknow}

③ Bring the calculated X_{std} into the master model to directly obtain the final predicted value of the spectrum, so as to realize the standardization of the near-infrared spectrum based on DS.

2.6.3 Piecewise Direct Standardization (PDS) Algorithm

The PDS algorithm operates on the same principles as the DS algorithm, with the only difference that PDS separates continuous bands from the spectrum, calculates the transformation coefficient in each window, establishes a spectral standardization matrix based on this coefficient, and uses the transfer matrix to correct the spectrum measured on the target instrument (the difference between DS and PDS algorithms is shown in Fig. 5.4). The coefficient b_i can be calculated from the MLR, PCA, PLS [195]. The specific steps are as

follows:

① Select the absorbance vector $X_{m,i}$ at the i wavelength point on the master spectrometer standard sample set arbitrarily, and also select the i wavelength point in the spectrum of the target spectrometer standard sample set, and select a window width as the $(2\omega + 1)$ of the spectral bands form a matrix $X_{t,i}$, correlate $X_{m,i}$ with $X_{t,i}$, then obtain the conversion coefficient b_i at the i wavelength point,

$$X_{m,i} = X_{t,i} b_i \qquad \text{Eq.45}$$

$X_{m,i}$: The spectral matrix of the standard set of the master spectrometer at wavelength i

$X_{t,i}$: Spectral matrix of the $(2\omega + 1)$ band range of the i wavelength of the standard sample set spectrum of the target spectrometer

b_i: Conversion coefficient at wavelength i

② Loop i to find all b_i, $i = 1\ 2\ 3 \ldots\ldots k$,

k: Wavelength points

③ A transition matrix F is established according to the transform coefficients of each band window,

$$F = diag(b_1^T, b_2^T, \ldots b_i^T, \ldots, b_k^T) \qquad \text{Eq.46}$$

F: Transfer matrix

④ Transform the unknown spectrum measured from the instrument by means of the transfer matrix F,

$$X_{std} = X_{unknow} F \qquad \text{Eq.47}$$

X_{unknow}: Unknown sample spectral matrix measured on the target spectrometer

X_{std}: Spectral matrix after normalizing X_{unknow}

⑤ The final predicted value of the spectrum can be directly obtained by bringing the calculated X_{std} into the master model, thereby realizing the standardization of the spectrum

based on PDS.

2.6.4 Simple Linear Regression Direct Standardization (SLRDS) Algorithm

The SLRDS algorithm assumes that the absorbance between different wavelength points is independent of each other, and uses simple linear regression to correct the spectrum from the target spectrometer [196]. Specific steps are as follows:

① Let the absorbances of the i sample measured on the master spectrometer and the target spectrometer at the j wavelength point be $x_m(i,j)$ and $x_t(i,j)$, respectively, and satisfy the following one linear regression equation:

$$x_m(i,j) = b_0(j) + b(j)x_t(i,j) = [1 \quad x_t(i,j)] \cdot \begin{bmatrix} b_0(j) \\ b(j) \end{bmatrix} \qquad \text{Eq.48}$$

$x_m(i,j)$: The absorbance of the i sample at the j wavelength point measured on the master spectrometer

$x_t(i,j)$: The absorbance of the i sample at the j wavelength point measured on the target spectrometer

$b_0(j)/b(j)$: Regression coefficient corresponding to any wavelength point $j(j \in 1 \ldots p)$

② The above Eq.48 is further written in matrix form, and the regression coefficients are obtained by least squares calculation:

$$x_m(i,j) = [1_{n\times 1} \quad x_t(i,j)] \cdot \begin{bmatrix} b_0(j) \\ b(j) \end{bmatrix} \qquad \text{Eq.49}$$

$$\begin{bmatrix} b_0(j) \\ b(j) \end{bmatrix} = [1_{n\times 1} \quad x_t(i,j)]^{+} \cdot x_m(i,j) \qquad \text{Eq.50}$$

$1_{n\times 1}$: $(n \times 1)$ column vector of all elements are 1

$[1_{n\times 1} \quad x_t(i,j)]^{+}$: The generalized inverse of $[1_{n\times 1} \quad x_t(i,j)]$

③ Calculate the unknown spectrum measured on the target spectrometer :

$$X_{std} = [1_{n\times 1} \quad x_{unknow}(i,j)] \cdot \begin{bmatrix} b_0(j) \\ b(j) \end{bmatrix} \qquad \text{Eq.51}$$

$x_{unknow}(i,j)$: The absorbance of the unknown i sample at the j wavelength point measured

on the target spectrometer

$X_{std}(i,j)$: Spectral matrix after normalizing $x_{unknow}(i,j)$

④ The final predicted value of the spectrum can be directly obtained by bringing the calculated X_{std} into the master model, so as to realize the standardization result of the near-infrared spectrum based on SLRDS.

2.7 Validation and Calibration

Calibration is performed during the development of an NIR spectroscopic analysis method with chemometric evaluation using the reference values and the spectral data of the measured calibration samples. Using linear regression, a statistical correlation is determined in the form of a calibration function. Validation is the process of determining whether an analytical method produces reproducible and reliable results that are accurate enough for the intended application [168, 182, 187].

The calibration models are validated by predicting the target quantities of "unknown" samples (with known reference values) that were not included in the calibration. The suitability of the chosen data pretreatment, wavelength range, and number of principal components of the multivariate calibration are all evaluated. The selection of reference analytics is critical for obtaining reliable calibration results. There are two types of validation: test set or external validation and cross or internal validation [169, 184].

External validation is carried out with the help of an independent test sample set, which does not have to be the same size as the calibration sample set. In this case, the target variables to be determined for both data sets are known (calibration as well as validation). This method necessitates the analysis of relatively large sample sets using a reference method, which is often regarded as a disadvantage due to the higher cost and time involved. For the cross-validation the model is created in the same way as in the test set method, with the exception that the test spectra are taken from the calibration set. When there are few calibration samples available, cross validation is the approach of choice. The samples

analyzed using the reference procedure are used for both calibration and validation of the model in this case by taking one sample at a time from the calibration set and using it as a test object for the variance calculation for calibration and validation. After calibration and prediction, this sample is returned to the calibration set, and another sample is chosen and removed. This process is repeated until each sample has been removed and predicted once from the calibration set. As calibration evaluation criteria, a mean method error RMSEP and a correlation coefficient r are calculated and reported. The number of principal components used for calibration is critical for accurate prediction and model robustness [169, 180, 184-185].

Over-prediction and under-prediction (too many or too few factors) are represented by larger and smaller prediction errors in the calibration data set, respectively, and result in a larger error in the validation or test set. Only the first factors are required to decompose a spectral data set's systematic properties; the following factors describe the calibration spectra's non-systematic properties (e.g., noise, random measurement errors) [180, 184]. The performance parameters listed below are used to assess the robustness of calibration as well as the certainty of future predictions [169, 180, 183, 184].

2.8 Performance Parameters

R (correlation coefficient) is a statistical measure of the strength of the relationship between the relative changes of two variables. It quantifies the relationship between the reference and predicted values. The closer the correlation coefficient is to one, the closer the measured values are to the regression line [131].

Standard deviation:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} \qquad \text{Eq.52}$$

Covariation between x and y:

$$\mathrm{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \qquad \text{Eq.53}$$

Correlation between x and y (correlation coefficient) is:

$$R = \frac{\mathrm{cov}(x, y)}{S_x S_y} \qquad \text{Eq.54}$$

RMSEC (Root Mean Square Error of Calibration) is defined as the calibration's root mean square error. The mean model error is interpreted using this parameter [131].

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad \text{Eq.55}$$

$\hat{y}_i$: Predicted value

y_i: Reference value

i: Sample number

n: Total number of samples in the calibration set

RMSECV (Root Mean Square Error of Cross Validation) is an internal method to 'test' a calibration to make sure it isn't badly skewed by a data point or if the model is over fitted, or if there are any outliers [131].

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad \text{Eq.56}$$

n: Total number of samples in the internal validation set

RMSEP (Root Mean Square Error of Prediction) is a measure of the average prediction error in the quantitative prediction of an unknown sample. This value is an important statistical parameter for assessing the accuracy of future predictions, and it can be calculated using the formula [131]:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad \text{Eq.57}$$

n: Total number of samples in the external validation set

The RMSEC, RMSECV and RMSEP prediction errors are given in the same units as the target variable y. According to Eq.58, the RMSEP can be used to detect data outliers with high SEP (Eq.60) as well as systematic errors in model construction and data structure due to high bias (Eq.61). Low RMSEP values are indicative of good models.

$$RMSEP^2 = SEP^2 + Bias^2 \quad \text{Eq.58}$$

SEC (Standard Error of Calibration) is the regression's standard error. The SEC is used in the calibration set to calculate the difference between the predicted and reference values of the target variable and to make a statement about how well the calibration equation fits the data. The SEC value should be slightly less than the SEP value [131].

$$SEC = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n - p - 1}} \quad \text{Eq.59}$$

p: Number of principal components

n: Total number of samples in the calibration set

SEP (Standard Error of Prediction) is a model inaccuracy measure defined as the error spread or standard deviation of the predicted values of the validation set [126, 131].

$$SEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i - Bias)^2}{n - 1}} \quad \text{Eq.60}$$

According to Eq. 61, the SEP is made up of a random error (SEC) and a systematic error (bias). When the bias in Eq. 58 approaches 0, the SEP is ideally equal to the RMSEP. The unit

of the SEC and SEP values is the same as that of the reference method data.

Bias represents the systematic error and indicates the extent to which the actual and predicted values of all samples in the validation set differ [126, 131].

$$Bias = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n} \qquad \text{Eq.61}$$

RPD (Relative Prediction Deviation) is defined as the standard deviation of observed values divided by the Root Mean Square Error or Prediction (RMSEP). The RPD takes both the prediction error and the variation of observed values into account, providing a metric of model validity that is more objective than the RMSEP and more easily comparable across model validation studies [231].

$$RPD = \frac{1}{\sqrt{1 - R^2}} \qquad \text{Eq.62}$$

Slope and **Offset**. The ordinate intercept (offset) and slope of the straight line from the calculated linear regression between the reference target values and the predicted values are denoted by these parameters. They are provided for both the calibration and test sets. In an ideal case, offset is 0 and slope is 1.

3 Experimental

The quality of flour directly affects the processing of the final flour products, their quality, and the taste or flavor of these products when they are eaten. The content of moisture, ash, starch, protein, and wet/dry gluten are important indicators that reflect the quality of flour. This chapter introduces the use of near-infrared spectroscopy to establish a rapid determination method for the content of moisture, protein, and wet gluten in flour. The purpose is to quickly evaluate flour quality and provide theoretical data for improving the processing quality and flavor characteristics of flour products.

3.1 Preparation of Test Samples and Chemical Value Parameters

3.1.1 Overview of Wheat Flour Samples

For the experiments and investigations described in this thesis, 50 wheat flour samples were acquired in Germany and 163 wheat flour samples were gathered from all over China. The wheat flour samples from Germany were purchased from DIGeFa GmbH (Detmolder Institut für Getreide- und Fettanalytik GmbH), and the reference values of the relevant parameters were also provided by DIGeFa GmbH; most of the Chinese wheat flour raw materials came from the three growing regions in China (as shown in Fig. 3.1).), and only two wheat flours were sourced from other countries (one sample from Kazakhstan and one from Russia).

Fig. 3.1 Distribution of wheat producing areas in China

The Chinese wheat flour samples come from different brands sold in various regions, including 4 types of low-/medium-/high-gluten wheat flour, and whole wheat flour. All 213 Chinese and German flour samples were tested by laboratories in China, and subsequently sealed and refrigerated for later use. The classification of all wheat flour samples is summarized in Tab. 3.1.

Tab.3.1 Classification of flour samples under investigation

Countries	Number of Samples	Classification of Samples	
CHN	163	Spring wheat region	37
		Northern winter wheat region	81
		Southern winter wheat region	45
GER	50	Weizenmehl Type 550	30
		Weizenvollkornmehl	10
		Weizenmehl Type 405	6
		Weizenmehl Type 1050	4

3.1.2 Comparison of the Analytical Methods used for the Determination of the Characteristic Flour Parameters in China and Europe

The determination of the contents of crude protein, wet gluten, moisture, ash and other indicators in wheat flour is an important step in the quality control of wheat flour by NIR spectroscopy. However, Europe and China have partly different analytical methods for the determination of the reference values which are the basis for the development of calibration models with NIR spectra.

3.1.2.1 Determination of Crude Protein

The determination method of crude protein is generally consistent internationally. Both the European standard (ICC 167:2000) [197] and the Chinese standard (GB/T 5009.5-2016) use the Dumas combustion method to determine the crude protein content of flour [198]. The only

difference is that different testing instruments are used for the two standards. The testing instrument used in the European standard is Dumatec™ 8000 (FOSS Analytical A/S, Foss A/S, Hillerød, Denmark), and the Chinese testing agency uses a D500 Dumas nitrogen analyzer (Hanon Future Technology Co., Ltd., Jinan, Shandong Province, China).

The principle of the Dumas combustion method for the determination of crude protein content is as follows:

The sample is heated and burned in the combustion tube and then converted into gas. Nitrogen-containing substances are converted into molecular nitrogen, and the interfering gas is absorbed and removed by a series of absorbents, and finally detected by a thermal conductivity detector. To obtain the crude protein content, the nitrogen content of the sample is multiplied by 6.25 [197].

3.1.2.2 Determination of Wet Gluten

The content of wet gluten is an important quality indicator of flour. Wet gluten is a viscoelastic substance that is mainly composed of the two protein components of wheat (glutenin and prolamin). Of the traditional methods for detecting wet gluten, hand washing is the most primitive. The European Standard (ICC 155:1994) from 1994 developed new methods for the detection of wet gluten using professional gluten instruments [199]. However, it was not until 2008, that China officially added the method of measuring wet gluten by a gluten instrument to the national standard (GB/T 5506.2-2008) [200].

Both standards use the same detection method but use different gluten instruments. The Glutomatic 2200 Gluten Instrument (Perten Instruments, Perkin Elmer, Waltham, Massachusetts, U.S.A.) was used in the European Standard, and the PG-2850 Gluten Measuring Instrument (Perkone Scientific Co., Ltd., Hangzhou, Zhejiang Province, China) was used in the Chinese Standard. Wet gluten is tested by adding a sample to a sodium chloride solution to make a dough, and subsequently letting it sit for a period of time. The purpose of this step is to form a network of gluten. The dough is then washed with a sodium

chloride solution to remove unwanted substances such as starch and excess water from the dough. After the gluten is separated, the wet gluten content is generally regarded as the percentage of the separated wet gluten mass in the wheat flour sample mass [199].

3.1.2.3 Determination of Moisture

The determination of moisture commonly uses a drying method. The basic principle is to measure the dry weight of the sample after drying at high temperature, including hygroscopic water, part of crystal water and substances that are volatile under this condition. Finally, the moisture content is calculated by weighing the values before and after drying.

However, there are two differences in the European Standard (ICC 110/1:1976) [201] and the Chinese Standard (GB/T 5009.3-2010) [202]. First, the two standards use different testing instruments, the Moisture Tester MT-CA (Brabender Messtechnik GmbH & Co. KG, Duisburg, NRW, Germany) is used in the European standard. In the Chinese standard, the RSD-252Z precision vacuum drying oven (Kunshan Rongshida Electronic Equipment Co., Ltd., Suzhou, Jiangsu Province, China) is used as testing instrument. Second, the settings of the instrument parameters in the two standards are different. The European standard sets the temperature conditions between 130 and 133 ℃, while the Chinese standard sets the temperature conditions between 101 and 105 ℃ [201].

3.1.2.4 Determination of Ash

The inorganic residues remaining after the flour is burned are called ash. Regarding the determination of ash, the European standard (ICC 104/1:1990) [203] and the Chinese standard (GB/T 5009.4-2016) [204] are consistent. According to the standard detection process, firstly, an appropriate amount of wheat flour sample needs to be weighed, and then 1 mL of magnesium acetate solution (240 g / L) is added to completely wet the sample. After standing for 10 minutes, the moisture of the wetted sample was evaporated to dryness. Heat the sample

on a hot plate with low heat to fully carbonize the sample until it is smoke-free. Immediately place the crucibles in a muffle furnace and raise the temperature to 900 ℃ ± 25 ℃. Keep this temperature until the carbon disappears, then reduce the temperature to about 200 ℃, and then place the sample in a desiccator for cooling before weighing [203].

3.2 Spectrometers used in the Experiments

3.2.1 Benchtop Spectrometers

Two benchtop spectrometers were selected for the measurement of the flour samples, and their basic parameters are shown in Tab. 3.2.

Tab. 3.2 Basic parameters of the two benchtop spectrometers used for the present work

Spectrometer Name	Company	Monochromator	Wavelength Range (nm)	Spectral Resolution	Signal/Noise Ratio
IAS 3100	Intelligent Analysis Service,Wuxi,China	MEMS + Grating Scan	950-1650	16 nm	8000:1
NIR Freespace	Shanghai Space-OE Technology Co.Ltd,Shanghai,China	Acousto Optic Tunable Filter	1100-2300	2-10 nm	8000:1

3.2.1.1 IAS 3100 Spectrometer

The IAS 3100 NIR spectrometer is an instrument for rapid analysis of small particles, powders, pastes and other solids by using a MEMS + grating scanning method. MEMS + grating as the core component of the instrument, uses the micro-mirror scanning method to achieve wavelength selection output, and requires only a single point detector to obtain the complete spectrum. The instrument measures solid samples by diffuse reflection, and introduces an automatic reference material (gold-plated reflector) and automatic wavelength calibration system, which can quickly self-test and calibrate the system.

The wavelength range of the spectrometer is 950-1650 nm, the resolution of the instrument is less than 16 nm, the wavelength interval is 1 nm, the signal-to-noise ratio is more than 8000:1.

The instrument has a built-in industrial control ARM processor, an embedded Linux system and self-developed spectral analysis software. The instrument is shown in Fig. 3.2.

Fig 3.2 IAS 3100 NIR Spectrometer

3.2.1.2 NIR-Freespace

The NIR-Freespace spectrometer uses an AOTF technology as measurement principle and has two InGaAs detectors to ensure the stability and synchronization of test data. In addition, the instrument has a built-in reference making it unnecessary to measure a separate reference spectrum before testing the sample. For each measurement, the spectrometer collects both the reference and sample spectra, and then uses an algorithm to obtain the reflectance/absorbance of the sample relative to the standard reference. The spectrometer is equipped with two rotating motors. The wavelength range of the NIR-Freespace is 1100-2300 nm, the resolution of the spectrometer is 2-10 nm, and the datapoint wavelength interval is 1 nm. The wavelength scan speed is > 4000 wavelength points/sec, and the S/N ratio is > 8000:1. A photo of the instrument is shown in Fig. 3.3.

Fig 3.3 The NIR-Freespace AOTF Spectrometer

3.2.2 Handheld Spectrometers

Four handheld spectrometers were selected for this project, and their basic parameters are summarized in Tab. 3.3.

Tab. 3.3 Instrumental parameters of the four handheld NIR spectrometers used for the present work

Spectrometer Name	Company	Monochromator	Wavelength Range (nm)	Spectral Resolution	Signal/Noise Ratio
MicroNIR	VIAVI Solutions Inc.,San Jose,USA	Linear Variable Filter	908-1676	〈12.5 nm	5324:1
NIRONE Sensor S 2.0	Spectral Engines Oy,Helsinki,Finland	Fabry Perot Tunable Filter	1550-1950	〈18 nm	9600:1
Neospectra Micro	Si-Ware Systems,Cairo,Egypt	MEMS FTNIR	1298-2606	8 nm/1550 nm	2820:1 （self-test）
C15511-01	Hamamatsu Photonics,Hamamatsu City,Japan	MEMS FTNIR	1100-2500	5.7 nm/1533 nm	1888:1 （self-test）

3.2.2.1 NeoSpectra Micro

The NeoSpectra Micro, a single-chip FT-IR spectrometer, uses its unique platform-Silicon Integrated Micro-Optics System Technology (SiMOST™), which allows the creation of multiple optical components on silicon. As a miniature, low-cost spectral sensor and scanner, the NeoSpectra Micro can be used to detect a wide range of materials and substances. The NeoSpectra Micro handheld spectrometer offers similar instrument performance parameters to those of benchtop laboratory NIR spectrometers, but the NeoSpectra Micro is much smaller and less costly. Fig. 3.4 shows a photo of the experimental set-up for flour measurements using the NeoSpectra Micro scanner in tandem with the IAS 3100 spectrometer. The NeoSpectra scanner is based on the MEMS technology with a wavelength range from 1298-2606 nm, a spectral resolution of 8 nm / 1550 nm, a wavelength datapoint interval of 9 nm, and a signal-to-noise ratio of 2820:1 (result of an in-house test). The optical head serves as the light source for sample illumination and collects radiation diffusely reflected by the

sample. A proprietary performance-optimized design with Application-Specific Integrated Circuits (ASICs) for system control and data processing aims to reduce the number of external components. The optical core module of the spectrometer uses a MEMS Michelson interferometer and a single InGaAs detector.

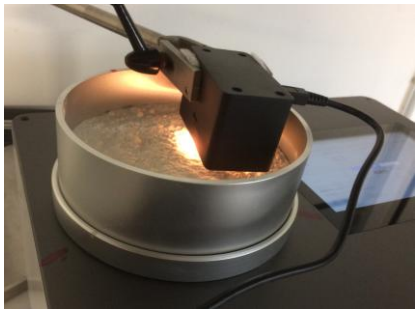


Fig 3.4 Photo of the experimental set-up of a flour measurement with the NeoSpectra Micro scanner in tandem with the IAS 3100 NIR spectrometer measurement

3.2.2.2 NIRONE Sensor S 2.0

The NIRONE Sensor series is manufactured by Spectral Engines oy (Finland). The NIRONE Sensor S 2.0 uses the patented MEMS of a Fabry-Perot interferometer, which is a fully programmable optical filter. The sensor can be driven over the entire wavelength range or can operate at a specific set wavelength range. Fig. 3.5 shows a photo of the experimental set-up used to measure flour simultaneously by the NIRONE Sensor S 2.0 and the IAS 3100 benchtop instrument. The NIRONE Sensor S 2.0 has a single point detector rather than a linear array, making it an economical and affordable solution for all applications. The use of a single detector and the Fabry-Perot interferometer technique allows the use of a larger detector area than a linear array. Its wavelength range is 1550 nm to 1950 nm, the instrument resolution is 18 nm, the wavelength data point interval is 5 nm, and the signal-to-noise ratio is 9600:1.

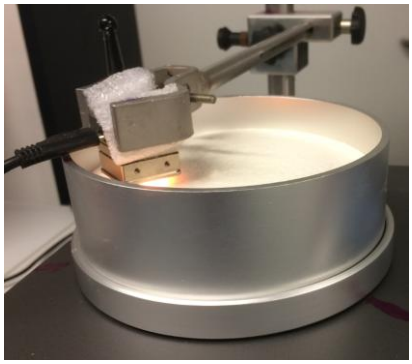


Fig 3.5 Photo of the experimental set-up of a flour measurement with the NIRONE Sensor S 2.0 in combination with the IAS 3100 NIR spectrometer

3.2.2.3 MicroNIR

The MicroNIR uses a LVF as the core dispersive element, and it also uses advanced thin-film coating design and manufacturing techniques. With its compact form factor and superior performance advantages, the MicroNIR spectrometer is a simple and practical spectrometer, which allows for wider adoption and greater flexibility in installation and scalability of NIR spectroscopy solutions. The spectrometer has a sapphire window, Anti-Reflection (AR)-coated on one side for maximum light throughput. Fig. 3.6 shows a photo of the experimental set-up using the MicroNIR and the IAS 3100 spectrometers in combination. The MicroNIR has a wavelength range from 908 nm to 1676 nm, its spectral resolution is 12.5 nm, the wavelength datapoint interval is 6 nm, and the signal-to-noise ratio is 5324:1(self-test result).

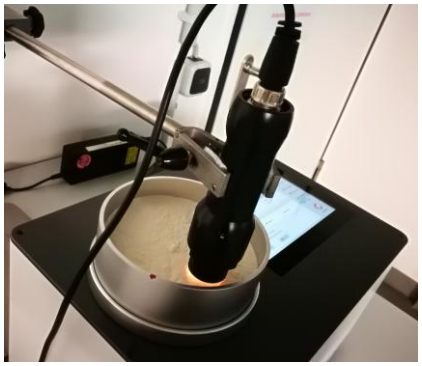


Fig 3.6 Photo of the experimental set-up of a flour measurement with the MicroNIR in combination with the IAS 3100 NIR spectrometer

3.2.2.4 Hamamatsu C15511-01

The Hamamatsu C15511-01 is a miniature FT-NIR spectrometer processed by using the MEMS technology with a Michelson optical interferometer as the core element. The optical interferometer has a built-in optical input section, a beam splitter, a fixed reflector, a movable reflector (Φ 3 mm), and a photodetector. The photodetector acquires the light intensity signals (interference information) that vary with the position of the movable reflector, and then processes (Fourier transforms) these light intensity signals to obtain the spectral signals. A semiconductor laser, Vertical Cavity Surface Emitting Laser (VCSEL), is built into the instrument to monitor the position of the movable reflector, enabling spectral measurements with high wavelength accuracy. The Hamamatsu C15511-01, like the NeoSpectra Micro, has a MEMS FT-NIR as the core of its spectrometer. It has a wavelength range from 1100 to 2500 nm, the instrument resolution is 5.7 nm/1533 nm, the datapoint interval is 10 nm, and the signal-to-noise ratio is 1888:1(self-test result). The individual components of the spectrometer are shown in Fig. 3.8.

3.3 Experimental Measurement and Evaluation Process

The process of spectra measurements and calibration model development for the analysis of the flour parameters is summarized in Fig. 3.7. The individual steps of collection and grinding of the raw material wheat to the NIR spectroscopic measurement of the flour samples, and finally the processing and analysis of spectra to develop calibration models are shown in the flow chart (See Fig. 3.7).

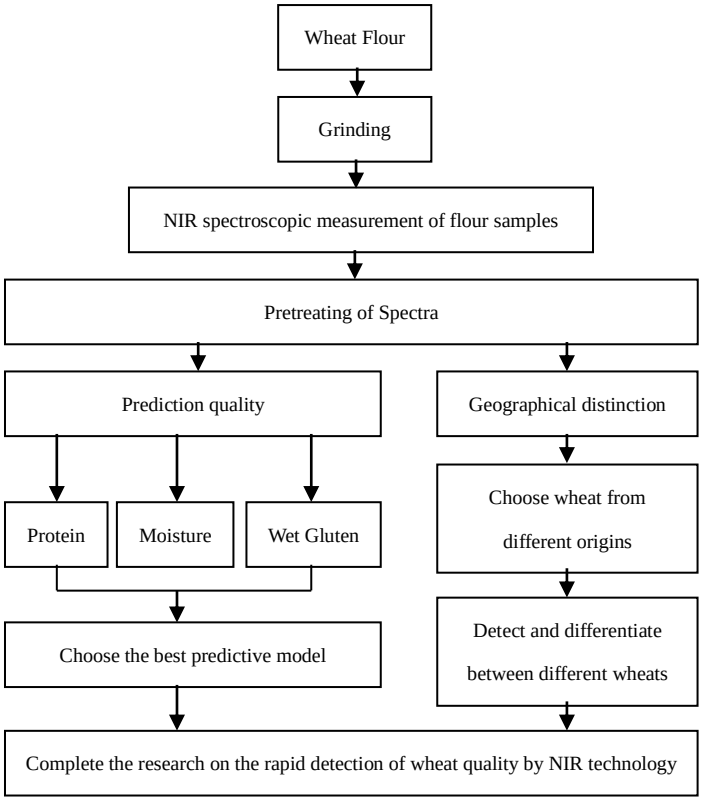


Fig. 3.7 Flow chart of the NIR spectroscopic measurement and calibration model development for the analysis of flour parameters

The whole experimental process was divided into three temporal stages: stage I was from July 2019 to September 2019 at the Department of Chemistry (University of Duisburg-Essen, Germany) and comprised the operation of 3 handheld spectrometers: NeoSpectra, NIRONE Sensor S 2.0, and MicroNIR and the IAS 3100 benchtop spectrometer for the measurement of 50 German flour samples. During stage II (September 2020), the NIR spectra of 50 German flour samples were measured with the Hamamatsu C15511-01 miniature FT-NIR spectrometer and the IAS 3100 benchtop spectrometer in the same location. In stage III a measurement campaign was run from October 2020 to December 2020 at the School of Food and Biological Engineering (Jiangsu University, China) by using four handheld spectrometers (NeoSpectra, NIRONE Sensor S 2.0, MicroNIR, and Hamamatsu C15511-01) for the 50 German and 163 Chinese flour samples, and the benchtop spectrometer (NIR-Freespace) for a total of 163 Chinese flour samples only.

3.3.1 Experimental Stage I

The stage I experiments were conducted using an IAS 3100 benchtop spectrometer as the basic operating station, in parallel with 3 handheld spectrometers (MicroNIR, NeoSpectra, NIRONE Sensor S 2.0), respectively, and additionally providing a rotation of the sample tray,. The experimental site was in the basement corridor of the Department of Chemistry, in order to maintain a relatively stable temperature and humidity environment for the NIR spectroscopic measurements, as it was very hot during the July summer month. Thereby, the experimental temperature environment was maintained between 22 ℃ and 23 ℃. The stage I experiment was divided into 3 parts, and a total of 3 measurement series (IAS 3100 and Neospectra Micro, IAS 3100 and MicroNIR, IAS 3100 and NIRONE Sensor S 2.0) were performed by using the IAS 3100 benchtop spectrometer simultaneously with the 3 handheld spectrometers, respectively. The flour samples were poured into the sample tray of the IAS 3100 spectrometer, and spread evenly in order to keep the surface as smooth as possible. Afterwards, the same sample was measured simultaneously with two spectrometers (the IAS 3100 benchtop spectrometer (bottom up) and the handheld spectrometer (top down)). The handheld spectrometer detection is started 30s after the IAS 3100 spectrometer detection process. Since the sample dish is rotated, the handheld spectrometer detects a moving sample area instead of a fixed point. Testing the same sample 3 times requires moving the sample tray two additional times, each time with a rotation angle of 120 °. This allows three sets of spectral data to be tested for each flour sample with subsequent storage for data classification, calibration and validation.

A total of 50 flour samples were tested. The Neospectra and MicroNIR sensors were mounted approximately 1 cm above the surface of the flour samples, while the NIRONE Sensor S 2.0 sensor was approximately 5 mm from the surface of the flour samples. In addition, the reference material used for the 3 handheld spectrometers was the reference plate included with the NeoSpectra NIR scanner. Since the IAS 3100 spectrometer has a circular rotating tray with a glass bottom above an instrument window facing upwards, the fitting of the handheld spectrometers with their sensor face down above the rotating sample surface

allowed to measure the same sample simultaneously with two spectrometers. The measurements of the 50 German flour samples were performed in three stages by combining the benchtop instrument successively with one of the three handheld instruments (NeoSpectra Scanner, MicroNIR and NIRONE Sensor S 2.0), respectively.

The IAS 3100 has a programmed measurement time of 1 min (with a preparation time of 30s and a scan time of 30s). To keep up with the IAS 3100 scan time, two handheld NIR spectrometers (Neospectra and MicroNIR) were set to a scan time of 30s and the NIRONE Sensor S 2.0 spectrometer was set to a scan time of 29s). Thus, after the 30s preparation time of the IAS 3100 spectrometer, the scan time of the three handheld spectrometers was started simultaneously with the scan time of the IAS 3100 spectrometer to measure the NIR spectra of the flour sample under investigation. The warm-up time before measurement start was 1h for the IAS 3100 spectrometer and 0.5h for the 3 handheld spectrometers.

Figs. 3.4-3.6 show the experimental set-up for the measurements of stage I.

3.3.2 Experimental Stage II

The IAS 3100 benchtop spectrometer was used as the basic operating station for this stage of the experiments. As described above, this spectrometer provides a rotating sample tray, and the Hamamatsu C15511-01 handheld spectrometer was mounted above the sample surface for simultaneous testing.

The difference between the Hamamatsu C15511-01 spectrometer and the previous three handheld spectrometers is that the Hamamatsu C15511-01 spectrometer currently provides only its spectrometer module. The module needs to be connected to an external light source, and the external light source has to be combined with one side of a Y-shaped optical fiber, and the other side of this optical fiber needs to be connected to the spectrometer module so that the optical signal of the sample can be transmitted to the spectrometer module. In comparison, the light source and spectrometer module of the other three handheld spectrometers have been integrated into one device. The experiments were conducted under similar temperature

conditions as for experimental stage I between 20-22 ℃. The measurement procedure was also the same as in section 3.3.1, and the reference material used for the Hamamatsu C15511-01 handheld spectrometer was the same reflection standard as used for the Neospectra scanner. The warm-up time was 1h for the IAS 3100 spectrometer and 0.5h for the Hamamatsu C15511-01 handheld spectrometer. Fig. 3.8 shows the experimental set-up of the combined measurements with the IAS 3100 benchtop and Hamamatsu C15511-01 handheld spectrometers.

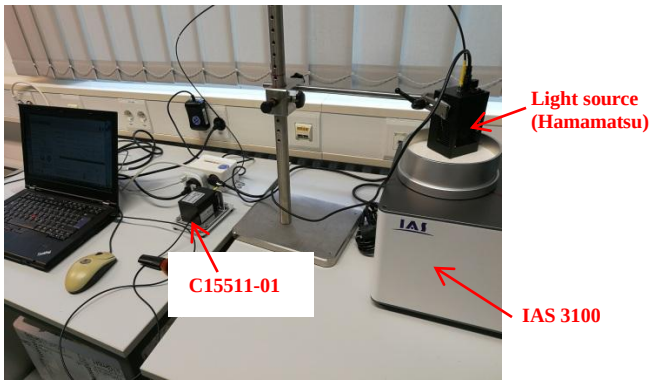


Fig 3.8 Photo of the experimental set-up of the combined measurements with the IAS 3100 benchtop and the Hamamatsu C15511-01 handheld spectrometers.

3.3.3 Experimental Stage III

In these experiments two benchtop spectrometers (IAS 3100 and NIR-Freespace) were used in combination with four handheld spectrometers to measure the Chinese flour samples in the third stage of this work. In 163 Chinese flour samples (only 154 Chinese flour samples were selected by NIR-Freespace because the sample volume of the other 9 samples was too small to meet the minimum measurement standard for sample volume in the detection tray) the three parameters protein, moisture and ash were analyzed. For wet gluten the reference values were only available for 159 samples.

The experimental work for stage III was conducted in the School of Food and Biological Engineering (Jiangsu University, Zhenjiang, Jiangsu Province, China). The environmental temperature was stabilized by air condition at about 22 ℃ in order to be consistent with the

previous measurement conditions in Germany.

In order to rationalize the measurements, the experimental part of stage III was divided into two parts. One benchtop/handheld spectrometer combination consisted of the IAS 3100 spectrometer and the two handheld instruments MicroNIR and NeoSpectra， the other combination comprised the NIR-Freespace spectrometer and the Hamamatsu C15511-01 and NIRONE Sensor S 2.0 handheld instruments. The handheld spectrometers were positioned top down over the sample trays of the benchtop spectrometers. The operating and sample presentation conditions in these experiments were analogous to the experimental conditions of stage I. For the four handheld spectrometers a poly(tetrafluoroethylene) (PTFE) diffuse reflection standard of the Hanon Company (Jinan, Shandong Province, China) was used.

For the NIR-Freespace benchtop instrument the operating time was set to 45s, (15s preparation time and 30s scan time). To match the NIR-Freespace`s scan time, both handheld NIR spectrometers (Hamamatsu C15511-01 and NIRONE Sensor S 2.0) were set to 29s measurement time. After the NIR-Freespace spectrometer's preparation time, both handheld spectrometers were started simultaneously to the NIR-Freespace spectrometer's scan time. Fig. 3.9 shows the experimental set-up of these measurements.

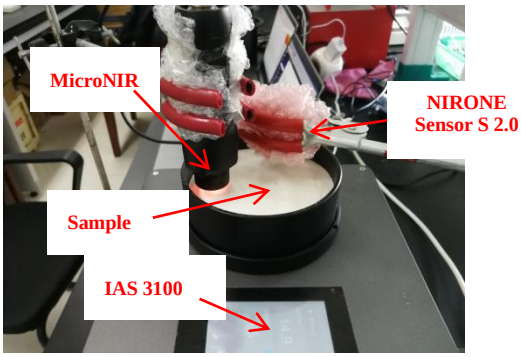


Fig 3.9 Photo of the experimental set-up of the combined measurements with the IAS 3100 benchtop and the MicroNIR and the NIRONE Sensor S 2.0 handheld spectrometers.

3.4 Acquisition of Spectral Data

Since the format of the spectral data obtained from the acquisition software of the different

instruments is not consistent, the format of the respective spectral data needs to be standardized before subsequent spectral data processing and calibration model development.

Each spectrometer has its own spectral acquisition software to complete the workflow of the spectra measurement. The first step is to acquire a background spectrum of the reflection standard, and then the spectrum of the sample is measured. From these data either a percent reflectance (%R) or absorbance (log 1/R) spectrum is calculated and stored in the computer. The spectral data collected by the benchtop spectrometers IAS 3100 and NIR-Freespace are automatically stored on a disk inside the instrument, and can be exported via USB or mobile flash drive for import into a computer.

The handheld spectrometers used in this work have different internal structures and due to their small size some of them do not have the ability to store data and the spectra collected by the instrument need to be transferred to a computer for storage and further processing. The common working situation is to install the software for the handheld spectrometer on a computer, tablet or cell phone and connect these with a data cable or via Bluetooth with the spectrometer.

3.5 Spectrum Processing Analysis Software

NIR spectroscopy is a so-called secondary analysis technique and therefore requires qualitative or quantitative calibration models to relate the NIR spectra of the samples to the chemical or physical reference values. The raw spectra measured by the spectrometers contain not only the physicochemical information of the sample, but also a large amount of interfering information, that is not related to the reference values of the sample under investigation. The data processing of NIR spectra requires software for spectral pretreatment and development of chemometric calibration models. Thereby the lack of “fingerprint character” of mid-IR spectra can be overcome and the purpose of qualitative and quantitative analysis of the sample is finally achieved.

The main spectral processing analysis software used in this project is the Unscrambler™ (Version 9.7, Camo Analytics AS, Aspen Technology, Inc., Bedford, Massachusetts, USA),

MATLAB (2016a, The MathWorks, Natick, Massachusetts, USA) and NIRSA (Version 5.8.8, NIR Research Laboratory of Jiangsu University/Automation Department of the School of Light Science and Engineering of Nanjing Forestry University, Zhenjiang/Nanjing, Jiangsu Province, China).

3.5.1 Unscrambler™ 9.7

The Unscrambler™ 9.7 is a multivariate data analysis software released by Camo Analytics AS, Inc. and provides users with a tool that meets the needs for qualitative and quantitative data analysis, thereby enhancing the ability to handle process data and providing the possibility for spectral calibration.

3.5.2 MATLAB R2016a

The MATLAB R2016a software allows visualization, numerical computation and programming to be centralized in a simple open environment with dimensionless matrices as its basic data unit. For NIR spectroscopy the MATLAB software can be used for chemometric-related algorithm writing and analysis of spectral data as well as their visualization.

3.5.3 NIRSA 5.8.8

The NIRSA 5.8.8 software integrates a variety of data processing algorithms and is mainly applied to the qualitative and quantitative analysis of NIR spectroscopic data. Since its development in 2004, this software has been improved and matured.

3.6 Processing Steps after Spectra Measurement

3.6.1 Selection of Pretreatment Methods for Raw Spectra

Since the acquired raw spectra are often accompanied by scatter effects, random noise, band overlap and baseline drifts, spectral pretreatment is an indispensable step in the calibration modeling process in order to eliminate external interferences. A reasonable pretreatment method can effectively filter out the noise information in the spectral signal and retain the effective information, thus solving the complexity problem of the calibration model. To optimize the prediction accuracy of the developed calibration models, this work used pretreatment methods like SNV, EMSC, Savitzky-Golay Smoothing Algorithm (SGSA) and First-Order Derivative (1^{st} Der.) to pretreat the original spectra.

3.6.2 Selection of Effective Spectral Range and Screening of Outliers

Since the raw spectral data contain a lot of interfering information, direct analysis using raw spectra can increase the calculation complexity and take a lot of time, and may contain redundant information that is not relevant to the chemical reference values. Therefore, in order to build fast and accurate calibration and prediction models, it is necessary to extract the valid spectral ranges related to the chemical reference values and to reject the outliers. In this work mainly the plot of leverage versus residuals ("influence plot") of the Unscrambler™ 9.7 software was used to detect outliers. Specific examples of analysis results are described in Sections 4.1.3 and 4.1.4.

3.6.3 Separation of Available Samples in Calibration and Prediction Set

The separation of the available samples into a calibration and a prediction set has an important impact on the quality of the calibration model. The basic requirement for selecting the calibration and prediction set samples is to ensure a representative distribution of the reference values in both sets. In this work, Sample Set Partitioning based on the joint X-Y distance (SPXY) algorithm of the NIRSA software is used to select 10% of the total number

of samples as the prediction set and the remaining 90% as the calibration set.

3.6.4 Calibration and Prediction Model Development

After completing the three steps described in Sections 3.6.1-3.6.3, the corresponding calibration models were built and the prediction sample sets were predicted by these models. For qualitative and quantitative analysis primarily two multivariate methods, PCA and PLS regression, respectively, were used in the present work.

4 Results and Discussion

After completing the experiments according to the test steps and procedures described in Chapter 3, this section presents the data analysis and discussion of the raw spectra obtained from the flour tests with all the spectrometers. The first step is the basic screening of the raw spectra with reference to the spectrometer hardware and performance followed by optimization of the spectra.

4.1 Screening and Optimization Process for the Spectra

4.1.1 Initial Screening of Spectrometers

The quality of the spectra is related to many factors, such as the different optical principles of the spectrometers, their different performance parameters (e.g., high or low signal-to-noise ratio, high or low spectral resolution, different spectral detection range, etc.), and changes of the investigated samples (e.g., high or low concentration range, changes in sample composition, etc.). All these factors have an impact on the NIR spectra, and even the changes in the measurement environment can lead to changes of the measured spectra. Thus, choosing the right spectrometer for flour testing is important to ensure the quality of the spectrum.

The signal-to-noise ratio, spectral resolution, and accuracy of absorbance values of handheld spectrometers are generally lower than those of benchtop spectrometers. The spectrum of a flour sample recorded over the whole NIR wavenumber range (800-2500 nm/12500-4000 cm^{-1}) is used as standard spectrum. The Chinese national standard for flour NIR spectra, for example, uses the full spectral range of the Lengguang S450 benchtop spectrometer (800-2500 nm/12500-4000 cm^{-1}) and is the basis of the comparison of the 2 benchtop spectrometers (IAS 3100, NIR Freespace) and the four handheld spectrometers (MicroNIR, Neospectra Micro, Hamamatsu C15511-01, NIRONE Sensor S 2.0) used in the present work. In Fig. 4.1 the NIR spectra measured of the same flour sample with the different spectrometers are shown. In this comparison it is clearly demonstrated, for example, that the NIRONE Sensor S 2.0

spectrometer has the narrowest spectral range of all instruments (spectrum marked in red). This narrow spectral range makes it difficult to ensure, that the most effective spectral information will definitely fall in that wavelength range, which can lead to a smaller amount of information obtained from the flour test object for effective qualitative and quantitative analysis.

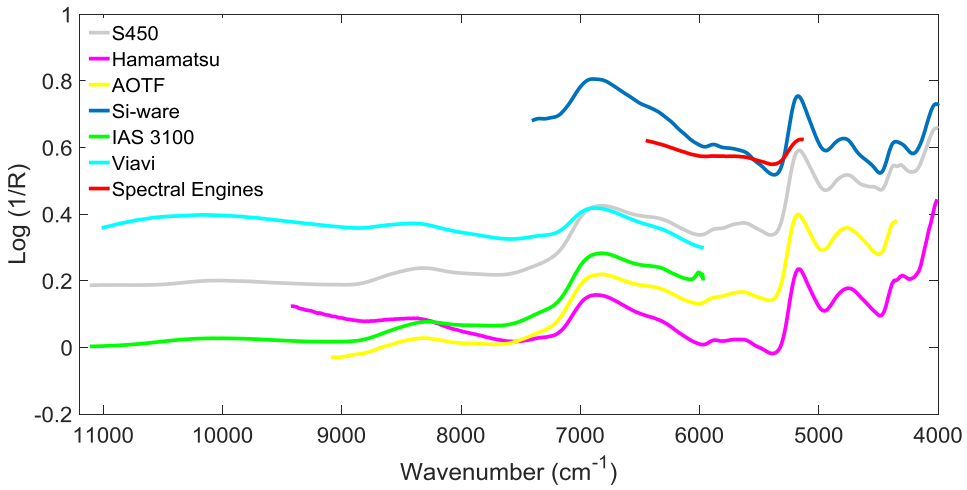
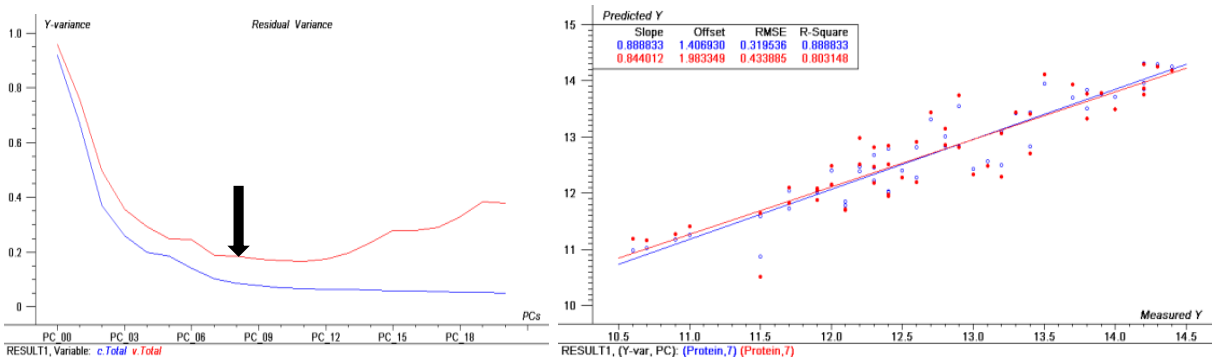


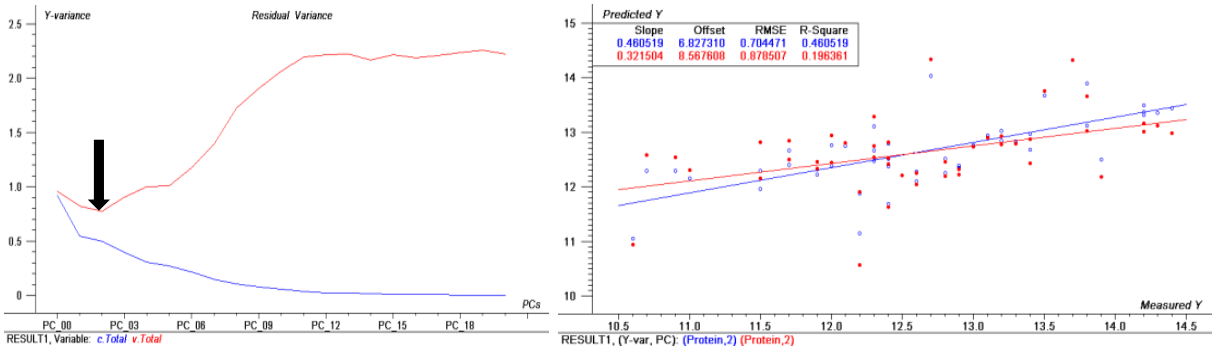
Fig. 4.1 Comparison of the NIR spectra of the same flour sample measured with the full NIR spectral range of the Lengguang S450 benchtop spectrometer (Grey spectrum), the IAS 3100 and NIR Freespace (AOTF) benchtop spectrometers and the four handheld spectrometers (MicroNIR , Neospectra Micro , Hamamatsu C15511-01, NIRONE Sensor S 2.0)

For the measurements in Germany (Stage 1) and China (Stage 3) different spectrometers of the same instrument type (benchtop and handheld) were used, because the instruments could not be exchanged due to customs restrictions. Furthermore, while German flour samples could be sent to China for test purposes, the Chinese flour samples could not be sent to Germany. To prove the low-quality of the NIR spectral data acquired with the NIRONE Sensor S 2.0 instruments, a comparison of calibrations was performed on the 50 German flour samples (which were available in both countries) using the different handheld spectrometers (NIRONE Sensor S 2.0 and MicroNIR) used in Germany and China, respectively. The PLS calibration results obtained for protein with the NIR spectra of both instruments after the same spectral pretreatment method (EMSC) are summarized in Fig. 4.2. To exclude errors due to sample presentation, both instruments were fixed above the same sample and the spectra were measured simultaneously (as shown in Fig. 3.9).

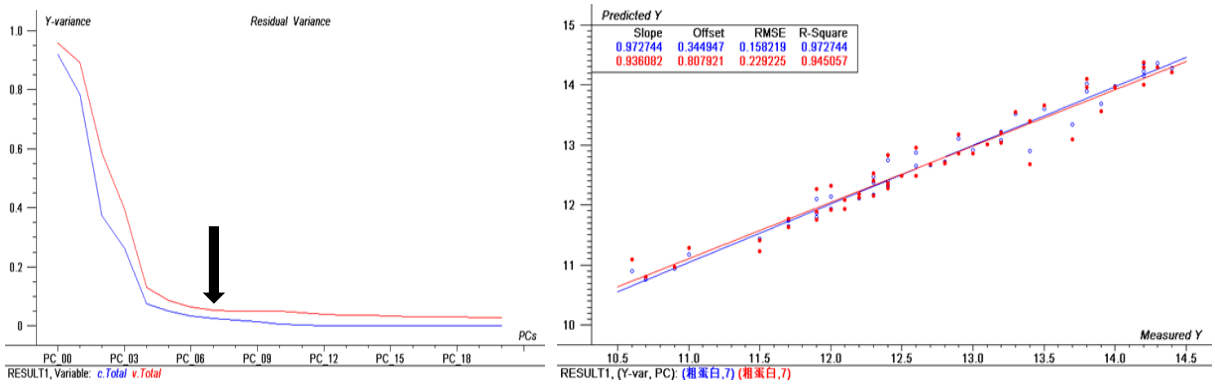
Fig. 4.2 reflects, that despite small differences the PLS results for protein derived from the MicroNIR spectra measured during Stage 3 are similar to those of Stage 1 (in terms of the number of factors and the RMSE and R^2 values), while the results achieved with the data of the NIRONE Sensor S 2.0 spectrometer during Stage 3 measurements are extremely different from Stage 1 results and of comparatively low quality for both measurement procedures. These discrepancies can be explained by the fact, that many spectroscopic footprints of proteins occur primarily in the wavenumber range 5000-4000 cm^{-1}, which is not available for the detection range of the NIRONE Sensor S 2.0 instrument (6452-5128 cm^{-1}).Therefore, it has been decided to exclude the NIRONE Sensor S 2.0 spectrometers from the further analysis process.



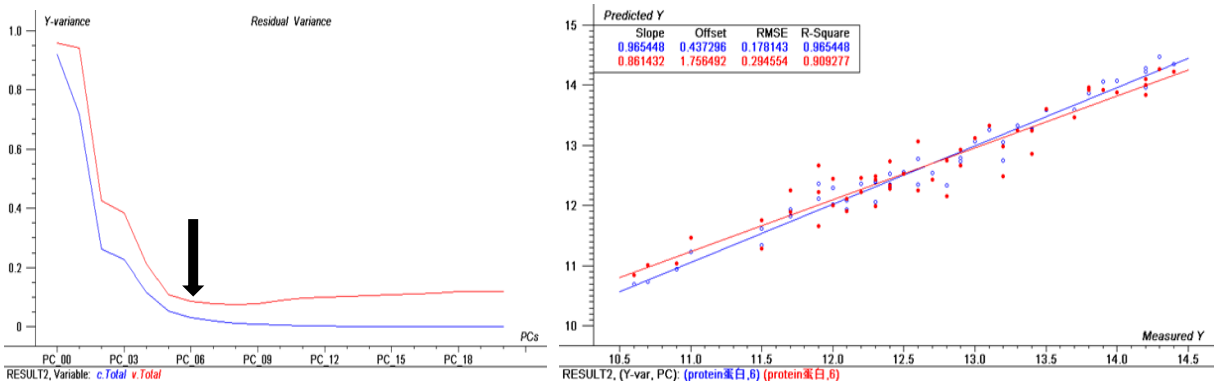
(a) Protein calibration results achieved during Stage 1 measurements of the German flour samples (measured in Germany) with the German NIRONE Sensor S 2.0 instrument.



(b) Protein calibration results achieved during Stage 3 measurements of the German flour samples (measured in China) with the Chinese NIRONE Sensor S 2.0 instrument.



(c) Protein calibration results achieved during Stage 1 measurements of the German flour samples (measured in Germany) with the German MicroNIR instrument.



(d) Protein calibration results achieved during Stage 3 measurements of the German flour samples (measured in China) with the Chinese MicroNIR instrument.

Fig. 4.2 Comparison of the PLS protein calibration results for the NIRONE Sensor S 2.0 and MicroNIR spectrometers of the same 50 German flour samples measured during Stage 1 (measurements in Germany) and Stage 3 (measurements in China)

4.1.2 Selection of Pretreatment Methods for Raw Spectra

Before calibration all spectra were divided into a calibration set (192) and a prediction set (21) at a ratio of 9:1 by the SPXY algorithm. For the development of protein calibration models with the spectra measured on the IAS 3100 benchtop instrument, the NIR spectra of 213 flour samples were collected and a PLS model with 191 calibration samples (removing one outlier) and 21 test set samples for external validation was developed. In order to fully exploit the available information in the spectra and to reduce unwanted interferences, the raw spectra of the samples were pretreated by different methods to minimize or eliminate irrelevant

information and noise. The best pretreatment method (EMSC) was selected on the basis of the PLS calibration and cross-validation parameters (number of factors, RMSE and R^2 values) and is highlighted in red in Tab. 4.1.

Tab. 4.1 Comparison of the protein PLS calibration results for different pretreatment methods applied to the spectra measured with the IAS 3100 benchtop spectrometer.

Method	Factor	Calibration		Cross Validation	
		R_c^2	*RMSEC*	R_c^2	*RMSECV*
None	7	0.9550	0.4003	0.9638	0.3631
SNV	6	0.9801	0.3745	0.9648	0.3582
EMSC	**5**	**0.9571**	**0.3907**	**0.9665**	**0.3493**
1st Derivative	8	0.8613	0.7023	0.8458	0.7498
Savitzky-Golay	7	0.9556	0.3974	0.9651	0.3563

4.1.3 Outlier Screening and Analysis of the Number of Factors for Calibration Development

An outlier is an error that clearly exceeds the expected statistical parameters, mainly due to some anomalies, and the source may be measurement error, spectral noise, or extreme sample properties. During the development of a calibration model, attention has to be paid to such outliers, which decrease the performance of the calibration. An outlier is a sample that deviates from the other calibration samples and may not belong to the same population, thereby negatively influencing the calibration model. For a PLS calibration, outliers can be detected by using score plots, plots of residuals, or the parameters of leverage and influence. When outliers are judged by the values of leverage and residuals, the larger the leverage and the smaller the residuals, the more important are the data for the corresponding calibration. By this control procedure, the outliers of the individual calibration sets for the different parameters have been identified and are eliminated before calibration model development. Fig. 4.3 shows all the spectra of an arbitrarily selected instrument for outlier screening. As can be

seen in the figure, the two samples marked as red circles outside the magenta threshold line belong to classical outlier spectra and need to be removed.

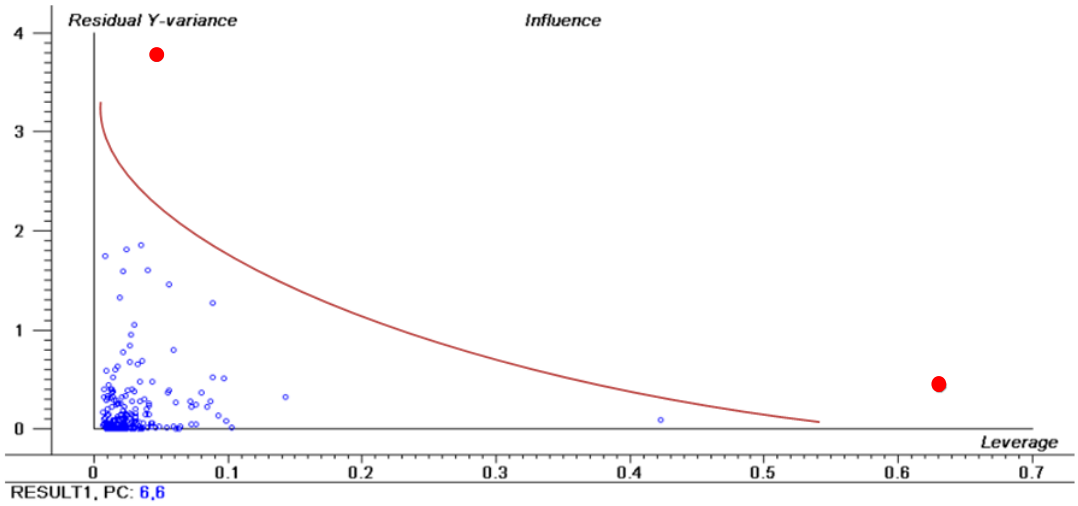


Fig.4.3 The influence plot of leverage and variance for the identification of outliers (red circles are marked as outliers)

Another parameter that has a significant effect on the performance of calibration is the number of chosen factors or principal components. Fewer factors lead to lower accuracy, though to more robust calibrations. Too many factors frequently induce overfitting with only an apparent improvement of predictive ability. In the present work, the optimal number of factors has been determined from the plot of the RMSE ́s versus the number of factors when the graphs reach a minimum or a plateau. Furthermore, in Tab. 4.24, the effect of the number of factors on the RMSEC and the RMSECV is demonstrated exemplarily for three flour parameters (wet gluten, moisture, and protein) calibrated with the spectra recorded on four different instruments. For example, for Neospectra Micro's principal component selection, although five factors are selected for protein, only two factors are selected for water. The number of factors selected for all parameters and instruments are summarized in Tab. 4.24. Apart from the optimal number of factors for the individual calibration models, the calibration/test set sample ratios, the RMSE ́s and the R^2 (split up in values for calibration (C), cross-validation (CV), and prediction (P)) are included.

4.1.4 Analysis of the Important Flour Parameters

The reference values of six parameters representing important quality indicators of flour were available for the flour samples of this project, namely protein, moisture, ash, wet gluten, sedimentation value and gluten index. These six parameters were determined using standard reference methods (see Chapter 3, Section 3.3). However, only the values of three parameters, protein, moisture and wet gluten, have been selected for spectroscopic data analysis and modeling and the other values were excluded, because the number of samples with reference values of sedimentation and gluten index was too small, accounting for only 18.8% of the total 213 samples. There are several reasons for not using the ash values for calibration development. First, because of the inconsistency between the national standards for ash testing in China and the ash testing standards in Germany, the different testing methods also lead to larger errors in the ash values; second, the accuracy of the ash values of Chinese flour samples tested in China is not high due to the imperfect testing technology of Chinese testing institutions for ash values.

For all samples parameter values of protein and moisture were available, and the coverage of wet gluten also reached 93.4%; thus, these three parameter values are more representative of the general applicability of the NIR analysis. In Tab. 4.2 the descriptive statistics of all parameter values for the investigated flours samples have been summarized.

Tab. 4.2 Descriptive statistical analysis of reference values of the flour samples measured in Germany and China

Reference	Number of Samples	Mean%(w/w)	Max%(w/w)	Min%(w/w)	Range%(w/w)	Std. %(w/w)
Protein	213	10.76	14.83	3.24	11.59	1.95
Moisture	213	13.06	19.35	7.07	12.28	1.13
Ash	213	0.65	1.75	0.37	1.38	0.24
Wet Gluten	199	29.38	42.79	18.70	24.09	4.29
Sedimentation	40	36.85	46	12	34	7.90
Gluten Index	40	90.35	96	74	22	4.02

4.2 Analysis Results of the Benchtop Spectrometers IAS 3100 and NIR-Freespace

4.2.1 PCA Results Obtained with Spectra of the IAS 3100 Spectrometer

The PCA method (see Chapter 2, Section 2.5.2.1) is applied here to highlight discriminative features in the spectral data of all 213 samples, and indeed some very interesting issues have been detected. Due to the larger wavelength ranges of the spectra measured on the benchtop spectrometers (the IAS 3100 spectral range is 11111-5970 cm^{-1} with 776 data points, and the NIR-Freespace spectral range is 9091-4348 cm^{-1} with 1201 data points), they contain information from several overtones and combination bands (see Fig. 4.28). Therefore, PCA can help to distinguish also materials with minor chemical differences. Not only can PCA determine the categories to which the samples belong by their scores in each factor space, but also the new variables can characterize the quality differences of the original samples in a more visual way.

As shown in Fig. 4.4(a), all samples are analyzed by a PCA model using the spectral data measured with the IAS 3100 instrument for the distribution of scores on the second and third principal components. In this graph the flours produced in China and Germany are clearly clustered and separated allowing an identification of the two species. The two outliers of Chinese flour samples can readily be identified and do not reduce the accuracy of sample classification. As shown in Fig. 4.4(b), due to the large distance between the two clusters, an approximate boundary line can be used to determine the clusters of Chinese and German flour. The boundary formula for this clustering discrimination is:

$$y = -1.66x + 0.05 \quad (x = \mathrm{PC2}, y = \mathrm{PC1}) \qquad \text{Eq.63}$$

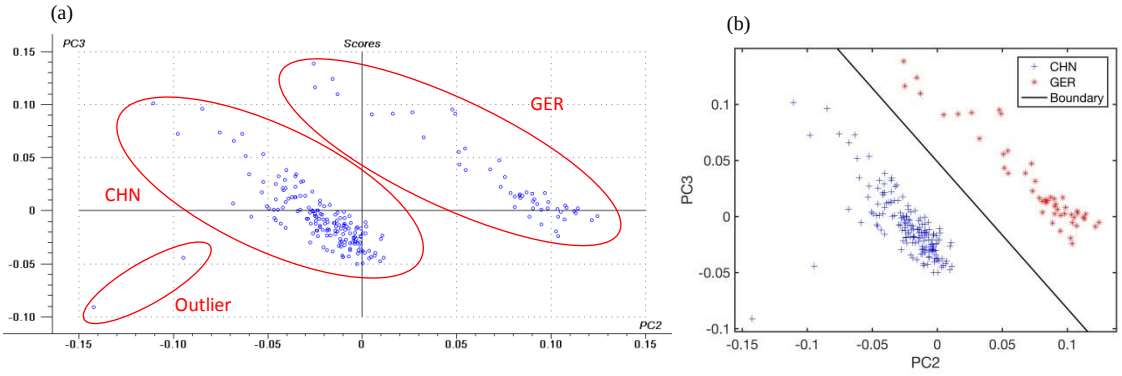


Fig. 4.4 PCA analysis results of 213 flour spectra measured with the IAS 3100 spectrometer ((a) two-dimensional score plot (b) boundary result)

In Fig. 4.5 the PCA 2D-score plot of the first and second principal components using the spectra of the 154 Chinese flour samples measured with the NIR-Freespace instrument is shown. Although two clusters are observable in the PCA score plot (Fig. 4.5), they are not based on the high and low gluten content required by the Chinese Flour Quality Classification Standard. They can also not be assigned to differences in origin, moisture, protein, ash and wet gluten. The outlier observed in Fig. 4.6 is not an ordinary flour sample, but a flour sample with admixed baking powder. The addition of a non-flour substance significantly affects the PCA classification result of this flour sample.

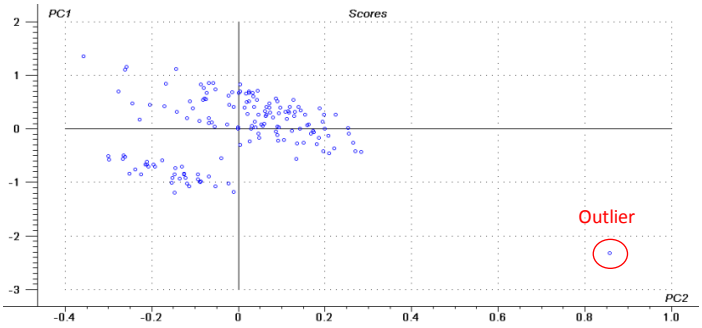


Fig. 4.5 PCA score plot of 154 Chinese flour spectra measured with the NIR-Freespace instrument

The above exemplary PCA analyses demonstrate that flours of different origin can be distinguished by PCA analysis.

4.2.2 PLS Results Obtained with Spectra of the IAS 3100 and NIR-Freespace Instruments

Due to the overlap of spectral information at various wavelength positions and the broad spectral peaks in the NIR spectra of flour, the qualitative and quantitative analysis of flour NIR spectra requires chemometric evaluation methods. For quantitative analysis PLS has proved an extremely versatile tool to build calibration models for various flour parameters. The spectra of flour samples from different batches measured with the IAS 3100 benchtop spectrometer were divided into a calibration set and a test set. A calibration model developed for practical use has to include samples with the whole range of content value variations in order to provide good prediction values of unknown samples not included in the calibration model. For the PLS calibrations developed in this work, 10 % of all spectra (after removal of outliers) were selected as test set and the remaining 90 % of samples were used to build the calibration model. The results of the descriptive statistical analysis of the flour parameters protein, moisture and wet gluten are summarized in Tab. 4.3.

Tab. 4.3 Descriptive statistical analysis of the investigated flour parameters of the German and Chinese flour samples measured with the IAS 3100 benchtop spectrometer.

Parameter	Data Set	Number of Samples	Mean%(w/w)	Max%(w/w)	Min%(w/w)	Range%(w/w)
Protein	Total	212	10.80	14.80	6.30	8.50
	Calibration	191	10.77	14.80	6.30	8.50
	Test	21	11.03	14.40	7.20	7.20
Moisture	Total	208	13.07	15.30	9.90	5.40
	Calibration	187	13.08	15.10	10.10	5.00
	Test	21	12.98	15.30	9.90	5.40
Wet Gluten	Total	198	29.38	42.80	18.70	24.10
	Calibration	178	29.28	42.20	18.70	23.50
	Test	20	30.30	42.80	20.60	22.20

Since with the NIR-Freespace benchtop spectrometer only 152 Chinese flour samples were tested, the total number of spectra in the calibration and test sets are lower than for the IAS

3100 instrument. The results of the descriptive statistical analysis of the three parameters are summarized in Tab. 4.4.

Tab. 4.4 Descriptive statistical analysis of the investigated flour parameters of the Chinese flour samples measured with the NIR-Freespace benchtop spectrometer

Parameter	Data Set	Number of Samples	Mean%(w/w)	Max%(w/w)	Min%(w/w)	Range%(w/w)
Protein	Total	152	10.24	14.80	6.70	8.10
	Calibration	137	10.25	14.80	6.70	8.10
	Test	15	10.01	12.20	7.30	4.90
Moisture	Total	150	12.97	14.3	10.20	4.10
	Calibration	135	12.96	14.3	10.20	4.10
	Test	15	13.11	14.2	12.40	1.80
Wet Gluten	Total	149	29.27	42.80	18.70	24.10
	Calibration	134	29.20	40.20	18.70	23.50
	Test	15	29.90	42.80	21.00	21.80

4.2.2.1 Calibration Models and their Prediction Results Obtained for the Spectra Measured with the IAS 3100 Benchtop Spectrometer

The spectral data measured on the IAS 3100 spectrometer have been analyzed quantitatively by developing PLS calibration models. Before the PLS modeling, the original spectra had to be pretreated in order to eliminate outliers according to the method described in Section 4.1.3, and remove noise (the specific spectral pretreatment process was according to section 4.1.2). Raw spectra and pretreated spectra are shown in Fig. 4.6.

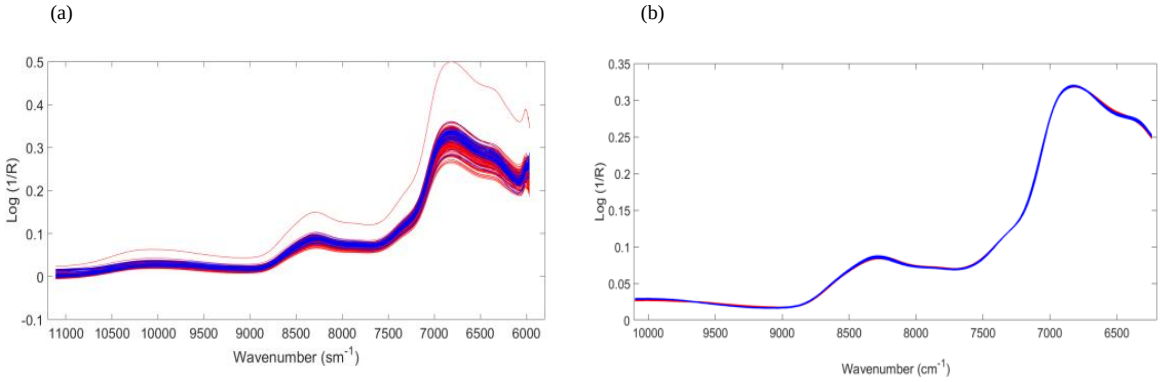
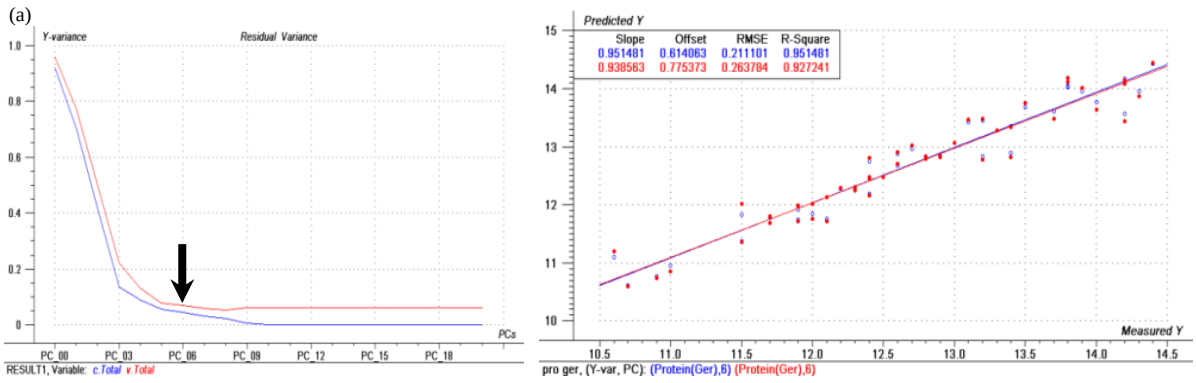


Fig. 4.6 Spectra of 50 German (red) and 163 Chinese (blue) flour samples measured with the IAS 3100 benchtop spectrometer. (a) raw spectra (11111-5970 cm^{-1}) and (b) EMSC pretreated and truncated spectra (10101-6238 cm^{-1}) after removal of one Chinese flour outlier sample.

The calibration results obtained for the three parameter values (protein, moisture, and wet gluten) are shown in the Figs. 4.7 to 4.9, respectively. The calibrations were performed in three steps: (a) modeling of German flour samples only, (b) modeling of Chinese flour samples only, and (c) developing a calibration model after merging the flour samples from both countries. In Fig. 4.7(c) to 4.9(c) the calibration set of the German flour samples is marked with black circles. From Fig. 4.7(c) it can be seen that the protein content interval of the German flours is limited to the region of high protein levels. In Tabs. 4.5 to 4.7, the predictions of the test sets corresponding to the calibrations shown in Figs. 4.7(c) to 4.9(c) are summarized.



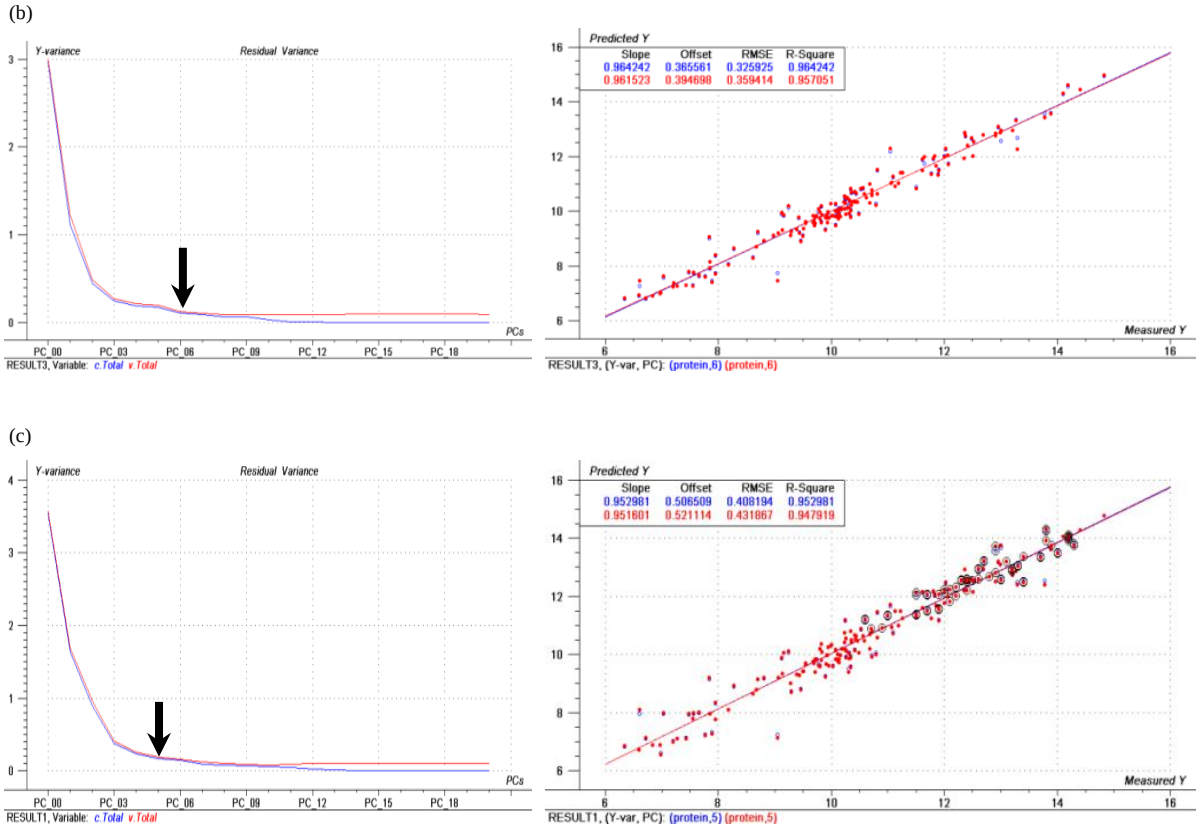


Fig. 4.7 Comparison of the protein calibrations obtained with the spectra of the individual and merged sample sets measured on the IAS 3100 benchtop spectrometer. (a) Calibration of 50 German samples (measured in Germany with German reference values); (b) calibration of 162 Chinese samples (measured in China with Chinese reference values); (c) calibration of merged samples (45 German and 146 Chinese calibration samples with respective reference values).

Tab. 4.5 Comparison of Protein reference and prediction values for five German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.7(c) (AE: absolute prediction error, MAE: mean absolute prediction error)

(a)

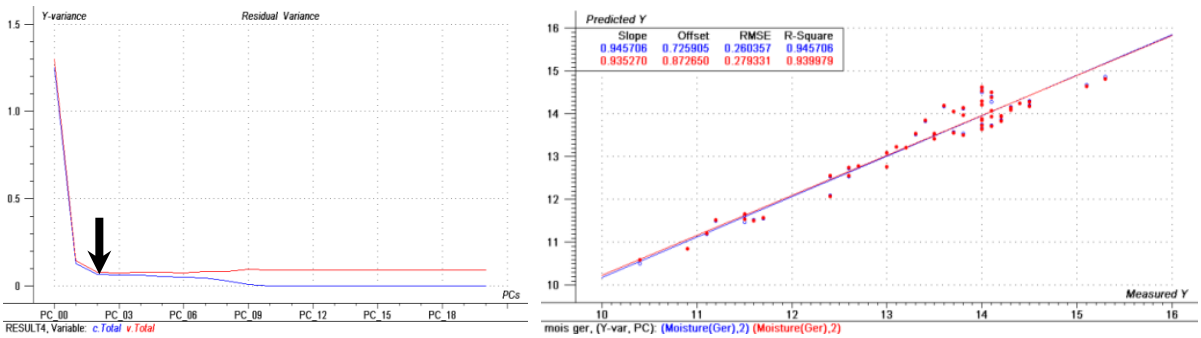
Samples #(GER)	11	16	37	38	49
Ref.	11.9	12.8	12.3	14.4	13.5
Pred.	12.10	12.30	12.44	15.01	14.14
AE (Ref.-Pred.)	0.20	0.50	0.14	0.61	0.64
MAE	0.42				

(b)

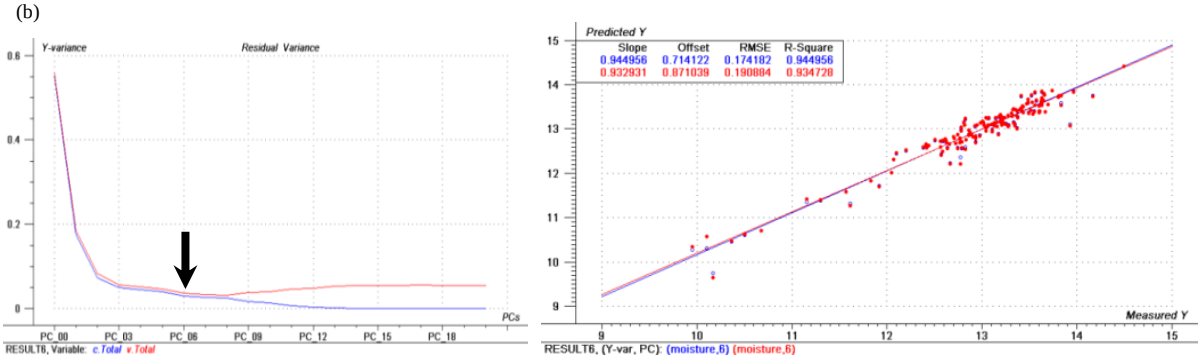
Samples #(CHN)	17	22	24	39	55	57	59	78
Ref.	9.7	10.3	10.2	9.9	10.4	11.1	14. 2	10.1
Pred.	9.12	10.55	10.12	9.87	10.10	11.19	14.58	9.63
AE (Ref.-Pred.)	0.58	0.25	0.08	0.03	0.30	0.09	0.38	0.47
Samples #(CHN)	81	87	93	97	109	127	129	135
Ref.	9.4	7.2	7.9	11.7	9.0	12.1	12.9	10.6
Pred.	9.88	7.61	8.20	12.50	9.28	11.79	12.71	10.82
AE (Ref.-Pred.)	0.48	0.41	0.30	0.80	0.28	0.31	0.19	0.22
MAE	0.38							

From the absolute difference of the reference and predicted protein values of the test samples, the lowest and highest Absolute Errors (AE) are 0.03 and 0.80, respectively. Mean Absolute Errors (MAE) of 0.42 and 0.38 are calculated for Germany and China, respectively.

(a)



(b)



(c)

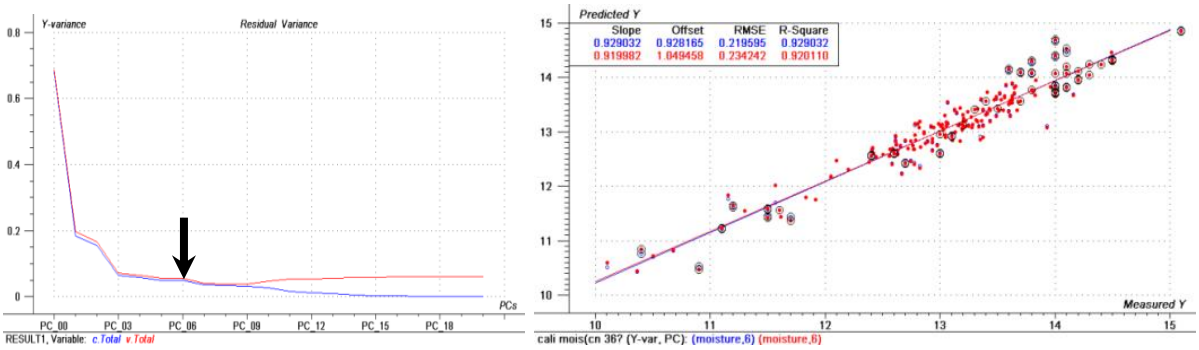


Fig. 4.8 Comparison of the moisture calibrations obtained with the spectra of the individual and merged sample sets measured on the IAS 3100 benchtop spectrometer. (a) Calibration of 50 German samples (measured in Germany with German reference values); (b) calibration of 158 Chinese samples (measured in China with Chinese reference values); (c) calibration of merged samples (45 German and 152 Chinese calibration samples) with respective reference values).

Tab. 4.6 Comparison of Moisture reference and prediction values for 5 German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.8(c).

(a)

Samples #(GER)	1	16	32	40	49
Ref.	13.5	14	15.3	14.1	12.4
Pred.	13.37	14.77	14.73	14.46	12.16
AE (Ref.-Pred.)	0.13	0.77	0.57	0.36	0.24
MAE	0.41				

(b)

Samples #(CHN)	9	12	19	22	30	35	77	88
Ref.	12.9	13.3	13.2	13.2	13.4	13.7	13.2	10.0
Pred.	12.83	13.47	13.72	13.13	13.4	13.8	13.04	10.35
AE (Ref.-Pred.)	0.07	0.17	0.52	0.07	0.00	0.10	0.16	0.35
Samples #(CHN)	94	102	129	136	145	148	155	156
Ref.	13.1	13.0	12. 6	12.7	12.8	12.1	13.5	10.2
Pred.	13.09	13.12	12.50	12.71	12.77	12.18	13.30	9.76
AE (Ref.-Pred.)	0.01	0.12	0.10	0.01	0.03	0.08	0.20	0.44
MAE	0.15							

From the absolute difference of the reference and predicted moisture values of the test

samples, the lowest and highest AE are 0.00 and 0.77, respectively. MAE of 0.41 and 0.15 are calculated for Germany and China, respectively.

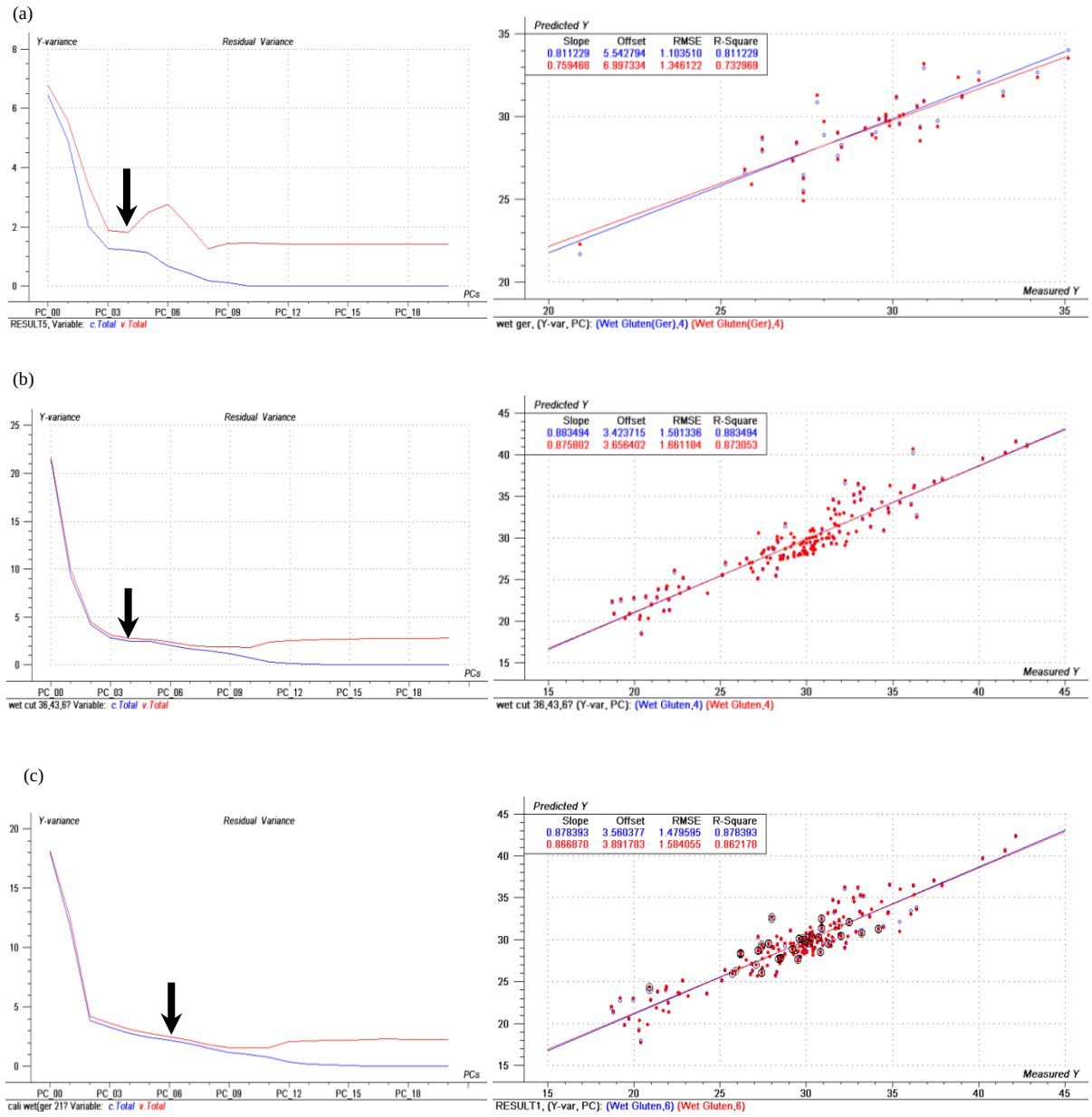


Fig. 4.9 Comparison of the wet gluten calibrations obtained with the spectra of the individual and merged sample sets measured on the IAS 3100 benchtop spectrometer. (a) Calibration of 40 German samples (measured in Germany with German reference values); (b) calibration of 158 Chinese samples (measured in China with Chinese reference values); (c) calibration of merged samples (36 German and 142 Chinese calibration samples with respective reference values).

Tab. 4.7 Comparison of Wet Gluten reference and prediction values for 4 German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.9 (c).

(a)

Samples #(GER)	1	6	41	46
Ref.	29.8	30.8	35.1	27.4
Pred.	29.50	27.70	33.30	29.40
AE (Ref.-Pred.)	0.3	3.1	1.8	2.0
MAE	1.80			

(b)

Samples #(CHN)	2	16	23	24	37	38	39	47
Ref.	26.7	36.2	22.3	28.3	42.8	30.3	30.5	20.6
Pred.	27.30	40.30	25.50	30.40	41.90	29.30	29.40	22.00
AE (Ref.-Pred.)	0.6	4.1	3.2	2.1	0.9	1.0	1.1	1.4
Samples #(CHN)	53	57	78	81	90	142	158	164
Ref.	31.7	33.7	28.9	27.7	31.3	29.6	32.6	29.9
Pred.	29.50	31.60	27.00	27.80	30.00	27.70	33.40	28.50
AE (Ref.-Pred.)	2.2	2.1	1.9	0.1	1.3	1.9	0.8	1.4
MAE	1.63							

From the absolute difference of the reference and predicted Wet Gluten values of the test samples, the lowest and highest AE are 0.1 and 4.1, respectively. MAE of 1.80 and 1.63 are calculated for Germany and China, respectively.

4.2.2.2 Calibration Models and their Prediction Results Obtained for the Spectra Measured with the NIR-Freespace Benchtop Spectrometer

The spectra measured on the NIR-Freespace benchtop instrument were also used to build PLS calibration models for the different flour parameters. Here too, before PLS model development, the original spectra had to be preprocessed and outliers had to be removed according to the method described in Section 4.1.3. It was found, that the noise in the spectra

is very low and therefore the entire wavelength range was used for further processing as shown in Fig. 4.10.

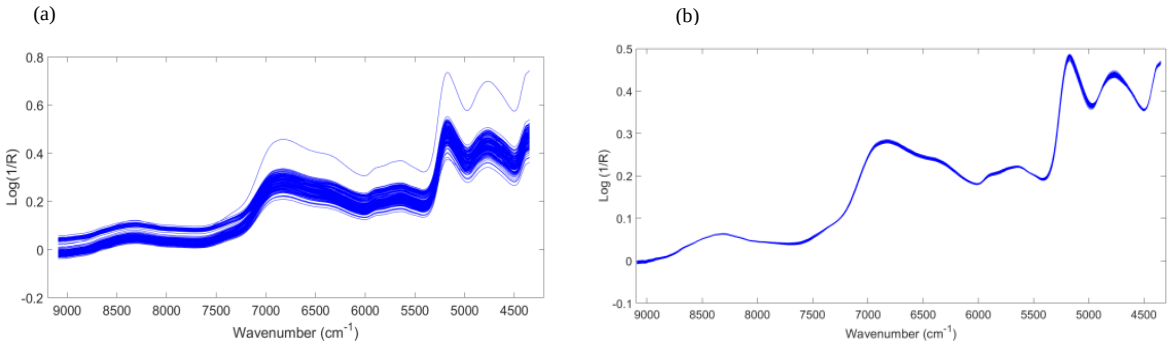


Fig. 4.10 Spectra of 154 Chinese flour samples measured with the NIR-Freespace benchtop spectrometer. (a) Raw spectra (9091-4348 cm^{-1}) and (b) EMSC pretreated spectra (9091-4348 cm^{-1}) after removal of 1 Chinese outlier sample

With the NIR-Freespace benchtop spectrometer only Chinese flour samples have been measured. Thus, no comparison with the calibration results obtained with German flour samples was possible. The PLS calibration results achieved for the three flour parameters under investigation are shown in the Figs. 4.11 to 4.13, and the test set prediction results are summarized in the Tabs. 4.8 to 4.10.

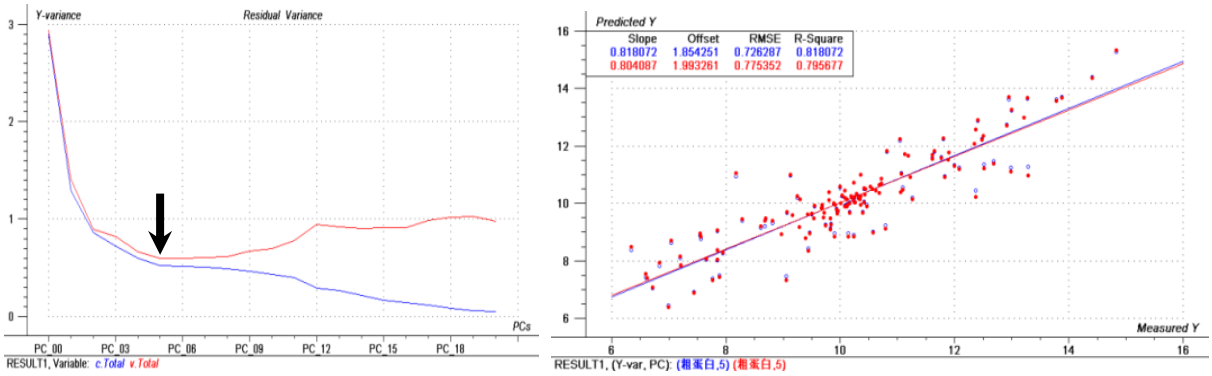


Fig. 4.11 Protein calibration results achieved with the spectra of 137 Chinese flour samples measured in China on the NIR-Freespace benchtop spectrometer by using the Chinese reference values.

Tab. 4.8 Comparison of protein reference and prediction values for 15 Chinese test samples based on the calibration of Fig. 4.11

Samples #(CHN)	5	8	14	19	35	41	54	59
Ref.	7.3	9.2	7.9	10.8	9.9	12.9	10.2	14. 2
Pred.	6.89	10.02	7.59	10.80	9.66	12.84	10.11	12.27
AE (Ref.-Pred.)	0.41	0.82	0.31	0.00	0.24	0.06	0.09	1.93
Samples #(CHN)	99	110	117	118	120	145	150	
Ref.	10.3	10.5	9.5	10.1	12.0	11.5	9.7	
Pred.	11.66	10.27	9.51	10.56	11.91	11.60	9.60	
AE (Ref.-Pred.)	1.36	0.23	0.01	0.46	0.09	0.10	0.10	
MAE	0.34							

From the absolute difference of the reference and predicted protein values of the test samples, the lowest and highest AE are 0.00 and 1.93, respectively, and a MAE of 0.34 is calculated.

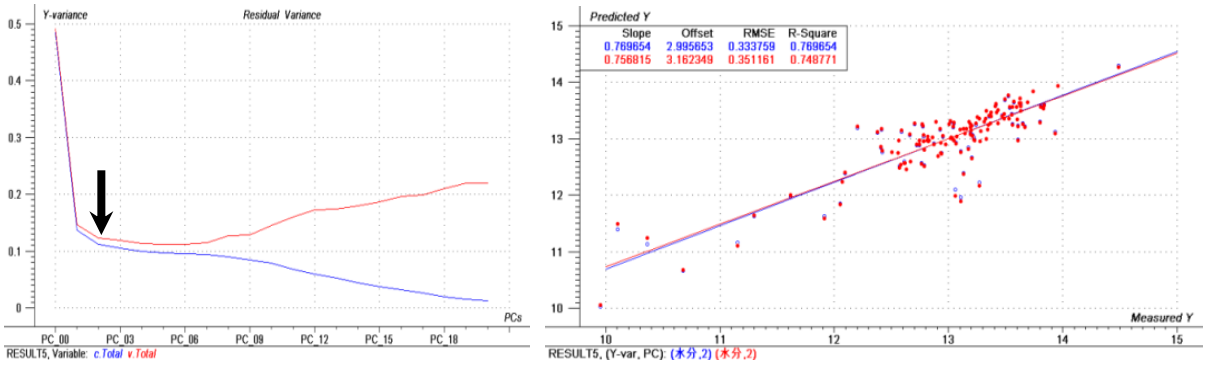


Fig. 4.12 Moisture calibration results achieved with the spectra of 135 Chinese flour samples measured in China on the NIR-Freespace benchtop spectrometer by using the Chinese reference values.

Tab. 4.9 Comparison of moisture reference and prediction values for 15 Chinese test samples based on the calibration of Fig. 4.12

Samples #(CHN)	4	6	24	28	34	42	51	55
Ref.	12.4	13.7	13.5	14.2	13.4	12.6	13.0	13.2
Pred.	12.45	13.63	13.30	13.70	13.25	12.52	13.21	13.16
AE (Ref.-Pred.)	0.05	0.07	0.20	0.50	0.15	0.08	0.21	0.04
Samples #(CHN)	104	106	122	123	139	140	141	
Ref.	12.9	13.3	13.6	13.2	13.1	12.8	11.8	
Pred.	13.08	13.19	13.61	13.40	13.07	12.31	12.40	
AE (Ref.-Pred.)	0.18	0.11	0.01	0.20	0.03	0.49	0.60	
MAE	0.19							

From the absolute difference of the reference and predicted moisture values of the test samples, the lowest and highest AE are 0.01 and 0.60, respectively, and a MAE of 0.19 is calculated.

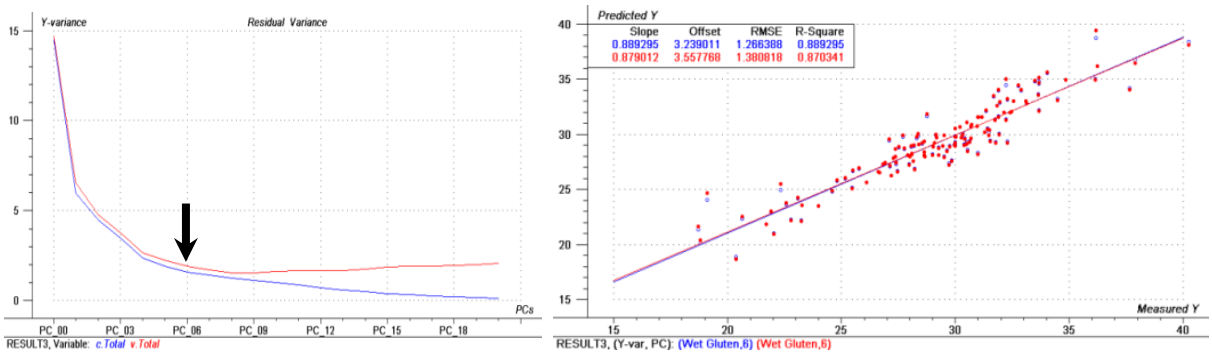


Fig. 4.13 Wet gluten calibration results achieved with the spectra of 134 Chinese flour samples measured in China on the NIR-Freespace benchtop spectrometer by using the Chinese reference values.

Tab. 4.10 Comparison of wet gluten reference and prediction values for 15 Chinese test samples based on the calibration of Fig. 4.13.

Samples #(CHN)	3	9	14	26	31	37	41	55
Ref.	21.0	33.2	22.8	29.1	29.6	42.8	30.0	32.2
Pred.	21.70	33.00	22.20	30.30	29.70	40.90	35.90	28.80
AE (Ref.-Pred.)	0.7	0.2	0.6	1.2	0.1	1.9	5.9	3.4
Samples #(CHN)	58	67	104	133	136	139	144	
Ref.	27.9	28.5	31.6	30.9	26.9	35.5	30.4	
Pred.	31.00	25.20	32.00	29.90	29.70	35.90	30.00	
AE (Ref.-Pred.)	3.1	3.3	0.4	1.0	2.8	0.4	0.4	
MAE	1.69							

From the absolute difference of the reference and predicted Wet Gluten values of the test samples, the lowest and highest AE are 0.1 and 5.9, respectively, and a MAE of 1.69 is calculated.

4.3 Analysis of Spectra Measured with the Handheld Spectrometers MicroNIR, Neospectra Micro and Hamamatsu C15511-01

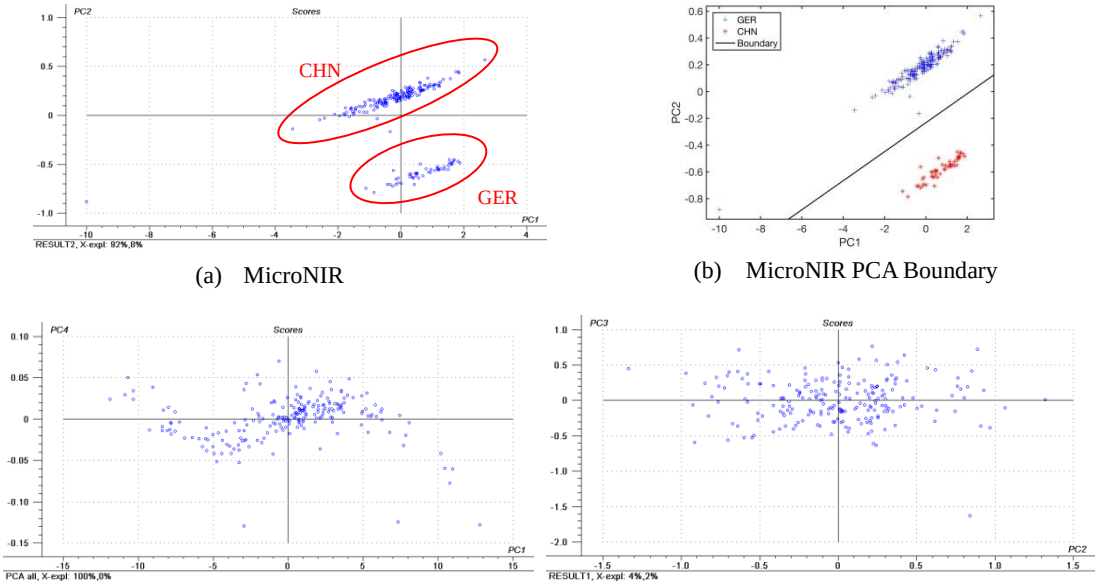
4.3.1 PCA Results for MicroNIR, Neospectra Micro and Hamamatsu C15511-01

With the spectra of the 213 samples measured with the handheld spectrometers PCA models were developed in analogy to the data of the benchtop spectrometers (see Section 4.2.1.). The amount of raw spectral information (in terms of data points) of the handheld spectrometers was much lower than that of the benchtop spectrometers; only 125 data points for the wavenumber range 11012-5966 cm^{-1} of the MicroNIR instrument, 257 data points for the wavenumber range 7407-3922 cm^{-1} of the Neospectra Micro spectrometer, and 359 data points for the 9991-3856 cm^{-1} wavenumber range of the Hamamatsu C15511-01 instrument. Since the performance of the three handheld spectrometers is lower than that of the benchtop

spectrometers, the results can be considered only as a preliminary comparison with the results of the benchtop spectrometers to verify whether PCA can effectively classify the flours of different origins and different gluten contents and characterize the flours with the same clustering effect as the PCA analysis of the benchtop spectrometers.

As shown in Fig. 4.14, in the 2D-score plot of the PCA analysis for the MicroNIR spectra, the flours from Germany and China are clearly clustered and can be separated in the score plot of the first and second principal components. However, the PCA results for the Neospectra Micro (first versus fourth principal component score plot) and Hamamatsu C15511-01 (second principal versus third principal component score plot), could not reach the same discrimination. Since the distance between the two clusters in Fig. 4.14 (b) is similar to the PCA result of the IAS 3100 instrument in Fig. 4.4 (b), an approximate boundary line can be defined to determine the separation of the China and Germany clusters. The equation for the boundary line of the clusters is:

$$y = 0.1x - 0.2 \quad (x = \mathrm{PC1}, y = \mathrm{PC2}) \qquad \text{Eq.64}$$



(a) MicroNIR (b) MicroNIR PCA Boundary

(c) Neospectra Micro (d) Hamamatsu C15511-01)

Fig. 4.14 PCA analysis results for the 213 flour spectra measured on the three handheld spectrometers ((a) MicroNIR, (b) MicroNIR PCA boundary (c) Neospectra Micro, (d) Hamamatsu C15511-01).

4.3.2 PLS Results for MicroNIR, Neospectra Micro and Hamamatsu C15511-01

With reference to Section 4.2.2, the spectral data obtained with the three handheld spectrometers were also analyzed by developing PLS calibration models for the different flour parameters. The spectra of the flour samples measured with the same type but different production batches of handheld spectrometers (MicroNIR, Neospectra Micro, Hamamatsu C15511-01) during different time periods of the experiment were combined and subsequently divided into calibration and test sets. After removing the outliers, 10% of all samples were selected as test set and 90% of the samples were used as calibration set to build the calibration model. The results of the descriptive statistical analysis of the investigated parameter values for the calibration set and test set are summarized in Tabs. 4.11 to 4.13.

Tab. 4.11 Descriptive statistical analysis of the investigated flour parameters for the sample sets measured with the MicroNIR spectrometer.

Parameter	Data Set	Number of Samples	Mean%(w/w)	Max%(w/w)	Min%(w/w)	Range%(w/w)
Protein	Total	210	10.78	14.80	6.30	8.50
	Calibration	189	10.75	14.40	6.30	8.10
	Test	21	11.07	14.80	7.20	7.60
Moisture	Total	206	13.07	15.3	9.90	5.40
	Calibration	186	13.05	15.1	9.90	5.20
	Test	20	13.25	15.3	11.60	3.70
Wet Gluten	Total	197	29.37	42.80	18.70	24.10
	Calibration	177	29.26	42.20	18.70	23.50
	Test	20	30.37	42.80	20.60	22.20

Tab. 4.12 Descriptive statistical analysis of the investigated flour parameters for the sample sets measured with the Neospectra Micro spectrometer.

Parameter	Data Set	Number of Samples	Mean%(w/w)	Max%(w/w)	Min%(w/w)	Range%(w/w)
Protein	Total	206	10.81	14.40	6.30	8.10
	Calibration	185	10.78	14.30	6.30	8.00
	Test	21	11.07	14.40	7.20	7.20
Moisture	Total	207	13.07	15.3	9.90	5.40
	Calibration	186	13.04	15.1	9.90	5.20
	Test	21	13.31	15.3	11.60	3.70
Wet Gluten	Total	195	29.38	42.80	18.70	24.10
	Calibration	175	29.27	42.20	18.70	23.50
	Test	20	30.42	42.80	20.60	22.20

Tab. 4.13 Descriptive statistical analysis of the investigated flour parameters for the sample sets measured with the Hamamatsu C15511-01 spectrometer.

Parameter	Data Set	Number of Samples	Mean%(w/w)	Max%(w/w)	Min%(w/w)	Range%(w/w)
Protein	Total	211	10.82	14.80	6.60	8.20
	Calibration	190	10.79	14.80	6.60	8.20
	Test	21	11.06	14.40	7.30	7.10
Moisture	Total	204	13.06	15.3	9.90	5.40
	Calibration	183	13.03	14.5	9.90	4.60
	Test	21	13.30	15.3	11.60	3.70
Wet Gluten	Total	198	29.43	42.80	18.70	24.10
	Calibration	178	29.32	42.20	18.70	23.50
	Test	20	30.38	42.80	20.80	22.00

4.3.2.1 Calibration Model and Prediction Results Obtained with the Spectra of the MicroNIR Spectrometer

Before building PLS calibration models for the parameter values based on the MicroNIR spectra, the outliers were removed according to the method described in Section 4.1.3. Then the noise in the spectra was removed by smoothing and scatter effects were eliminated by EMSC (Fig. 4.15)

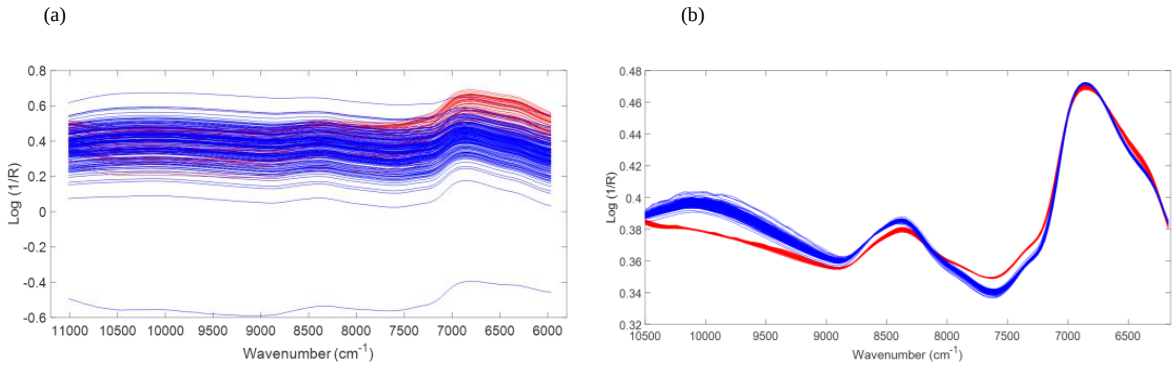


Fig. 4.15 Spectra of 50 German (red) and 163 Chinese (blue) flour samples measured with the MicroNIR handheld spectrometer. (a) Raw spectra (11012-5966 cm^{-1}) and (b) EMSC pretreated and truncated spectra (10510-6171 cm^{-1}) after removal of 3 Chinese outlier samples

Figs. 4.16-4.18 summarize the results of the PLS calibration models for protein, moisture, and wet gluten, respectively, based on the spectra measured with the MicroNIR instrument. A total of three operations are performed for the calibration set modeling: (a) modeling of German flour samples only, (b) modeling of Chinese flour samples only, and (c) modeling after merging the flour samples of both countries. In Figs. 4.16(c) to 4.18(c) the calibration results for the German flour samples are marked with black circles and in Tabs. 4.14 to 4.16 the prediction results of the test sets for the investigated parameters are summarized.

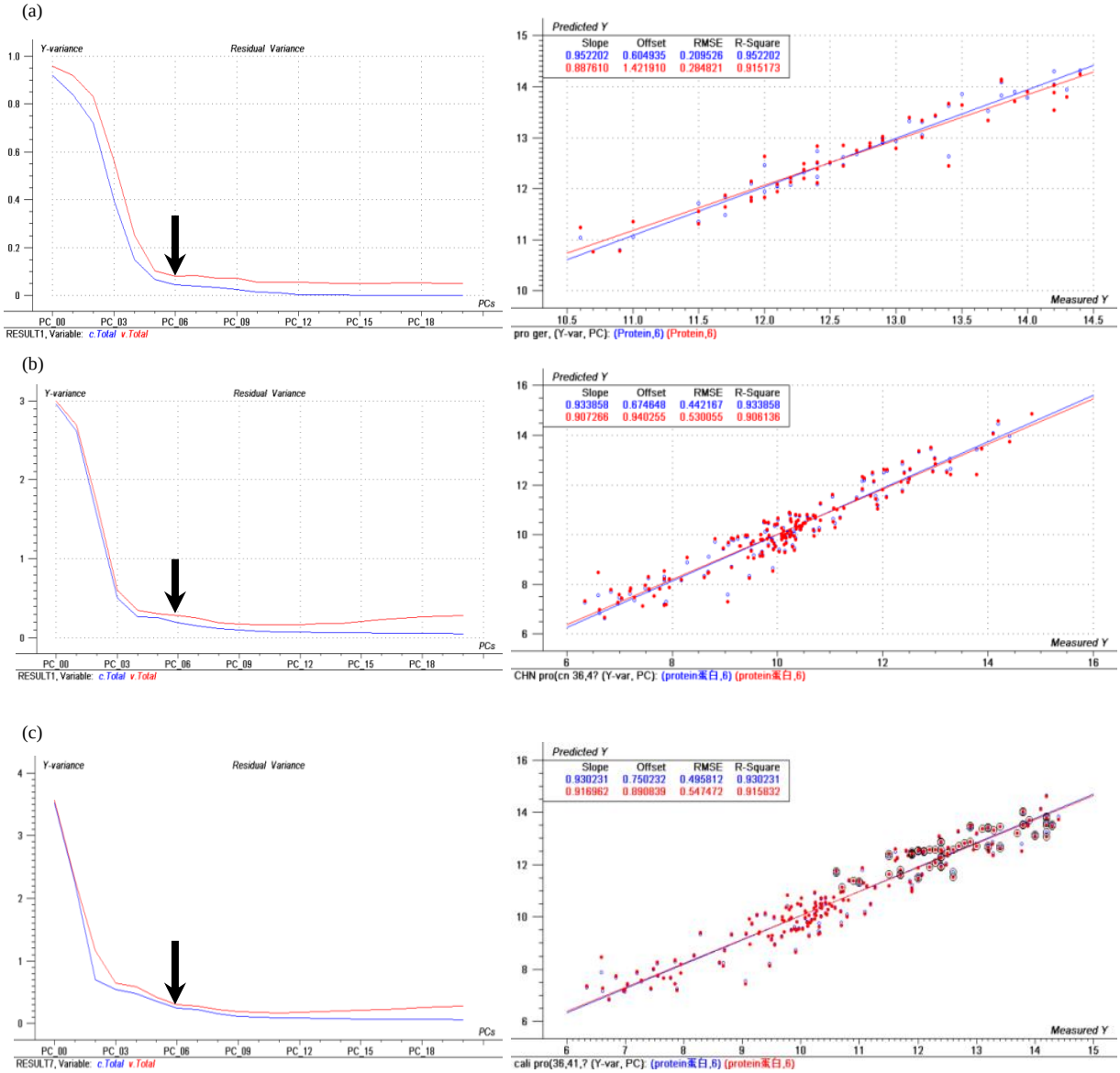


Fig. 4.16 Comparison of country-individual and merged protein calibrations achieved with the spectra measured on the MicroNIR spectrometer; (a) calibration of 50 German samples measured in Germany with German reference values, (b) calibration of 160 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (45 German and 144 Chinese samples) with respective reference values.

Tab. 4.14 Comparison of Protein reference and prediction values for 5 German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.16(c).

(a)

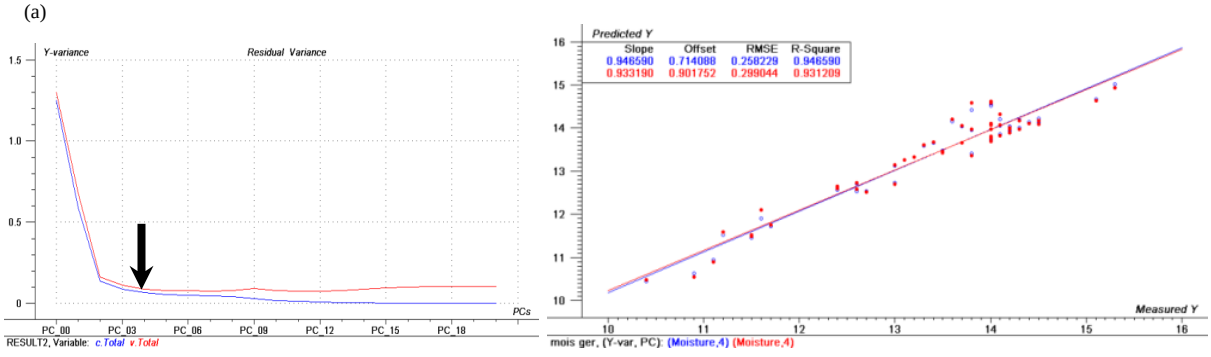
Samples #(GER)	11	16	37	38	49
Ref.	11.9	12.8	12.3	14.4	13.5
Pred.	12.15	12.65	12.54	14.55	12.91
AE (Ref.-Pred.)	0.25	0.15	0.24	0.15	0.59
MAE	0.28				

(b)

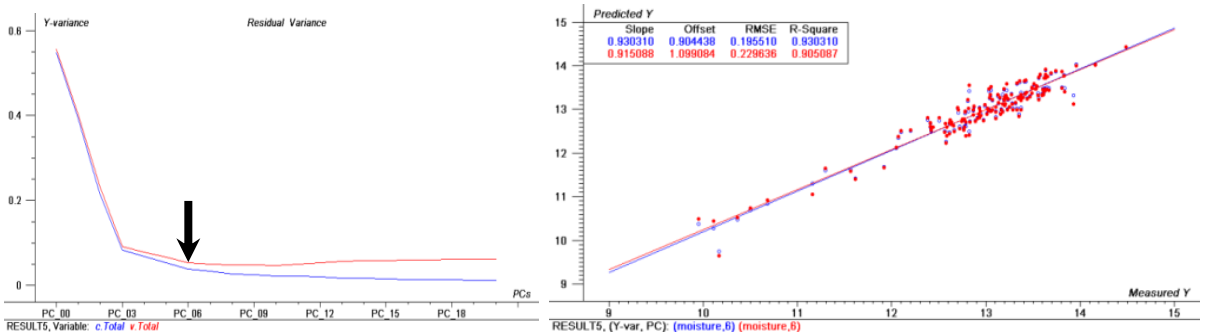
Samples #(CHN)	16	17	22	24	39	55	57	78
Ref.	14.8	9.7	10.3	10.2	9.9	10.4	11.1	10.1
Pred.	14.89	9.26	9.50	9.84	9.86	10.32	10.46	10.07
AE (Ref.-Pred.)	0.09	0.44	0.80	0.36	0.04	0.08	0.64	0.03
Samples #(CHN)	81	87	93	108	109	135	153	154
Ref.	9.4	7.2	7.9	11.7	9.0	10.6	11.8	12.3
Pred.	10.10	7.83	8.28	13.05	9.15	10.84	11.58	12.10
AE (Ref.-Pred.)	0.70	0.63	0.38	1.35	0.15	0.24	0.22	0.20
MAE	0.40							

From the absolute difference of the reference and predicted Protein values of the test samples, the lowest and highest AE are 0.03 and 0.80, respectively. MAE of 0.28 and 0.40 are calculated for Germany and China, respectively.

(a)



(b)



(c)

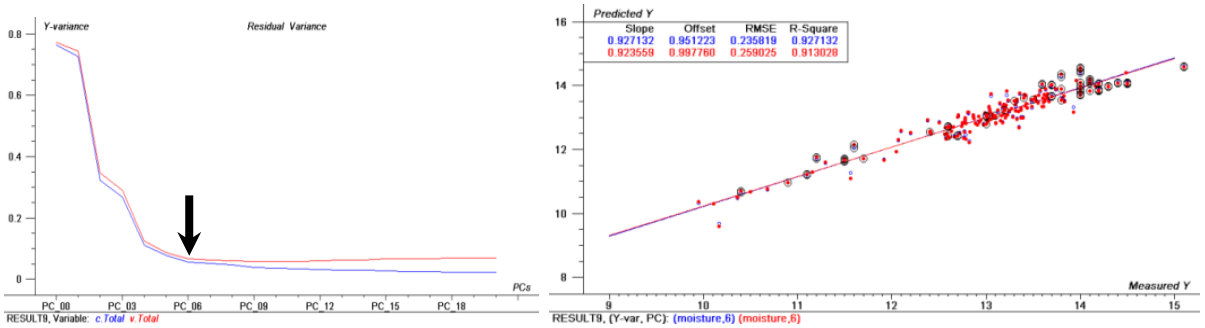


Fig. 4.17 Comparison of country-individual and merged moisture calibrations achieved with the spectra measured on the MicroNIR spectrometer; (a) calibration of 50 German samples measured in Germany with German reference values, (b) calibration of 156 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (45 German and 141 Chinese samples) with respective reference values.

Tab.4.15 Comparison of Moisture reference and prediction values for 5 German (a) and 15 Chinese (b) test samples based on the calibration of Fig. 4.17(c).

(a)

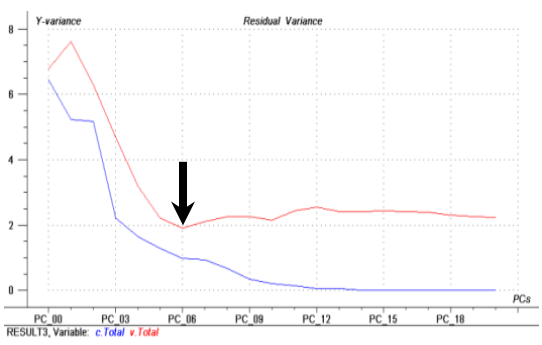
Samples #(GER)	1	18	32	40	49
Ref.	13.5	14	15.3	14.1	12.4
Pred.	13.42	13.75	14.86	14.09	12.65
AE (Ref.-Pred.)	0.08	0.25	0.44	0.01	0.25
MAE	0.21				

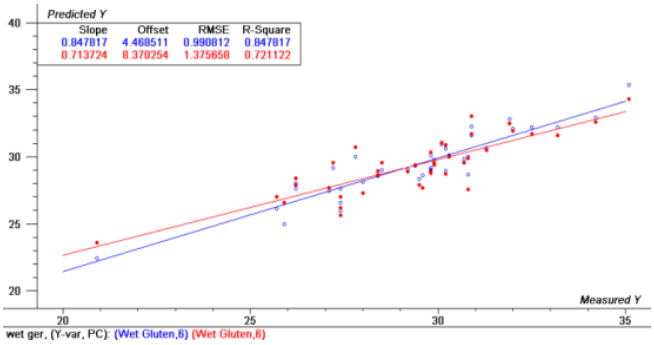
(b)

Samples #(CHN)	4	29	39	50	57	63	67	97
Ref.	12.4	12.9	13.5	13.6	13.2	12.8	13.6	11.6
Pred.	12.39	13.07	13.31	13.67	13.22	12.54	13.15	11.46
AE (Ref.-Pred.)	0.01	0.17	0.19	0.07	0.02	0.26	0.45	0.14
Samples #(CHN)	98	109	118	124	135	144	160	
Ref.	12.6	13.0	13.4	13.1	13.8	13.3	12.8	
Pred.	12.71	12.99	13.23	13.23	13.88	13.19	13.67	
AE (Ref.-Pred.)	0.11	0.01	0.17	0.13	0.08	0.11	0.87	
MAE	0.19							

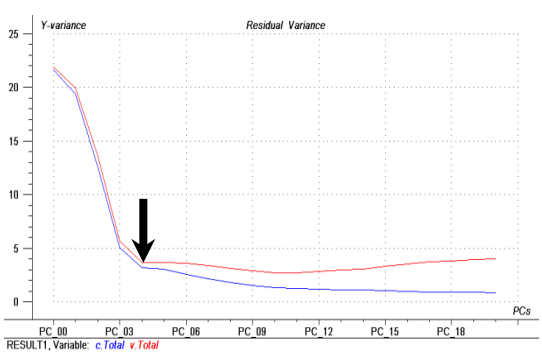
From the absolute difference of the reference and predicted Moisture values of the test samples, the lowest and highest AE are 0.01 and 0.87, respectively. MAE of 0.21 and 0.19 are calculated for Germany and China, respectively.

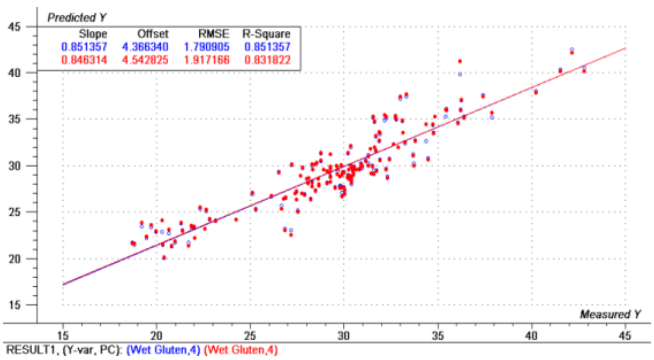
(a)





(b)





(c)

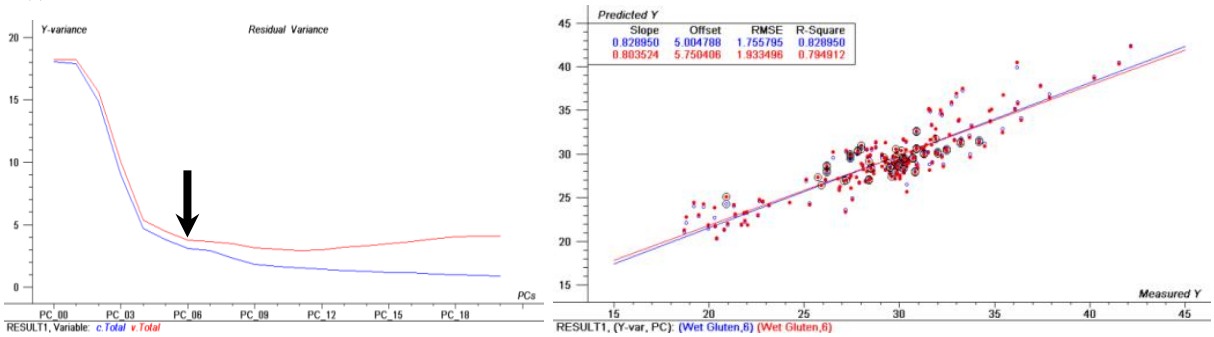


Fig. 4.18 Comparison of country-individual and merged wet gluten calibrations achieved with the spectra measured on the MicroNIR spectrometer; (a) calibration of 40 German samples measured in Germany with German reference values, (b) calibration of 157 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (36 German and 141 Chinese samples) with respective reference values.

Tab.4.16 Comparison of Wet Gluten reference and prediction values for 5 German (a) and 15 Chinese (b) test samples based on the calibration of Fig. 4.18(c).

(a)

Samples #(GER)	6	15	25	41
Ref.	30.8	29.8	27.4	35.1
Pred.	28.60	30.10	27.30	32.00
AE (Ref.-Pred.)	2.2	0.3	0.1	3.1
MAE	1.43			

(b)

Samples #(CHN)	2	21	23	24	37	47	62	78
Ref.	26.7	30.5	22.3	28.3	42.8	20.6	31.4	28.9
Pred.	26.60	30.60	25.20	29.60	42.20	22.70	29.30	29.10
AE (Ref.-Pred.)	0.1	0.1	2.9	1.3	0.6	2.1	21	0.2
Samples #(CHN)	81	102	130	142	148	152	154	164
Ref.	27.7	36.3	34.4	29.6	30.3	32.8	31.9	29.9
Pred.	28.30	36.90	32.50	29.50	28.90	34.10	33.10	28.70
AE (Ref.-Pred.)	0.6	0.6	1.9	0.1	1.4	1.3	1.2	1.2
MAE	1.11							

From the absolute difference of the reference and predicted Wet Gluten values of the test samples, the lowest and highest AE are 0.1 and 3.1, respectively. MAE of 1.43 and 1.11 are calculated for Germany and China, respectively.

4.3.2.2 Calibration Model and Prediction Results Obtained with the Spectra of the Neospectra Micro Spectrometer

Prior to PLS calibration model development of the spectra measured with the Neospectra Micro instrument, outliers were removed according to the method described in Section 4.1.3 and then the spectra were smoothed and scatter effects were eliminated by EMSC (Fig. 4.19).

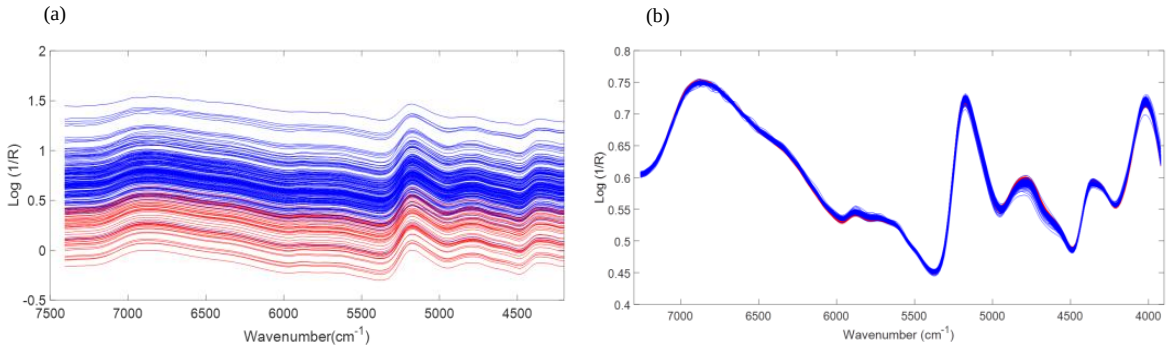


Fig. 4.19 Spectra of 50 German (red) and 163 Chinese (blue) flour samples measured with the Neospectra Micro handheld spectrometer. (a) Raw spectra (7407-3922 cm^{-1}) and (b) EMSC pretreated and truncated spectra (7258-3922 cm^{-1}) after removal of 2 Chinese outlier samples.

In the Fig. 4.20-4.22 the results of the PLS calibration models for protein, moisture, and wet gluten, respectively, are shown. The data were modeled in three steps: (a) modeling of German samples only, (b) modeling of Chinese samples only, and (c) calibration development after merging the samples from both countries. In Fig. 4.20(c)-4.22(c) the calibration set of the German samples is accentuated by black circles and in the Tab. 4.17-4.19 the prediction results of the test sets are summarized for the three different flour parameters.

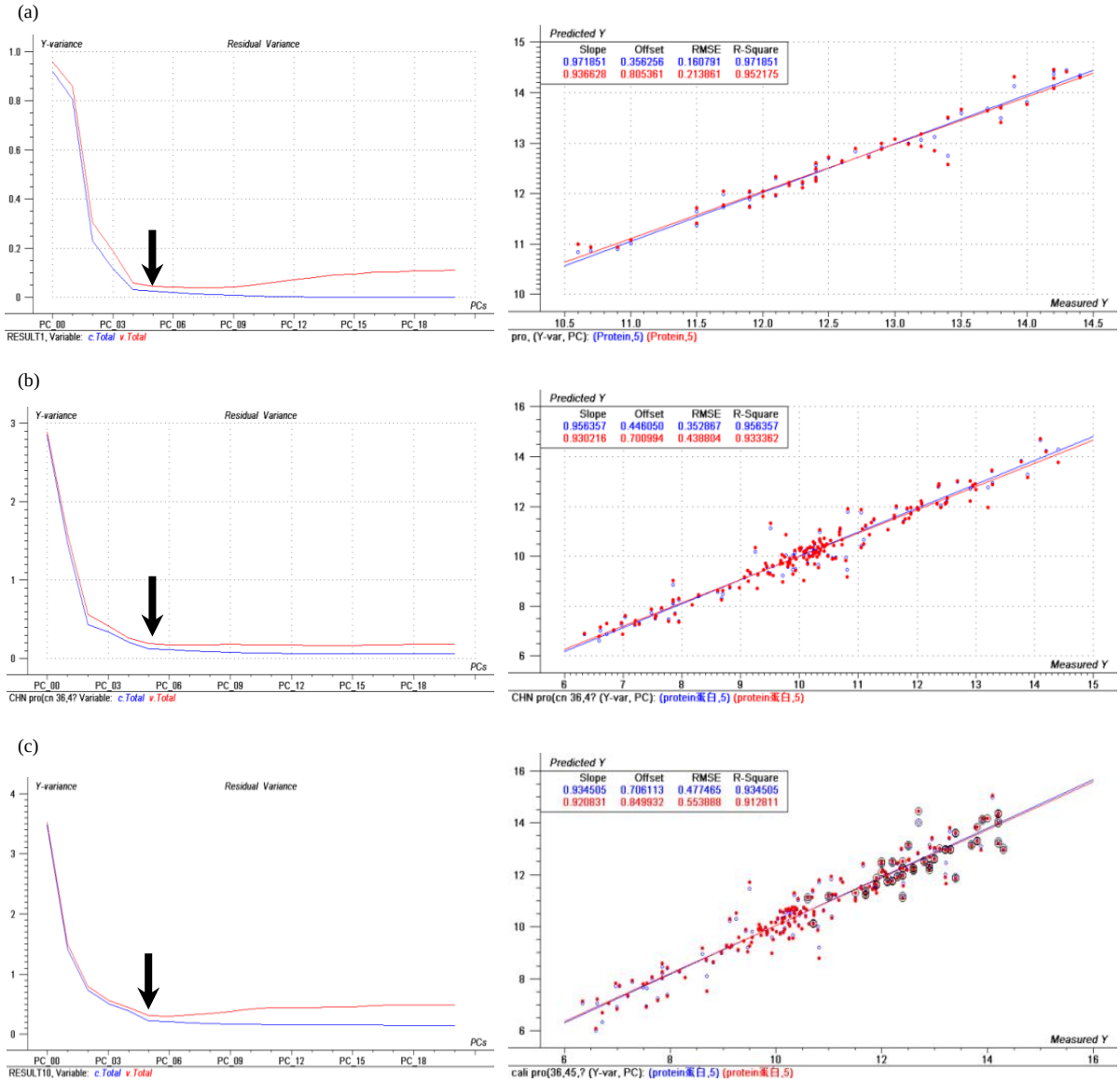


Fig. 4.20 Comparison of country-specific and merged protein calibrations achieved with the spectra measured on the Neospectra Micro spectrometer; (a) calibration of 50 German samples measured in China with German reference values, (b) calibration of 156 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (45 German and 140 Chinese samples) with respective reference values.

Tab. 4.17 Comparison of protein reference and prediction values for five German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.20 (c).

(a)

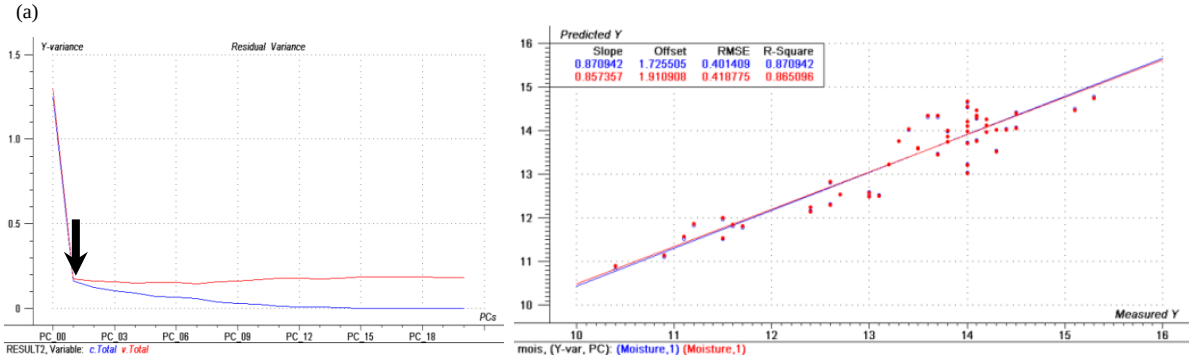
Samples #(GER)	11	16	37	38	49
Ref.	11.9	12.8	12.3	14.4	13.5
Pred.	12.66	11.70	12.15	13.97	14.19
AE (Ref.-Pred.)	0.76	1.10	0.15	0.43	0.69
MAE	0.63				

(b)

Samples #(CHN)	1	3	6	37	62	64	85	87
Ref.	9.1	7.9	9.7	14.4	10.1	9.9	10.7	7.2
Pred.	7.22	8.38	11.14	14.79	10.32	10.38	11.25	7.52
AE (Ref.-Pred.)	1.88	0.48	1.44	0.39	0.22	0.48	0.55	0.32
Samples #(CHN)	89	100	101	108	131	133	148	149
Ref.	10.3	12.4	11.8	13.0	9.4	11.1	10.4	10.2
Pred.	10.58	12.82	11.26	12.63	9.56	10.89	10.65	10.13
AE (Ref.-Pred.)	0.28	0.42	0.54	0.37	0.16	0.21	0.25	0.07
MAE	0.50							

From the absolute difference of the reference and predicted protein values of the test samples, the lowest and highest AE are 0.07 and 1.88, respectively. MAE of 0.63 and 0.50 are calculated for Germany and China, respectively.

(a)



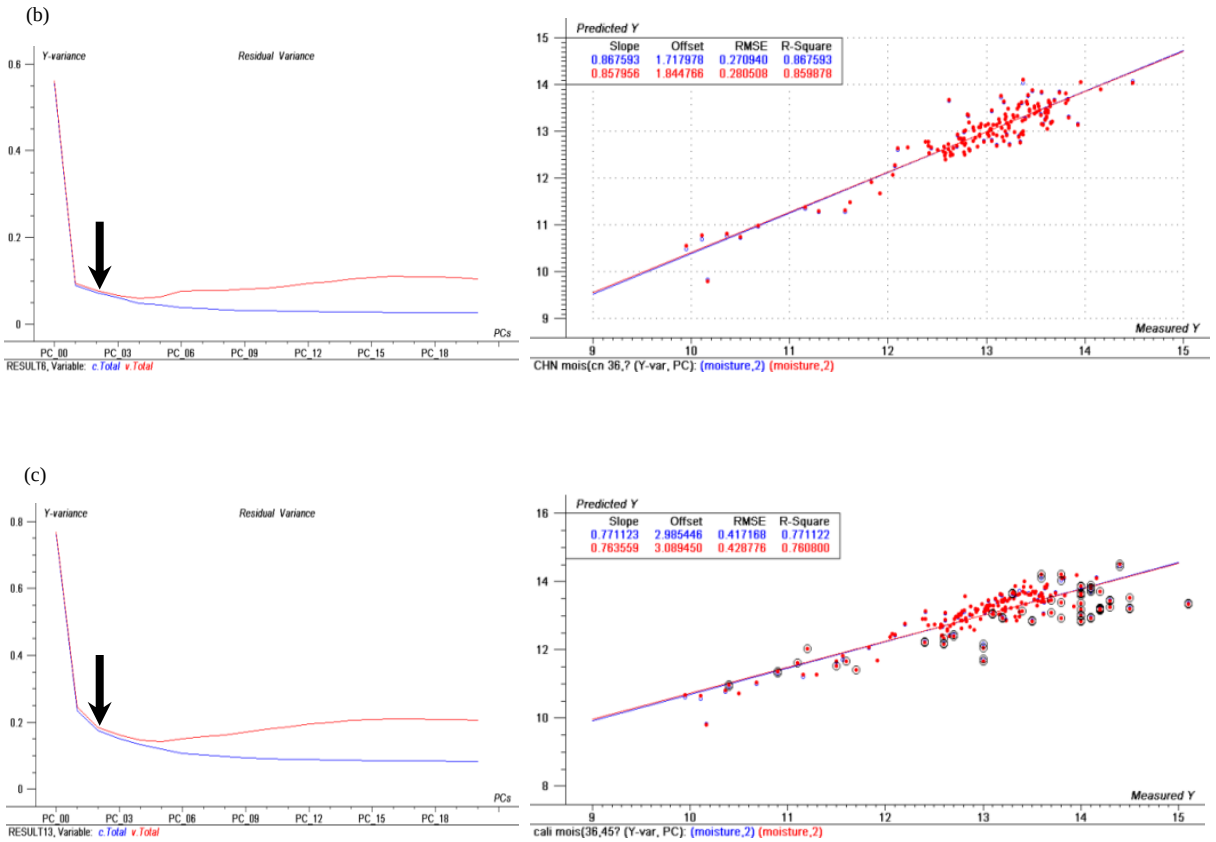


Fig. 4.21 Comparison of country-individual and merged moisture calibrations obtained with the spectra measured on the Neospectra Micro spectrometer; (a) calibration of 50 German samples measured in China with German reference values, (b) calibration of 156 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (45 German and 141 Chinese samples) with respective reference values.

Tab. 4.18 Comparison of moisture reference and prediction values for five German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.21(c).

(a)

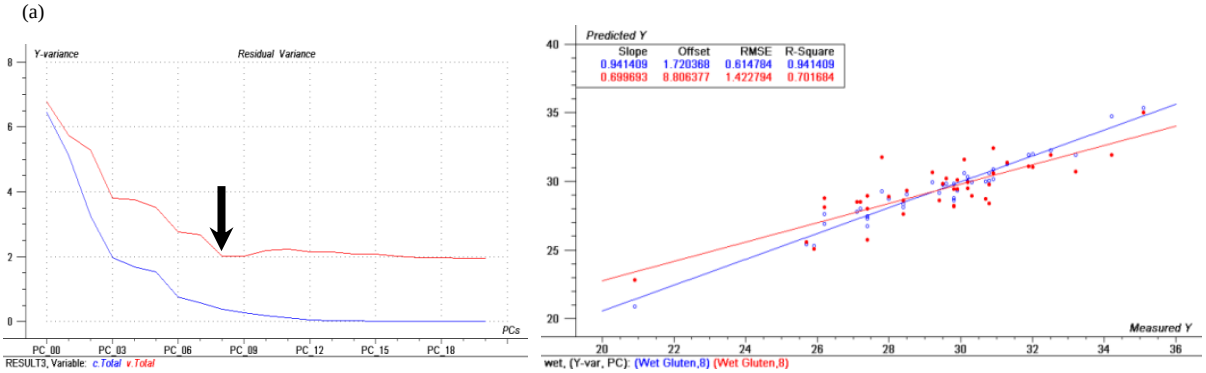
Samples #(GER)	1	18	32	40	49
Ref.	13.5	14	15.3	14.1	12.4
Pred.	13.61	13.72	14.58	13.66	12.58
AE (Ref.-Pred.)	0.11	0.28	0.72	0.44	0.18
MAE	0.35				

(b)

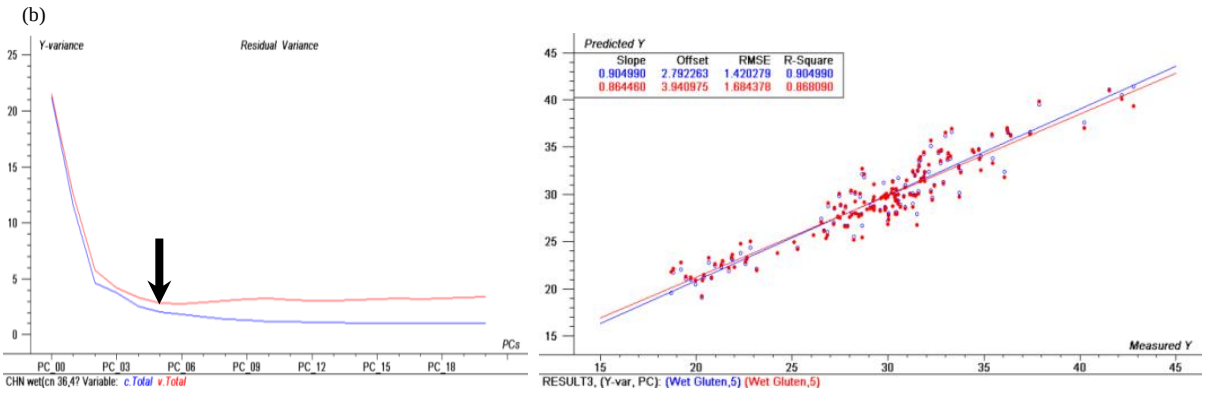
Samples #(CHN)	17	29	57	62	63	64	92	97
Ref.	12.6	12.94	13.2	13.8	12.8	14.5	12.4	11.6
Pred.	13.88	13.03	13.27	13.88	12.79	15.12	12.43	11.30
AE (Ref.-Pred.)	1.28	0.09	0.07	0.08	0.01	0.62	0.03	0.30
Samples #(CHN)	109	118	124	137	144	147	158	160
Ref.	13.0	13.4	13.1	13.6	13.3	13.6	13.4	12.8
Pred.	13.22	13.63	13.30	14.09	13.39	13.73	13.37	13.24
AE (Ref.-Pred.)	0.22	0.23	0.20	0.49	0.09	0.13	0.03	0.44
MAE	0.27							

From the absolute difference of the reference and predicted Moisture values of the test samples, the lowest and highest AE are 0.01 and 1.28, respectively. MAE of 0.35 and 0.27 are calculated for Germany and China, respectively.

(a)



(b)



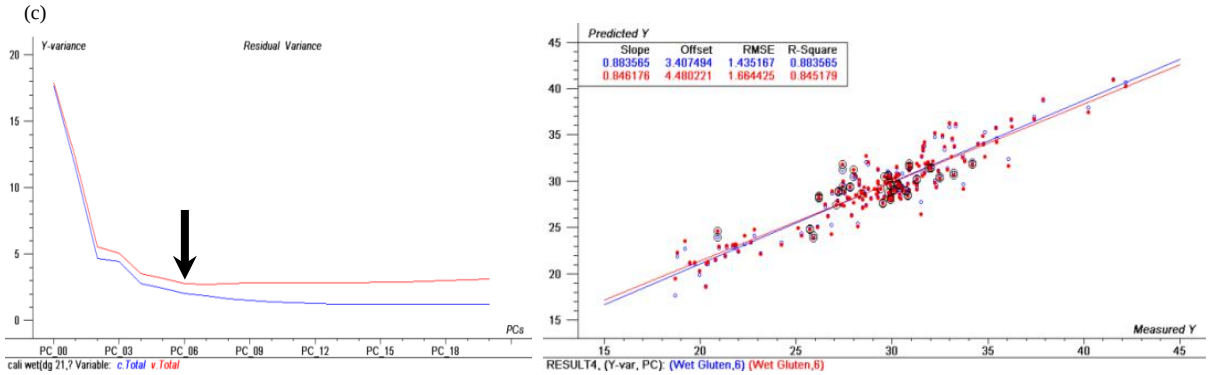


Fig. 4.22 Comparison of country-specific and combined wet gluten calibrations achieved with the spectra measured on the Neospectra Micro spectrometer; (a) calibration of 40 German samples measured in Germany with German reference values, (b) calibration of 156 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (36 German and 139 Chinese samples) with respective reference values.

Tab. 4.19 Comparison of wet gluten reference and prediction values for four German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.22(c).

(a)

Samples #(GER)	6	15	25	41
Ref.	30.8	29.8	27.4	35.1
Pred.	28.70	29.70	26.20	34.10
AE (Ref.-Pred.)	2.1	0.1	1.2	1.0
MAE	1.10			

(b)

Samples #(CHN)	31	35	37	47	56	60	62	65
Ref.	29.6	29.1	42.8	20.6	27.7	22.6	31.4	28.3
Pred.	29.60	27.80	39.30	22.00	26.60	21.50	29.90	28.00
AE (Ref.-Pred.)	0.0	1.3	3.5	1.4	1.1	1.1	1.5	0.3
Samples #(CHN)	70	84	106	112	130	149	152	154
Ref.	36.4	30.0	26.8	30.3	34.4	30.6	32.8	31.9
Pred.	36.30	27.10	25.40	29.30	34.40	29.20	24.80	33.90
AE (Ref.-Pred.)	0.1	2.9	1.4	1.0	0.0	1.4	8.0	2.0
MAE	1.69							

From the absolute difference of the reference and predicted wet gluten values of the test samples, the lowest and highest AE are 0.0 and 8.0, respectively. MAE of 1.10 and 1.69 are calculated for Germany and China, respectively.

4.3.2.3 Calibration Model and Prediction Results Obtained with the Spectra Measured on the Hamamatsu C15511-01 Spectrometer

To develop PLS calibration models with the Hamamatsu C15511-01 spectra, in a first step outliers in the original raw spectra were eliminated (see Section 4.1.3), followed by smoothing, EMSC scatter correction and truncation (Fig. 4.23).

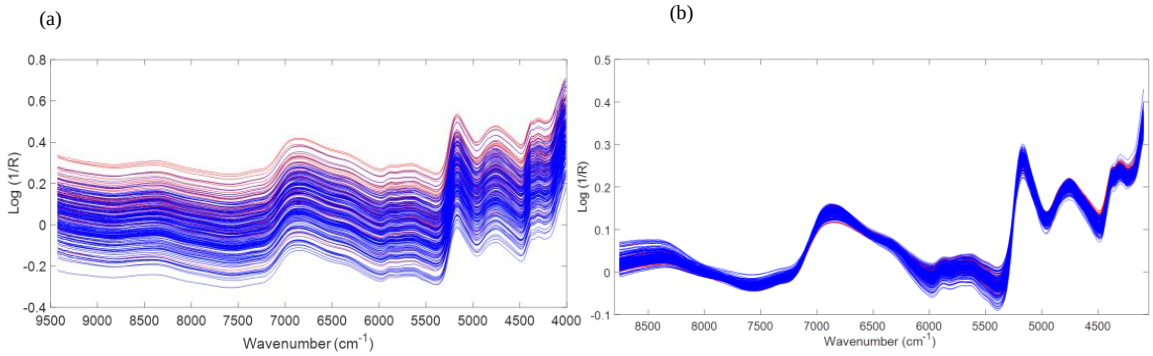


Fig. 4.23 Spectra of 50 German (red) and 163 Chinese (blue) flour samples measured with the Hamamatsu C15511-01 spectrometer. (a) Raw spectra (9460-4000 cm^{-1}) and (b) EMSC pretreated and truncated spectra (8757-4096 cm^{-1}) after removal of two Chinese outlier samples

The results of the PLS calibration models based on Hamamatsu C15511-01 spectra for protein, moisture, and wet gluten are shown in Figs. 4.24-4.26, respectively. Calibration model development is performed in three steps: (a) modeling for German samples only, (b) modeling of Chinese samples only, and (c) calibration model development after merging the samples of both countries. In Figs. 4.24(c)-4.26(c) the German calibration samples are marked with black circles. The predictions of the test set samples for the different flour parameters are summarized in Tab. 4.20-4.22.

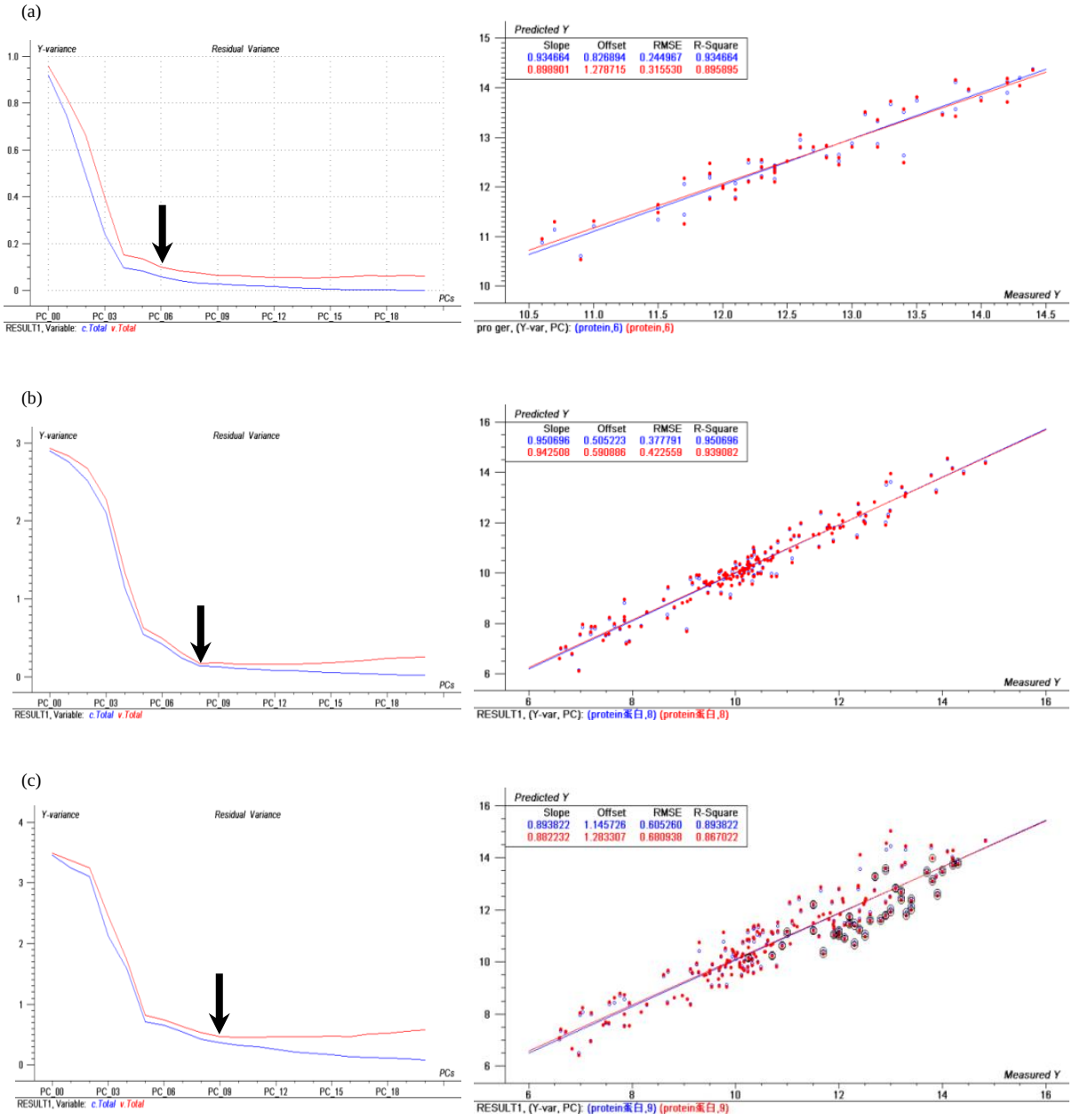


Fig. 4.24 Comparison of country-specific and merged protein calibrations achieved with the spectra measured on the Hamamatsu C15511-01 spectrometer; (a) calibration of 50 German samples measured in Germany with German reference values, (b) calibration of 161 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (45 German and 144 Chinese samples) with respective reference values.

Tab. 4.20 Comparison of protein reference and prediction values for five German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.24(c).

(a)

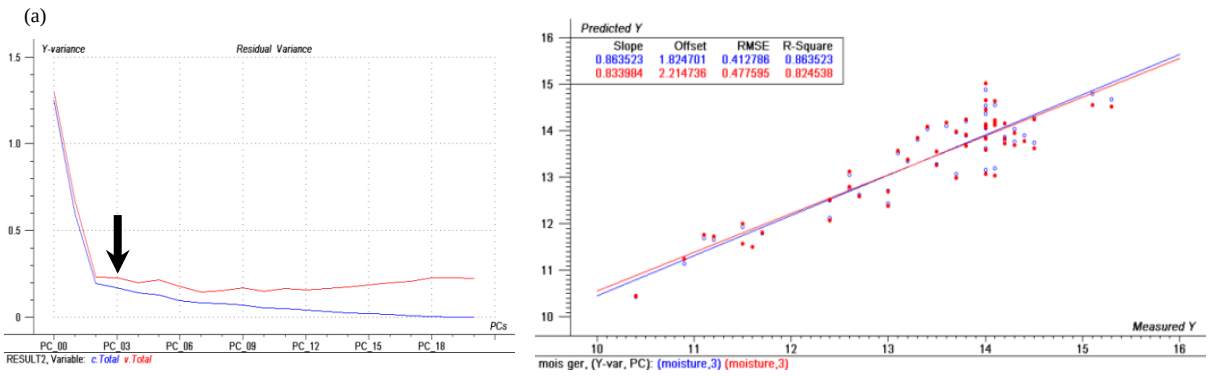
Samples #(GER)	11	16	37	38	49
Ref.	11.9	12.8	12.3	14.4	13.5
Pred.	11.84	11.28	12.19	14.25	14.01
AE (Ref.-Pred.)	0.06	1.52	0.11	0.15	0.51
MAE	0.47				

(b)

Samples #(CHN)	1	3	5	6	29	37	38	42
Ref.	9.1	7.9	7.3	9.7	13.0	14.4	10.1	10.2
Pred.	6.88	7.07	7.63	8.89	12.75	14.01	9.68	10.21
AE (Ref.-Pred.)	2.22	0.83	0.33	0.81	0.25	0.39	0.42	0.01
Samples #(CHN)	53	63	64	111	131	153	154	156
Ref.	10.7	10.4	9.9	11.1	9.4	11.8	12.3	10.3
Pred.	10.03	10.74	8.82	11.52	10.12	12.71	12.48	12.25
AE (Ref.-Pred.)	0.67	0.34	1.08	0.42	0.72	0.91	0.18	1.95
MAE	0.72							

From the absolute difference of the reference and predicted protein values of the test samples, the lowest and highest AE are 0.01 and 2.22, respectively. MAE of 0.47 and 0.72 are calculated for Germany and China, respectively.

(a)



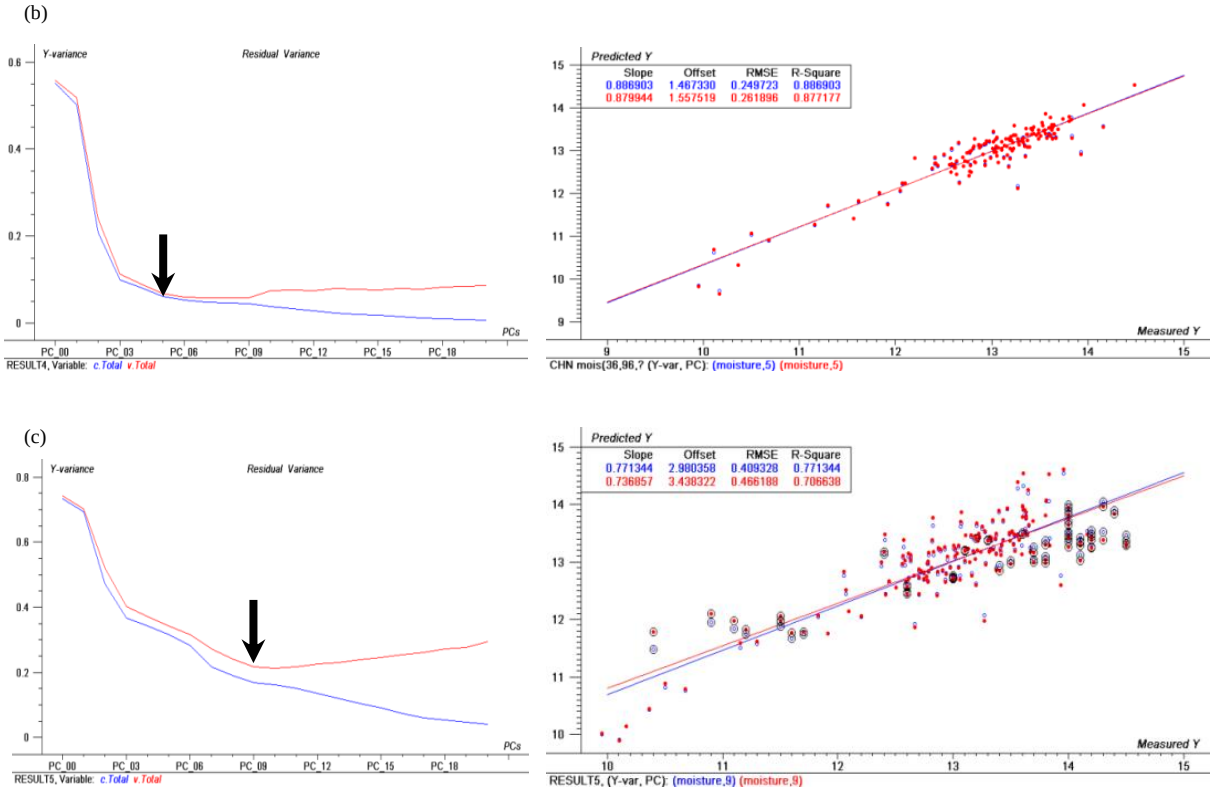


Fig. 4.25 Comparison of country-individual and merged moisture calibrations achieved with the spectra measured on the Hamamatsu C15511-01 instrument; (a) calibration of 50 German samples measured in Germany with German reference values, (b) calibration of 161 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (43 German and 140 Chinese samples) with respective reference values.

Tab. 4.21 Comparison of moisture reference and prediction values for five German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.25(c).

(a)

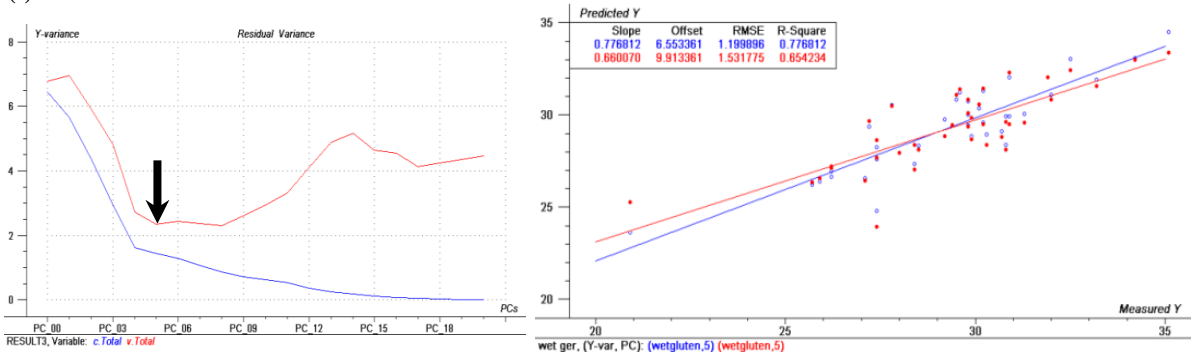
Samples #(GER)	1	18	32	40	49
Ref.	13.5	14	15.3	14.1	12.4
Pred.	13.21	13.30	14.79	12.26	12.15
AE (Ref.-Pred.)	0.29	0.70	0.51	1.84	0.25
MAE	0.72				

(b)

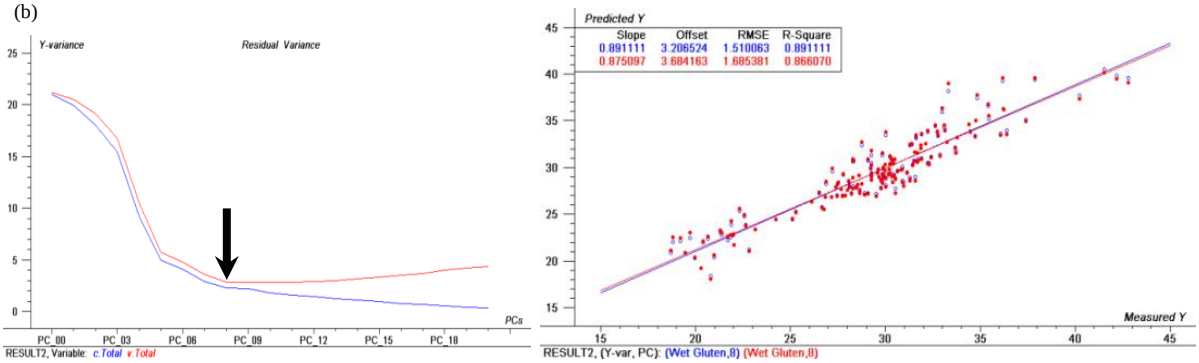
Samples #(CHN)	1	17	20	32	35	41	45	49
Ref.	13.1	12.6	13.20	13.355	13.7	13.4	12.8	13.3
Pred.	12.82	12.70	13.41	13.54	13.71	13.00	13.23	13.52
AE (Ref.-Pred.)	0.28	0.10	0.21	0.185	0.01	0.40	0.43	0.22
Samples #(CHN)	52	64	73	83	92	97	108	143
Ref.	13.6	14.5	12.8	13.0	12.4	11.6	13.6	12.9
Pred.	13.69	14.90	13.14	13.06	12.06	11.37	13.24	13.00
AE (Ref.-Pred.)	0.09	0.40	0.34	0.06	0.34	0.23	0.36	0.10
MAE	0.23							

From the absolute difference of the reference and predicted moisture values of the test samples, the lowest and highest AE are 0.01 and 1.84, respectively. MAE of 0.72 and 0.23 are calculated for Germany and China, respectively.

(a)



(b)



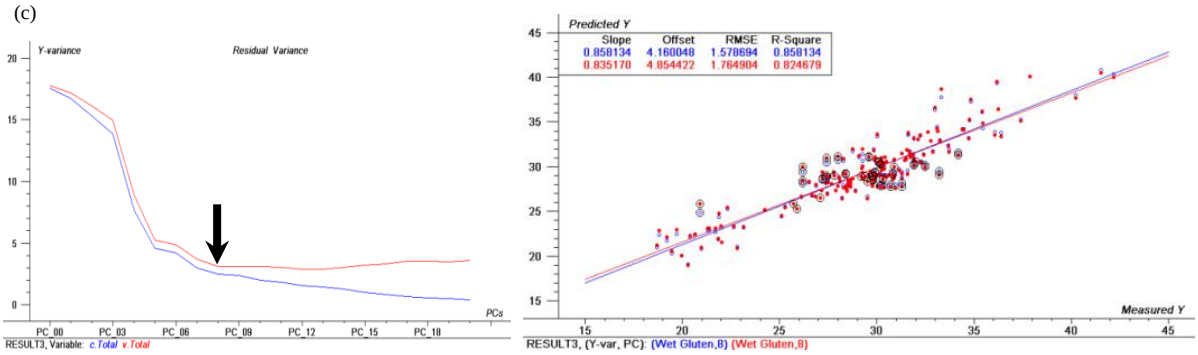


Fig. 4.26 Comparison of country-specific and merged wet gluten calibrations achieved with the spectra measured on the Hamamatsu C15511-01 spectrometer; (a) calibration of 40 German samples measured in Germany with German reference values. (b) Calibration of 159 Chinese samples measured in China with Chinese reference values, and (c) calibration of merged samples (36 German and 143 Chinese samples) with respective reference values.

Tab. 4.22 Comparison of wet gluten reference and prediction values for four German (a) and 16 Chinese (b) test samples based on the calibration of Fig. 4.26(c).

(a)

Samples #(GER)	6	15	25	41
Ref.	30.8	29.8	27.4	35.1
Pred.	28.70	29.10	24.70	34.00
AE (Ref.-Pred.)	2.1	0.7	2.7	1.1
MAE	1.65			

(b)

Samples #(CHN)	20	21	22	35	37	56	60	65
Ref.	36.2	30.5	30.0	29.1	42.8	27.7	22.6	28.3
Pred.	36.40	27.40	30.80	27.40	39.70	27.30	25.00	31.20
AE (Ref.-Pred.)	0.2	3.1	0.8	1.7	3.1	0.4	2.4	2.9
Samples #(CHN)	106	111	118	119	120	148	157	160
Ref.	26.8	31.9	29.6	31.4	32.7	30.3	20.8	33.8
Pred.	25.60	32.30	30.10	32.70	34.30	29.90	17.30	32.70
AE (Ref.-Pred.)	1.2	0.4	0.5	1.3	1.6	0.4	3.5	1.1
MAE	1.54							

From the absolute difference of the reference and predicted wet gluten values of the test samples, the lowest and highest AE are 0.2 and 3.5, respectively. MAE of 1.65 and 1.54 are calculated for Germany and China, respectively.

4.4 Optimization of the PCA , PLS-DA and PLS Calibration Models

4.4.1 PCA Calibration Models

(1) According to the classification results represented in Figs. 4.5 and 4.15 (a) and (b), it can be derived, that the PCA score-plot classification effects observed for the benchtop IAS 3100 and the handheld MicroNIR spectrometer for wheat flour samples from different countries are excellent. The reason can be primarily assigned to the fact that China and Germany are geographically far apart, and there are large differences in both wheat varieties due to natural conditions, cultivation methods, field management patterns, and flour processing techniques and technologies. These distinctive features lead to characteristic differences in their vibrational spectra (mid- and near-infrared), that may not be obvious from visual inspection, but are accentuated by the multivariate, chemometric evaluation techniques. Thus, NIR spectroscopy with handheld instrumentation provides a rapid, non-invasive detection method to trace the origin of different flour species.

(2) From Figs. 4.15 (c) and (d), it is found that PCA has a poor identification effect for the two FT-NIR handheld spectrometers (NeoSpectra Micro and Hamamatsu C15511-01). In the 2D score plot the spectra of the flour samples from China and Germany are mixed together, and no clustering can be observed. Thus, it is necessary to find another discriminant analysis method to try to solve the problem of FT-NIR spectral classification.

(3) Comparing the PCA score plots used for the identification of geographical origin, it is recognized that the principal components used for the respective score plots are different for different instruments. Thus, PC2 and PC3 provided the best separation for the data measured with the IAS 3100 instrument, whereas PC1 and PC2 performed best for the MicroNIR data. Other choices of principal components for the score plots are possible, but are often

accompanied by a decrease of discrimination and identification performance. The example of the IAS 3100 instrument (PC2 and PC3) shows, that the first principal components of the PCA analysis is not necessarily the best choice for a score plot, because higher factors may contain more important information for the discrimination. In conclusion, the optimum combination of principal components for a PCA score plot will have to be adjusted from case to case to the specific analytical discrimination problem.

4.4.2 PLS-DA Calibration Models

Since the origin of flour cannot be well discriminated by the FT-NIR spectrometer (NeoSpectra Micro and Hamamatsu C15511-01) using PCA, the PLS-DA discriminant analysis method was applied to verify sections 4.2.1 and 4.3.1 and these results were compared with the PCA results (since only Chinese flour samples were measured with the NIR-freespace spectrometer, the PLS-DA method was not used to analyze these spectra).

Tab. 4.23 Number of different latent variables selected for the PLS-DA models of the four spectrometers

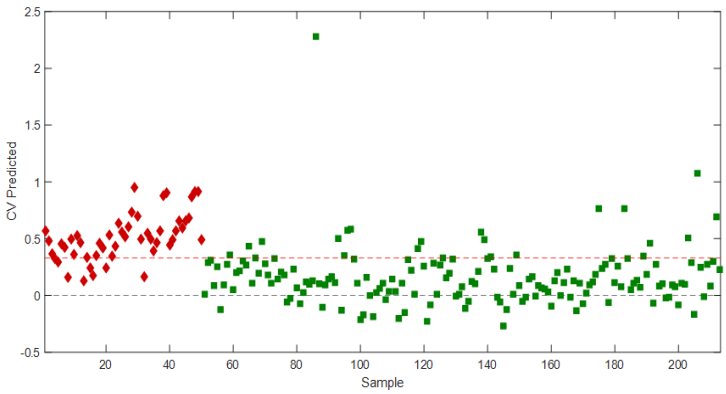
NIR spectrometers	Number of different latent variables
Hamamatsu C15511-01	4 LVs
NeoSpectra Micro	5 LVs
MicroNIR	2 LVs
IAS 3100	1 LVs

In fact, in NIR spectroscopy applications, PLS-DA is designed to predict sample class membership via the Y-axis from spectral data contained in the X-axis. The X-axis represents the number of samples, and the Y-axis is the predicted result of data cross-validation.

First, the 213 flour spectra from all spectrometers were preprocessed using the EMSC method, after which the amount of latent variables (LVs) was determined by leave-one-out cross-validation, as shown in Tab. 4.23. In PLS-DA, the dimensionality reduction transformation results in LV, which is a linear combination of spectral variables trying to explain the maximum covariance between X and Y. The smaller the number of latent variables

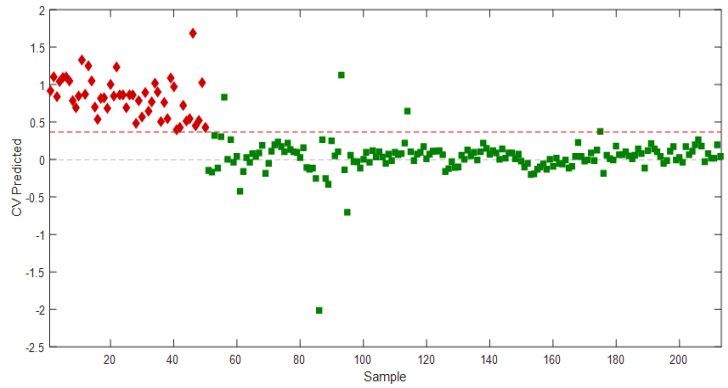
selected after cross-validation, the better the classification effect of PLS-DA is. The next step is to draw a mean-centered line for all data classes (Fig. 4.27).

(a) Hamamatsu C15511-01



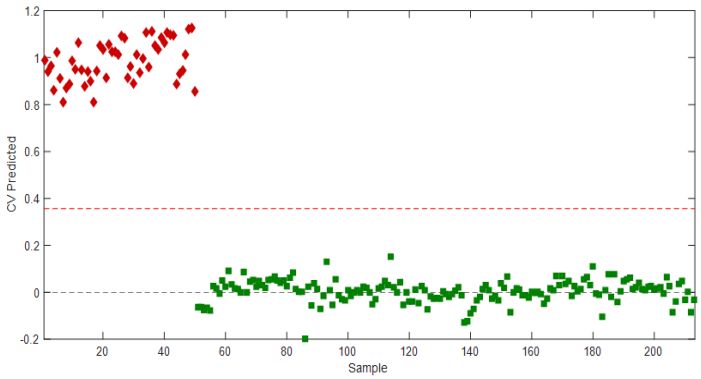
	Class 1	Class 2	Accuracy (%)
Parameters			
Sensitivity Cal (%)	88.0	86.5	87.3
Sensitivity CV (%)	82.0	86.5	84.3
Specificity Cal (%)	86.5	88.0	87.3
Specificity CV (%)	86.5	82.0	84.3

(b) NeoSpectra Micro



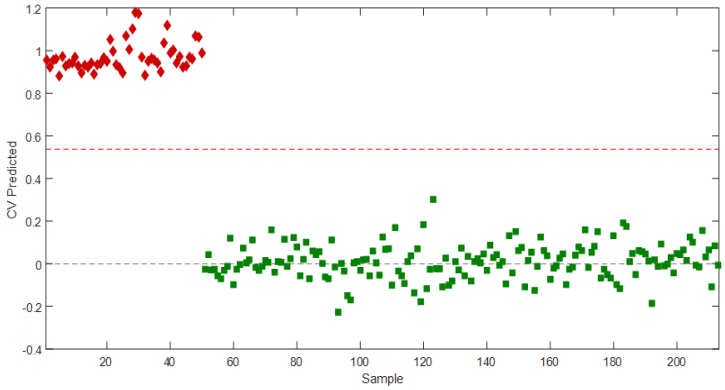
	Class 1	Class 2	Accuracy (%)
Parameters			
Sensitivity Cal (%)	100.0	98.2	99.1
Sensitivity CV (%)	100.0	97.5	98.8
Specificity Cal (%)	98.2	100.0	99.1
Specificity CV (%)	97.5	100.0	98.8

(c) MicroNIR



	Class 1	Class 2	Acurracy
Parameters			
Sensitivity Cal (%)	100	100	100
Sensitivity CV (%)	100	100	100
Specificity Cal (%)	100	100	100
Specificity CV (%)	100	100	100

(d) IAS 3100



	Class 1	Class 2	Acurracy
Parameters			
Sensitivity Cal (%)	100	100	100
Sensitivity CV (%)	100	100	100
Specificity Cal (%)	100	100	100
Specificity CV (%)	100	100	100

Fig. 4.27 PLS-DA scatter plot and discrimination parameters for the spectra of the German (red) and Chinese (green) samples measured with the four spectrometers under investigation
(a) Hamamatsu C15511-01; (b) NeoSpectra Micro; (c) MicroNIR; (d) IAS 3100

In the tables included in Fig. 4.27, the second column (Class 1) and the third column (Class 2) represent the sample set of German flour and the sample set of Chinese flour, respectively.

As shown in Figure 4.27 (c) and (d), after selecting two LVs and one LV, respectively, the spectral data of the MicroNIR instrument and the spectral data of the IAS 3100 spectrometer can be clearly divided into Class 1 and Class 2. The accuracy of both spectrometers is 100%, which is expected and consistent with the results of the MicroNIR and IAS 3100 spectrometers obtained by PCA (see Section 4.2.1 and 4.3.1). However, for the results of the two FT-NIR spectrometers (Hamamatsu C15511-01 and Neospectra Micro), as shown in Figure 4.27 (a) and (b), their prediction results are not as accurate as those of the other two spectrometers (MicroNIR and IAS 3100). Compared with the previous classification results of PCA (see Section 4.3.1), the prediction accuracy of Neospectra Micro (Si-Ware) and Hamamatsu C15511-01 using PLS-DA is greatly improved, and the prediction accuracy of their cross-validation is 84.3% and 98.8%, respectively. The prediction results of Neospectra Micro (Si-Ware) are slightly better than Hamamatsu C15511-01, which is also consistent with the previous trend of the corresponding PCA classification results.

4.4.3 Optimization of PLS Modeling Results

According to Tab. 4.24, it is possible to compare and classify the calibration results for each parameter, thereby ranking the calibration performance for the benchtop instrument (IAS 3100) and the three handheld spectrometers (Neospectra Micro (Si-Ware), MicroNIR (VIAVI), C15511-01 (Hamamatsu)).

Tab. 4.24 PLS calibration results for the three flour parameters under investigation obtained with the merged spectra sets (213 spectra) for the German and Chinese samples and recorded with the four different NIR instruments.

Calibration parameters		Flour parameters			Calibration parameters		Flour parameters		
		Protein	Moisture	Wet Gluten			Protein	Moisture	Wet Gluten
Outliers (Cal.Set/ Test.Set)	IAS 3100	1/0	5/0	1/0	R^2_P	IAS 3100	0.9554	0.9430	0.8346
	Neospectra	7/0	6/0	4/0		Neospectra	0.8761	0.6434	0.9062
	MicroNIR	3/0	7/0	2/0		MicroNIR	0.9638	0.8846	0.8759
	C15511-01	2/0	9/0	1/0		C15511-01	0.7922	0.6096	0.8420
Number of Factors	IAS 3100	5	6	6	R^2_C	IAS 3100	0.9530	0.9290	0.8784
	Neospectra	5	2	6		Neospectra	0.9345	0.7711	0.8836
	MicroNIR	6	6	6		MicroNIR	0.9302	0.9271	0.8290
	C15511-01	9	9	8		C15511-01	0.8938	0.7713	0.8581
RMSEC (% (w/w))	IAS 3100	0.4082	0.2196	1.4796	R^2_{CV}	IAS 3100	0.9479	0.9201	0.8622
	Neospectra	0.4775	0.4172	1.4352		Neospectra	0.9128	0.7608	0.8452
	MicroNIR	0.4958	0.2358	1.7558		MicroNIR	0.9158	0.9130	0.7949
	C15511-01	0.6053	0.4093	1.5788		C15511-01	0.8670	0.7066	0.8247
RMSECV (% (w/w))	IAS 3100	0.4319	0.2342	1.5841	RPD	IAS 3100	3.3862	3.0049	1.8154
	Neospectra	0.5539	0.4288	1.6644		Neospectra	2.0741	1.3063	2.3649
	MicroNIR	0.5475	0.2590	1.9335		MicroNIR	3.7506	2.1443	2.0726
	C15511-01	0.6809	0.4662	1.7649		C15511-01	1.6386	1.2615	1.8536
RMSEP (% (w/w))	IAS 3100	0.4032	0.2778	1.9225					
	Neospectra	0.6820	0.4894	1.4653					
	MicroNIR	0.3784	0.2624	1.6864					
	C15511-01	0.8785	0.5120	1.8691					

(1) The comparison of the calibration results for the three flour parameters presented in Tab. 4.24 shows, that protein produced the best results among all four instruments, followed by moisture and wet gluten. Proteins are polymers with amino acid repetition units and include CH, NH, OH, and amide (CONH) functionalities, that lead to strong absorption bands in the NIR spectra. The availability of these characteristic spectroscopic signatures certainly contributes to the superior protein calibration models. Tab. 4.24 contains the important RMSE parameters for calibration/cross-validation and test-set prediction of the PLS calibrations for protein, moisture and wet gluten obtained with the spectra measured on the four instruments under investigation. The PLS calibrations developed with the spectra measured on the IAS 3100 instrument had very similar RMSECV and RMSEP values for the protein and moisture parameters, but the corresponding RMSE values for wet gluten showed similar differences of up to 0.35 %(w/w) as the RMSECV and RMSEP values for the other three handheld spectrometers for all three parameters.

(2) Wet gluten is an index that is characteristic of the adhesion of dough after the addition of water to the flour. For the determination of wet gluten, the dough is prepared from the wheat sample with sodium chloride buffer. In the next step, starch, sugar, cellulose, and soluble protein are isolated from the dough by washing with sodium chloride buffer, and after removal of the excess washing solution, the remaining gelatinous material is representative of wet gluten. Therefore, the content of wet gluten correlates with the protein content, viz. the higher the protein content, the higher the wet gluten value. Water has strong OH-specific absorption bands at 5167 cm^{-1} and 6855 cm^{-1}, however, the absorption bands of the OH-functionalities of starch and proteins can interfere with the determination of water and lead to a lower calibration performance of the moisture parameter. Another source of error is the dependence of moisture content on the time and storage conditions between reference and NIR measurements.

(3) Theoretically, the calibration results for moisture should be better than those for wet gluten, because water has very strong and characteristic absorption bands (Fig. 4.28). The result of the moisture calibration actually depends on the time interval between the determination of

the moisture reference values and the acquisition of the NIR spectra of the corresponding flour sample: the shorter the time interval, the lower the calibration error and the higher the accuracy of the modeling. However, the actual calibration results are only slightly better than those for wet gluten (refer to sections 4.2 and 4.3 for a comparison of the PLS calibration results for moisture and wet gluten). The main reason for this is that the total time span of the three stages of the project was very long: the time interval between the determination of the German flour moisture parameter values and the completion of the 2nd stage of the test being more than 15 months. The time interval between the acquisition of the Chinese flour moisture parameter values and the completion of the 3rd stage of testing is also more than one year, so that the moisture calibration results are not optimal.

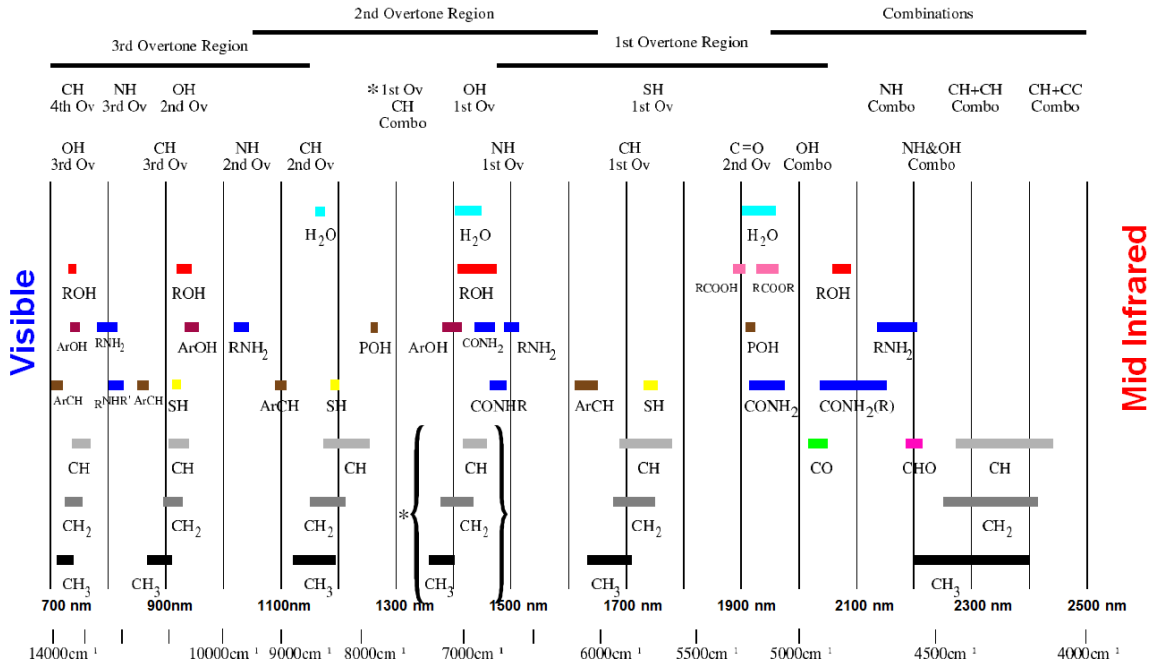


Fig. 4.28 Band assignments of functional groups in the NIR spectrum.

4.4.3 Influence of Spectrometer Conditions on Calibration Performance

(1) The influence of and competition between available wavenumber range and signal/noise (S/N) ratio for instrument performance becomes relevant in the comparison of the four spectrometers used for this project. Although in the 4000-6000 cm^{-1} wavenumber range the

Neospectra Micro instrument contains several additional intense combination and overtone absorption bands of the above-mentioned functionalities compared to the wavenumber range of the MicroNIR (see also Figs. 4.19 and 4.15), surprisingly, the calibration performance of the Neospectra Micro spectrometer is at best equivalent. Actually, the calibration for protein is slightly better for the Neospectra Micro spectrometer compared to the MicroNIR instrument (RMSEC values of 0.4775 and 0.4958 for Neospectra Micro and MicroNIR spectrometers, respectively, but RMSEP values of 0.6820 and 0.3784, respectively). For moisture, on the other hand, the MicroNIR spectrometer has much better calibration parameters. Thus, for these parameters, the higher S/N ratio obviously compensates the disadvantage of the narrower wavenumber range.

(2) The best calibration models for the three flour parameters are achieved with the IAS 3100 instrument. This is mainly due to the fact, that the instrument uses the monochromator principle of Texas Instrument ś digital mirror device (DMD™), which leads to a high signal-to-noise ratio. Furthermore, because it is a benchtop instrument with more internal space which dissipates heat better, high-power light sources can be used, thereby improving the intensity of the signals and the stability of the instrument.

(3) Although both, Neospectra Micro and Hamamatsu C15511-01 instruments, are Fourier-Transform NIR instruments, use MEMS technology, and operate almost over the entire NIR spectral range thereby covering more spectroscopic information on organic compounds, the performance of their calibration models is not as good as that of the MicroNIR instrument with LVF technology. Additionally, the Hamamatsu C15511-01 spectrometer is built with an external light source (see photo of Hamamatsu C15511-01 instrument in Chapter 3, Fig. 3.7), the measurement procedure is subject to more influencing factors and the spectral stability obtained is not as good as that of the Neospectra Micro spectrometer, which has been realized as an integrated monolithic spectrometer (see photo of Neospectra Micro spectrometer in Chapter 3, Fig. 3.6). The positions of the light source, interferometer and detector of the Neospectra Micro instrument are fixed, thereby leading to a higher spectral stability. With the progress of MEMS technology, it is believed that this kind

of FT-NIR technology will be widely used in various industries in the near future.

(4) Miniaturization of the spectrometer leads to a better portability which is the basis for its application as in-the-field and on-site measurement system. Among the three handheld NIR spectrometers used in this project, the MicroNIR and Neospectra Micro instruments have the better handiness and are therefore easier to use. Nevertheless, of these two handheld spectrometers, the MicroNIR instrument has the higher S/N ratio and is therefore more widely used despite the disadvantage of a higher price.

4.5 Chapter Summary

(1) In this PhD project, five NIR spectrometers (two benchtop and three handheld systems) are used to measure the NIR spectra of 50 German and 163 Chinese flour samples. Based on the near-infrared spectral characteristics of the samples, PCA models (Figs. 4.4 and Fig. 4.14) were established, in which IAS 3100 and MicroNIR (VIAVI) achieved 100% correct classification of German and Chinese flour.

(2) PLS calibration models were developed for protein, moisture and wet gluten for the samples of both countries, and the number of samples used for the calibrations of these three parameters are summarized in Tab. 4.24. The calibration models for the three flour parameters were developed with the spectra measured on two benchtop spectrometers and three portable instruments. The described experimental results prove that accurate calibrations can be achieved with the benchtop spectrometers with better instrument performance as well as with the handheld spectrometers with more flexible measurement operation. In comparison, the calibration models developed with spectra measured on the benchtop spectrometers are superior to those of spectra acquired with the handheld spectrometers. Within the handheld spectrometer systems the LVF-type MicroNIR instrument performed best. Notwithstanding some deficiencies, all instruments are able to satisfy the accuracy requirements for industrial flour quality and process control.

(3) Because handheld FT-NIR instruments based on the MEMS technology cover almost the whole NIR spectral range, the spectra of these instruments contain a large amount of

structural information of organic substances. However, due to the small size of the FT-NIR handheld instruments they suffer of poor heat dissipation and their signal-to-noise ratio still needs to be improved for further optimization of qualitative and quantitative calibration models.

5 Model Transfer

5.1 Introduction

5.1.1 General Process of Model Transfer

Modern near-infrared spectroscopy analysis is an indirect analysis technology, which combines spectroscopic measurement technology with multivariate evaluation techniques [206-209]. It has the advantages of being nondestructive and fast and can simultaneously detect multi-components. Furthermore, it can be applied for online detection in industrial processes [210-214]. At present, NIR spectroscopy is widely used in food, medicine, agriculture, petrochemical industry and other fields. However, in the application process of NIR spectroscopy, due to differences in spectroscopic systems (e.g. light source, detector and other components), or their assembly process, the same sample may exhibit different spectral features when analyzed by different instruments. Thus, when a calibration model developed for one instrument is applied to the spectra of test samples measured on another instrument, the test results may exhibit large deviations or the model may be nonfunctional. Because it takes a lot of manpower and material resources to develop a calibration model, the realization of model transfer between different instruments is of great significance for practical application and for the promotion of the NIR spectroscopy analysis technology [215-220].

In recent years, many researchers in China and other countries have made in-depth studies on model transfer for NIR spectroscopy and achieved great progress [221-224]. The general process of model transfer is shown in Fig. 5.1. The methods to realize model transfer can be divided into three categories: The first is to correct parameters of the prediction model, such as two-step partial least squares method. The second is to correct prediction results of the model, such as slope/deviation algorithm. The third is to correct spectral data, such as the direct standardization algorithm [225]. To correct parameters of the prediction model is to enhance the predictive effect of the model by adding a series of sample spectra measured under the new environmental conditions and/or with new instruments, so that the model can

adapt to the sample spectra collected under these new testing conditions. To correct the predicted results of the model assumes, that the predicted results of the master spectrometer and the target spectrometer are linearly related, and the model transfer is realized by reducing the systematic error of the prediction results. However, the above assumption is not valid in most cases. Therefore, this approach has a poor effect in model transfer and is only applicable under special conditions. To correct spectral data is to standardize spectra, which makes spectra of the same sample collected by different instruments and under different testing conditions as consistent as possible [226, 227]. Thus, the differences between spectra are reduced, so that the established model can be shared between different instruments. The first two methods are generally called transfer of calibration model, and the latter realizes the sharing of a calibration model.

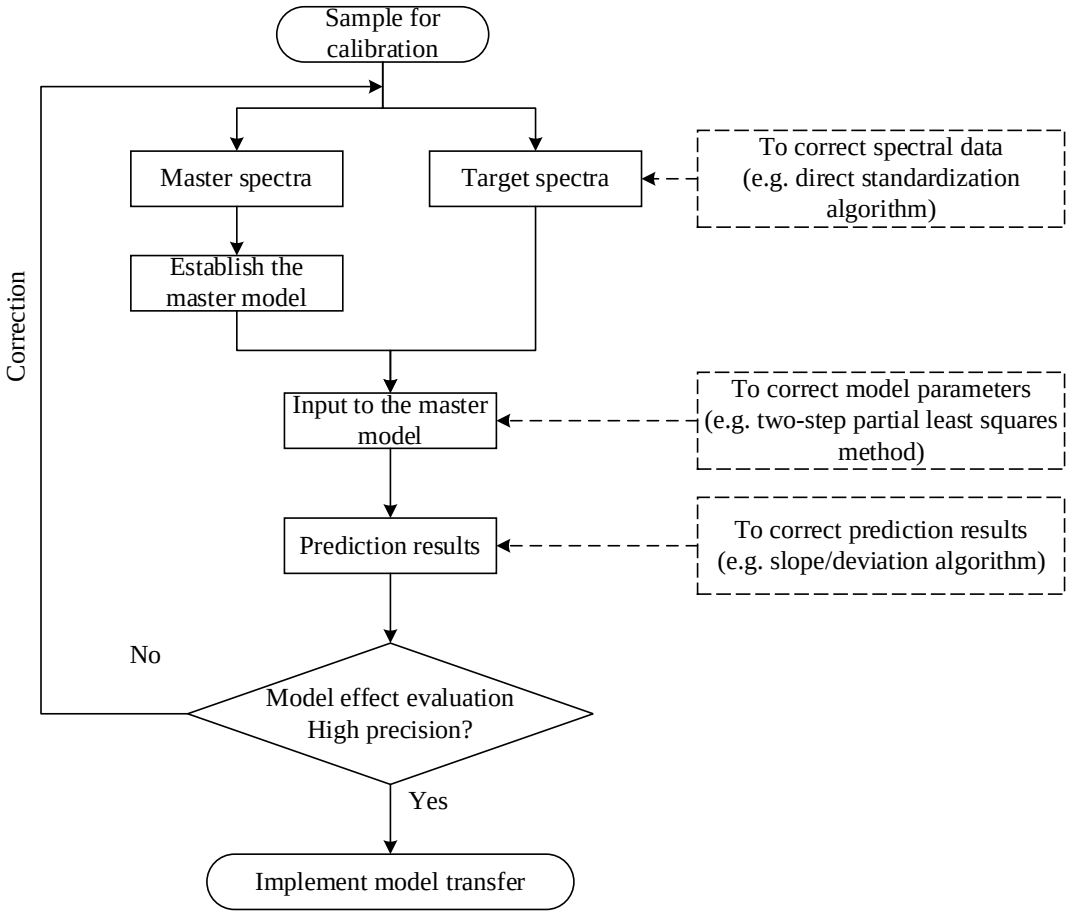


Fig. 5.1 Flow chart of the discussed model transfer methods

In this experiment, crude protein content of wheat flour was taken as detection index, and two NIR spectrometers with different spectroscopic principles, namely Grating Scanning type (IAS 3100) and LVF type (MicroNIR), were used to collect NIR spectral data of wheat flour samples. DS, PDS and SLRDS algorithms were employed to standardize sample spectra. After analyzing the error rate of spectral standardization, the error rate of principal component score was put forward to quantitatively describe spectra differences between master and target spectrometers of the same sample before and after standardization. The smaller the error rate, the smaller the spectra difference of samples. Therefore, the prediction error of the calibration model shared among different instruments can be minimized.

5.1.2 Model Transfer Method based on Principal Component Score Difference Correction

In NIR calibration models established by the PLS method, principal component decomposition of the spectral data matrix and variable matrix is required. After obtaining the loading and score matrices, the number of selected principal components is determined according to the principle of minimum predictive residual error sum of squares (PRESS) cross-verified by the leave-one-out method, based on which the calibration model is established [228]. Therefore, the effect of spectral standardization can be evaluated by the difference in principal components for the same group of samples after standardization treatment to spectra collected by master and target spectrometers. Specifically, the standardized spectra of the target spectrometers are put into the principal component matrix of the calibration model for calculating the scores of each principal component. Then, the obtained principal component scores are compared with those of the master spectrometer. The smaller the difference, the better the effect of spectral standardization to the target spectrometer (or vice versa). On this basis, the principal component score error rate (PCSER) (described in 5.2.3.2) was proposed to quantitatively evaluate the difference of principal components, so as to realize the sharing of a wheat flour protein model among different instruments. The main algorithm flow is shown in Fig. 5.2.

(1) On the basis of a proper spectral pretreatment, PLS is used to establish the master model

and determine the principal component number, loading and score matrices.

(2) Representative samples are selected from the master and target sample calibration sets as standard sample sets for elaboration of spectral standardization methods (DS, PDS, SLRDS algorithms, etc.). The optimal standard sample number is decided according to $SSER_{ave}$ minimum principle of master and target sample calibration sets.

(3) A variety of spectral standardization methods are used to standardize the target spectral data by using the master spectra of calibration sample sets not involved in spectral standardization as standard.

(4) The standardized spectra of the target spectrometer are put into the principal component matrix of the established calibration model, and the principal component scores are calculated. The differences of principal component scores between the corrected target spectra and the master spectra are evaluated. The similarity of the spectral score matrix of the master and target spectra is quantitatively evaluated by using the principal component score error rate as evaluation index. If the error is fairly large, the standard sample set should be re-selected and the spectral data re-corrected to enable selection of the optimal standardization method with minimum error for correcting the spectra of target prediction set.

Finally, the standardized target spectra are put into the master model for prediction and evaluation, in order to demonstrate the sharing of the calibration model established on the master spectrometer with other different instruments.

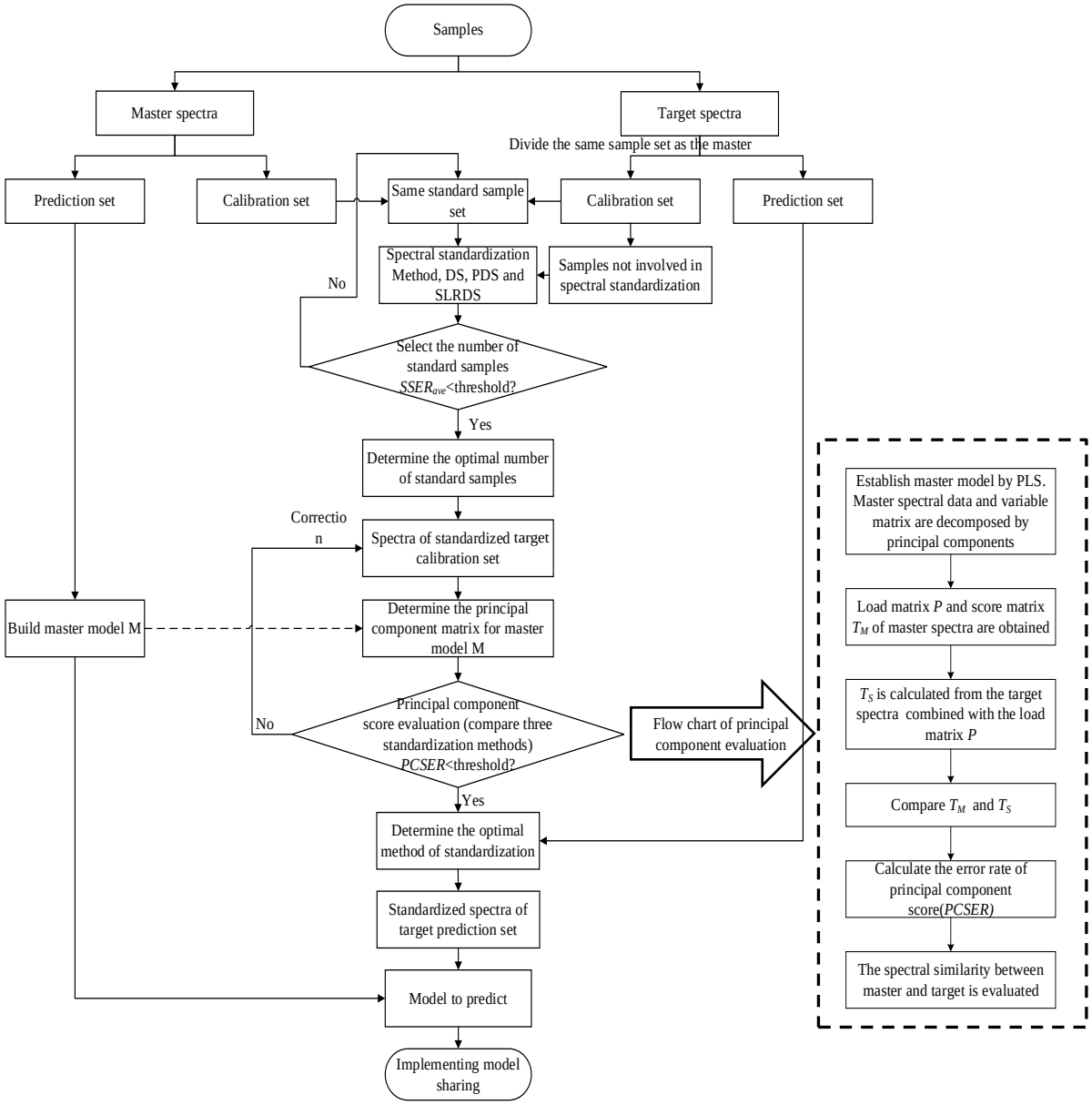


Fig. 5.2 The flow chart of model transfer based on principal component score evaluation

5.2 Materials and Methods

5.2.1 Samples

The samples were divided into a calibration set (153) and a prediction set (51) in a ratio of 3:1 by SPXY (sample set partitioning based on joint X-Y distance) algorithm. The test results of crude protein content in wheat flour of each set are shown in Tab. 5.1.

Tab.5.1 Distribution of crude protein content in wheat flour of the different sample sets

Sample Set	Number	Range/% (w/w)	Average/% (w/w)	Standard Deviation	Coefficient of Variation
Total set	211	6.34~14.83	10.79	1.89	0.18
Calibration set	153	6.34~14.83	10.77	2.04	0.19
Prediction set	51	7.84~14.00	10.75	1.30	0.12

5.2.2 Instruments and Test Methods

Absorbance measurements were performed using an IAS 3100 NIR spectrometer as master instrument (denoted as M) and a MicroNIR (VIAVI) spectrometer as target instrument (denoted as T). The wavelength range and sampling interval of the NIR spectrometers are shown in Tab. 5.2.

Tab. 5.2 Spectral wavelength range and sampling interval of NIR spectrometers

NIR spectrometers	Wavelength/nm
IAS 3100	900 - 1675 nm
MicroNIR	908 - 1676 nm

In order to ensure the consistency of wavelength range in subsequent tests, the spectral data for the 908-1670 nm wavelength range of the IAS 3100 spectrometer were selected for future studies of spectral standardization methods. The average spectra of samples collected by the master and the target spectrometers are shown in Fig. 5.3.

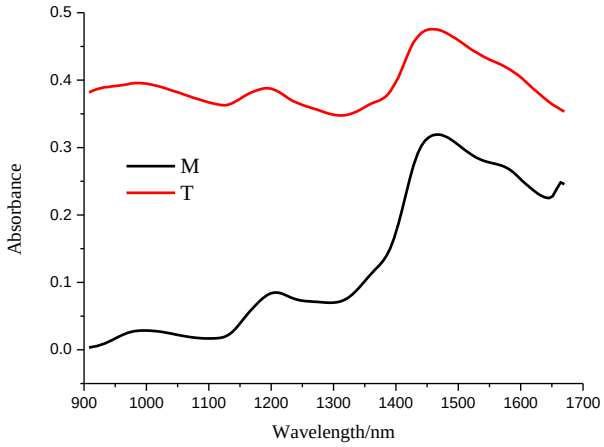


Fig. 5.3 Average spectra of all samples collected by the master and target spectrometers

5.2.3 Analysis Method

5.2.3.1 Spectral Standardization Methods

DS, PDS and SLRDS algorithms were used to standardize spectra for realizing model transfer among different NIR spectrometers. The DS algorithm is based on the mathematical relationship between the spectra of the master standard sample set and the spectra of the target standard sample set to establish the spectral standardization transfer matrix, which is then used to correct the spectra collected from the target spectrometers and reduce spectral differences of the same sample measured by different instruments. The principle of PDS algorithm is similar to DS as shown in Fig. 5.4.

Yet, PDS separates continuous wavelengths from spectra, calculates transformation coefficients in each wavelength window, establishes a spectral standardization transfer matrix according to transformation coefficients of each wavelength window, and uses the transfer matrix to correct spectra of the target spectrometers to achieve maximum similarity between master and target spectra. The width of the left and right wavelength window region (ω) was set to 3 in the test. The SLRDS algorithm assumes that the absorbance of different wavelength points is independent of each other, and uses linear regression to correct spectra measured on the target spectrometers.

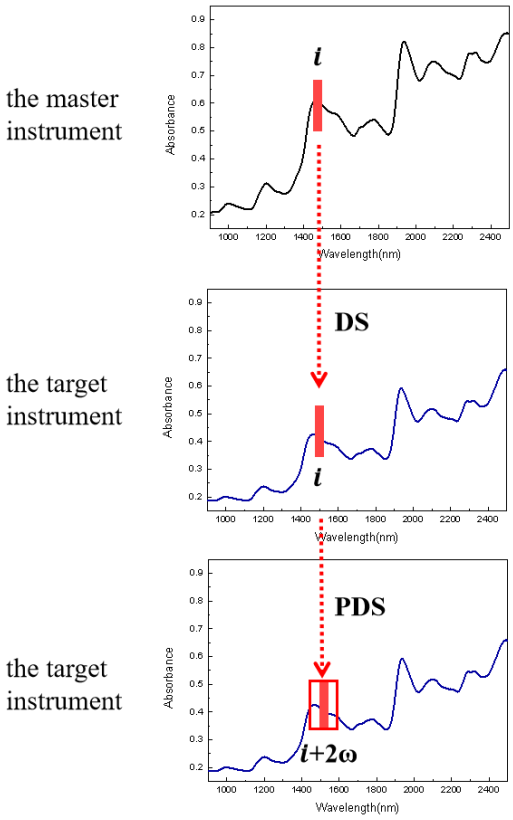


Fig. 5.4 Difference between DS and PDS algorithms

5.2.3.2 Evaluation Methods of Spectral Differences

（1）Euclidean Distance

The spectral differences between instruments were quantitatively evaluated by the Euclidean Distance (*D*) between spectra. The larger the value D, the more obvious the spectral difference between instruments. The formula for calculating D is:

$$D = \sqrt{\sum_{k=1}^{n}(A_{ik} - \overline{A}_k)^2} \qquad \text{Eq.65}$$

In this formula, k is the wavelength point; A_{ik} is the absorbance of the spectrum collected on the target instrument at the k wavelength point and the i spectrum, and $\overline{A}_k$ is the absorbance of the standard spectrum (master instrument) at the k wavelength point.

（2）Difference of Spectral Data

The spectral standardization error rate (SSER) was used to characterize the accuracy of

spectra standardization between different instruments, and to quantitatively describe spectral differences between the target spectra after standardization and the master spectra of the same sample.

The spectral standardization error rate for a sample is defined as:

$$SSER_i = \sum_{j=1}^{K} \frac{|M_{ij} - T_{ij}|}{|M_{ij} + T_{ij}|} \times 100\% \quad \text{Eq.66}$$

in which $\{M_{ij}, i=1,\ldots,N,\ j=1,\ldots,K\}$ is the sample spectral matrix of the master spectrometer; $\{T_{ij}, i=1,\ldots,N,\ j=1,\ldots,K\}$ is the sample spectral matrix of the target spectrometer after standardization; N is the number of samples; K is the data point in the spectra.

For all sample sets, the average error rate $SSER_{ave}$ and maximum error rate $SSER_{max}$ are defined as:

$$SSER_{ave} = \frac{1}{N} \sum_{i=1}^{N} SSER_i \quad \text{Eq.67}$$

$$SSER_{max} = max(SSER_i) \quad \text{Eq.68}$$

（3） Error of Principal Component Scores

The principal component score error rate (PCSER) was used to characterize the similarity of principal component score matrices. A lower PCSER means that the principal component score matrix of master and target spectra is more similar and the spectral difference is smaller. The calibration model based on principal component analysis or partial least square method has then a better sharing effect.

The PCSER formula of the first n principal component scores between master and target spectra of a sample is as follows:

$$PCSER = \frac{1}{n} \sum_{i=1}^{n} W_i \sqrt{(T_{m,i} - T_{t,i})^2} \quad \text{Eq.69}$$

in which $T_{m,i}$ is the score rate of the i^{th} principal component of the master spectrum; $T_{t,i}$ is the score rate of the i^{th} principal component of the target spectrum after standardization of the corresponding spectrum; W_i is the contribution rate of the i^{th} principal component.

For all sample sets, the average error rate $PCSER_{ave}$ and maximum error rate $PCSER_{max}$ are defined as:

$$PCSER_{ave}=\frac{1}{N}\sum_{i=1}^{N} PCSER_i \quad \text{Eq.70}$$

$$PCSER_{max}=max(PCSER_i) \quad \text{Eq.71}$$

5.2.3.3 Evaluation of Model Performance

In the process of model establishment, the correlation coefficient of calibration (R_c), RMSEC and RMSECV were used to evaluate the performance of the model, and the optimal calibration model was developed. After establishment of the model, the prediction performance of the model is evaluated comprehensively using indicators such as correlation coefficient of prediction (R_p), RMSEP and RPD. The smaller RMSEC, RMSECV and RMSEP are and the closer R_c and R_p are to 1, the better the stability and prediction performance of the established model are. RPD is used to evaluate the accuracy of the model. When RPD < 1.75, the prediction accuracy of the model is too low and this means, that the model is not applicable. For RPD $>$ 3, the prediction accuracy is high enough to use the model.

5.2.4 Data Processing and Analysis

The NIRSA 5.8.8 system (computer software copyright registration number of 2007SR06801), IBM SPSS Statistics 25 and Excel 2016 was adopted for data analysis.

5.3 Results and Discussion

5.3.1 Model Development for the Master Spectrometer

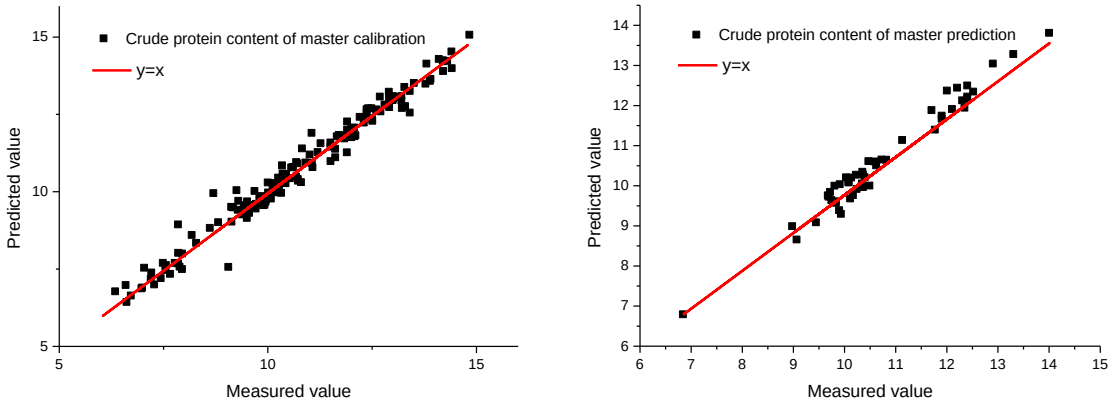
The PLS method was used to establish a calibration model for the correlation between crude protein content of 153 calibration samples and their NIR spectra collected on the master spectrometer. In order to fully extract effective information from the spectra, various pretreatment methods were employed to process the original spectra for the elimination of irrelevant and interference information such as noise in the spectral data. The optimal

pretreatment method was selected based on the predictive effect of the established PLS analysis model. Evaluation results of the calibration model under different pretreatment methods are shown in Tab. 5.3.

Tab 5.3 Evaluation of the PLS calibration models for protein under different pretreatment methods

Pretreatment method	Number of principal components	Calibration set		Prediction set		RPD
		R_c^2	RMSEC	R_p^2	RMSEP	
None	7	0.9594	0.3653	0.9661	0.3802	5.4312
SNV	**6**	**0.9615**	**0.3559**	**0.9712**	**0.3507**	**5.8928**
Normalization	6	0.9567	0.3833	0.9684	0.3792	5.6306
MSC	6	0.9612	0.3719	0.9695	0.3610	5.7260
Normalization +MSC	6	0.9614	0.3720	0.9698	0.3613	5.7544
1st derivative	13	0.9548	0.3857	0.9739	0.3340	6.1900

It can be clearly seen from Tab. 5.3 that, after comparing the modeling effects of different pretreatment methods, when the number of principal components is 6, the crude protein calibration model after SNV pretreatment has the best effect. At this condition, the PLS model has an R_c^2 of 0.9615, an RMSEC of 0.3559, an R_p^2 of 0.9712, an RMSEP of 0.3507, and an RPD of 5.8928, meaning that the model has a high prediction accuracy. The prediction effect of master crude protein model is shown in the Fig. 5.5.



(a) Calibration (153) (b) Prediction (51)

Fig. 5.5 Scatter plot of IAS 3100 master calibration (SNV) and prediction for protein content

5.3.2 Spectral Standardization

DS, PDS and SLRDS algorithms belong to supervised algorithms. Hence, a standard sample set needs to be selected first, and the selection of the number of samples in the standard sample set has an important impact on the effect of spectral standardization. Too few samples will lead to insufficient information, while too many samples will increase difficulty of data processing, resulting in illusion of over-fitting. Using the Kennard-Stone (K-S) algorithm, 10, 20, 30, 40, 50, 60, 70 and 80 samples were selected from the master and target calibration sets respectively as the standard sample set for spectral standardization, and the standardized transfer matrix was established. The three spectral standardization methods were used to calibrate spectra from the target calibration sets, and the $SSER_{ave}$ of the master and target spectral data after calibration were calculated. Under the three algorithms, the relationship between the number of standard samples and the $SSER_{ave}$ value is shown in Fig. 5.6 for the three algorithms.

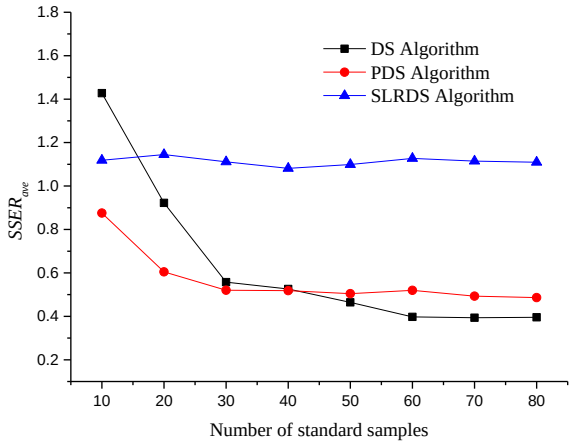


Fig. 5.6 Variation of the $SSER_{ave}$ of the target spectrometer with the number of standard samples for the three spectral standardization methods.

With the increase of the number of standard samples, the effective information contained in the standard sample set increases as well. As can be seen from Fig. 5.6, the $SSER_{ave}$ value decreases, for the PDS and DS algorithms meaning that spectral difference between master and target decreases as well.

For the DS algorithm, $SSER_{ave}$ reaches a minimum when the number of standard samples is

60. For the PDS algorithm, $SSER_{ave}$ reaches a plateau when the standard sample number is less than 30. Similarly, when the SLRDS algorithm is used, $SSER_{ave}$ changes very little with the increase of the standard sample number, and reaches the minimum when the standard sample number is 40.

5.3.3 Analysis of Spectral Differences

Fifty samples were selected as optimum alternative to constitute the standard sample set for all three algorithms (DS, PDS and SLRDS) and were used to establish a transfer matrix between the spectra of the standard sample set collected by the master and the target spectrometers. With the help of the transfer matrix, spectra of the calibration sample set measured on the target spectrometer were standardized. SSER and PCSER of the master and target spectra before and after standardization were calculated, and the average and maximum error rate SSER and PCSER, respectively, were compared (Tab. 5.4).

Tab. 5.4 Difference between the target and master spectra before and after standardization

Algorithm	*SSER*		*PCSER*	
	$SSER_{ave}$	$SSER_{max}$	$PCSER_{ave}$	$PCSER_{max}$
None	72.0133	83.7803	218.1008	323.3647
DS	0.3973	3.5975	1.3583	5.3757
PDS	0.5203	1.0622	2.0735	6.0494
SLRDS	1.0812	2.0744	4.6754	13.0615

As can be seen from Tab. 5.4, SSER and PCSER values of untreated target spectra are high, indicating that the master and target spectra have great differences. This is due to the different NIR spectrometers selected for the test, which leads to significantly different spectra from the same sample. After standardization with DS, PDS and SLRDS algorithms, SSER and PCSER of the target and master spectra are largely reduced, however to different degrees. For the target spectrometer, the standardization effect of the DS algorithm is the best.

5.3.4 The Process of Model Transfer

Using the spectral standardization methods and parameters identified in section 3.2, the spectra of the target prediction set samples (51) were calibrated. The spectra before and after standardization were put into the established optimal master model of crude protein (Section 3.1) for prediction. Fig. 5.7 shows the process of model normalization after using the calibration set (153) of the target instrument as the prediction set of the master instrument. The predictive scatter diagram of each method is shown in Fig. 5.7(b), and the predictive effect of model transfer is shown in Tab. 5.5.

It can be clearly seen from Fig. 5.7(a) that, compared with Y=X, there is a big difference in the intercept of the original spectra predicted by the master model for the target spectra, which means that there is a large systematic error in the predictive result. As can be seen from Fig. 5.7(b), the predictive effect of the model is improved after the standardized (DS and PDS) target spectra are input to the master model.

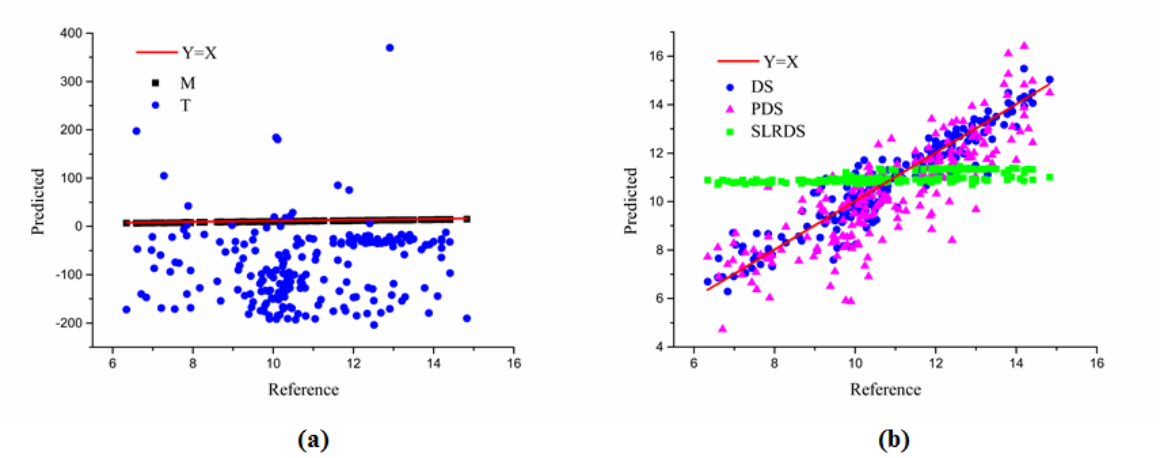


Fig. 5.7 Predictive scatter plot of crude protein calibration model
(a) before standardization；(b) after standardization

Tab. 5.5 Effect of Model Transfer for target MicroNIR spectrometer

Algorithm	Model transfer		
	R_p	*RMSEP*	*RPD*
None	0.0925	119.764	1.0043
DS	0.9426	0.643	2.9947
PDS	0.8299	1.2631	1.7924
SLRDS	0.7567	1.1729	1.2558

As shown in Tab. 5.5, when target spectra without standardization are put into the master model, the RMSEP is high and RPD < 1.75, which indicates that the master model has a poor predictive effect on original target spectra and cannot be directly applied for prediction of target spectra. Instead, when target spectra after standardization by DS and PDS algorithms are subjected to the master model, the prediction correlation coefficients increase (both above 0.8). However, from the parameter values it can be derived that the standardization of the SLRDS algorithm does not work, RMSEPs decrease sharply, and RPDs also have a certain improvement. This shows that the spectral standardization algorithm greatly reduces the spectral difference between the master and target spectrometers. Among the three algorithms, the predictive effect of the target spectra after standardization by DS algorithm is the best, being consistent with the conclusion from section 3.2. The predicted scatter plot of crude protein content of target prediction based on the DS algorithm is shown in the Fig. 5.8. These results demonstrate that the proposed two evaluation indexes (i.e. SSER and PCSER) can effectively analyze spectral differences, accurately evaluate performance of various spectral standardization methods, and greatly facilitate model transfer between different spectrometers.

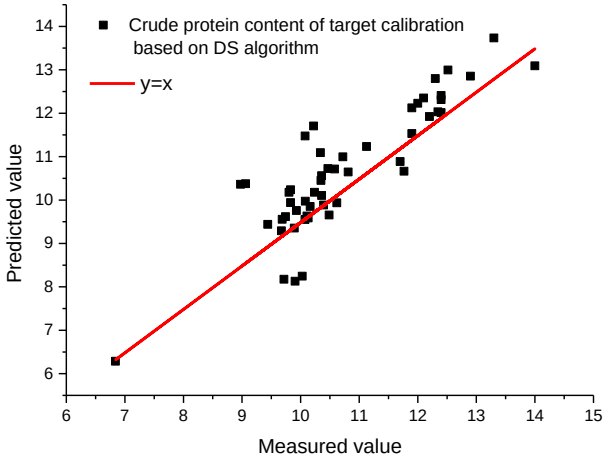


Fig. 3.8 Scatter plot of target protein content prediction crude protein content based on DS algorithm

5.4 Chapter Summary

Taking wheat flour as sample and the NIR spectral calibration model of crude protein as example, this study explores spectral standardization methods between different NIR

spectrometers, seeks the best spectral standardization method, and aims to realize the sharing of calibration models among different instruments. The main conclusions are as follows:

(1) Spectral standardization and principal component score error rate were proposed for evaluating the effect of spectral standardization, and enabled the quantitative assessment of spectral differences and the improvement of accuracy in spectral standardization.

(2) DS, PDS and SLRDS algorithms all belong to supervised spectral standardization algorithms. With the increase of sample number, the effective information contained in the standard sample set increases, and the $SSER_{ave}$ values of two algorithms (DS and PDS) show a downward trend. Yet, too many samples could lead to over-fitting.

(3) The SLRDS algorithm is ineffective as standardization method, but by the other two algorithms (DS and PDS), the spectral differences between the target and the master spectrometers are significantly reduced. After standardization by the DS algorithm, the error rate of the target spectrometers was the lowest, and the master model had the best effect. Its prediction accuracy was greatly improved compared with that before standardization.

6 Conclusions

Near-infrared spectroscopy is a technology that integrates analytical chemistry, applied optics, applied mathematics and computer science. Because this technology is fast, efficient, non-destructive, and practical in on-line detection, it has been chosen by many industries.

In this project, five NIR spectrometers (2 benchtop and 3 handheld) were selected to measure the NIR spectra of 50 flour samples from Germany and 163 flour samples from China. The objective of the project is to analyze the indicators of flour, such as crude protein, moisture and wet gluten, in order to solve two problems: 1. how to develop calibration models based on the NIR spectra of these samples for each reference value of flour; 2. how to solve the transfer of spectral models across different spectrometers.

6.1 How to Build the NIR Spectral Model for each Reference Value of the Flour

In the process of modeling the NIR spectra for each reference value of flour, the number of outliers rejected varies from one spectral sample to another. Although there are various algorithms for judging outliers, in the actual operation of spectral analysis, outliers are not only related to spectra, but also to the accuracy of reference values. Therefore, when analyzing sample data, two factors need to be considered separately, and then the selection criteria for outliers in the sample data will be set.

At the stage of determining the pretreatment method for the data, the first step is to find the best spectra pretreatment method for calibration development of all flour reference values. The selected pretreatment method in this thesis is EMSC, which aims to balance the best results of the pretreatment by scatter correction of the raw flour spectra obtained by NIR spectrometers with different optical principles. Finding the best pretreatment method requires further analysis and optimization steps by trial and error.

After pretreatment of the data, several major reference values of wheat flour were quantitatively modeled by using PLS regression. The NIR spectral models for crude protein,

moisture and wet gluten of flour were developed, and the RMSEPs of the best flour models were 0.3784% (MicroNIR), 0.2624% (MicroNIR) and 1.4653% (Neospectra Micro), respectively. The R_p^2 were 0.7922 (Hamamatsu C15511-01) for the best crude protein model, 0.6096 (Hamamatsu C15511-01) for the best moisture model, and 0.8346 (IAS 3100) for the best wet gluten model, which showed that of the proposed calibration models the most accurate was for the prediction of moisture which was superior to protein and wet gluten. Generally, the results showed that the analysis of the main reference values of wheat flour by NIR spectroscopy was feasible and good prediction models were obtained.

The identification of wheat flour origin was studied by developing different identification models from the measured spectra: PCA, PLS-DA, and significance of difference analysis of reference values, respectively. Through the analysis of NIR spectra, it was clearly observed that there were differences in the spectra between wheat varieties of different origins. The differences in the spectral regions of 1900-2200 nm（5263-4545 cm^{-1}）is more obvious, because there is a large amount of N-H and C-O bond information of proteins in this region, which can more accurately reveal the differences in chemical information between wheat flour of different origins. Using the PCA method, German and Chinese flours could be well identified and classified with the raw flour spectra of spectrometers other than FT-NIR spectrometers (NeoSpectra Micro and Hamamatsu C15511-01). However, using the PLS-DA analysis method, the assignment of the origin of the flour models from FT-NIR spectrometers can be significantly and effectively improved. The experimental results also demonstrate that NIR spectroscopy can be effectively used for the rapid quantification of the investigated wheat flour parameters. However, the next stage of cross-spectrometer sharing of calibration models can be achieved only, when the problem of spectral standardization and spectral differences between different instruments is solved.

6.2 How to Solve the Problem of Sharing Spectral Models across Spectrometers

Firstly, two problems need to be solved: optimization of the spectral standardization method

and evaluation of the spectral standardization effect. For this purpose, in Section 5.2.3.2 the goodness-of-fit evaluation indexes, SSER and PCSER have been proposed after spectral standardization.

In order to achieve the purpose of model transfer for the most important reference value crude protein, 163 flour samples from China and 50 flour samples from Germany were collected. After comparing various methods of spectral pretreatment and calibration models, the IAS 3100 spectrometer was selected as the master and the MicroNIR spectrometer as the target. The standard sample set for spectral standardization was optimized using the fit of the spectra as the final evaluation index. The different effects of three spectral standardization methods, DS, PDS and SLRDS, on the sharing of calibration models across instruments are investigated. After the three standardization methods were tested, the DS and PDS methods significantly reduced the inter-instrument spectral data variability and improved the prediction accuracy of the calibration model for wheat flour reference value proteins, while the result of the SLRDS method was invalid. Among them, the DS algorithm for crude protein prediction was the best, with R_p improved to 0.9426 and RMSEP of 0.643. The results showed, that the PLS regression model with SNV pretreatment in the wavelength range of 1750-2150 nm worked best.

This experiment basically achieved the cross-spectrometer sharing of the crude protein correction model. The results demonstrate that the spectral standardization error rate can effectively describe the differences between spectra quantitatively, and the principal component score error rate can characterize the similarity between the principal component score matrices. Traditionally, the error of the cross-spectrometer prediction of the correction model is solved in order to evaluate the effectiveness of spectral standardization, which ultimately improves the efficiency of model transfer.

6.3 Future Prospects

This thesis investigates and discusses the application of near-infrared spectroscopy for the development of calibration models for the quantitative determination of flour parameters and

discrimination of geographical origin of different flours, and applies the spectral standardization method to the calibration model transfer for flour parameters. However, there are still some aspects that need further improvement and refinement.

1. The investigations only established three reference value models for moisture, wet gluten, and crude protein of wheat flour samples, and the study of model transfer was only applied to the reference value of crude protein. In fact, other reference values such as starch, fat, ash and sedimentation value of flour samples were not investigated in depth. Future studies should be extended to other reference values of wheat flour to establish further calibration models so that the purpose of the complete parameter profile for wheat flour can be achieved.

2. Although the most popular spectral pretreatment method for scatter correction - EMSC - and the most frequently used calibration methods for qualitative and quantitative purposes - PCA and PLS, respectively - have been applied in the present thesis, additional approaches could be tested.

3. The main factors affecting the sharing of calibration models across spectrometers are the differences in working principles and performance indexes of the master and target instruments. Follow-up research can further expand the types of spectrometers beyond the scope of this thesis, thereby generalizing the approach of calibration model transfer across instruments.

References

[1] Akhtar, S., Anjum, F. M., & Anjum, M. A. (2011). Micronutrient fortification of wheat flour: Recent development and strategies. Food Research International, Vol. 44(3), 652-659.

[2] Yang, X., Wu, L., Zhu, Z., Ren, G., & Liu, S. (2014). Variation and trends in dough rheological properties and flour quality in 330 Chinese wheat varieties. The Crop Journal, Vol. 2(4), 195-200.

[3] Atwell, W. A., & Finnie, S. (2016). Wheat flour. Elsevier.

[4] Posner, E. S., & Hibbs, A. N. (2005). Wheat flour milling (No. Ed. 2). American Association of Cereal Chemists, Inc.

[5] Goesaert, H., Brijs, K., Veraverbeke, W. S., Courtin, C. M., Gebruers, K., & Delcour, J. A. (2005). Wheat flour constituents: how they impact bread quality, and how to impact their functionality. Trends in food science & technology, Vol. 16(1-3), 12-30.

[6] Hoseney, R. C., & Rogers, D. E. (1990). The formation and properties of wheat flour doughs. Critical Reviews in Food Science & Nutrition, Vol. 29(2), 73-93.

[7] Kumar, P., Yadava, R. K., Gollen, B., Kumar, S., Verma, R. K., & Yadav, S. (2011). Nutritional contents and medicinal properties of wheat: a review. Life Sciences and Medicine Research, Vol. 22(1), 1-10.

[8] Contreras-Jiménez, B., Torres-Vargas, O. L., & Rodríguez-García, M. E. (2019). Physicochemical characterization of quinoa (Chenopodium quinoa) flour and isolated starch. Food chemistry, Vol. 298, Article 124982.

[9] Coudray, C., Levrat-Verny, M. A., Tressol, J. C., Feillet-Coudray, C., Horcajada-Molteni, N. M., Demigné, C., Rayssiguier,Y. & Rémésy, C. (2001). Mineral supplementation of white wheat flour is necessary to maintain adequate mineral status and bone characteristics in rats. Journal of trace elements in medicine and biology, Vol. 15(2-3), 131-137.

[10] Seibel, W. (1992). Leitsätze für Brot und Kleingebäck. Getreide, Mehl und Brot (1972), Vol. 46(1), 25-27.

[11] DIN 10355:1991-12, Mahlerzeugnisse aus Getreide; Anforderungen, Typen, Prüfung

[12] Kirsch, B. & Odenthal, A. (2003). Fachkunde Müllereitechnologie. Werkstoffkunde. Ein Lehrbuch über die Zusammensetzung, Untersuchung, Bewertung und Verwendung von Getreide und Getreideprodukten. 5. Auflage, Bayerischer Müllerbund, München.

[13]https://www.kitchenstories.com/de/stories/mehl-type-405-550-und-1050-alle-mehlsorte-auf-einen-blick

[14]https://www.egle.de/blog/mehltypen-und-ihre-verwendung

[15]https://lebensmittel-warenkunde.de/lebensmittel/getreideprodukte/getreide/roggen/roggen-mehl-typ-1150.html

[16]https://lebensmittel-warenkunde.de/lebensmittel/getreideprodukte/getreide/roggen/roggen-mehl-typ-815.html

[17]https://lebensmittel-warenkunde.de/lebensmittel/getreideprodukte/getreide/roggen/roggen-mehl-typ-997.html

[18]https://www.edeka.de/ernaehrung/lebensmittelwissen/kuechen-1x1/dinkelmehl-typen.jsp

[19] Zhao, N. X., Wang, L. K., Cheng, A. H., Lan, J. & Dai, C. J. (2003). Correlation between Bread Baking Quality and Wheat Quality Traits. Journal of Wheat Crops, Vol. 3, 33-36.

[20] Goesaert, H., Brijs, K., Veraverbeke, W. S., Courtin, C. M., Gebruers, K., & Delcour, J. A. (2005). Wheat flour constituents: how they impact bread quality, and how to impact their functionality. Trends in food science & technology, Vol. 16(1-3), 12-30.

[21] Jiang, W. L., Sun, H. & Lin, J. Y. (2009). Study on the relationship between physical and chemical quality indicators of wheat flour and food processing quality. Journal of the Chinese Cereals and Oils Association, Vol. 3, 12-16.

[22] Sun, H., Wu, C. R., Li, Y. & Jiang, W. L. (2010). Current Situation and Development Direction of China's Wheat Standard System. Cereal & Feed Industry, Vol. 4, 13-17.

[23] Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. Longman Scientific and Technical, Harlow , Essex, England. 49-78.

[24] Williams, P., & Norris, K. (1987). Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists, Inc.. St. Paul, Minnesota, USA.

[25] Osborne, B. G. (2006). Near-infrared spectroscopy in food analysis. Encyclopedia of

analytical chemistry: applications, theory and instrumentation. John Wiley & Sons. Inc., Hoboken, New Jersey, USA.

[26] Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest biology and technology,Vol. 46(2), 99-118.

[27] O'Brien, N., Hulse, C. A., Pfeifer, F., & Siesler, H. W. (2013). Near infrared spectroscopic authentication of seafood. Journal of Near Infrared Spectroscopy, Vol. 21(4), 299-305.

[28] Prieto, N., Roehe, R., Lavín, P., Batten, G., & Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. Meat science, Vol. 83(2), 175-186.

[29] Kamruzzaman, M., ElMasry, G., Sun, D. W., & Allen, P. (2013). Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. Food Chemistry, Vol. 141(1), 389-396.

[30] Prevolnik, M., Candek-Potokar, M., & Skorjanc, D. (2004). Ability of NIR spectroscopy to predict meat chemical composition and quality: A review. Czech Journal of Animal Science-UZPI (Czech Republic). Vol. 49 no. 11, 500-510.

[31] Wang, Y. J., Li, T. H., Li, L. Q., Ning, J. M., & Zhang, Z. Z. (2020). Micro-NIR spectrometer for quality assessment of tea: Comparison of local and global models. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol. 237, Article 118403.

[32] Jirsa, O., Hrušková, M., & Švec, I. (2007). Bread features evaluation by NIR analysis. Czech Journal of Food Sciences, Vol. 25(5), 243.

[33] Sweeney, R. A., & Rexroad, P. R. (1987). Comparison of LECO FP-228" nitrogen determinator" with AOAC copper catalyst Kjeldahl method for crude protein. Journal of the Association of Official Analytical Chemists, Vol. 70(6), 1028-1030.

[34] Camuffo, D. (2018). Standardization activity in the evaluation of moisture content. Journal of Cultural Heritage, Vol. 31, S10-S14.

[35] Shi, H., & Yu, P. (2017). Comparison of grating-based near-infrared (NIR) and Fourier

transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. Food Control, Vol. 82, 57-65.

[36] Lu, C. X., Jiang, X. P., Zhang, Y. Q., & Zhang, X. (2016). Variable selection based near infrared spectroscopic quantitative analysis on wheat crude protein content. Trans. Chin. Soc. Agric. Mach, Vol. 47, 340-346.

[37] Sun, L. X., Wang, L. K., Qian, H. B. (2011). Determination of wheat protein content based on near-infrared transmission spectroscopy analysis technology. Chinese Agricultural Science Bulletin, Vol. 03, 35-39.

[38] Barton, F. E., Shenk, J. S., Westerhaus, M. O., & Funk, D. B. (2000). The development of near infrared wheat quality models by locally weighted regressions. Journal of Near Infrared Spectroscopy, Vol. 8(3), 201-208.

[39] Zheng, Y. M., Zhang, T. Q., Zhang, J., Chen, X. D., & Shen, X. G. (2004). Influence of smooth, 1st derivative and baseline correction on the near-infrared spectrum analysis with PLS. Guang pu xue yu Guang pu fen xi= Guang pu, Vol. 24(12), 1546-1548.

[40] Agelet, L. E., & Hurburgh Jr, C. R. (2010). A tutorial on near infrared spectroscopy and its calibration. Critical Reviews in Analytical Chemistry, Vol. 40(4), 246-260.

[41] Zhu, Z. H., Wang, W. Z., Liu, S. C.(2006). Application of Near Infrared Diffuse Reflectance Spectroscopy in the Identification of Quality Traits of Crop Germplasm Resources. Modern Scientific Instruments, Vol. 1, 63-66.

[42] Sohn, M., Barton, F. E., McClung, A. M., & Champagne, E. T. (2004). Near-infrared spectroscopy for determination of protein and amylose in rice flour through use of derivatives. Cereal chemistry, Vol. 81(3), 341-344.

[43] Sinelli, N., Benedetti, S., Bottega, G., Riva, M., & Buratti, S. (2006). Evaluation of the optimal cooking time of rice by using FT-NIR spectroscopy and an electronic nose. Journal of Cereal Science, Vol. 44(2), 137-143.

[44] Villareal, C. P., De La Cruz, N. M., & Juliano, B. O. (1994). Rice amylose analysis by near-infrared transmittance spectroscopy. Cereal Chemistry, Vol. 71(3), 292-296.

[45] Kweon, M., Slade, L., & Levine, H. (2011). Solvent retention capacity (SRC) testing of wheat flour: Principles and value in predicting flour functionality in different wheat -

based food processes and in wheat breeding - A review. Cereal Chemistry, Vol. 88(6), 537-552.

[46] Létang, C., Piau, M., Verdier, C., & Lefebvre, L. (2001). Characterization of wheat-flour–water doughs: a new method using ultrasound. Ultrasonics, Vol. 39(2), 133-141.

[47] Miralbes, C. (2004). Quality control in the milling industry using near infrared transmittance spectroscopy. Food Chemistry, Vol. 88(4), 621-628.

[48] Che, W., Sun, L., Zhang, Q., Zhang, D., Ye, D., Tan, W., & Dai, C. (2017). Application of visible/near‐infrared spectroscopy in the prediction of azodicarbonamide in wheat flour. Journal of food science, Vol. 82(10), 2516-2525.

[49] Perez-Mendoza, J., Throne, J. E., Dowell, F. E., & Baker, J. E. (2003). Detection of insect fragments in wheat flour by near-infrared spectroscopy. Journal of stored products research, Vol. 39(3), 305-312.

[50] Hrušková, M., & Faměra, O. (2003). Prediction of wheat and flour Zeleny sedimentation value using NIR technique. Czech journal of food sciences, Vol. 21(3), 91.

[51] Cocchi, M., Durante, C., Foca, G., Marchetti, A., Tassi, L., & Ulrici, A. (2006). Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. Talanta, Vol. 68(5), 1505-1511.

[52] Jiang, H., Liu, T., & Chen, Q. (2020). Quantitative detection of fatty acid value during storage of wheat flour based on a portable near-infrared (NIR) spectroscopy system. Infrared Physics & Technology, Vol. 109, Article 103423.

[53] Wesley, I. J., Larroque, O., Osborne, B. G., Azudin, N., Allen, H., & Skerritt, J. H. (2001). Measurement of gliadin and glutenin content of flour by NIR spectroscopy. Journal of Cereal Science, Vol. 34(2), 125-133.

[54] Yan, L. H., Wang, J. H., Jin, H. H.(2011). Establishment of non-destructive testing model of flour moisture based on near-infrared spectroscopy technology. Modern Food Science and Technology, Vol. 27(2), 235-238.

[55] http://www.fao.org/faostat/en/#search/wheat%20production%20and%20utilization%20globally

[56] Siah, S., & Quail, K. J. (2018). Factors affecting Asian wheat noodle color and time‐dependent discoloration - A review. Cereal Chemistry, Vol. 95(2), 189-205.

[57] Reddecliffe, T. M., Moot, D. J., Wilson, D. R., & Scott, W. R. (2000). Grain yield and quality of two durum wheat cultivars grown in Canterbury. Agronomy New Zealand, Vol. 30, 77-82.

[58] Mao, Y., & Flores, R. A. (2001). Mechanical starch damage effects on wheat flour tortilla texture. Cereal chemistry, Vol. 78(3), 286-293.

[59] Morgan, J. E., & Williams, P. C. (1995). Starch damage in wheat flours: a comparison of enzymatic, iodometric, and near-infrared reflectance techniques. Cereal Chemistry, Vol. 72(2), 209-212.

[60] Test, S. D. S. (1979). Note on the Sodium Dodecyl Sulfate Test of Breadmaking Quality:Comparison with Pelshenke and Zeleny Tests. Cereal Chemistry, Vol. 56(6), 582-584.

[61] Robertson, G. H., Cao, T. K., & Ong, I. (1999). Wheat Gluten Swelling and Partial Solubility with Potential Impact on Starch‐from‐Gluten Separation by Ethanol Washing. Cereal chemistry, Vol. 76(6), 843-845.

[62] Wieser, H. (2007). Chemistry of gluten proteins. Food microbiology, Vol. 24(2), 115-119.

[63] Du, C., Yang, X. J., & Liu, G. R. (2002). Studies on the relations between starch properties of wheat flour and baking quality. Hebei Nongye Daxue Xuebao (China). Vol. 25, 29-33.

[64] Liu, C., Li, L., Hong, J., Zheng, X., Bian, K., Sun, Y., & Zhang, J. (2014). Effect of mechanically damaged starch on wheat flour, noodle and steamed bread making quality. International journal of food science & technology, Vol. 49(1), 253-260.

[65] Xiao, Z. S., Park, S. H., Chung, O. K., Caley, M. S., & Seib, P. A. (2006). Solvent retention capacity values in relation to hard winter wheat and flour properties and straight‐dough breadmaking quality. Cereal chemistry, Vol. 83(5), 465-471.

[66] Michel, S., Löschenberger, F., Ametz, C., Pachler, B., Sparry, E., & Bürstmayr, H. (2019). Combining grain yield, protein content and protein quality by multi-trait genomic selection in bread wheat. Theoretical and Applied Genetics, Vol. 132(10),

2767-2780.

[67] Pistón, F., Gil-Humanes, J., Rodríguez-Quijano, M., & Barro, F. (2011). Down-regulating γ-gliadins in bread wheat leads to non-specific increases in other gluten proteins and has no major effect on dough gluten strength. PLOS ONE, Vol. 6(9), e24754.

[68] Bouachra, S., Begemann, J., Aarab, L., & Hüsken, A. (2017). Prediction of bread wheat baking quality using an optimized GlutoPeak®-Test method. Journal of Cereal Science, Vol. 76, 8-16.

[69] Luo, C., Branlard, G., Griffin, W. B., & McNeil, D. L. (2000). The effect of nitrogen and sulphur fertilisation and their interaction with genotype on wheat glutenins and quality parameters. Journal of Cereal Science, Vol. 31(2), 185-194.

[70] Hrušková, M., & Faměra, O. (2003). Prediction of wheat and flour Zeleny sedimentation value using NIR technique. Czech journal of food sciences, Vol. 21(3), 91.

[71] Goesaert, H., Brijs, K., Veraverbeke, W. S., Courtin, C. M., Gebruers, K., & Delcour, J. A. (2005). Wheat flour constituents: how they impact bread quality, and how to impact their functionality. Trends in food science & technology, Vol. 16(1-3), 12-30.

[72] Gélinas, P., Morin, C., Reid, J. F., & Lachance, P. (2009). Wheat cultivars grown under organic agriculture and the bread making performance of stone‐ground whole wheat flour. International journal of food science & technology, Vol. 44(3), 525-530.

[73] http://www.pyyfy.net/0730wy/product/2018-5-5/268.html

[74] Ring, S. G., Gee, J. M., Whittam, M., Orford, P., & Johnson, I. T. (1988). Resistant starch: its chemical form in foodstuffs and effect on digestibility in vitro. Food chemistry, Vol. 28(2), 97-109.

[75] Shewry, P. R., Halford, N. G., Belton, P. S., & Tatham, A. S. (2002). The structure and properties of gluten: an elastic protein from wheat grain. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, Vol. 357(1418), 133-142.

[76] Guan, L., Seib, P. A., Graybosch, R. A., Bean, S., & Shi, Y. C. (2009). Dough rheology and wet milling of hard waxy wheat flours. Journal of agricultural and food chemistry, Vol. 57(15), 7030-7038.

[77] Moiraghi, M., Vanzetti, L., Bainotti, C., Helguera, M., León, A., & Pérez, G. (2011). Relationship between soft wheat flour physicochemical composition and cookie‐making performance. Cereal Chemistry, Vol. 88(2), 130-136.

[78] O'Brien, C. M., Mueller, A., Scannell, A. G. M., & Arendt, E. K. (2003). Evaluation of the effects of fat replacers on the quality of wheat bread. Journal of Food Engineering, Vol. 56(2-3), 265-267.

[79] Henry, R., & Kettlewell, P. (Eds.). (2012). Cereal grain quality. Springer Science & Business Media, Berlin, Germany.

[80] Marlett, J. A., McBurney, M. I., & Slavin, J. L. (2002). Position of the American Dietetic Association: health implications of dietary fiber. Journal of the American Dietetic Association, Vol. 102(7), 993-1000.

[81] Ozbek, N., & Akman, S. (2016). Method development for the determination of calcium, copper, magnesium, manganese, iron, potassium, phosphorus and zinc in different types of breads by microwave induced plasma-atomic emission spectrometry. Food chemistry, Vol. 200, 245-248.

[82] Torre, M., Rodriguez, A. R., & Saura‐Calixto, F. (1991). Effects of dietary fiber and phytic acid on mineral availability. Critical Reviews in Food Science & Nutrition, Vol. 30(1), 1-22.

[83] DellaPenna, D., & Pogson, B. J. (2006). Vitamin synthesis in plants: tocopherols and carotenoids. Annu. Rev. Plant Biol., Vol. 57, 711-738.

[84] Kumar, P., Yadava, R. K., Gollen, B., Kumar, S., Verma, R. K., & Yadav, S. (2011). Nutritional contents and medicinal properties of wheat: a review. Life Sciences and Medicine Research, Vol. 22(1), 1-10.

[85] https://m.cqn.com.cn/cpkkxbg/content/2008-05/26/content_819988.htm

[86] Ahn, J. Y., Kil, D. Y., Kong, C., & Kim, B. G. (2014). Comparison of oven-drying methods for determination of moisture content in feed ingredients. Asian-Australasian journal of animal sciences, Vol. 27(11), 1615.

[87] Misra, N. N., Kaur, S., Tiwari, B. K., Kaur, A., Singh, N., & Cullen, P. J. (2015). Atmospheric pressure cold plasma (ACP) treatment of wheat flour. Food Hydrocolloids, Vol. 44, 115-121.

[88] Basman, A., & Yalcin, S. (2011). Quick-boiling noodle production by using infrared drying. Journal of Food Engineering, Vol. 106(3), 245-252.

[89] Kabaha, K., Taralp, A., Cakmak, I., & Ozturk, L. (2011). Accelerated hydrolysis method to estimate the amino acid content of wheat (Triticum durum Desf.) flour using microwave irradiation. Journal of agricultural and food chemistry, Vol. 59(7), 2958-2965.

[90] Rains, D. M., & Holder, J. W. (1981). Ethylene dibromide residues in biscuits and commercial flour. Journal of the Association of Official Analytical Chemists, Vol. 64(5), 1252-1254.

[91] Corpaş, L., Hădărugă, N. G., David, I., Pîrşan, P., Hădărugă, D. I., & Isengard, H. D. (2014). Karl Fischer Water Titration− Principal Component Analysis Approach on Wheat Flour. Food Analytical Methods, Vol. 7(6), 1353-1358.

[92] Hester, R. E., & Quine, D. E. (1977). Quantitative analysis of food products by pulsed NMR rapid determination of oil and water in flour and feedstuffs. Journal of the Science of Food and Agriculture, Vol. 28(7), 624-630.

[93] Poliszko, N., Kowalczewski, P. Ł., Rybicka, I., Kubiak, P., & Poliszko, S. (2019). The effect of pumpkin flour on quality and acoustic properties of extruded corn snacks. Journal of Consumer Protection and Food Safety, Vol. 14(2), 121-129.

[94] Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., & Uhlén, M. (2016). Gene-specific correlation of RNA and protein levels in human cells and tissues. Molecular systems biology, Vol. 12(10), 883.

[95] Osborne, B. G., Douglas, S., & Fearn, T. (1982). The application of near infrared reflectance analysis to rapid flour testing. International Journal of Food Science & Technology, Vol. 17(3), 355-363.

[96] Williams, P. C. (1961). The determination of proteins in whole wheatmeal and flour by the biuret procedure. Journal of the Science of Food and Agriculture, Vol. 12(1), 58-61.

[97] Wang, S., & Copeland, L. (2012). Effect of alkali treatment on structure and function of pea starch granules. Food Chemistry, Vol. 135(3), 1635-1642.

[98] Takasaki, S., Kato, Y., Murata, M., Homma, S., & Kawakishi, S. (2005). Effects of peroxidase and hydrogen peroxide on the dityrosine formation and the mixing

characteristics of wheat-flour dough. Bioscience, biotechnology, and biochemistry, Vol. 69(9), 1686-1692.

[99] McDonald, C. E., & Chen, L. L. (1965). The Lowry modification of the Folin reagent for determination of proteinase activity. Analytical biochemistry, Vol. 10(1), 175-177.

[100] Chen, W. Z., & Hoseney, R. C. (1995). Wheat flour compound that produces sticky dough: isolation and identification. Journal of food science, Vol. 60(3), 434-437.

[101] Eynard, L, Guerrieri, N, & Cerletti, P. (1994). Determination of wheat proteins in solution by dye binding in flour, dough, and bread crumb. Cereal Chemistry, Vol. 71, 434-434.

[102] Keran, H., Salkić, M., Odobašić, A., Jašić, M., Ahmetović, N., & Šestan, I. (2009). The importance of determination of some physical–chemical properties of wheat and flour. Agriculturae Conspectus Scientificus, Vol. 74(3), 197-200.

[103] Ming, Y., & Bing, L. (1998). Determination of rare earth elements in human hair and wheat flour reference materials by inductively coupled plasma mass spectrometry with dry ashing and microwave digestion. Spectrochimica Acta Part B: Atomic Spectroscopy, Vol. 53(10), 1447-1454.

[104] Létang, C., Piau, M., Verdier, C., & Lefebvre, L. (2001). Characterization of wheat-flour–water doughs: a new method using ultrasound. Ultrasonics, Vol. 39(2), 133-141.

[105] Sinelli, N., Casiraghi, E., & Downey, G. (2008). Studies on proofing of yeasted bread dough using near-and mid-infrared spectroscopy. Journal of agricultural and food chemistry, Vol. 56(3), 922-931.

[106] Peruchi, L. C., Nunes, L. C., de Carvalho, G. G. A., Guerra, M. B. B., de Almeida, E., Rufini, I. A., & Krug, F. J. (2014). Determination of inorganic nutrients in wheat flour by laser-induced breakdown spectroscopy and energy dispersive X-ray fluorescence spectrometry. Spectrochimica Acta Part B: Atomic Spectroscopy, Vol. 100, 129-136.

[107] Lang, V., Bellisle, F., Oppert, J. M., Craplet, C., Bornet, F. R., Slama, G., & Guy-Grand, B. (1998). Satiating effect of proteins in healthy subjects: a comparison of egg albumin, casein, gelatin, soy protein, pea protein, and wheat gluten. The American journal of clinical nutrition, Vol. 67(6), 1197-1204.

[108] Cesevičienė, J., & Butkutė, B. (2012). Comparison of gluten parameters when gluten is determined by the glutomatic and hand washing methods. In Proceedings 6th International Congress of Flour - Bread & 8th Croatian Congress of Cereal Technologists. Opatija, Croacia. 180-189.

[109] Schober, T. J., Clarke, C. I., & Kuhn, M. (2002). Characterization of functional properties of gluten proteins in spelt cultivars using rheological and quality factor measurements. Cereal Chemistry, Vol. 79(3), 408-417.

[110] Wieser, H., Antes, S., & Seilmeier, W. (1998). Quantitative determination of gluten protein types in wheat flour by reversed - phase high - performance liquid chromatography. Cereal Chemistry, Vol. 75(5), 644-650.

[111] Czaja, T., Mazurek, S., & Szostak, R. (2016). Quantification of gluten in wheat flour by FT-Raman spectroscopy. Food Chemistry, Vol. 211, 560-563.

[112] Clements, D. L. (2014). Infrared Astronomy–Seeing the Heat: from William Herschel to the Herschel Space Observatory. CRC Press. Marcel Dekker, Inc., New York, USA.

[113] Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. Journal of food Engineering, Vol. 87(3), 303-313.

[114] Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., & Choi, K. H. (2003). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. Transactions of the ASAE, Vol. 46(6), 1721.

[115] Hindle, P. H. (2007). Historical development. In Handbook of near-infrared analysis (pp. 21-24). CRC Press. Marcel Dekker, Inc., New York, USA.

[116] McClure, W. F. (1994). Near-infrared spectroscopy: the giant is running strong. Analytical Chemistry, Vol. 66(1), 42A-53A.

[117] Günzler, H., & Gremlich, H. U. (2012). IR-Spektroskopie: Eine Einführung. Wiley-VCH, Weinheim, Germany.

[118] Gottwald, W., & Wachter, G. (1997). IR-Spektroskopie für Anwender. Wiley-VCH-Verlag.

[119] Schrader, B. (Ed.). (2008). Infrared and Raman spectroscopy: methods and applications. John Wiley & Sons. Inc., Hoboken, New Jersey, USA.

[120] Banwell, C. N. (1972). Fundamentals of molecular spectroscopy. McGraw-Hill, Inc., New York, USA.

[121] Hollas, J. M. (2002). Basic atomic and molecular spectroscopy. Royal Society of Chemistry, London, U.K. Vol. 11

[122] Bellamy, L. J. (1975). The IR spectra of complex molecules. Chapman and Hall, London, U.K. Vol. I.

[123] Hummel, D. O. (1968). Ultrarot-Spektrum und chemische Konstitution. Zeitschrift für Physikalische Chemie, Vol. 58(5/6), 330-331.

[124] Socrates, G. (1980). Infrared Characteristic Group Frequencies, Wiley - Interscience, Chichester, U.K.

[125] Hediger, H. (1957). Physical principles and possibilities of application of infrared spectroscopy. Acta histochemica, Vol. 4(5-8), 197-199.

[126] Fearn, T. (2002). Assessing calibrations: SEP, RPD, RER and R^2. NIR news, Vol. 13(6), 12-13.

[127] Burgess, C., & Hammond, J. (2009). Exploring the "forgotten" region - an update on NIR spectrometry. Spectroscopy Europe, Vol. 21, 28-31.

[128] Siesler, H. W., & Paul, C. (1989). NIR-Spektroskopie. Nachr. Chem. Tech. Lab, Vol. 37, 172-174.

[129] K. Molt (1998). NIR-Spektrometrie - ein Füllhorn für die Instrumentelle Analytik? GIT Laborzeitschrift, Band 42, 386-387.

[130] Salzer, R. (1986). Schwingungsspektroskopie im Nahen Infrarot. Zeitschrift für Chemie, Vol. 26(8), 275-284.

[131] Siesler, H. W., Ozaki, Y., Kawata, S., & Heise, H. M. (Eds.). (2008). Near-infrared spectroscopy: principles, instruments, applications. John Wiley & Sons. Inc., Hoboken, New Jersey, USA.

[132] Burns, D. A., & Ciurczak, E. W. (Eds.). (2007). Handbook of near-infrared analysis. CRC Press, Marcel Dekker, Inc., New York, USA.

[133] Workman Jr, J. J. (1996). Interpretive spectroscopy for near infrared. Applied Spectroscopy Reviews, Vol. 31(3), 251-320.

[134] Workman Jr. J. J., & Weyer, L. (2012). Practical guide to interpretive near-infrared

spectroscopy. CRC Press. Boca Raton, Florida, USA.

[135] Kessler, R. W. (Ed.). (2012). Prozessanalytik: Strategien und Fallbeispiele aus der industriellen Praxis. Wiley-VCH, Weinheim, Germany.

[136] Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., & De La Pezuela, C. (1998). Near-infrared spectroscopy in the pharmaceutical industry. Critical review. Analyst, Vol. 123(8), 135R-150R.

[137] K. Molt (1992). Grundlagen und Anwendungen der modernen NIR-Spektrometrie, Teil I. GIT Laborzeitschrift, Band 36, 107-113.

[138] K. Molt (1992), Grundlagen und Anwendungen der modernen NIR-Spektrometrie, Teil II. GIT Laborzeitschrift, Band 36, 353-362.

[139] Kortüm, G. (1969). Reflexionsspektroskopie.(In German.) Springer-Verlag, Berlin,Germany.

[140] Heller, W., & Naumer, H. (1986). Untersuchungsmethoden in der Chemie. Wiley-VCH, Weinheim, Germany.

[141] Hecht, E. (2018). Optik. Walter de Gruyter GmbH & Co KG,Berlin, Germany.

[142] Kubelka, P., & Munk, F. (1931). An article on optics of paint layers. Z. Tech. Phys, Vol. 12(593-601), 259-274.

[143] Böcker, J. (1997). Spektroskopie: Instrumentelle Analytik mit Atom-und Molekülspektrometrie. Vogel Buchverlag, Würzburg, Germany.

[144] Olinger, J. M., & Griffiths, P. R. (1988). Quantitative effects of an absorbing matrix on near-infrared diffuse reflectance spectra. Analytical Chemistry, Vol. 60(21), 2427-2435.

[145] Olinger, J. M., & Griffiths, P. R. (1992). Theory of diffuse reflectance in the NIR region.Practical Spectroscopy Series, Academic Press, New York, USA. Vol.13, 13-13.

[146] Ulbricht, R. (1900). Photometer for mean spherical candle-power. Elektrotech. Z., Vol. 21, 595.

[147] Bellamy, L. J. F. C. (2013). The infra-red spectra of complex molecules. Springer Science & Business Media, Berlin, Germany.

[148] Stark, E., Luchter, K., & Margoshes, M. (1986). Near-infrared analysis (NIRA): A technology for quantitative and qualitative analysis. Applied Spectroscopy Reviews, Vol. 22(4), 335-399.

[149] Brenchley, J. M., Horchner, U. W. E., & Kalivas, J. H. (1997). Wavelength selection characterization for NIR spectra. Applied Spectroscopy, Vol. 51(5), 689-699.

[150] Ozaki, Y. (2012). Near-infrared spectroscopy - its versatility in analytical chemistry. Analytical Sciences, Vol. 28(6), 545-563.

[151] Blanco, M., & Villarroya, I. N. I. R. (2002). NIR spectroscopy: a rapid-response analytical tool. TrAC Trends in Analytical Chemistry, Vol. 21(4), 240-250.

[152] Ozaki, Y., & Kawata, S. (1996). Near-infrared spectroscopy. Gakkai Shuppan Center, Tokyo, Japan.

[153] McClure, W. F. (2003). 204 years of near infrared technology: 1800–2003. Journal of Near Infrared Spectroscopy, Vol. 11(6), 487-518.

[154] Norris, K. H. (1996). History of NIR. Journal of Near Infrared Spectroscopy, Vol. 4(1), 31-37.

[155] Griffiths, P. R. (2010). Introduction to the theory and instrumentation for vibrational spectroscopy. Applications of Vibrational Spectroscopy in Food Science, Part 1, 31-46.

[156] Dzsaber, S., Negyedi, M., Bernáth, B., Gyüre, B., Fehér, T., Kramberger, C., & Simon, F. (2015). A Fourier-transform Raman spectrometer with visible laser excitation. Journal of Raman Spectroscopy, Vol. 46(3), 327-332.

[157] Stark, E., & Luchter, K. (2005). NIR instrumentation technology. NIR news, Vol. 16(7), 13-16.

[158] Huck, C. W. (2021). New Trend in Instrumentation of NIR Spectroscopy - Miniaturization. Near-Infrared Spectroscopy, Springer, Singapore, 193-210.

[159] Plans, M., Simó, J., Casañas, F., Sabaté, J., & Rodriguez-Saona, L. (2013). Characterization of common beans (Phaseolus vulgaris L.) by infrared spectroscopy: comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. Food Research International, Vol. 54(2), 1643-1651.

[160] Beć, K. B., Grabska, J., & Huck, C. W. (2021). Principles and applications of miniaturized near‐infrared (NIR) spectrometers. Chemistry–A European Journal, Vol. 27(5), 1514-1532.

[161] O'Brien, N. A., Hulse, C. A., Friedrich, D. M., Van Milligen, F. J., von Gunten, M. K., Pfeifer, F., & Siesler, H. W. (2012). Miniature near-infrared (NIR) spectrometer engine

for handheld applications. In Next-generation spectroscopic technologies V. The International Society for Optics and Photonics, Bellingham, Washington, USA. Vol. 8374, 837404.

[162] Huang, J., Wen, Q., Nie, Q., Chang, F., Zhou, Y., & Wen, Z. (2018). Miniaturized NIR spectrometer based on novel MOEMS scanning tilted grating. Micromachines, Vol. 9(10), 478.

[163] Mayr, S., Schmelzer, J., Kirchler, C. G., Pezzei, C. K., Beć, K. B., Grabska, J., & Huck, C. W. (2021). Theae nigrae folium: Comparing the analytical performance of benchtop and handheld near-infrared spectrometers. Talanta, Vol. 221, 121165.

[164] Rajendran, V., Chang, H., & Liu, R. S. (2019). Recent progress on broadband near-infrared phosphors-converted light emitting diodes for future miniature spectrometers. Optical Materials: X, Vol. 1, 100011.

[165] Minotto, A., Haigh, P. A., Łukasiewicz, Ł. G., Lunedei, E., Gryko, D. T., Darwazeh, I., & Cacialli, F. (2020). Visible light communication with efficient far-red/near-infrared polymer light-emitting diodes. Light: Science & Applications, Vol. 9(1), 1-11.

[166] Jiang, G. L. (2020). Comparison and application of non-destructive NIR evaluations of seed protein and oil content in soybean breeding. Agronomy, Vol. 10(1), 77.

[167] Reinig, P., Grüger, H., Knobbe, J., Pügner, T., & Meyer, S. (2018). Bringing NIR spectrometers into mobile phones. In MOEMS and Miniaturized Systems XVII. The International Society for Optics and Photonics, Bellingham, Washington, USA. Vol. 10545, 97-104.

[168] Eriksson, L., Byrne, T., Johansson, E., Trygg, J., & Vikström, C. (2013). Multi-and megavariate data analysis basic principles and applications. Umetrics Academy, Umea, Sweden, Vol. 1.

[169] Martens, H., & Naes, T. (1992). Multivariate calibration. John Wiley & Sons. Hoboken, New Jersey, USA.

[170] Chaminade, P., Baillet, A., & Ferrier, D. (1998). Data treatment in near infrared spectroscopy. Analusis, Vol. 26(4), 33-37.

[171] Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. Longman scientific and technical, London,

U.K.

[172] Esquerre, C., Gowen, A. A., Burger, J., Downey, G., & O'Donnell, C. P. (2012). Suppressing sample morphology effects in near infrared spectral imaging using chemometric data pre-treatments. Chemometrics and Intelligent Laboratory Systems, Vol. 117, 129-137.

[173] Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. Applied Spectroscopy, Vol. 39(3), 491-500.

[174] Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy, Vol. 43(5), 772-777.

[175] "PLS-IQTM: User's Guide", (2002).Thermo Galactic, Salem, New Hampshire, USA.

[176] Martens, H., & Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. Journal of pharmaceutical and biomedical analysis, Vol. 9(8), 625-635.

[177] Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. Analytical Chemistry, Vol. 75(3), 394-404.

[178] Kohler, A., Kirschner, C., Oust, A., & Martens, H. (2005). Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. Applied spectroscopy, Vol. 59(6), 707-716.

[179] Thennadil, S. N., Martens, H., & Kohler, A. (2006). Physics-based multiplicative scatter correction approaches for improving the performance of calibration models. Applied Spectroscopy, Vol. 60(3), 315-321.

[180] Esbensen, K. H., Guyot, D., Westad, F., & Houmoller, L. P. (2002). Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design. Multivariate Data Analysis. Cengage Learning EMEA, Boston, Massachusetts, USA.

[181] “The Unscrambler User Manual”.(2006), Camo Analytics AS, Aspen Technology, Inc., Bedford, Massachusetts, USA.

[182] Otto, M. (2016). Chemometrics: statistics and computer application in analytical chemistry. John Wiley & Sons, Hoboken, New Jersey, USA.

[183] Martens, H., & Naes, T. (1989). Multivariate Calibration; John Willey & Sons. Inc., New York, USA.

[184] Kessler, W. (2011). Multivariate Datenanalyse: für die Pharma, Bio-und Prozessanalytik. Wiley-VCH, Weinheim, Germany.

[185] Danzer, K., Hobert, H., Fischbacher, C., & Jagemann, K. U. (2013). Chemometrik: Grundlagen und Anwendungen. Springer-Verlag, Berlin, Germany.

[186] Godoy, J. L., Vega, J. R., & Marchetti, J. L. (2014). Relationships between PCA and PLS-regression. Chemometrics and Intelligent Laboratory Systems, Vol. 130, 182-191.

[187] Conzen, J. P. (2005). Multivariate Kalibration: ein praktischer Leitfaden zur Methodenentwicklung in der quantitaven Analytik. Bruker Optik GmbH, Ettlingen, Germany.

[188] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin philosophical magazine and journal of science, Vol. 2(11), 559-572.

[189] Joliffe, I. T., & Morgan, B. J. T. (1992). Principal component analysis and exploratory factor analysis. Statistical methods in medical research, Vol. 1(1), 69-95.

[190] Fredericks, P. M., Lee, J. B., Osborn, P. R., & Swinkels, D. A. (1985). Materials characterization using factor analysis of FT-IR spectra. Part 1: Results. Applied spectroscopy, Vol. 39(2), 303-310.

[191] Fredericks, P. M., Lee, J. B., Osborn, P. R., & Swinkels, D. A. (1985). Materials characterization using factor analysis of FT-IR spectra. Part 2: Mathematical and statistical considerations. Applied spectroscopy, Vol. 39(2), 311-316.

[192] Beebe, K. R., & Kowalski, B. R. (1987). An introduction to multivariate calibration and analysis. Analytical Chemistry, Vol. 59(17), 1007A-1017A.

[193] Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. Analytica chimica acta, Vol. 185, 1-17.

[194] Ji, N. Y., Li, M., Lü, W. B., Liu, R., Zhang, Y. Y., & Han, D. H. (2017). Study on NIR model transfer between similar kinds of fruits based on Slope/Bias algorithm. Guang pu xue yu Guang pu fen xi= Guang pu, Vol. 37(1), 227-231.

[195] Swierenga, H., De Groot, P. J., De Weijer, A. P., Derksen, M. W. J., & Buydens, L. M. C. (1998). Improvement of PLS model transfer ability by robust wavelength selection. Chemometrics and Intelligent Laboratory Systems, Vol. 41(2), 237-248.

[196] Yang, H., Zhang, X., Fan, Y., Xie, P., & Chu, X. (2014). Near Infrared Spectroscopic Model Transfer Based on Simple Linear Regression. Chinese Journal of Analytical Chemistry, Vol. 12, 1229-1234.

[197] Standard No, I. C. C. 167: 2000. Determination of Crude Protein in Grain and Grain Products for Food and Feed by the Dumas Combustion Principle. International Association for Cereal Science and Technology, Vienna, Austria.

[198] Liu, Q., & Wang, S. (2019). Effects of various radio frequency treatment protocols on physicochemical properties and sensory quality of packaged milled rice. Lebensmittel-Wissenschaft & Technologie, Vol. 113, Article 108269.

[199] Standard No, I. C. C. 155: 1994. Determination of wet gluten quantity and quality (Gluten Index ac. to Perten) of whole wheat meal and wheat flour (Triticum aestivum). International Association for Cereal Science and Technology, Vienna, Austria.

[200] Pan, Z., Tian, P., Huang, Z., Wang, N., Suo, B., & Ai, Z. (2017). Effects of flour characteristics of different wheat cultivars on quality of frozen cooked noodles. Transactions of the Chinese Society of Agricultural Engineering, 33(3), 307-314.

[201] Standard No, I. C. C. 110/1:1976. Determination of moisture content of cereals and cereal products (Practical method). International Association for Cereal Science and Technology, Vienna, Austria.

[202] Liu, F., He, C., Wang, L., & Wang, M. (2018). Effect of milling method on the chemical composition and antioxidant capacity of Tartary buckwheat flour. International Journal of Food Science & Technology, Vol. 53(11), 2457-2464.

[203] Standard No, I. C. C. 104/1:1990. Determination of ash in Cereals and Cereal Products. International Association for Cereal Science and Technology, Vienna,

Austria.

[204] Sun, X., Liu, C., & Wang, Y. (2020). Influence of Na_2CO_3 on the quality of dough with rice wine sourdough and steamed bread. International Journal of Food Science & Technology, Vol. 55(5), 2261-2270.

[205] Chevallier, S., Bertrand, D., Kohler, A., & Courcoux, P. (2006). Application of PLS‐DA in multivariate image analysis. Journal of Chemometrics: A Journal of the Chemometrics Society, Vol. 20(5), 221-229.

[206] da Silva, A. L. D., Alves Filho, E. G., Silva, L. M. A., Tavares, O. C. H., Pereira, M. G., de Campos, T., & da Silva, L. M. (2021). Near infrared spectroscopy to rapid assess the rubber tree clone and the influence of maturation and disease at the leaves. Microchemical Journal, Vol. 168, Article 106478.

[207] Qin, F. L., Wang, X. C., Ding, S. R., Li, G. S., & Hou, Z. C. (2021). Prediction of Peking duck intramuscle fat content by near-infrared spectroscopy. Poultry Science, Vol. 100(8), Article 101281.

[208] Tian, W. F., Chen, G. J., Zhang, G. R., Wang, D. H., Tilley, M., & Li, Y. H. (2021). Rapid determination of total phenolic content of whole wheat flour using near-infrared spectroscopy and chemometrics. Food Chemistry, Vol. 344, Article 128633.

[209] Chapman, J., Elbourne, A., Truong, V. K., Newman, L., Gangadoo, S., Pathirannahalage, P. R., Cheeseman, S., & Cozzolino, D. (2019). Sensomics-from conventional to functional NIR spectroscopy-shining light over the aroma and taste of foods. Trends in Food Science & Technology, Vol. 91, 274-281.

[210] Pissard, A., Marques, E. J. N., Dardenne, P., Lateur, M., Pasquini, C., Pimentel, M. F., Pierna, J. A. F., & Baeten, V. (2021). Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. Postharvest Biology and Technology, Vol. 172, Article 111375.

[211] Tao, L., & Chen, S. (2018). Authenticity identification and classification of Rhodiola species in traditional Tibetan medicine based on Fourier transform near-infrared spectroscopy and chemometrics analysis. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol. 204, 131-140.

[212] Jia, C., Zhu, S. P., & Zhao, G. H. (2017). Rapid determination of total protein and wet

gluten in commercial wheat flour using siSVR-NIR. Food Chemistry, Vol. 221, 1939-1946.

[213] Grassi, S., & Alamprese, C. (2018). Advances in NIR spectroscopy applied to process analytical technology in food industries. Current Opinion in Food Science, Vol. 22, 17-21.

[214] Sánchez, M. T., Pérez-Marín, D., Torres, I., Gil, B., Garrido-Varo, A., & De la Haba, M. J. (2017). Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production. Postharvest Biology and Technology, Vol. 125, 122-128.

[215] Salguero-Chaparro, L., Palagos, B., Peña-Rodríguez, F., & Roger, J. M. (2013). Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. Computers and electronics in agriculture, Vol. 96, 202-208.

[216] Shi, Y. Y., Li,J. Y., & Chu, X. L. (2019). Progress and applications of multivariate calibration model transfer methods. Chinese Journal of Analytical Chemistry, Vol. 47(4), 479-487.

[217] Pu, Y. Y., Sun, D. W., Riccioli, C., Buccheri, M., Grassi, M., Cattaneo, T. M., & Gowen, A. (2018). Calibration transfer from micro NIR spectrometer to hyperspectral imaging: a case study on predicting soluble solids content of bananito fruit (Musa acuminata). Food analytical methods, Vol. 11(4), 1021-1033.

[218] Wang, J. X., Qu, J., Li, H., Han, X., & Xu, G. (2012). Application of GA-DS to calibration transfer of aviation fuel density in near infrared spectroscopy. Petroleum science and technology, Vol. 30(19), 1975-1980.

[219] Skotare, T., Nilsson, D., Xiong, S., Geladi, P., & Trygg, J. (2019). Joint and unique multiblock Analysis for integration and calibration transfer of NIR instruments. Analytical chemistry, Vol. 91(5), 3516-3524.

[220] Hacisalihoglu, G., Gustin, J. L., Louisma, J., Armstrong, P., Peter, G. F., Walker, A. R., & Settles, A. M. (2016). Enhanced single seed trait predictions in soybean (Glycine max) and robust calibration model transfer with near-infrared reflectance spectroscopy. Journal of agricultural and food chemistry, Vol. 64(5), 1079-1086.

[221] Huang, Z., Sanaeifar, A., Tian, Y., Liu, L., Zhang, D. Y., Wang, H., Ye, D. P., & Li, X. L.

(2021). Improved generalization of spectral models associated with Vis-NIR spectroscopy for determining the moisture content of different tea leaves. Journal of Food Engineering, Vol. 293,Article 110374.

[222] Liu, Y., Cai, W., & Shao, X. (2014). Standardization of near infrared spectra measured on multi-instrument. Analytica chimica acta, Vol. 836, 18-23.

[223] Chen, Y. Y., & Wang, Z. B. (2019). Cross components calibration transfer of NIR spectroscopy model through PCA and weighted ELM-based TrAdaBoost algorithm. Chemometrics and Intelligent Laboratory Systems, Vol. 192, Article 103824.

[224] Padarian, J., Minasny, B., & McBratney, A. B. (2019). Transfer learning to localise a continental soil vis-NIR calibration model. Geoderma. Vol. 340, 279-288.

[225] Sun, X. T., Zhang, P., Shen, Z. W., Yuan, H. F., Song, C. F., Han, X. L., Xiu, Y., Lv, Z., & Hu, A. Q. (2021). Investigation on spectral standardization among multi-channel of an on-line near-infrared spectrometer. Vibrational Spectroscopy, Vol. 113, Article 103206.

[226] Zhang, J., Guo, C., Cui, X., Cai, W., & Shao, X. (2019). A two-level strategy for standardization of near infrared spectra by multi-level simultaneous component analysis. Analytica Chimica Acta, Vol. 1050, 25-31.

[227] Wang, X., Mao, D. Z., & Yang, Y. J. (2021). Calibration transfer between modelled and commercial pharmaceutical tablet for API quantification using backscattering NIR, Raman and transmission Raman spectroscopy (TRS). J Pharm Biomed Anal, Vol. 194, Article 113766.

[228] Peerapattana, J., Shinzawa, H., Otsuka, K., Hattori, Y., & Otsuka, M. (2013). Partial Least Square Discriminant Analysis of Mangosteen Pericarp Powder by near Infrared Spectroscopy. Journal of Near Infrared Spectroscopy. Vol. 21(3), 195-202.

[229] Oliveri, P. (2017). Class-modelling in food analytical chemistry: development, sampling, optimisation and validation issues–a tutorial. Analytica Chimica Acta, Vol. 982, 9-19.

[230] Neves, M. D. G., Poppi, R. J., & Breitkreitz, M. C. (2022). Authentication of plant-based protein powders and classification of adulterants as whey, soy protein, and wheat using FT-NIR in tandem with OC-PLS and PLS-DA models. Food Control, Vol.

132, Article 108489.

[231] Williams, P. (2014). The RPD statistic: A tutorial note. NIR news, Vol. 25(1), 22-26.

Abbreviations index

Symbol Description

% (w/w)	- % (weight by weight)
1st Der.	- first-order derivative
AC	- amylose content
AE	- absolute error
AOTF	- acousto-optic tuneable filter
ASICs	- application-specific integrated circuits
ASTM	- American Society for Testing and Materials
AR	- anti-reflection
ARM	- advanced RISC machines
ATR	- attenuated total reflection
C	- calibration
CARS	- competitive adaptive reweighted sampling
cm	- centimeter
CV	- cross validation
D	- euclidean distance
DA	- discriminant analysis
DIGeFa GmbH	- Detmolder Institut für Getreide-und Fettanalytik GmbH
DMD™	- digital mirror device
DS	- direct standardization
ELM	- extreme learning machine
EMSC	- extended multiplicative scatter correction
FAO	- food and agriculture organization
FDA	- food and drug administration
FT-MIR	- fourier-transform mid-infrared
FT-NIR	- fourier-transform near-infrared
g	- gram
GB/T	- 推荐性国家标准 “Recommended Chinese National Standard”
ICC	- International Association for Cereal Chemistry
IPC	- industrial personal computer
IR	- infrared
K-S	- Kennard-Stone
L	- liter
LED	- light-emitting diode
log	- logarithm
LVF	- linear variable filter
LVs	- latent variables
MAE	- mean absolute error
MEMS	- micro-electro-mechanical systems
MIR	- mid-infrared

mL - milliliters
MLR - multiple linear regression
mol - mole
MPLS - modified partial least square
MSC - multiplicative scatter correction
NIPALS - nonlinear-iterative partial least squares
NIR - near-infrared
NIRS - near-infrared spectroscopic
nm - nanometer
P - prediction
PCA - principal component analysis
PCR - principal component regression
PCSER - principal component score error rate
PDS - piecewise direct standardization
PLS - partial least squares
PLS-DA - partial least squares discriminant analysis
PRESS - predictive residual error sum of squares
PTFE - Poly(tetrafluoroethylene)
R - correlation coefficient
R_c - correlation coefficient of calibration
R_{cv} - correlation coefficient of cross validation
R_p - correlation coefficient of prediction
R^2 - R squared
RISC - reduced instruction set computer
RMSEC - root mean square error of calibration
RMSECV - root mean square error of cross validation
RMSEP - root mean square error of prediction
RPD - relative prediction deviation
RPi - raspberry Pi
S/B - slope/bias
SEC - standard error of calibration
SE-HPLC - size-exclusion high performance liquid chromatography
SEP - standard error of prediction
SGSA - Savitzky-Golay smoothing algorithm
SIMCA - soft independent modeling of class analogies
SiMOST™ - Silicon Integrated Micro-Optics System Technology
SLRDS - simple linear regression direct standardization
S/N - signal/noise
SNV - standard normal variate
SPXY - sample set partitioning based on the joint X-Y distance
SRC - solvent retention capacity
SSER - spectral standardization error rate
UV - ultraviolet
VCPA - variable combined population analysis

VCSEL	- vertical cavity surface emitting laser
WILMA	- wavelet interface to linear modelling analysis
WT	- wavelet transform

Lebenslauf

Aus datenschutzrechtlichen Gruenden ist der Lebenslauf in der Online-Dissertation nicht enthalten