E- Learning Lesson on

Predictive Modeling of Crop Nutritional Traits via Near Infrared Reflectance Spectroscopy (NIRS) and WinISI: A step-by-step Guide

Publication No: ICARNEH/PME-TU/Pub/2023/129.

Prepared by:

Simardeep Kaur

Naseeb Singh

Rakesh Bhardwaj

ICAR Research Complex for North Eastern Hill Region

Umiam- 793103, Meghalaya (India)

Citation: Kaur S, Singh N, and Bhardwaj R (2024) Predictive Modeling of Crop Nutritional Traits via Near Infrared Reflectance Spectroscopy (NIRS) and WinISI: A step-by-step Guide. Publication No: ICARNEH/PME-TU/Pub/2023/129.

TABLE OF CONTENTS

S. No.	Topic	Page No.
1	Introduction	1-2
2	Aim and methodology of the module	2
3	Principle of NIR-spectroscopy	3
4	Development of NIRS-based predictive models using WinISI	3-5
5	Step 1: Collection and preparation of samples	6
6	Step 2: NIR-spectra acquisition	6-8
7	Step 3: Designing calibration and validation sets	9
9	Step 4: Development of calibration equations with full-length spectra	9-13
8	Step 5: External validation of the model and parameters to assess the robustness of the developed model	13-15
9	Step 6: Selection of best-fit model based on global metrics	15-18
10	Step 7: Statistical analysis	18-20
11	Conclusion	20
12	Acknowledgements	21
13	Suggested references and contacts	21-22

Predictive Modeling of Crop Nutritional Traits via Near Infrared Reflectance Spectroscopy (NIRS) and WinISI: A step-by-step Guide

1. Introduction

Concerning global food, nutritional, and ecological security, both major and minor crops play a key role. Major crops, such as wheat, rice, and maize form the backbone of the global food supply and provides essential calories and nutrients to billions of people. These staple crops are important for sustaining large populations and are integral to the agricultural economies of many countries. Minor crops, often referred to as underutilized or potential or orphan crops, include a diverse range of species such as millets, pulses, oilseeds, and traditional vegetables. Despite their vast potential, these crops are often overlooked in mainstream agriculture. However, they hold significant promise for enhancing nutritional security due to their rich nutrient profiles and adaptability to diverse agro-ecological conditions.

India, with its vast and diverse agricultural landscape is a treasure trove of germplasm for both staple and underutilized crops. The country's rich heritage includes a wide variety of crop species and genetic diversity which are maintained in national gene bank (located at ICAR- NBPGR, New Delhi) and research institutions. These collections are invaluable resources for plant breeders and researchers aiming to develop nutritionally rich and resilient crop varieties. Before such potential crops can be mainstreamed into agricultural systems, it is essential to conduct comprehensive nutritional and anti-nutritional profiling. This profiling is critical for identifying varieties with superior nutritional qualities and low levels of anti-nutritional factors, thereby ensuring that new cultivars meet the dietary needs of the population. Nutritional profiling involves assessing the levels of essential nutrients such as proteins, carbohydrates, vitamins, and minerals, as well as identifying the presence of anti-nutritional compounds like phytic acid, saponins, and oxalates. These anti-nutritional factors can interfere with the absorption of nutrients and reduce the overall nutritional value of the crops. Therefore, a detailed understanding of both the beneficial and detrimental components in crop germplasm is necessary to guide breeding programs aimed at enhancing the nutritional quality of crops.

However, traditional methods for nutritional profiling are often labor-intensive, time-consuming, and costly. They require extensive technical expertise and sophisticated laboratory equipment, which can be a significant barrier, especially when dealing with large germplasm collections. The scale of screening required for global repositories is immense, thus making it

necessary to find more efficient and cost-effective methods. To address these challenges, there is a growing need for alternative approaches that are rapid, non-destructive, and cost-effective. Near Infrared Reflectance Spectroscopy (NIRS) has emerged as a promising tool in this regard. NIRS is a non-destructive analytical technique that measures the reflectance of near-infrared light by a sample. It provides a quick and accurate way to assess the nutritional and anti-nutritional composition of crops without the need for extensive sample preparation or chemical analysis. This approach not only reduces the time and cost associated with traditional methods but also enables the high-throughput screening of large germplasm collections. By facilitating the rapid identification of nutrient-rich genotypes, NIRS can accelerate the development of nutritionally superior crop varieties. This, in turn, contributes to global food and nutritional security by promoting the cultivation of crops that are both high-yielding and rich in essential nutrients. Chemometrics involves the application of mathematical and statistical methods to design or select optimal measurement procedures and experiments and to provide maximum chemical information by analyzing chemical data. The use of WinISI software enhances the ability to analyze spectral data and build robust predictive models that can be applied to large datasets. Combining NIRS with advanced chemometric techniques, implemented through software such as WinISI, allows for the development of predictive models that can accurately estimate the nutritional traits of crop germplasm.

2. Aim and methodology of the lesson:

The aim of this lesson is to provide a comprehensive understanding of how to develop and validate NIR-based predictive models using WinISI software for assessing the nutritional traits of any crop germplasm. To better illustrate the process of prediction modeling, we will use a case study of rice bean germplasm as an example throughout this module, focusing on protein content as our desired trait for model development. We will cover the entire process, beginning with sample preparation for NIR spectra acquisition and moving through the acquisition of the spectra itself. The module will discuss the various parameters used to assess the accuracy of the models, including calibration and validation techniques. Additionally, we will explore the applicability of these models in practical scenarios, as well as the statistical analysis and measures necessary to ensure robust and reliable results.

3. Principle of NIR-spectroscopy

Near-Infrared (NIR) spectroscopy is a technique that measures the absorption of near-infrared light by molecular bonds in a sample. The principle is based on the fact that when NIR light is directed at a sample, specific wavelengths are absorbed by the sample's chemical bonds (primarily C-H, N-H, S-H, and O-H bonds). This absorption results in overtones and combinations of fundamental vibrations that can be measured and analyzed. The resulting spectrum provides a unique molecular fingerprint of the sample, reflecting its composition. NIR spectroscopy is particularly useful in predicting the nutritional content of crops because it is a non-destructive, rapid, and cost-effective method. By calibrating NIR models with known reference samples, robust predictive models can be developed. These models correlate the spectral data with specific nutritional parameters such as protein, moisture, fat, and fiber content. Once calibrated, the models can be used to quickly and accurately predict the nutritional content of unknown crop samples, thus facilitating quality control and breeding programs aimed at improving crop nutritional profiles.

4. Development of NIRS-based predictive models using WinISI

NIR spectroscopy faces challenges due to interference from factors such as spectral molecular vibration, mathematical treatments, and statistical methods during signal acquisition, leading to disturbances in spectra including baseline shifts and non-linearities. Therefore, pre-processing of spectral data is important for developing reliable prediction models, enhancing signal-to-noise ratio, increasing signal variation, and eliminating irrelevant sources unrelated to the property of interest. Empirical methods such as derivatives, multiplicative scatter correction (MSC), and standard normal variate (SNV) are commonly used for spectral pre-processing. Derivatives effectively enhance spectral resolution but may lead to overfitting of models. MSC and SNV methods help eliminate both additive and multiplicative impacts within spectra. SNV involves centering and scaling each spectrum by its standard deviation to reduce multiplicative effects of scattering. MSC establishes a reference spectrum, typically the mean spectrum of calibration data, to rectify baseline and amplification effects concerning the reference spectrum for each individual spectrum.

Following pre-processing, various regression algorithms can be used to develop predictive models for nutritional content estimation using NIR spectroscopy. Principal

Component Regression (PCR), Partial Least Squares (PLS), and Modified Partial Least Squares (MPLS) are among the commonly used regression techniques. PCR is based on decomposing the spectral data into orthogonal components, followed by regression analysis on these components. PLS, on the other hand, utilizes both spectral and reference data to identify latent variables that explain the maximum covariance between them, thereby building a predictive model. MPLS is a variation of PLS that incorporates modifications to improve model performance, such as considering the first derivative of the spectra or using weighted spectra. These regression algorithms utilise the relationship between spectral data and reference values to accurately predict the nutritional content of crops, thus offering valuable tools for quality control and breeding programs in agricultural research and industry.

Our research group has successfully developed and reported Near-Infrared Spectroscopy (NIRS) based prediction models for various crops, including rice, wheat, chickpea, pearl millet, amaranth, mung bean, cowpea and potato (John et al., 2022; Kondal et al, 2024; Priyadarshi et al, 2023; Tomar et al., 2021; Shruti et al., 2023; Bartwal et al., 2023; Padhi et al., 2022, Chaukhande et al., 2024). These models have demonstrated significant potential in accurately predicting key nutritional and antinutritional traits, facilitating rapid screening processes, and are currently being used to screen large germplasm collections in the national gene bank. Presently, we are focussing on developing robust NIRS prediction models for underutilized crops from the North Eastern Hill (NEH) Region of India. These crops, including Perilla, Buckwheat, Rice bean, Lablab bean, and French bean, possess immense potential due to their nutritional value and adaptability to the region's unique agro-climatic conditions.

In this lesson, we will consider a specific case study on the development of a robust NIRS-based prediction model for the rapid determination of protein content in rice bean germplasm. This case study will illustrate the step-by-step process of model development, from sample preparation to model validation. The methodology used in the present case study is depicted in Figure 1.

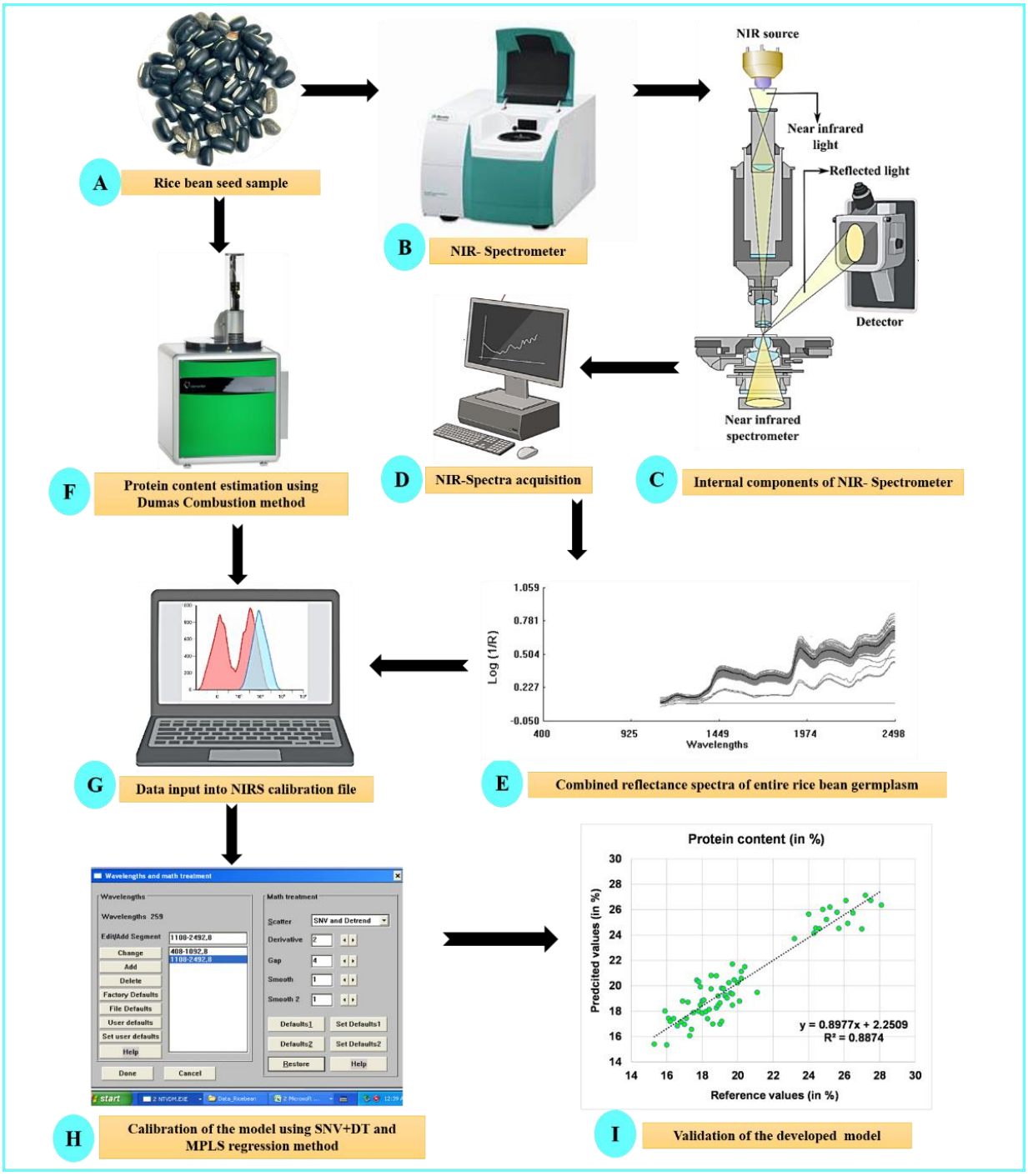

Fig. 1. Illustration of workflow employed in the present study for predicting protein content (in %) in rice bean germplasm Near-Infrared Spectroscopy (NIRS). (A) Collection of rice bean samples; (B) Loading of homogenized samples into a Near Infrared Spectrometer (NIRS); (C) Internal components of a typical NIRS; (D) NIR-spectra acquisition from 400nm to 2498 nm wavelength; (E) Presentation of the average combined reflectance spectra of entire rice bean germplasm; (F) Estimation of protein content of homogenized rice bean samples using Dumas combustion method; (G) Input of data into the NIRS calibration file; (H) Development of NIRS-based models using modified partial least square (MPLS); (I) Validation of the

developed model using test data, depicted by accuracy through scatter plots between the reference and predicted values.

Step 1: Collection and preparation of samples

- ✓ Obtain a collection of at least 100 diverse germplasm (here we are using 207 rice bean accessions), with varying shapes, sizes, and colors.
- ✓ Process each sample by grinding, homogenizing, and sieving through a 1 mm sieve using a non-contaminating cyclone mill (Foss Cyclotec™ or UDY Sample Mill).
- ✓ Store the resulting flour in airtight polyethylene terephthalate (PET) containers at 4°C until further NIR-spectra and wet chemistry analysis.
- ✓ Conduct wet lab analyses to determine the protein content using the Dumas combustion method (you may also choose method based on available resources and equipments).

Step 2: NIR-spectra acquisition

- ✓ Stabilize the homogenized rice bean samples at room temperature (25°C) for 6 hours to standardize temperature and moisture levels.
- ✓ Calibrate the NIR spectrometer by scanning a 100% white reference tile before scanning and at 30-minute intervals thereafter to ensure accurate measurements.
- ✓ Weigh approximately 4-5 g of the flour samples and load them into a circular ring cup with a quartz window (3.8 cm in diameter and 1 cm in thickness).
- ✓ Here we used FOSS NIRS DS3 spectrometer.
- ✓ Record each spectrum as an average of 32 scans across the spectral range of 400–2498 nm at 0.5 nm intervals**,** logging the data as log (1/R), where R represents relative reflectance. Your file will be saved as ".nir" file.

  equipped with Win ISI Project Manager Software version 1.61, to acquire reflectance spectra of the samples

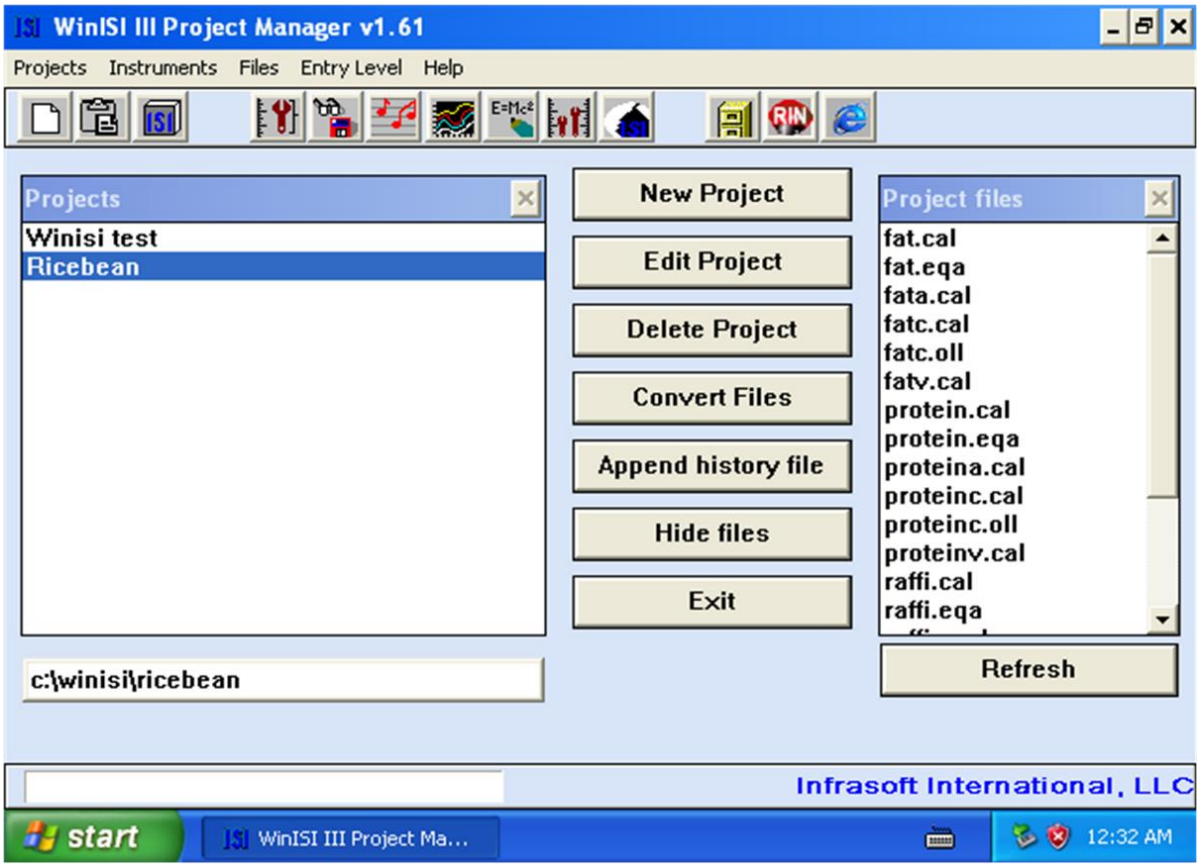

Fig. 2 Snapshot of Win ISI Project Manager Software version 1.61

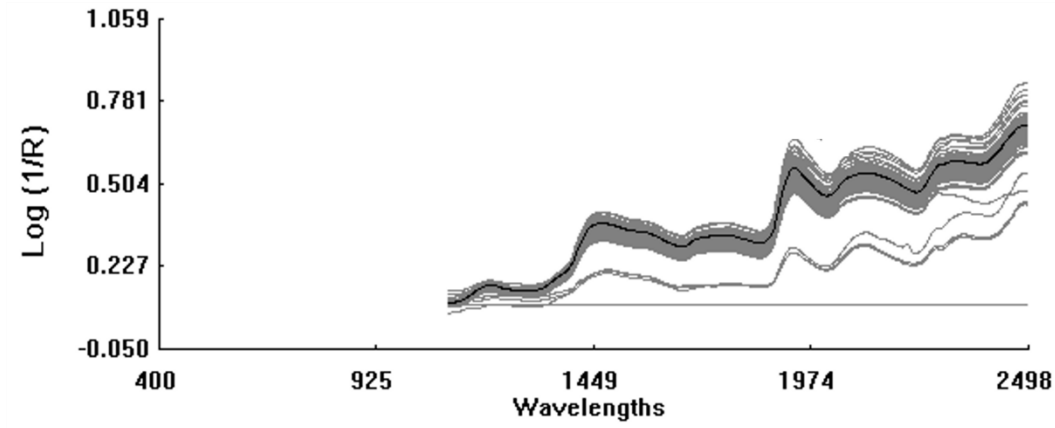

Fig. 3 A combined plot of the reflectance of entire rice bean germplasm.

Outlier identification

The technique of outlier detection was employed to determine the 'uniqueness' of a sample using Generalized H (GH) and Neighborhood H (NH) distances. According to Shenk and Westerhaus (1991b), NH is used to estimate the proximity of each sample to every other sample in the population, while GH assesses the ability of the calibration model to predict accuracy on an unknown sample and facilitates the removal of redundant spectra from the calibration population. NH measures the similarity between samples, whereas GH evaluates the predictive accuracy and helps to exclude irrelevant spectra by ranking them according to their H distance from the average spectrum. The algorithm defines spectral boundaries to eliminate outliers, with GH > 2.5 and NH < 0.6 (Fig. 4). Consequently, the final number of samples varied for each parameter, depending on the spectral and chemical variability within the sample population used for NIRS estimation.

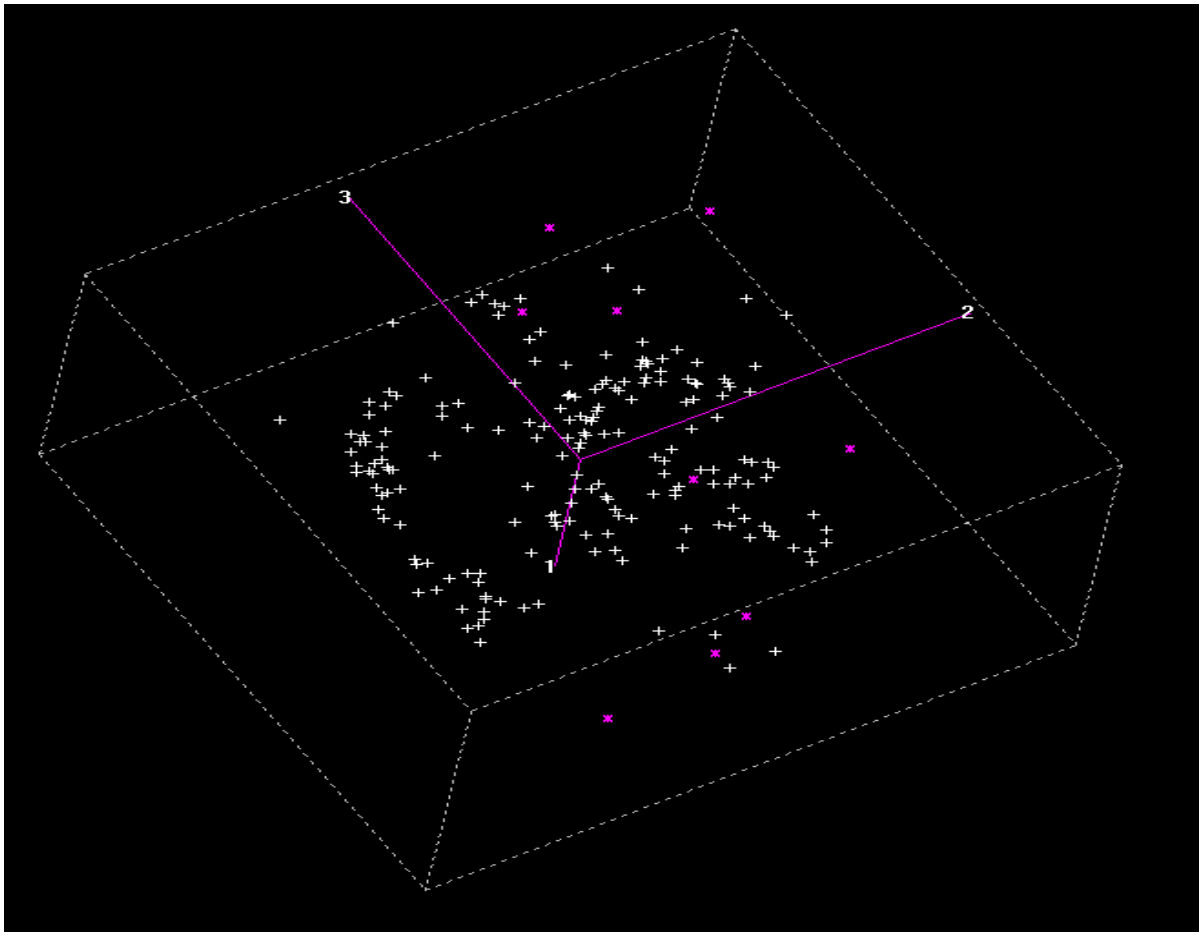

Fig. 4 Three-dimensional (3D) view showing outlier samples

Step 3: Designing calibration and validation sets

- ✓ Divide the dataset into a calibration/training set comprising ~60% of the samples and an external validation/testing set containing ~40% of the samples.
- ✓ Categorize the samples into these sets by considering the variability in protein content.
- ✓ Use Excel to sort the values, ensuring that both sets exhibit nearly equal variability and have comparable minimum, median, and maximum values to facilitate robust model development and validation.

Step 4: Development of calibration equations with full-length spectra

- ✓ Select the calibration (training) set and the validation set to span the entire concentration range. You may save your calibration file as "protein.cal", validation file as "protein.val", and equation file as "protein.equ". We selected 136 samples in calibration set and 71 samples in validation set.
- ✓ Select the function "global equations" and then "Develop equations with full spectrum" (Fig. 5).

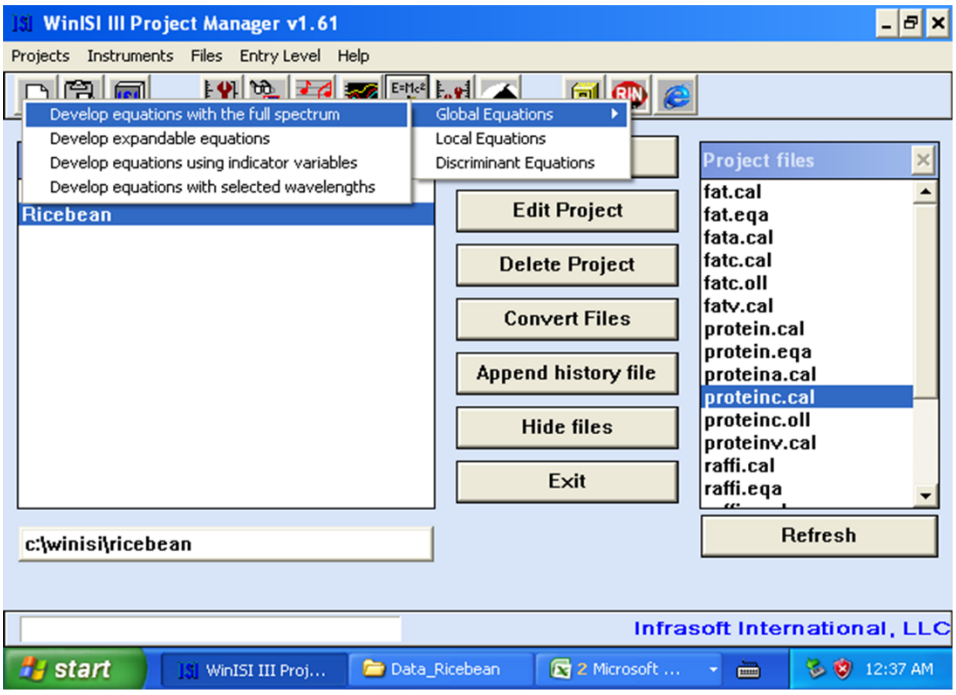

Fig. 5 Development of equations with full length rice bean NIR- spectra.

- ✓ Use spectral pre-processing techniques to address complexities from light scattering and path length variations. Although various scatter correction techniques are available (e.g., Standard MSC, Weighted MSC, Inverse MSC), we chose a combination of Standard Normal Variate (SNV) and Detrend (DT) for this study (Fig. 6).

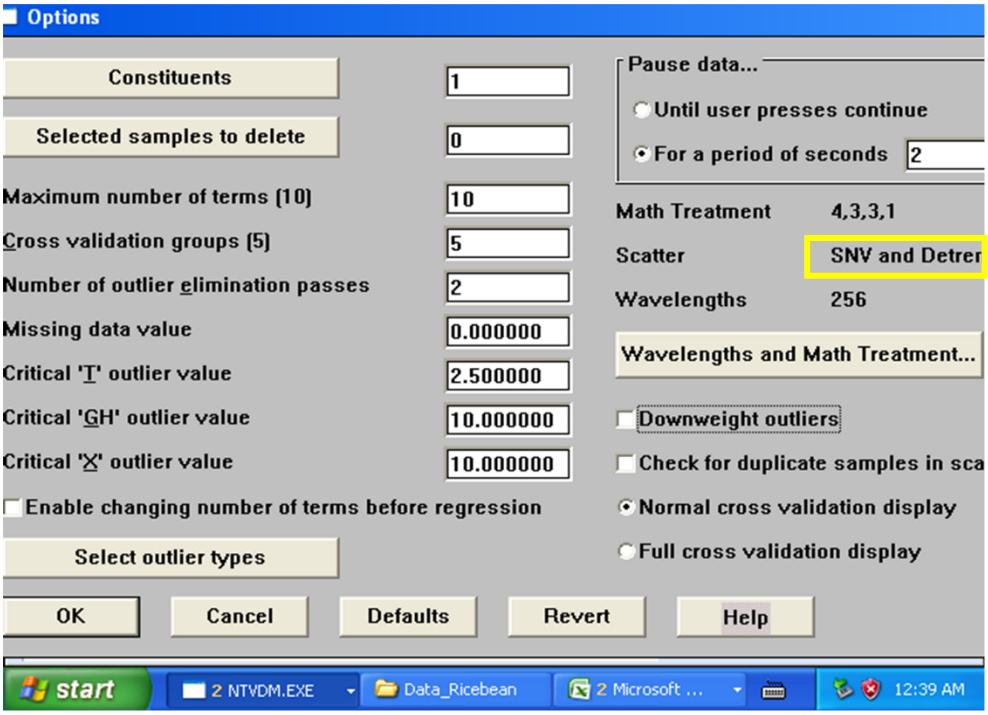

Fig. 6 Scatter correction method selection- Here we choose combination of SNV and DT.

- ✓ Apply SNV to center each spectrum around zero, effectively diminishing the multiplicative effects of particle size and scattering, reducing differences in global signal intensities. Use Detrend (DT) to fit a polynomial function to all data points, removing specific trending variations and large background interference, enhancing the determination of signal attributes related to analyte concentration.
- ✓ Combine SNV and DT pre-processing techniques to circumvent curvilinearity and noise in the NIRS signal baseline, preparing the spectral data for accurate analysis.
- ✓ Compute spectral derivatives to eliminate overlapping absorption bands and baseline shift effects.

- ✓ Create NIRS calibrations for the 400–2500 nm spectral region through trial and error. Incorporate mathematical treatments involving various combinations of derivatives to optimize predictive accuracy. For example, in the combination (2,4,2,2) (Fig. 7):
  - ➢ "2" represents the second derivative, addressing overlapping absorption bands and baseline shifts.
  - ➢ "4" signifies the gap, specifying four data points calculated by the second derivative.
  - ➢ "2" and "2" denote the number of data points in the first and second smoothing processes, respectively.
- ✓ Post scatter correction, use various regression algorithms such as Partial Least Squares (PLS), Modified Partial Least Squares (MPLS), and Principal Component Regression (PCR); we chose MPLS due to its superior accuracy and stability, its ability to address multicollinearity, improve robustness against outliers, and optimize spectral data treatment (Fig. 8).

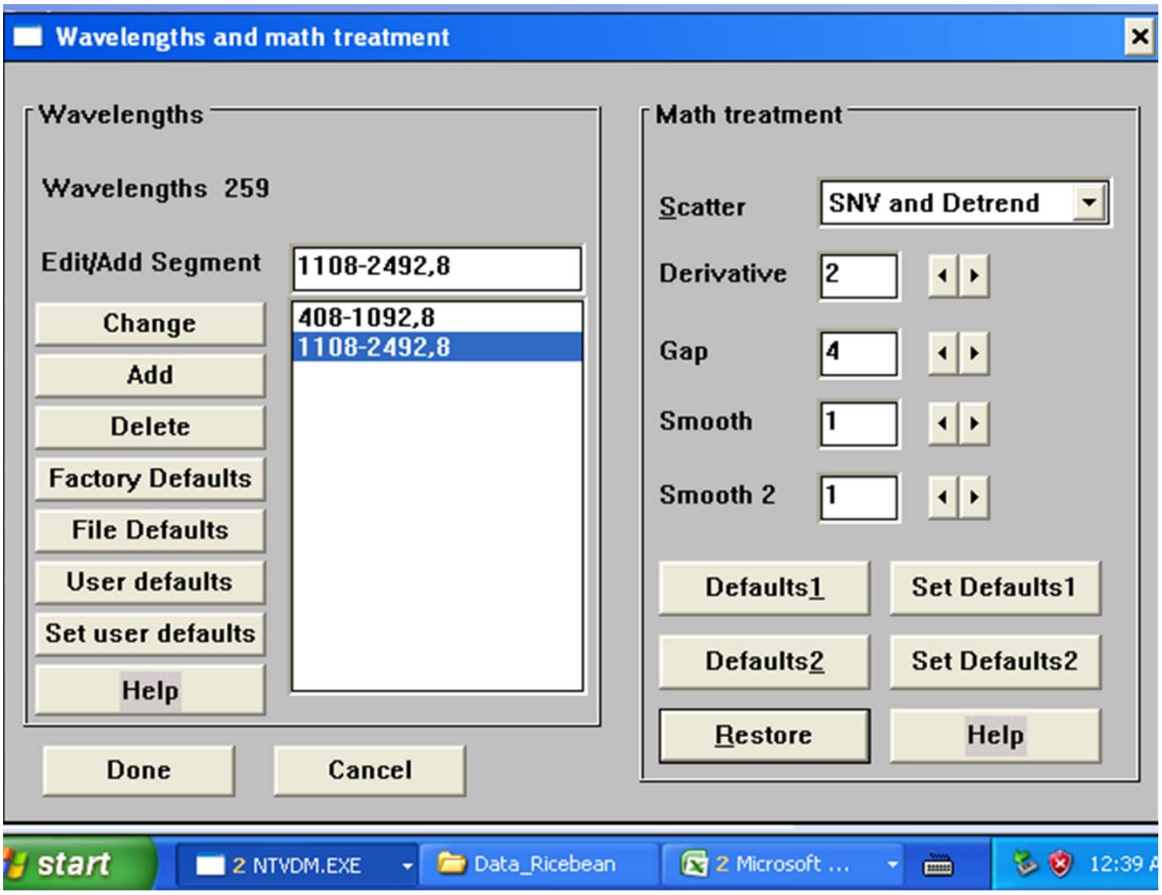

Fig. 7 Selection of wavelengths and mathematical treatments to develop the equations.

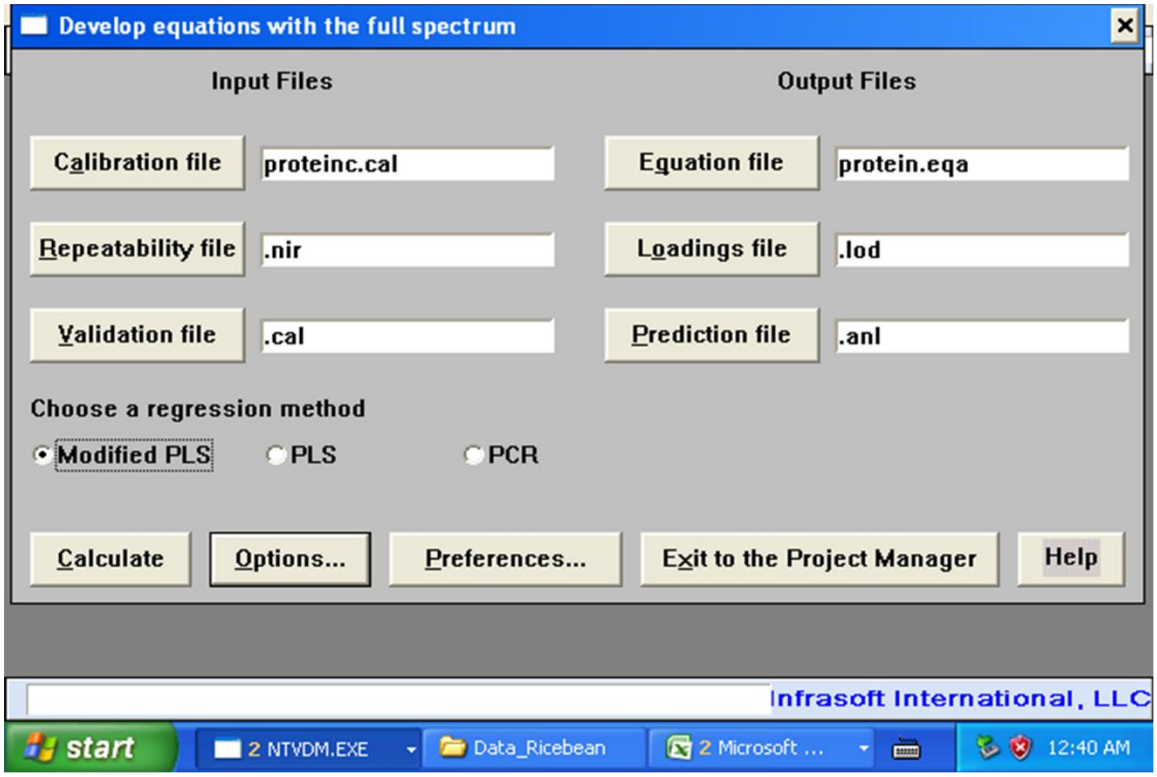

Fig. 8 Selection of regression method (here we chose MPLS method).

- ✓ In MPLS, compute and standardize residuals at each wavelength to minimize the impact of irrelevant spectroscopic variations, maximizing the capture of variation in spectroscopic data by incorporating reference values.
- ✓ Perform cross-validation to prevent overfitting, using SNV and detrend scatter correction techniques.
- ✓ Evaluate statistical parameters such as range, standard deviation (SD), standard error of calibration (SEC), and coefficient of determination in internal validation (RSQ internal) using Win ISI Project Manager Software version 1.61. Aim for lower SEC and higher RSQ values for superior models (Fig. 9).

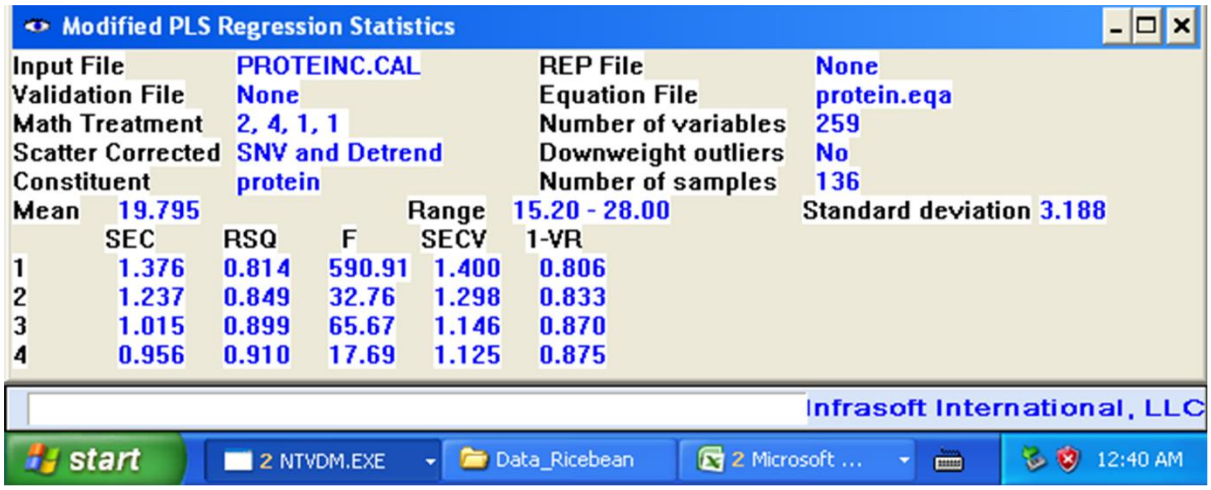

Fig. 9 MPLS Regression statistics of 136 samples of rice bean germplasm using SNV+DT with mathematical treatment of 2,4,4,1.

- ✓ Further evaluate the models using the standard error of cross-validation (SEC(V)) and 1 minus variance ratio (1-VR) to assess error and cross-validation accuracy.
- ✓ Refine mathematical processing treatments through trial and error to minimize SEC(V) and maximize 1-VR during cross-validation. Adjust the combination of derivatives, gaps, and smoothing processes to achieve the best calibration model.

The detailed statistics of the best calibrated model for rice bean protein content are summarized in Table 1.

Table 1 Statistics of laboratory values and calibration set of protein content in rice bean germplasm.

Trait	Laboratory values			Calibration of NIRS-model							
	Range	Mean	SD	Range	Mean	SD	Math Treatment	SEC	SEC(V)	$RSQ_{internal}$	1-VR
Protein	15.3-28.1	20.07	3.37	15.2-28.0	19.87	3.23	2,4,4,1	0.943	1.136	0.915	0.875

*SD= Standard deviation; RSQ (internal)= coefficient of determination in internal validation; SEC= standard error of calibration; SEC(V)= standard error of cross-validation; 1-VR= complement of variance.

Step 5: External validation of the model and parameters to assess the robustness of the developed models

- ✓ Compare the validation file (having independent dataset) with equation file to compute the prediction values for protein content in independent samples (Fig. 10).

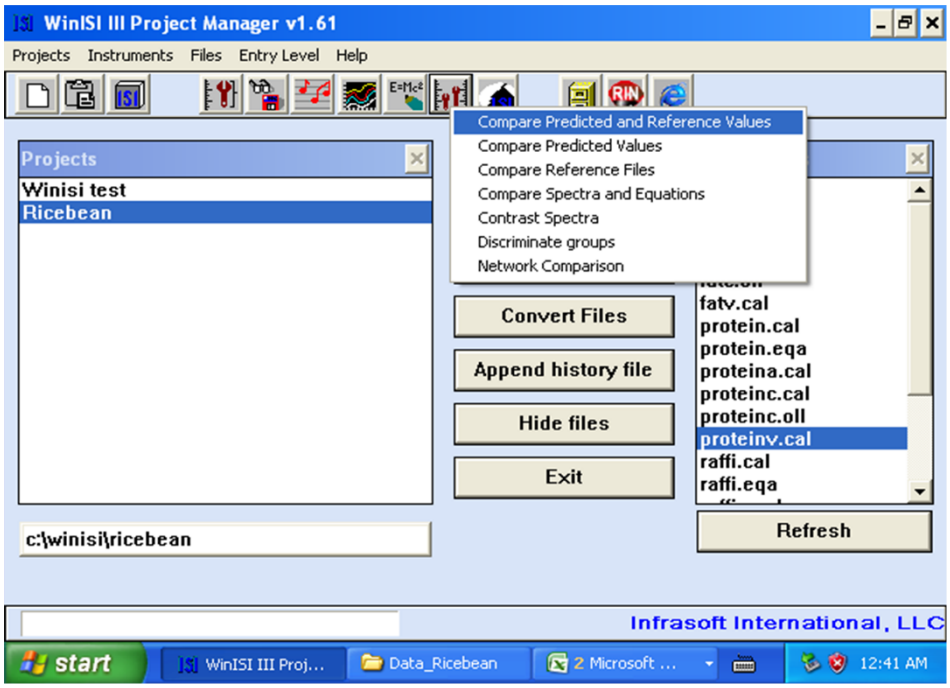

Fig. 10 Validation of the developed model by comparing predicted and reference values.

- ✓ The performance evaluation or validation of the calibrated equation is conducted using comprehensive global statistical metrics, including RSQexternal (Coefficient of determination for external validation), bias, SEP(C) (Corrected standard error of prediction), and RPD (Residual predicted deviation).
- ✓ You may choose the following metrics to confirm the robustness of your developed models.

$$RSQ_{internal/external} = \frac{\sum(y_{calculated} - y_{predicted})^2}{\sum(y_{calculated} - y_{mean})^2} \quad Eq.(1)$$

$$Bias\ (b) = \frac{1}{n}\sqrt{\sum(y_{predicted} - y_{Calculated})^2} \quad Eq.(2)$$

$$SEP\ (C)\sqrt{\frac{\sum(y_{predicted} - y_{calculated} - b)^2}{n}} \quad Eq.(3)$$

$$RPD = \frac{SD}{SEP(C)} \quad Eq.(4)$$

Here,

$RSQ_{internal/external}$: Coefficient of determination for calibration (internal) and validation (external) sample set.

$Y_{calculated}$: The actual evaluated value of the parameter

$Y_{predicted}$: The predicted value of the parameter through a regression line

Y_{mean} : Arithmetic mean of y values

n : Number of samples

SEP (C): Corrected standard error of prediction

SD: Standard deviation

RPD: Residual predicted deviation

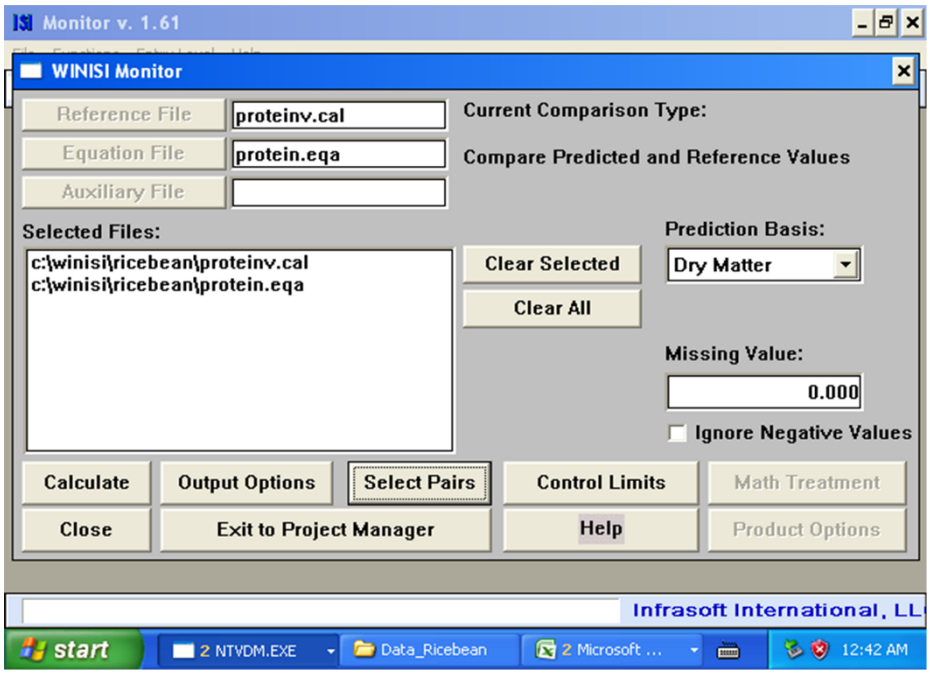

Fig. 11 Comparing reference file (containing independent dataset) with the developed equation.

Step 6: Selection of best-fit model based on global metrics

- ✓ The coefficient of determination (R-squared or RSQ) measures the proximity of the fitted regression line to the actual data by comparing the sum of squared differences between the predicted and observed values to the sum of squared differences between the observed values and their mean. A higher R-squared value, approaching 1,

signifies greater accuracy in how the regression model aligns with the actual data values.

- ✓ The slope signifies how predicted values change in response to a one-unit shift in reference values. An ideal slope is 1, and values approaching this indicate model accuracy.
- ✓ Bias, measured through the mean squared difference between predicted and calculated values, serves as a crucial metric for evaluating the agreement between reference and predicted values. An ideal scenario is characterized by a bias value of zero, indicating perfect alignment between the reference and predicted values. Positive bias values signify model overestimation, while negative values suggest model underestimation of the evaluated parameter.
- ✓ Corrected standard error of prediction (SEP(C)) is calculated using the square root of the mean squared difference between the predicted and calculated values. A smaller SEP indicates that the predicted values are closer to the calculated values on average.
- ✓ Residual prediction deviation (RPD) expresses the relationship between the variability in the observed data (measured by the standard deviation) and the average magnitude of errors in predictions (measured by the corrected standard error of prediction). This provides a matric to authenticate the validity of the model, which is more precise than SEP(C) and can be easily compared across various model validation studies. If RPD values are:

i) RPD < 1.5; the model is not reliable

ii) RPD lies in between 1.5-2.0, it indicates the capacity of a model to differentiate high and low values.

iii) RPD lies in between 2-2.5, it indicates approximate quantitative prediction.

iv) RPD lies in between 2.5-3.0, it indicates good quality prediction.

v) RPD > 3.0, it indicates excellent prediction.

WINISI project manager software is required to calculate all the above-mentioned statistical parameters. The models which have high RSQ, low SEP(C), nearly ideal slope, zero bias, and high RPD values are superior (Table 2). Scatter plot of reference and predicted values for protein content (%) is given in Fig. 13.

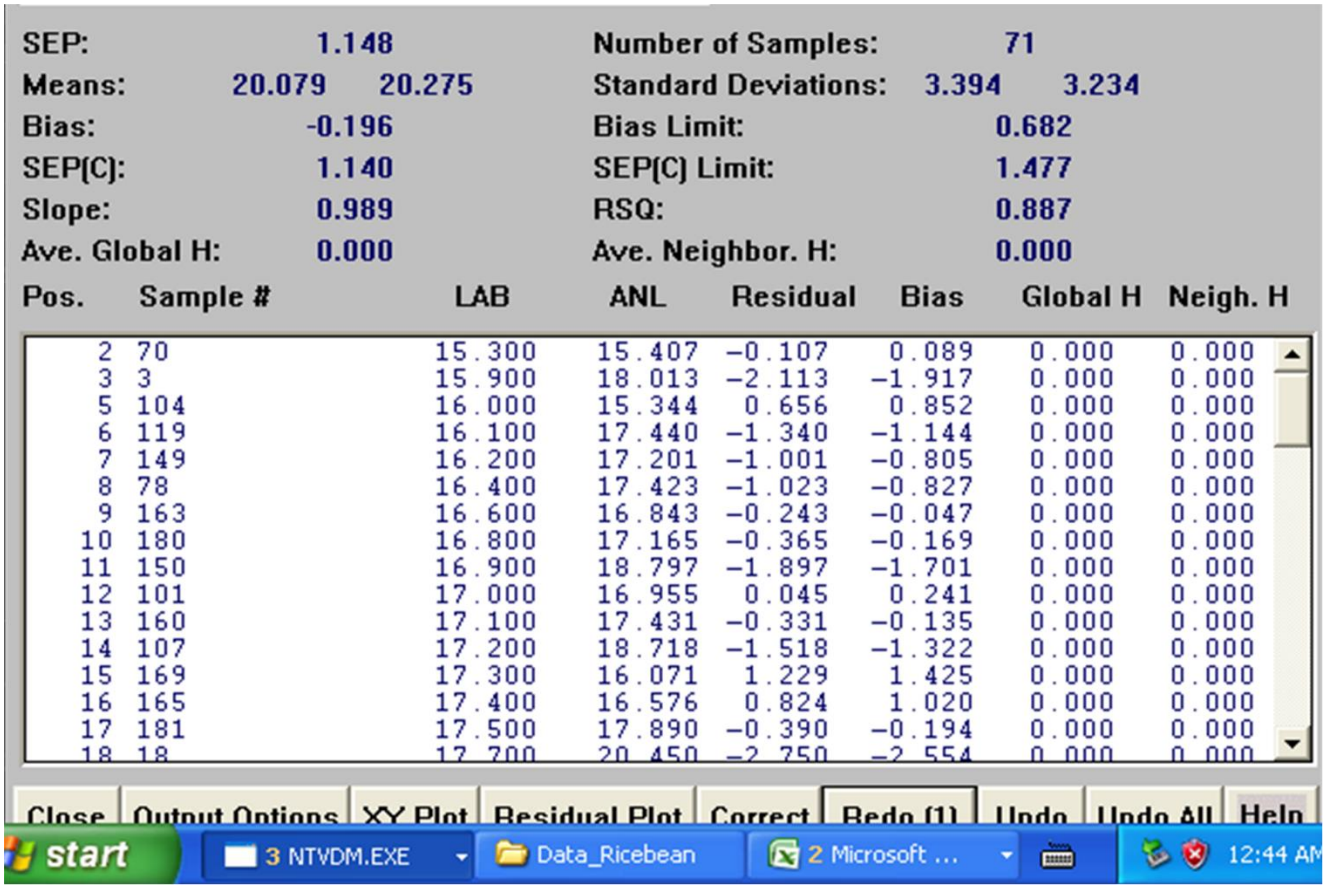

Fig. 12 NIRS-based model statistics for external validation set of rice bean germplasm

Table 2 NIRS-based model statistics for external validation set of rice bean germplasm

Trait	N	$RSQ_{external}$	Slope	Bias	SEP(C)	RPD
Protein content	71	0.887	0.989	-0.196	1.140	2.83

*N= Number of samples; RSQ external = coefficient of determination for external validation; SEP(C) = corrected standard error of prediction; RPD = residual prediction deviation

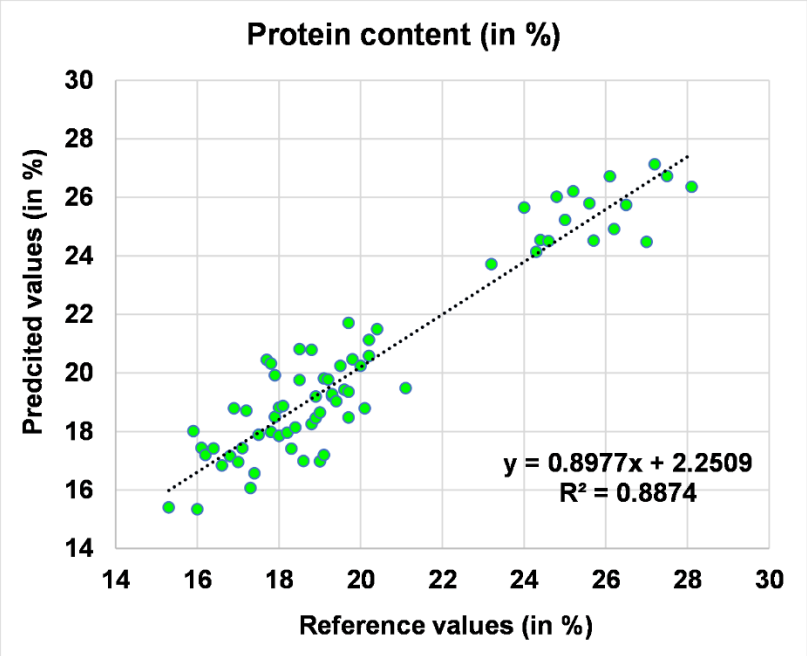

Fig. 13 Scatter plot of reference and predicted values for protein content (%) in rice bean.

Step 7: Statistical analysis

- ✓ **Reference data (protein content in this lesson) generation:** Conduct all analyses in triplicate and report results on a dry weight basis. Use suitable standards and reagent blanks to ensure accuracy during method validation and recovery checks for protein
- ✓ **Statistical analysis**: Perform a one-way analysis of variance (ANOVA) to evaluate the differences between groups. And you may use Duncan's test to describe means with 95% confidence ($p < 0.05$).
- ✓ **Calibration and prediction models:** Utilize WinISI III Project Manager software version 1.50 for calibrations and predictions. And apply various mathematical treatments to both spectral and analyzed data to enhance model accuracy.
- ✓ **Prediction accuracy assessment:** Use a paired t-test at a 95% confidence interval to assess the prediction accuracy of all developed models. You may use IBM SPSS Statistics 21 for this statistical analysis. The resulting p-value exceeded the significance threshold of 0.05, suggesting the robust accuracy and reliability of the model (Table 3, Fig. 14).

Table 3. Paired sample t-test at 95% confidence interval.

Comparison	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Reference vs Predicted values of protein content	-.19596	1.13960	.13525	-.46570	.07378	-1.449	70	.152

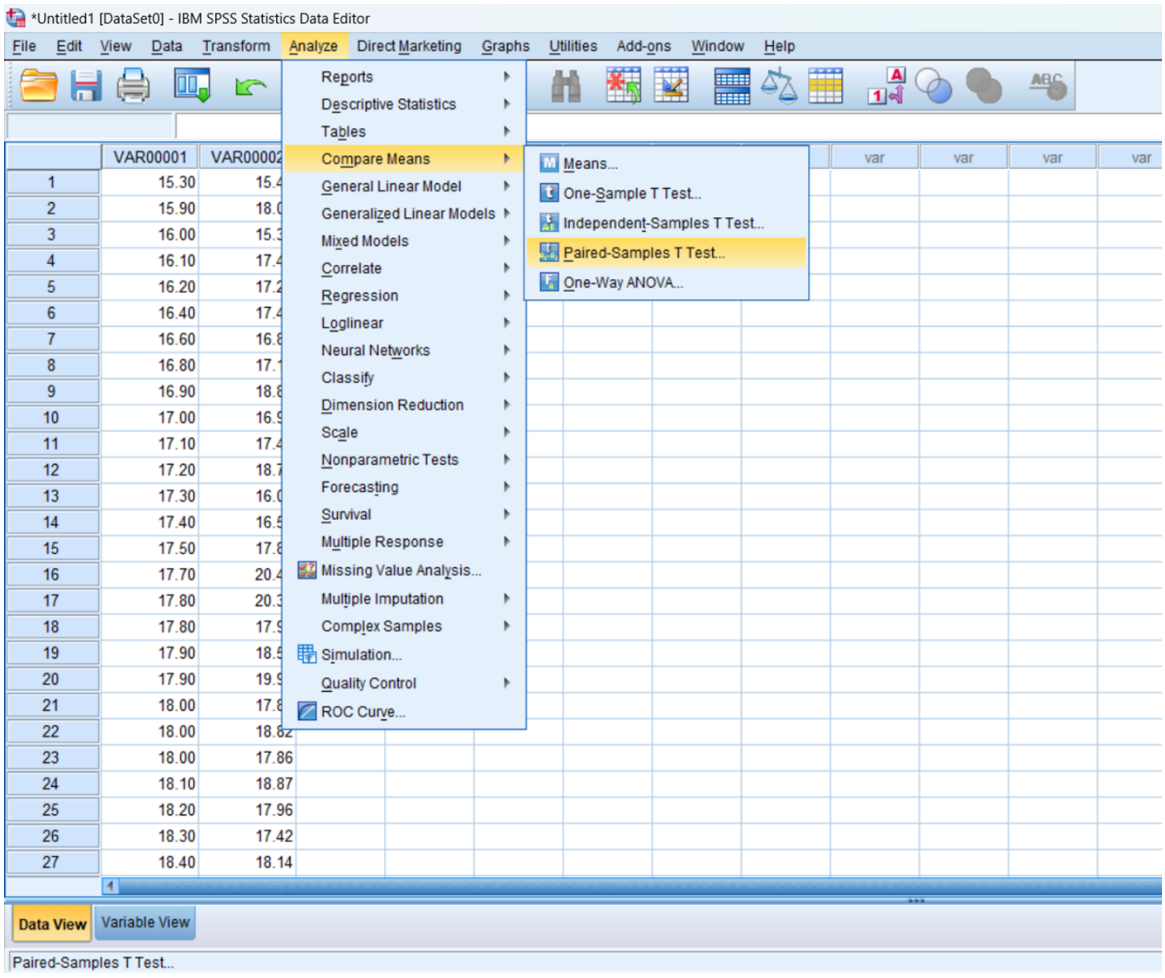

Fig. 14 Paired sample t-test at 95% confidence interval analysis using IBM SPSS software.

- ✓ **Reliability and correlation analysis:** Conduct a stringent parallel reliability (unbiased) test using IBM SPSS Statistics 21. Perform correlation analysis between reference and predicted values of biochemical traits to validate the model's accuracy.
- ✓ For instance, the scale demonstrated a high reliability with a coefficient of 0.970. Additionally, the correlation between the reference and predicted values was 0.942, indicating the models' strong accuracy in predicting protein content (Fig. 15).

.

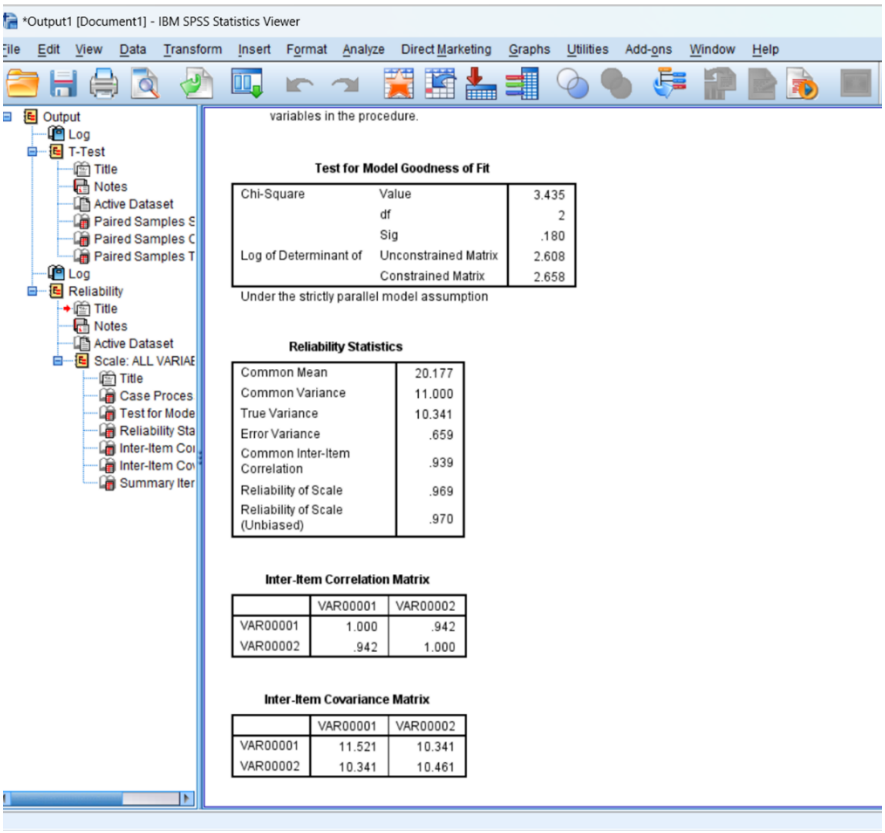

Fig.15 Reliability and correlation analysis between reference and predicted value of protein content using IBM SPSS software.

5. Conclusion

The developed model characterized by its high accuracy, precision, and reliability provide an innovative and rapid approach for screening protein-rich large rice bean germplasm present in global gene banks. By accelerating the identification of specific, desirable germplasms and the elimination of less favorable ones, NIRS-based models facilitate the selection of optimal chemotypes across diverse genetic backgrounds. This significantly enhances global crop

improvement programs, contributing to the development of nutritionally superior and resilient crop varieties.

6. Acknowledgments

The authors are thankful to ICAR RC for NEH Region, Umiam, Meghalaya, India and ICAR NBPGR, New Delhi, India for providing the necessary facilities and funding for conducting NIRS-based prediction modeling in diverse crop germplasm under in-house projects of all the authors of this module.

Suggested readings

Bartwal, A., John, R., Padhi, S. R., Suneja, P., Bhardwaj, R., Wankhede, D. P., & Archak, S. (2023). NIR spectra processing for developing efficient protein prediction model in mungbean. Journal of Food Composition and Analysis, 116, 105087. https://doi.org/10.1016/j.jfca.2022.105087

Chaukhande, P., Luthra, S. K., Patel, R. N., Padhi, S. R., Mankar, P., Mangal, M., Ranjan, J. K., Solanke, A. U., Mishra, G. P., Mishra, D. C., Singh, B., Bhardwaj, R., Tomar, B. S., & Riar, A. S. (2024). Development and validation of near-infrared reflectance spectroscopy prediction modeling for the rapid estimation of biochemical traits in potato. Foods, 13(11), 1655. https://doi.org/10.3390/foods13111655

Field, A. (2013). Discovering statistics using IBM SPSS statistics. SAGE. Retrieved from http://repo.darmajaya.ac.id/5678/1/Discovering%20Statistics%20Using%20IBM%20SPSS%20Statistics%20%28%20PDFDrive%20%29.pdf

Fontaine, J., Schirmer, B., & Horr, J. (2002). Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran, and sorghum. Journal of Agricultural and Food Chemistry, 50(15), 3902-3911. https://doi.org/10.1021/jf011637k

John, R., Bartwal, A., Jeyaseelan, C., Sharma, P., Ananthan, R., Singh, A. K., Singh, M., Gayacharan, Rana, J. C., & Bhardwaj, R. (2023). Rice bean-adzuki bean multitrait near infrared reflectance spectroscopy prediction model: A rapid mining tool for trait-specific germplasm. Frontiers in Nutrition, 10, 1224955. https://doi.org/10.3389/fnut.2023.1224955

John, R., Bhardwaj, R., Jeyaseelan, C., Bollinedi, H., Singh, N., Harish, G. D., Singh, R., Nath, D. J., Arya, M., Sharma, D., & Singh, S. (2022). Germplasm variability-assisted near infrared reflectance spectroscopy chemometrics to develop multi-trait robust prediction models in rice. Frontiers in Nutrition. https://doi.org/10.3389/fnut.2022.1001551

Kamboj, U., Guha, P., & Mishra, S. (2016). Characterization of chickpea flour by near infrared spectroscopy and chemometrics. Analytical Letters, 50(11), 1754-1766. https://doi.org/10.1080/00032719.2016.1247163

Kaur, S., Singh, N., Dagar, P., Kumar, A., Jaiswal, S., Singh, B K., Bhardwaj, R., Rana, J C., and Riar, A. (2024). Comparative analysis of modified partial least squares regression and hybrid deep learning models for predicting protein content in Perilla (Perilla frutescens L.) seed meal using NIR spectroscopy. Food Bioscience, 61, 104821. https://doi.org/10.1016/j.fbio.2024.104821.

Kondal, V., Jain, A., Garg, M., Kumar, S., Singh, A. K., Bhardwaj, R., & Singh, G. P. (2024). Gap derivative optimization for modeling wheat grain protein using near-infrared transmission spectroscopy. Cereal Chemistry. https://doi.org/10.1002/cche.10795

Padhi, S. R., John, R., Bartwal, A., Tripathi, K., Gupta, K., Wankhede, D. P., Mishra, G. P., Kumar, S., Rana, J. C., Riar, A., & Bhardwaj, R. (2022). Development and optimization of NIRS prediction models for simultaneous multi-trait assessment in diverse cowpea germplasm. Frontiers in Nutrition, 9, 1001551. https://doi.org/10.3389/fnut.2022.1001551

Padhi, S. R., John, R., Tripathi, K., Wankhede, D. P., Joshi, T., Rana, J. C., Riar, A., & Bhardwaj, R. (2024). A comparison of spectral pre-processing methods and their effects on nutritional traits in cowpea germplasm. Legume Science, 6(2), e2977. https://doi.org/10.1002/leg3.229

Priyadarshi, M. B., Sharma, A., Chaturvedi, K. K., Bhardwaj, R., Lal, S. B., Farooqi, M. S., Kumar, S., Mishra, D. C., & Singh, M. (2023). Comparing various machine learning algorithms for sugar prediction in chickpea using near-infrared spectroscopy. Legume Research-An International Journal, 46(2), 251-256. https://doi.org/10.18805/LR-4884

Shukla, A., Rahman, S. S., Suneja, P., Yadav, R., Hussain, Z., Singh, R., Yadav, S. K., Rana, J. C., Yadav, S., & Bhardwaj, R. (2023). Developing an NIRS prediction model for oil, protein, amino acids and fatty acids in amaranth and buckwheat. Agriculture, 13(2), 469. https://doi.org/10.3390/agriculture13020469

Tomar, M., Bhardwaj, R., Kumar, M., Singh, S. P., Krishnan, V., Kansal, R., Verma, R., Yadav, V. K., Ahlawat, S. P., Rana, J. C., & Satyavathi, C. T. (2021). Development of NIR spectroscopy-based prediction models for nutritional profiling of pearl millet (Pennisetum glaucum (L.) R. Br): A chemometrics approach. LWT, 111813. https://doi.org/10.1016/j.lwt.2020.111813

****

For any queries:

Please feel free to reach out at:

Simardeep.kaur@icar.gov.in

Naseeb.singh@icar.gov.in

Rakesh.bhardwaj1@icar.gov.in