



A calibration transfer optimized single kernel near-infrared spectroscopic method

DOI:

[10.1016/j.saa.2019.05.003](https://doi.org/10.1016/j.saa.2019.05.003)
[10.1016/j.saa.2019.05.003](https://doi.org/10.1016/j.saa.2019.05.003)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Xu, Z., Fan, S., Liu, J., Liu, B., Tao, L., Wu, J., Hu, S., Zhao, L., Wang, Q., & Wu, Y. (2019). A calibration transfer optimized single kernel near-infrared spectroscopic method. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. <https://doi.org/10.1016/j.saa.2019.05.003>, <https://doi.org/10.1016/j.saa.2019.05.003>

Published in:

Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Accepted Manuscript

A calibration transfer optimized single kernel near-infrared spectroscopic method

Zhuopin Xu, Shuang Fan, Jing Liu, Binmei Liu, Liangzhi Tao, Jin Wu, Shupeng Hu, Liping Zhao, Qi Wang, Yuejin Wu



PII: S1386-1425(19)30479-2

DOI: <https://doi.org/10.1016/j.saa.2019.05.003>

Reference: SAA 17098

To appear in: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*

Received date: 30 August 2018

Revised date: 25 February 2019

Accepted date: 5 May 2019

Please cite this article as: Z. Xu, S. Fan, J. Liu, et al., A calibration transfer optimized single kernel near-infrared spectroscopic method, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, <https://doi.org/10.1016/j.saa.2019.05.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A calibration transfer optimized single kernel near-infrared spectroscopic method

Zhuopin Xu^{1,2,†}, Shuang Fan^{1,2,†}, Jing Liu¹, Binmei Liu¹, Liangzhi Tao¹, Jin Wu^{1,2}

Shupeng Hu³, Liping Zhao³, Qi Wang^{1,*}, Yuejin Wu^{1,*}

¹*Key Laboratory of High Magnetic Field and Ion Beam Physical Biology, Hefei
Institutes of Physical Science, Chinese Academy of Sciences, 350 Shushanhu Road,
Hefei, Anhui 230031, People's Republic of China*

²*University of Science and Technology of China, No. 96 Jinzhai Road, Hefei, Anhui
230026, People's Republic of China*

³*School of Computer Science, University of Manchester, Manchester, Britain*

[†] These authors contributed equally to this work and should be considered co-first authors.

^{*} Corresponding authors. Tel.: +86-551-65591382; Fax: +86-551-65592270.

E-mail addresses: wangqi@ipp.ac.cn (Q.W.), yjwu@ipp.ac.cn (Y.W.).

ABSTRACT: Single kernel near-infrared spectroscopy (SKNIRS) could aid in the quality screening of early-generation seeds, to improve the efficiency of seed breeding. However, the application of SKNIRS is limited due to the irregular physical characteristics, the heterogeneous constituent distributions of individual seeds, and the insufficient detection accuracy of the reference method. The reported near-infrared detection results of single seeds are often less accurate than those of dehusked seeds and seed flour. In this paper, a calibration transfer-optimized single kernel near-infrared spectroscopic method is proposed. This method aims to accurately detect the chemical composition of single seeds by using the calibration model of the corresponding dehusked seeds or seed flour. The proposed method was applied to the analysis of the protein content of a single rice kernel. The near-infrared transmission spectra of three forms of rice (single rice kernel (SRK), single brown rice kernel (SBK) and rice flour (RF)) of 201 individual rice seeds and the corresponding protein content values were obtained. By comparing different pretreatment methods and spectral ranges, the spectral range 950–1250 nm, the standard normal variate transformation (SNV) pretreatment, and 9 PLS factors were selected to construct the optimal partial least squares (PLS) regression models. Then, the protein content of single rice kernels were determined through two different methods: (i) the direct method, in which single rice kernels were analyzed using the single rice kernel model directly; and (ii) the proposed method, in which the spectra of single rice kernels were transferred into the spectra of single brown rice kernels and rice flours with a calibration transfer algorithm, spectral space transformation (SST), and were analyzed

by the respective calibration models. The external validation coefficient correlation (R) value of the direct method was 0.971, and the R values of the proposed method were 0.962 (SBK) and 0.975 (RF). The root mean square error of prediction (RMSEP) value of the direct method was 0.423, and the RMSEP of the proposed method were 0.480 (SBK) and 0.401 (RF). In addition, the transfer results among the spectra of three forms of rice were compared. By comparison, the results of the proposed method are fairly close to the results of the direct method. The results indicate that the spectra generated from one individual rice seed can be transferred freely among the three forms by means of calibration transfer. The proposed method is a promising way to overcome the challenges associated with analyzing individual seeds and to improve SKNIRS.

KEYWORDS: single kernel near-infrared spectroscopy; calibration transfer; spectral space transformation; protein content

1. Introduction

Breeding new plant varieties for specific traits is a constant goal of the seed industry. Analyzing the quality traits of individual seeds helps breeders to understand the genetic characteristics of crop quality and to screen and isolate seeds with desired qualities [1]. Near-infrared spectroscopy (NIRS) is a rapid, non-destructive, and multi-component analysis technique [2], and it can be used to improve the efficiency of seed breeding and lead to the development of new seed breeding technology. In order to select seeds of specific traits in early generations, NIRS should be implemented non-destructively at the single kernel level. However, there are some restrictions to the development of single kernel near-infrared spectroscopy (SKNIRS). Due to the irregular physical characteristics (for example, the seed size and morphology) and the heterogeneous constituent distributions (including the interference of glumes) of individual seeds, the reported spectral analysis results of single seeds are often less accurate than those of dehusked seeds and seed flour [3,4,5]. Moreover, considering most of the reference methods are designed for bulk samples, individual seeds are too small to analyze; therefore, applying such reference methods might lead to poor results [5].

In recent years, different methods have been proposed to improve SKNIRS. For example, (1) choosing a suitable trace detection method as the single seed reference method [6]; (2) using bulk reference values and mean spectra of single seeds to predict single seed constituents [7]; (3) correcting the weight interference of single seeds [8,9]; or (4) using chemometrics methods to correct background absorbance

[10]. The continuous development of seed breeding and the continuously evolving demands for crop quality require researchers to explore SKNIRS methods.

Calibration transfer [11] is a model standardization technique, and it is used to calibrate the spectral differences or prediction result deviations when the test conditions or instrument parameters change [12]. In a study published by Pereira et al. [13], calibration transfer was used to transfer the spectra recorded for the same samples in different physical forms. Inspired by this work, a single kernel near-infrared spectroscopic method optimized by calibration transfer was proposed. Using the calibration transfer algorithm, spectral space transformation (SST) [14], the near-infrared individual seed spectra were reconstructed to the new spectra, which were similar to the spectral profiles of single dehusked seed or seed flour. Then, the reconstructed spectra were analyzed by the calibration model of single dehusked seed or seed flour. This method should be advantageous, because the sample to be analyzed from an individual seed is replaced by that of a single dehusked seed or seed flour, the irregular physical characteristics and the heterogeneous constituent distributions of individual seeds could be ignored in the model. This enables us to obtain a more precise calibration model. In addition, since the applied calibration sample can be seed flour, especially a large amount of seed flour, the reference methods designed for bulk samples are applicable. Therefore, it is now possible to detect the constituents of individual seeds (for example, the oil content of individual rice seeds) that were previously difficult to analyze due to the limitations of the reference method.

In this paper, the protein content of single rice kernel (SRK) was used as a model

constituent to evaluate the proposed method. Since the protein content is closely related to rice yield [15] and quality [16], it is an important index in rice breeding. Protein is mainly distributed in the aleurone layer, the endosperm and the embryo of rice, and the presence of glume [17] interferes with its analysis. In previous studies, the protein content of rice flour (RF) [4,18], bulk brown rice [19], and single brown rice kernel (SBK) [20] were successfully analyzed by NIRS, but the protein content of single rice kernels has seldom been reported. In this work, 123 individual rice seeds (73 seeds for calibration and 50 seeds for validation) with different protein contents were collected, and the NIR transmission spectra of three forms of rice (single rice kernel, single brown rice kernel, and rice flour) and the corresponding reference values were obtained. Optimal models of the three forms of rice were established, and the protein contents of single rice kernels were predicted by two methods: the direct method (the tradition NIR method) and the proposed method. The direct method involved the analysis of single rice kernels directly via the single rice kernel model. The proposed method involved the transfer of single rice kernel spectra into single brown rice and rice flour spectra, followed by analysis using single brown rice and rice flour models. The results of the two methods were compared. In addition, the transfer results among the spectra of three forms of rice were compared. Ultimately, we aimed to achieve the following objectives: (i) to identify the optimal protein content by comparing different pretreatment methods and different spectral ranges; (ii) to discuss the validity of the proposed method by comparing the prediction results of single rice kernel protein content obtained from the proposed method and the direct

method; and (iii) to evaluate the feasibility of calibration transfer among the spectra of three forms of rice.

2. Materials and methods

2.1. Single rice kernel protein content spectral analysis via the two methods

The flow chart of single rice kernel protein content NIR analysis via the two methods is shown in Fig. 1. In the direct method, the SRK spectra were obtained for calibration and validation. Conversely, in the proposed method, the SBK and RF spectra were obtained for calibration, and the SRK spectra were obtained for validation. Compared with the direct method, the proposed method included a calibration transfer step, in which the SRK spectra were transferred into SBK spectra (the SRK-trans-SBK spectra) and RF spectra (the SRK-trans-RF spectra). In addition, to validate the calibration transfer performance of SRK and to compare the transfer results among the spectra of all three rice forms, the SBK and RF spectra were obtained for validation as well (Figure not shown).

2.2. Sample preparation

In total, 201 single rice kernels including 149 kernels for calibration and 52 kernels for external validation were used for near infrared analysis. The calibration samples were from 25 rice mutant varieties with different protein content (about 4–7 rice kernels per rice mutant variety). These rice mutant varieties were derived from the rice agronomic traits mutant library constructed by our laboratory. The mutant library was generated by low energy and heavy ion beam irradiation on rice variety ‘9311’ [21]. The external validation samples were selected from the 2018 national

crop variety regional test materials, with a total of 21 rice varieties (about 1–4 rice kernels per rice variety).

2.3. Sample pretreatment and spectra acquisition

The SRK samples were placed in an incubator for 24 h at 25°C and relative humidity (r.h.) of 50% before spectra acquisition. The NIR transmission spectra were obtained by MPA Fourier transform near-infrared spectrometer (Bruker, Germany) in the full spectral range (800–1726 nm), with a spectral resolution of 1 nm. The sketches of near-infrared diffuse transmission spectra acquisition are reported in Fig.

2. For the three different rice forms, the pretreatment steps were as follows:

(i) A circular aluminum sheet with a 2 mm diameter hole in the center was fixed in the detection window of the spectrometer. Each of the SRK samples was placed on the sheet horizontally while scanning. The averaged spectra were calculated after scanning the sample from the front side and reverse side, one time each (Fig. 2a).

(ii) The SBK samples were obtained by removing each rice kernel glume manually. Other pretreatment steps were the same as above (Fig. 2b).

(iii) The RF samples were obtained by grinding each SBK sample manually with a small mortar and pestle. The milled rice was passed through a 0.15 mm screen. Each RF sample was pressed flat into a 6 mm diameter glass vessel that was fixed on the detection window, equipped with a Bruker iris aperture for spectra collection (Fig. 2c). Each RF sample was scanned once.

2.4. Chemical analysis for protein content

After the spectra acquisition of three forms of rice, the protein contents of single

rice kernels were analyzed using the Dumas combustion method. This method had been successfully applied in single seed spectral analysis [22], for the minimum sampling weight were only a few milligrams. Each rice flour sample was weighed to 4.0 ± 0.2 mg using the electronic scale (Mettler Toledo, Switzerland) and wrapped in foil into a ball shape, and approximately 5 mg of pure benzene sulfonic acid was weighed and wrapped in foil, as the standard. All samples were analyzed in an elemental analyzer (Elementar, Germany). The protein content was calculated according to the instrument output of nitrogen ($N\% \times 5.95$).

2.5. Spectral Space Transformation

Spectral Space Transformation (SST) is a calibration transfer method that relies on standard samples. It is often used to transfer the spectra of the same sample set, measured on one instrument (or one test condition), to another. In this work, it was used to transfer the spectra from one form of rice to another. We assumed the matrices X_1 and X_2 are the spectral matrices of two different rice forms of the same set of standard rice samples. For example, if X_2 stands for the SRK spectral matrix, then X_1 stands for the corresponding SBK or RF spectral matrix of the same sample set. Assuming the augmented matrix $X_{comb}=[X_1, X_2]$, by singular value decomposition, the matrix X_{comb} can be expressed as follows:

$$X_{comb} = [U_s, U_n] \begin{bmatrix} \sum_s & 0 \\ 0 & \sum_n \end{bmatrix} [V_s, V_n]^T = T_s P_s^T + E = T_s [P_1^T, P_2^T] + E \quad (1)$$

where the superscript 'T' denotes the transpose, and the subscripts 's' and 'n' represent the corresponding factors of spectral information and noise, respectively.

$T_s = U_s * \sum_s$, $P_s = V_s$, $P_s^T = [P_1^T, P_2^T]$, $E = U_n \sum_n V_n^T$. P_1 and P_2 are the loadings of X_1 and X_2 ,

and E is the residual matrix. When matrix E is ignored, the matrix X_{comb} can be expressed as the product of the score matrix T_s and the loading matrix $[P_1^T, P_2^T]$. In addition, the number of SST factors is a key parameter, which corresponds to the column number of matrix V_s^T , the column number of U_s , or the row number of Σ_s . It determines how much spectral signals (i.e., the right singular matrix $[V_s, V_n]^T$) are taken as valid information (V_s^T), while the rest will be treated as noise (V_n^T).

When transferring the test spectra, the transferred spectra can be calculated according to the following formula:

$$X_{\text{trans}} = X_{\text{test}} (P_2^T)^+ P_1^T + X_{\text{test}} - X_{\text{test}} (P_2^T)^+ P_2^T \quad (2)$$

where X_{test} represents the spectral matrix of test samples, and X_{trans} represents the transferred test spectral matrix. The superscript '+' denotes the Moore-Penrose generalized inverse.

2.6. Multivariate analysis

After the selection of data pretreatment and spectral range, the calibration models were established through Partial Least Squares (PLS) regression. The coefficient of determination (R^2) and root mean square error of cross validation (RMSECV) were used to evaluate the performance of calibration models. The correlation coefficient (R) and root mean square error of prediction (RMSEP) were used to evaluate external validation results. Paired t-test (95% confidence) was used to evaluate the statistical significance between predicted values and reference values. Data pretreatments and calibration transfer were performed on matlab software version 2015b (The Mathworks, USA). The model construction and validation were performed on The

Unscrambler X software version 10.4 (CAMO, Norway). Statistic analysis and plotting were performed on OriginPro software version 9.1 (OriginLab Corp., USA).

3. Results and discussion

3.1. Protein contents

The descriptive statistics for protein content measured by the combustion method are shown in Table 1. The protein content of the calibration set was distributed between 6.0% and 13.1%, covering most of the protein content of validation set. The means, standard deviations (SD), and standard errors (SE) between calibration set and external validation set indicated that, the protein distribution of the validation samples was well represented by the calibration set.

3.2. Original spectra of three forms of rice samples

The NIR original diffuse transmission spectra of three forms of rice are shown in Fig. 3. As shown, the baseline drift was obvious in the spectra of each rice form. The main reasons lie in (i) the spectral absorption of glumes, (ii) the morphological differences of SRK and SBK samples, and (iii) the thickness differences of RF samples. The spectral absorption of SRK was higher than that of SBK, while the spectral baseline drift of SRK was lower than that of SBK. One explanation is that SBK was smaller than SRK. As a result, smaller seeds reflected less energy back to the sensor than larger seeds, and thus contributed to the variation in spectra. Smaller seeds also allowed more energy to pass through the kernels, and thus again contributed to variation in the spectra. Another possibility is that SRK had two layers of glumes. Light was reflected and refracted between glumes, which reduced the

energy reflected back to the sensor, and thus increased the spectral absorption of SRK.

3.3. Model optimization

The combinations of four pretreatments (no pretreatment, standard normal variate transformation (SNV), first derivative (1st der., default set to 17 point smoothing), and second derivative (2st der., default set to 17 point smoothing)) and three spectral ranges (full spectrum, the short wavelength region 950–1250 nm, the long wavelength region 1400–1700 nm) were used to screen the optimum models. Each combination selects the model with the maximum cross validation R^2 and the minimum RMSECV for comparison. In particular, SNV, first derivative, and second derivative are commonly used to eliminate particle shape interference and baseline drift. The selected short wavelength region (950–1250 nm) mainly contains the protein or peptide groups related absorption bands of 975–1015 nm (N-H stretch second overtone), 1159 nm (CH_2 antisymmetric stretching second overtone) and 1189 nm (CH_2 symmetric stretching second overtone); the long wavelength region (1400–1700 nm) mainly contains the protein-related absorption bands of 1480–1550 nm (N-H stretch first overtone) and 1620–1700 nm (aromatic C-H stretch first overtone) [23]. The comparison results are shown in Table 2.

Comparing the four pretreatment methods, the SNV can accurately correct the size and morphology interference of SRK and SBK. It also corrects the optical path difference caused by inconsistent RF thickness. The first derivative and the second derivative can also reduce the background interference to a certain extent, but the overall results were less accurate than that of the SNV pretreatment.

Comparing the three spectral ranges, the cross validation results of the spectral range of 950–1250 nm were better than that of the other two spectral ranges, because the former have higher R^2 values and lower RMSECV values. Rittiron et al. [20] have detected the protein content of SBK by NIRS. Under the pretreatment of second derivative, the long wavelength region (1300–1800 nm) was better than the short wavelength region (850–1000 nm) in modeling. Similarly, as shown in Table 2, under the same pretreatment, the long wavelength region (1400–1700 nm) was also better than the short wavelength region (950–1250 nm) in SBK modeling. However the results were the opposite for SRK. The findings indicated that the interference absorption of glume [17, 23] could not be eliminated well at 1400–1700 nm.

Comparing the three forms of rice samples, the RF samples that benefit from the uniformity of particle size and distribution have the best performance in modeling, followed by SRKs and SBKs. Based on the previous observations of Fig. 3, more spectral information and lower baseline drift may be the reason why the SRK model has better calibration results than the SBK model.

To summarize, the spectral range of 950–1250 nm, the SNV pretreatment, and 9 PLS factors were chosen to construct the optimum PLS models of all three forms of rice. Fig. 4 provides the PLS regression coefficients of SRK, SBK, and RF models under the above parameters. The high regression coefficients were mainly distributed at 975, 1000, 1156, 1185, and 1200 nm nearby. These bands are close to the reported protein-related absorption bands [23], which proves that the optimized PLS models are appropriate.

3.4. Spectral space transformation for single rice kernel spectra

In order to ensure the transfer quality, SST is based on the preprocessed spectra (i.e., spectral range of 950–1250 nm, SNV pretreatment), rather than the original spectra, to minimize background absorption interference and baseline drift. Two steps are required for optimization prior to calibration transfer: the selection of standard spectra and the determination of the optimum SST factor number.

The number and the accuracy of the standard spectra determine the transfer result of SST. After screening the pretreated SRK spectra with Kennard–Stone (KS) algorithm [24], 48 spectra with the largest differences were determined as the standard spectra of SRK. Similarly, the spectra of corresponding numbers are selected from SBK and RF spectra of the respective data set, as the standard spectra of SBK and RF.

The appropriate SST factor number ensures that the process of calibration transfer does not lead to underfitting or overfitting. The SST factor number can be at most the same as the number of standard spectra. To determine the optimum SST factor number, the calibration SRK spectra were transferred into the corresponding SBK and RF spectra, and the RF and SBK models were used to predict the transferred spectra under different SST factor numbers, respectively. The corresponding SST factor number of the minimum root mean square error of calibration (RMSEC) was chosen as the best factor number (Fig. 5).

As shown in Fig. 5, when the SST factor number was less than 10, the RMSEC value fluctuated greatly; but as the SST factor number increased, the RMSEC value gradually decreased and tended to be stable. When the SST factor numbers were 28

and 24, respectively, the RMSEC of the calibration SRK-trans-SBK spectra and SRK-trans-RF spectra were the lowest (0.394 and 0.394, respectively). As a result, the same SST factor numbers were used to transfer the validation set spectra.

After the transfer of the SRK spectra of the validation set, the spectral differences before and after the transfer were examined through validation sample No. 97 (reference value of 11.2931%) and No. 11 (reference value of 7.5268%). Fig. 6a shows the external validation spectra of SRK, SRK-trans-SBK and SBK of these two samples, and Fig. 6b shows the external validation spectra of SRK, SRK-trans-RF and RF.

From Fig. 6, the spectral profiles between SRK and SBK, and between SRK and RF, were different. After calibration transfer, the spectral profiles between SRK-trans-SBK and SBK, and between SRK-trans-RF and RF, were similar.

3.5. External validation results for protein content of single rice kernels via two methods

The SBK model and RF model were used to predict the external validation SRK-trans-SBK spectra and external validation SRK-trans-RF spectra (the proposed method), as the experimental group. The SRK model was used to predict the external validation SRK spectra (the direct method), as the positive control group. The SBK model and RF model were used to predict the external validation SRK spectra, as the negative control group. The results of external validation and paired t-test are shown in Table 3.

When the SBK model and RF model were used to predict SRK spectra directly,

the RMSEP was large, and a statistically significant difference was found between the predicted values and the reference values. However, after the SRK spectra were transferred into SBK and RF spectra, the RMSEP was reduced, and no statistically significant difference was found between the predicted values and the reference values. In particular, the RMSEP of RF model (0.401) was slightly lower than that of the SRK model (0.423), which may be due to the higher accuracy of RF model. This indicates that SST can effectively correct the spectral differences between SRK and SBK, and between SRK and RF. Compared with the results of previous studies, the prediction results of both the direct method and the proposed method were less accurate than that of single brown rice kernels reported by Rittiron [20] and of the rice flour reported by Delwiche [18] and Xie [4]. However, the results were better than that of bulk brown rice that reported by Bagchi [19], for the latter had a higher SEP value (0.872). This implies that it is feasible to analyze single rice kernel protein content accurately using either the direct method or the proposed method.

3.6. External validation of protein content of single brown rice kernel and rice flour via two methods

We confirmed the transferred SRK spectra were successfully predicted by the models of the other two rice forms. Therefore, to understand whether the transferred SBK and RF spectra could be successfully predicted by the models of the other two rice forms, the external validation SBK spectra were transferred into the corresponding SRK and RF spectra, and the external validation RF spectra were transferred into the corresponding SRK and SBK spectra, respectively. The data sets

were subsequently predicted by corresponding models as experimental groups. In addition, the SBK model (and RF model) was used to predict the external validation SBK spectra (and RF spectra), as controls. The results of the external validation and paired t-test are shown in Table 4.

Similar to the SRK spectra, the SBK and RF spectra were transferred into the spectra of the other two rice forms successfully. There was no statistically significant difference between the predicted values and the reference values. The experimental groups have good correlation coefficients and RMSEP values, which were only slightly worse than those of the control groups.

The spectral difference between SRK and SBK largely originates from the interference of glumes, including the spectral absorption of glume and the change in the optical path caused by the reflection and refraction of diffuse light in the upper and lower glume layers. The spectral differences between SBK and RF are mainly due to the morphology and composition distribution of caryopsis, whereas the spectral differences between SRK and RF spectra are a combination of the above effects. From Tables 3 and 4, the SST algorithm can largely correct these differences, so the three spectra of rice forms can be transferred freely among each other. Therefore, by using the proposed method, the spectra of all three rice forms can be predicted by the models of all three rice forms. In particular, the calibration transfer results between SBK and RF ($\text{RMSEP}_{\text{SBK-trans-RF}}=0.446$, $\text{RMSEP}_{\text{RF-trans-SBK}}=0.384$) are slightly better than the results between SRK and RF ($\text{RMSEP}_{\text{RF-trans-SRK}}=0.455$, $\text{RMSEP}_{\text{SRK-trans-RF}}=0.401$), and those between SRK and SBK

($RMSEP_{SBK-trans-SRK}=0.527$, $RMSEP_{SRK-trans-SBK}=0.480$). The results indicated that the interferences from morphology and composition distribution are easier to correct by calibration transfer than other interferences.

4. Conclusion

In this study, a calibration transfer optimized single kernel near-infrared spectroscopic method was proposed, and the use of this method was illustrated in a case study of single rice kernel protein content spectral analysis. The method mainly consists of three steps: (i) Identify the optimal NIR models of SBK and RF. (ii) Use the SST algorithm to transfer the SRK spectra into the spectra of SBK and RF. The selection of standard spectra and appropriate SST factors determine the effect of the transfer. (iii) Use the SBK model and RF model to predict the transferred spectra. In addition, we confirmed that the spectra of all three rice forms could be freely transferred among each other. Some predicted results of the proposed method were better than those of the direct method. Therefore, when difficulties arise in the spectral analysis of crop seeds, such as the irregular physical characteristics and the heterogeneous constituent distribution of single seed, and the insufficient detection accuracy of reference method, it is possible that they are overcome by transferring the spectra of one seed form into the spectra of other forms of these seeds. In conclusion, this study provides a promising approach to improve SKNIRS, which can be applied to the analysis of more seed constituents and more crop seeds.

Acknowledgement

This work was financially supported by the National Natural Science Foundation

of China (Grant No. 31500300) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA08040107).

Conflict of interest

The authors declare no competing financial interest.

ACCEPTED MANUSCRIPT

References

- [1] L. Silvela, R. Rodgers, A. Barrera, D.E. Alexander, Effect of Selection Intensity and Population-Size on Percent Oil in Maize, *Zea-Mays-L*, *Theor Appl Genet*, 78 (1989) 298–304.
- [2] C. Pasquini, Near infrared spectroscopy: A mature analytical technique with new perspectives – A review, *Anal Chim Acta*, 1026 (2018) 8–36.
- [3] J.G. Wu, C.H. Shi, Prediction of grain weight, brown rice weight and amylose content in single rice grains using near-infrared reflectance spectroscopy, *Field Crops Res*, 87 (2004) 13–21.
- [4] L.H. Xie, S.Q. Tang, N. Chen, J. Luo, G.A. Jiao, G.N. Shao, X.J. Wei, P.S. Hu, Optimisation of near-infrared reflectance model in measuring protein and amylose content of rice flour, *Food Chem*, 142 (2014) 92–100.
- [5] L.E. Agelet, C.R. Hurburgh, Limitations and current applications of Near Infrared Spectroscopy for single seed analysis, *Talanta*, 121 (2014) 288–299.
- [6] L.E. Agelet, P.R. Armstrong, I.R. Clariana, C.R. Hurburgh, Measurement of Single Soybean Seed Attributes by Near-Infrared Technologies. A Comparative Study, *J Agr Food Chem*, 60 (2012) 8314–8322.
- [7] P.R. Armstrong, J.G. Tallada, C. Hurburgh, D.F. Hildebrand, J.E. Specht, Development of Single-Seed near-Infrared Spectroscopic Predictions of Corn and Soybean Constituents Using Bulk Reference Values and Mean Spectra, *T Asabe*, 54 (2011) 1529–1535.
- [8] S.R. Delwiche, Single Wheat Kernel Analysis by near-Infrared Transmittance -

Protein-Content, *Cereal Chem*, 72 (1995) 11–16.

[9] T. Bramble, F.E. Dowell, T.J. Herrman, Single-kernel near-infrared protein prediction and therole of kernel weight in 14 hard red winter wheat, *Appl Eng Agric*, 22 (2006) 945–949.

[10] G. Spielbauer, P. Armstrong, J.W. Baier, W.B. Allen, K. Richardson, B. Shen, A.M. Settles, High-Throughput Near-Infrared Reflectance Spectroscopy for Predicting Quantitative and Qualitative Composition Phenotypes of Individual Maize Kernels, *Cereal Chem*, 86 (2009) 556–564.

[11] C.F. Pereira, M.F. Pimentel, R.K.H. Galvao, F.A. Honorato, L. Stragevitch, M.N. Martins, A comparative study of calibration transfer methods for determination of gasoline quality parameters in three different near infrared spectrometers, *Anal Chim Acta*, 611 (2008) 41–47.

[12] R.N. Feundale, N.A. Woody, H.W. Tan, A.J. Myles, S.D. Brown, J. Ferre, Transfer of multivariate calibration models: a review, *Chemometr Intell Lab*, 64 (2002) 181–192.

[13] L.S.A. Pereira, M.F. Carneiro, B.G. Botelho, M.M. Sena, Calibration transfer from powder mixtures to intact tablets: A new use in pharmaceutical analysis for a known tool, *Talanta*, 147 (2016) 351–357.

[14] W. Du, Z.P. Chen, L.J. Zhong, S.X. Wang, R.Q. Yu, A. Nordon, D. Littlejohn, M. Holden, Maintaining the predictive abilities of multivariate calibration models by spectral space transformation, *Anal Chim Acta*, 690 (2011) 64–70.

[15] C.F. Jenner, T.D. Ugalde, D. Aspinall, The Physiology of Starch and Protein

Deposition in the Endosperm of Wheat, *Aust J Plant Physiol*, 18 (1991) 211–226.

[16] M. Martin, M.A. Fitzgerald, Proteins in rice grains influence cooking properties, *J Cereal Sci*, 36 (2002) 285–294.

[17] Y. F. Huang, S. L. Lo, 19 - Utilization of rice hull and straw, in: J. Bao (Ed.) *Rice* (Fourth Edition), AACC International Press 2019, 627–661.

[18] S.R. Delwiche, M.M. Bean, R.E. Miller, B.D. Webb, P.C. Williams, Apparent Amylose Content of Milled Rice by near-Infrared Reflectance Spectrophotometry, *Cereal Chem*, 72 (1995) 182–187.

[19] T.B. Bagchi, S. Sharma, K. Chattopadhyay, Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran, *Food Chem*, 191 (2016) 21–27.

[20] R. Rittiron, S. Saranwong, S. Kawano, Useful tips for constructing a near infrared-based quality sorting system for single brown-rice kernels, *J Near Infrared Spec*, 12 (2004) 133–139.

[21] W. M. Cheng, B. M. Liu, Y. F. Ye, L. Z. Tao, Y. J. Wu, Screening of amylose and protein mutants and correlation analysis of agronomic character in rice. *Acta Laser Biology Sinica*, 25 (2016) 356–361.

[22] N. Caporaso, M.B. Whitworth, I.D. Fisk, Protein content prediction in single wheat kernels using hyperspectral imaging, *Food Chem*, 240 (2018) 32–42.

[23] J. Workman, – Spectra Numbers 1–560, in: J. Workman (Ed.) *The Handbook of Organic Compounds*, Academic Press, Burlington, (2001) 170–176.

[24] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments,

Technometrics, 11 (1969) 137–148.

ACCEPTED MANUSCRIPT

Figure captions:

Fig. 1. Flow chart of single rice kernel protein content near-infrared spectral analysis via two methods.

Fig. 2. Sketches of near-infrared diffuse transmission spectra acquisition for single rice kernel (a), single brown rice kernel (b), and rice flour (c).

Fig. 3. Near-infrared diffuse transmission original spectra of single rice kernels (a), single brown-rice kernels (b), and rice flours (c).

Fig. 4. Regression coefficients of PLS models for single rice kernel (black), single brown-rice kernel (red), and rice flour (blue) at 950–1250 nm and SNV pretreatment

Fig. 5. RMSEC for calibration SRK-trans-SBK spectra (a) and calibration SRK-trans-RF spectra (b) at different SST factor numbers.

Fig. 6. The transfer of two external validation samples (No. 11 and No. 97) from SRK spectra to SBK spectra (a) and RF spectra (b).

Table 1

Descriptive statistics for protein content of single rice kernels measured by combustion.

Parameter	Calibration samples					External validation samples				
	N ^a	range	mean	SE ^b	SD ^c	N ^a	range	mean	SE ^b	SD ^c
% protein	149	6.0–13.1	9.18	0.114	1.394	52	5.6–12.8	8.72	0.243	1.751

^aSample size. ^bStandard error. ^cStandard deviation.

Table 2

Cross-validation statistics of three forms of rice protein content PLS prediction models under four pretreatment methods and three spectral ranges

Sample	Data pretreatment	Spectral range (nm)			R^2 , RMSECV, PLS factors
		Full wavelength	950–1250	1400–1700	
Single rice kernel	None	0.787, 0.646, 13	0.842, 0.556, 8	0.687, 0.782, 9	
	SNV	0.854, 0.535, 11	0.931, 0.366, 9	0.757, 0.689, 12	
	1st der.	0.851, 0.540, 12	0.841, 0.558, 5	0.809, 0.612, 11	
	2st der.	0.847, 0.547, 9	0.832, 0.573, 3	0.732, 0.724, 9	
Single brown rice kernel	None	0.730, 0.729, 11	0.388, 1.094, 5	0.756, 0.691, 8	
	SNV	0.903, 0.436, 13	0.914, 0.409, 9	0.788, 0.644, 7	
	1st der.	0.810, 0.609, 9	0.714, 0.748, 6	0.765, 0.678, 7	
	2st der.	0.822, 0.590, 8	0.725, 0.733, 5	0.809, 0.611, 8	
Rice flour	None	0.837, 0.565, 10	0.789, 0.643, 6	0.838, 0.562, 8	
	SNV	0.899, 0.444, 10	0.944, 0.332, 9	0.895, 0.453, 8	
	1st der.	0.875, 0.495, 10	0.839, 0.561, 6	0.868, 0.509, 8	
	2st der.	0.870, 0.504, 10	0.840, 0.559, 4	0.877, 0.490, 9	

The results of the optimum models are boldfaced.

Table 3

External validation and paired *t*-test (at 5% level of significance) for protein content of single rice kernels.

Groups	Calibration model	External validation spectra	SST factor number	R	RMSEP	T value	Significance
SRK experimental group	SBK	SRK-trans-SBK	28	0.962	0.480	0.242	NS
	RF	SRK-trans-RF	24	0.975	0.401	1.635	NS
SRK control group	SRK	SRK	/	0.971	0.423	1.433	NS
	SBK	SRK	/	0.956	1.384	15.106	S
	RF	SRK	/	0.971	0.765	10.883	S

Table 4External validation and paired *t*-test (at 5% level of significance) for protein content of single brown rice kernels and rice flour

Groups	Calibration models	External validation spectra	SST factor number	R	RMSEP	T value	Significance
SBK experimental group	SRK	SBK-trans-SRK ^a	21	0.958	0.527	1.053	NS
	RF	SBK-trans-RF ^b	26	0.973	0.446	0.145	NS
SBK control group	SBK	SBK	/	0.975	0.385	0.011	NS
RF experimental group	SRK	RF-trans-SRK ^c	9	0.978	0.455	1.933	NS
	SBK	RF-trans-SBK ^d	20	0.977	0.384	0.920	NS
RF control group	RF	RF	/	0.986	0.368	1.457	NS

^aSingle brown rice kernel spectra transferred into single rice kernel spectra. ^bSingle brown rice kernel spectra transferred into rice flour spectra.^cRice flour spectra transfer into single rice kernel spectra. ^dThe rice flour spectra transferred into the single brown rice kernel spectra.

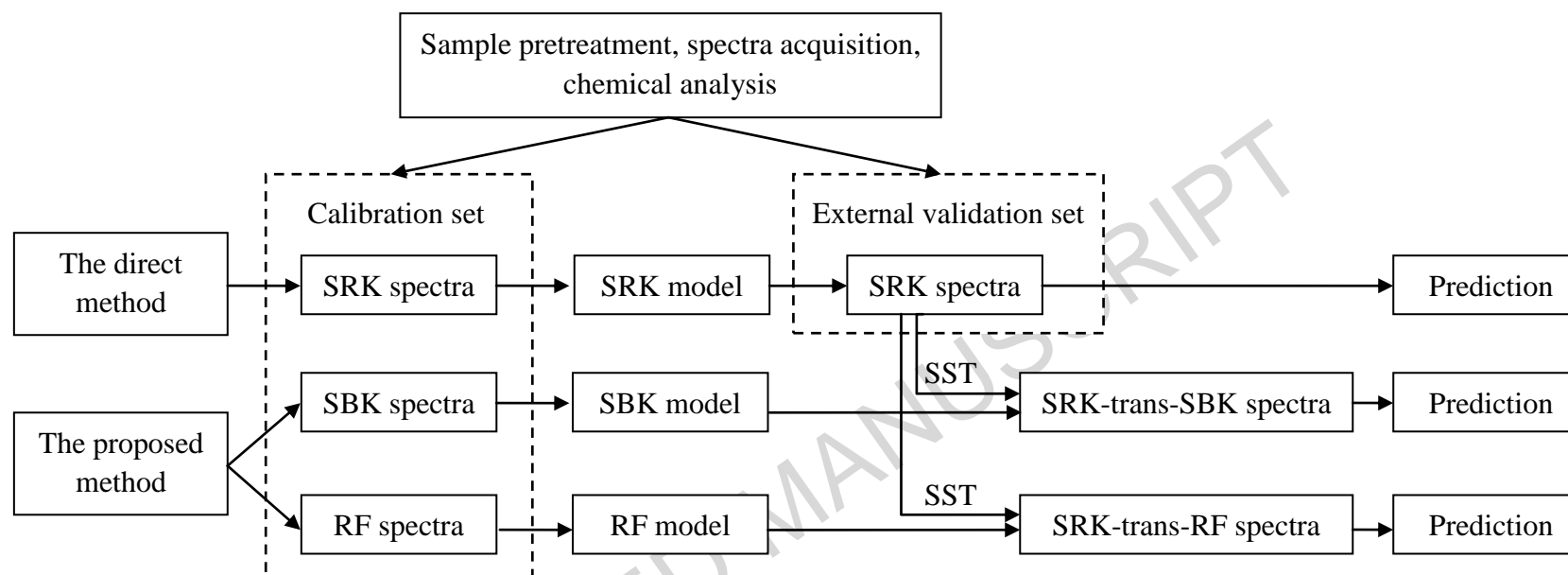


Fig. 1. Flow chart of single rice kernel protein content near-infrared spectral analysis via two methods.

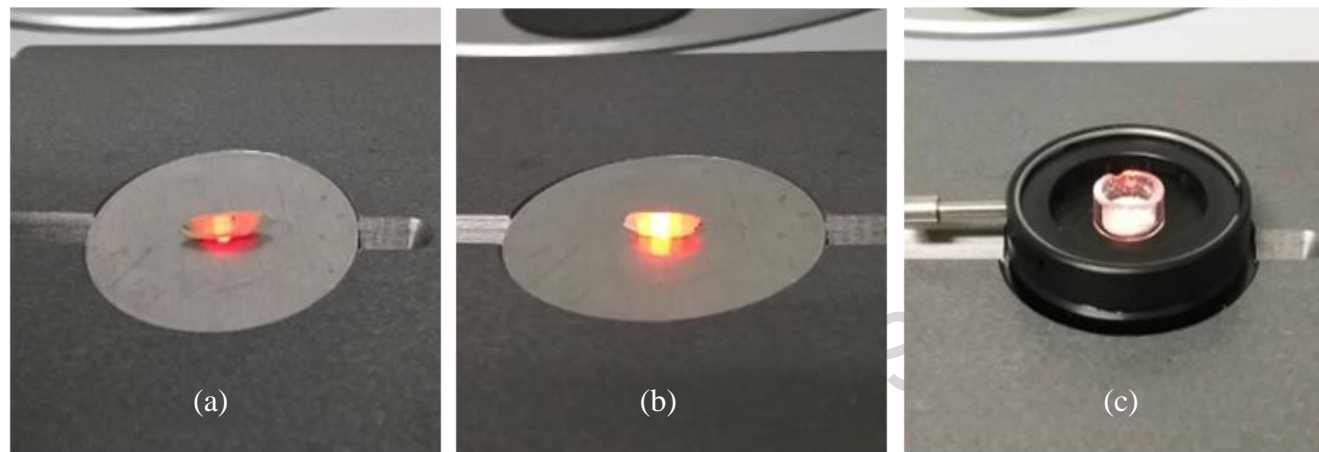


Fig. 2. Sketches of near-infrared diffuse transmission spectra acquisition for single rice kernel (a), single brown rice kernel (b), and rice flour (c).

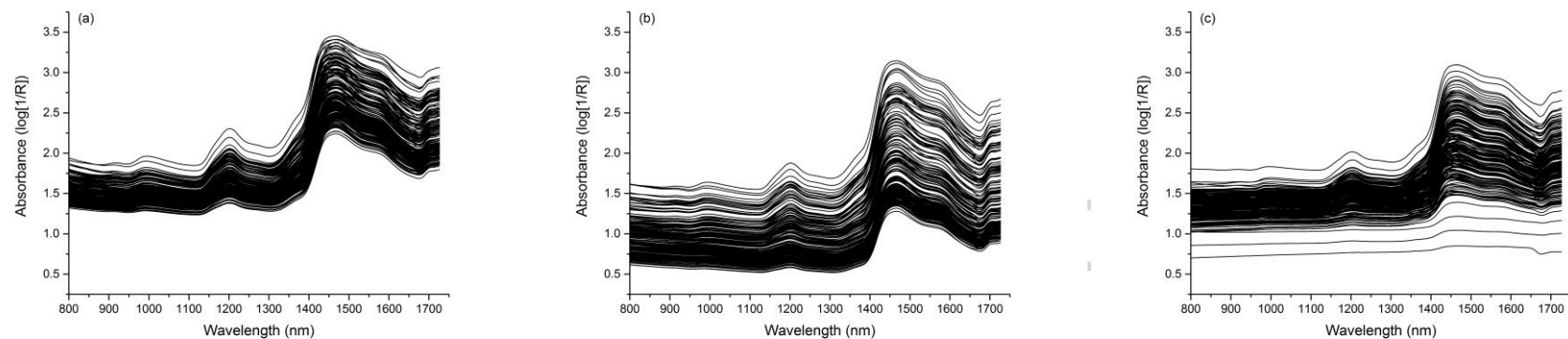


Fig. 3. Near-infrared diffuse transmission original spectra of single rice kernels (a), single brown-rice kernels (b), and rice flours (c).

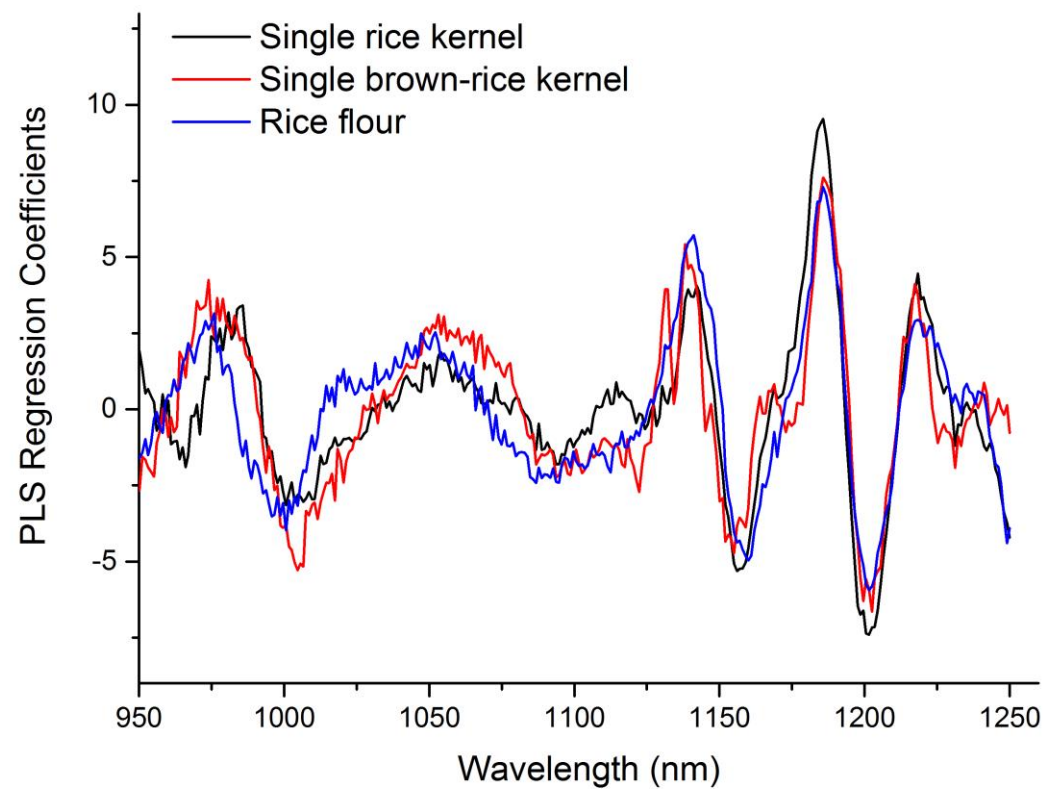


Fig. 4. Regression coefficients of PLS models for single rice kernel (black), single brown-rice kernel (red), and rice flour (blue) at 950–1250 nm and SNV pretreatment

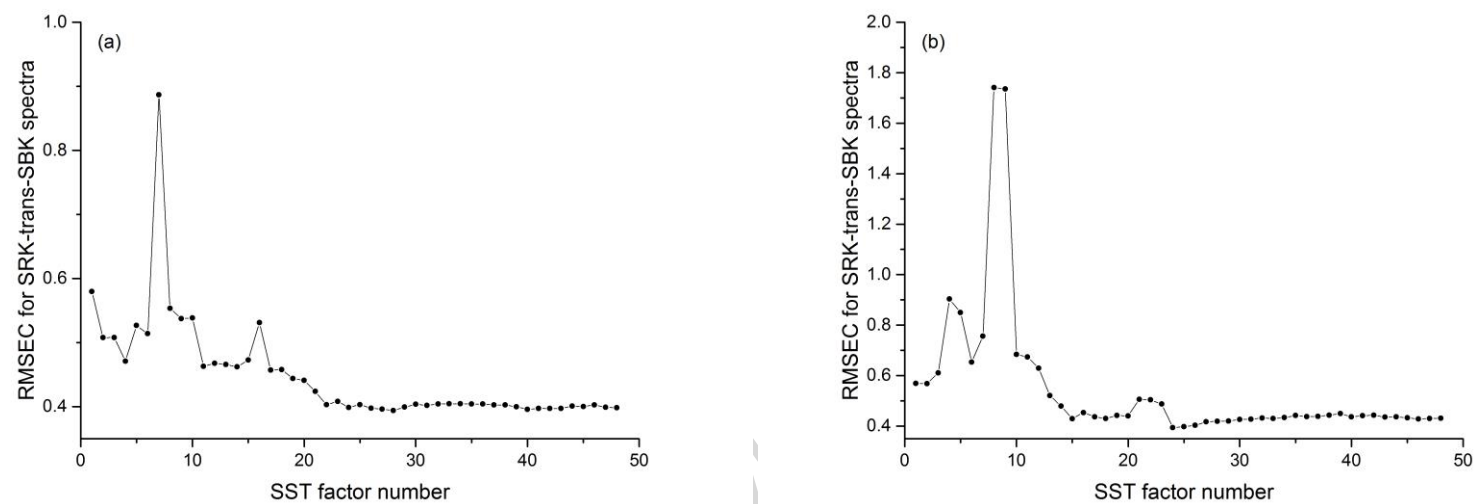


Fig. 5. RMSEC for calibration SRK-trans-SBK spectra (a) and calibration SRK-trans-RF spectra (b) at different SST factor numbers.

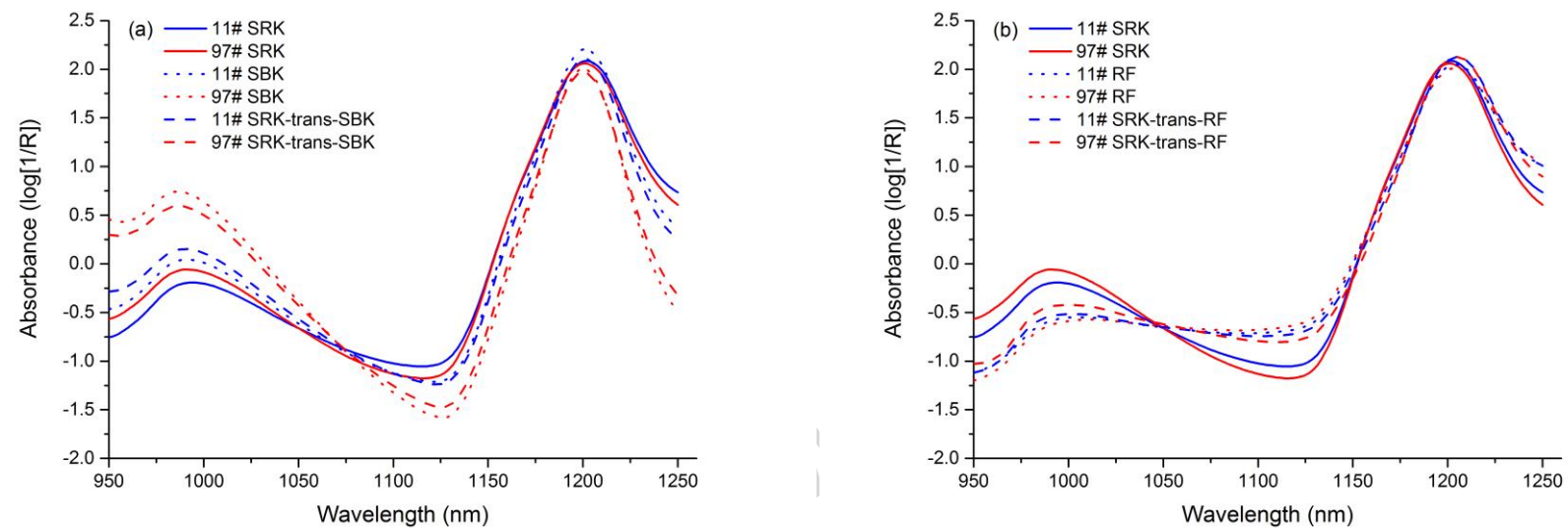
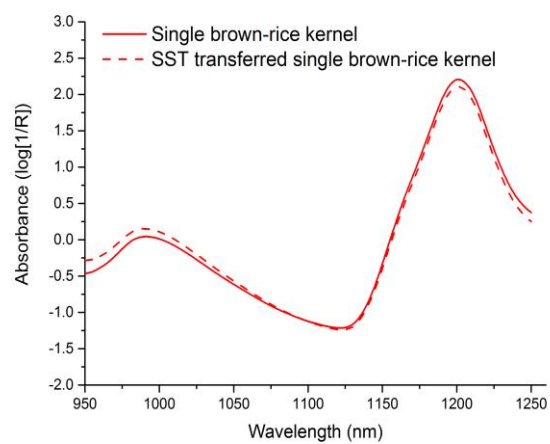
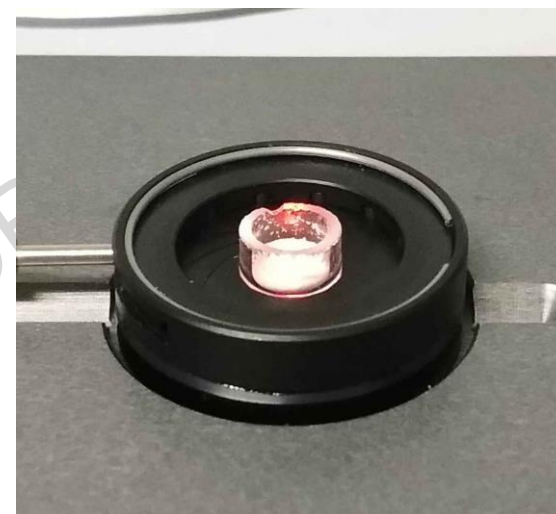
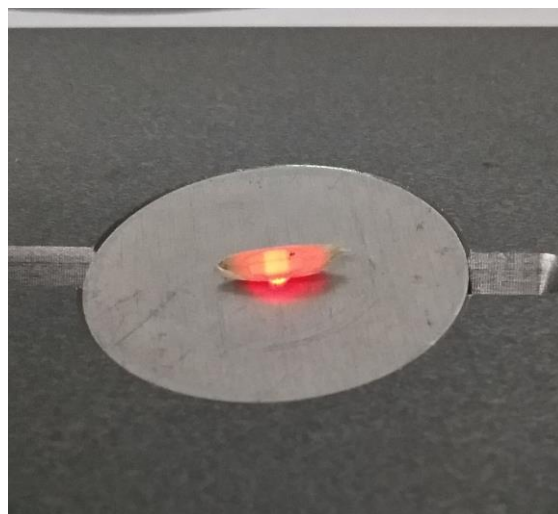
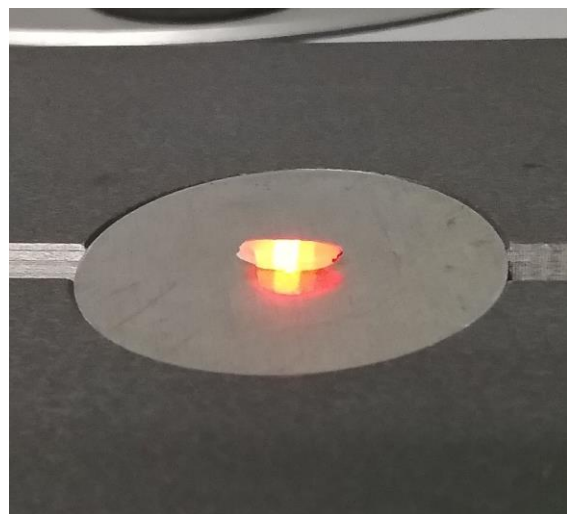
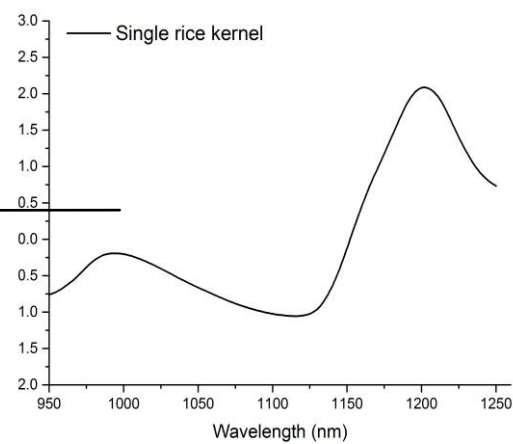


Fig. 6. The transfer of two external validation samples (No. 11 and No. 97) from SRK spectra to SBK spectra (a) and RF spectra (b).

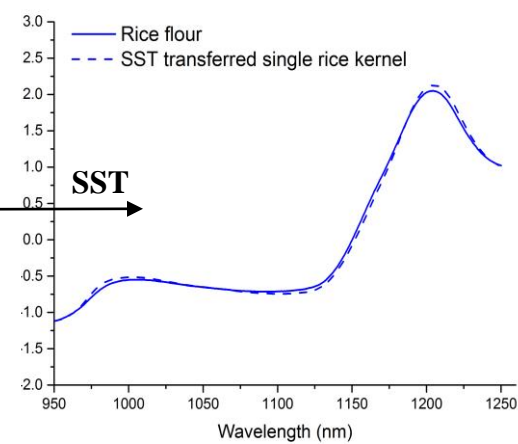
Graphical Abstract



SST



SST



Highlights

- A calibration transfer optimized NIR method was used to analyze individual seeds.
- The protein contents of single rice kernels were evaluated via two NIR methods.
- Spectra of individual seeds in different forms can be transferred among each other.