Turkish Journal of Agriculture - Food Science and Technology

Near Infrared Reflectance Spectroscopy and Multivariate Analyses for Fast and Non-Destructive Prediction of Corn Seed Germination

Nafiz Çeliktaş[1,a,*], Ömer Konuşkan[1,b]

[1]Departmant of Field Crops, Faculty of Agriculture, Mustafa Kemal University, 31034 Antakya-Hatay, Turkey
*Corresponding author

A R T I C L E I N F O	A B S T R A C T
	The application of near-infrared reflectance spectroscopy (NIRS) and multivariate analysis for determining the seed germination rate of corn genotypes was assessed. Seed samples about 90 gr belong to commercial and local corn varieties at various ages were scanned with FT-NIRS on the reflectance mode from 1000 to 2500 nm wavelength. Filter paper technique showed the seed germination rates varied between 18-100% depending on the genotypes after 7 days at ±25°C. Partial least squares regression (PLSR) was applied to the reference values corresponding to the spectra. The best statistical results obtained from the pre-treatment combinations of Smooth Savitzky-Golay 9 Points (sg9), MSC full and normalization to unit length (nle). The regression coefficient of calibration (R2C) and prediction (R2P) of the created NIRS calibration via chemometric software NIRCal are realized 0.97 and 0.98 respectively for the property of corn germination rate. The standard error of both calibration (SEC) and prediction (SEP) were almost overlapping (4.17%, 4.61% respectively). The prediction accuracy of the final NIRS model was quite reasonable with the acceptable root mean standard error of prediction (RMSEP) as 8.88%. According to the residual predictive deviation (RPD) index (4.18), the accuracy of the NIRS model regarded as in the best category. Therefore, the NIRS model developed here is sufficient to predict the corn seed germination rate very fast and non-destructively without using any regents.

[a] *nafizcel@hotmail.com* *https://orcid.org/0000-0002-0467-1034* [b] *okonuskan@mku.edu.tr* *https://orcid.org/0000-0003-1135-2346*

Introduction

In terms of planning the agricultural production and the input outputs, it is an absolute necessity to know the seed viability. Seed viability is significantly under control by both ecological factors and internal seed biochemical metabolism. Long-term storage in unfavorable conditions is an important factor in decreasing viability. International seed testing association (ISTA) sorts the classical seed viability or germination tests as tetrazolium, conductivity, immunoassay tests and germination tests (Zhang et al., 2018). However, all these techniques are destructive, requiring intensive labor and time. On the other hand, determining the physical properties of the seed may vary depending on the technicians performing the seed quality testing. In this case, the experience and the sensitivity of the technicians performing the assessments becoming more important. Therefore, there is a need for methods that are not dependent on the technician, laboratory conditions, sophisticated equipment and the chemicals (Rahman and Cho, 2016). New generation technologies developed for these purposes have become widespread in recent years.

These technologies provide rapid, nondestructive, low-cost and environment-friendly measurements since due to lack of chemical waste. Techniques such as machine vision, spectroscopy, hyperspectral imaging, X-ray imaging, and thermal imaging can quickly specify the seed quality as well as its chemical quality, genetic purity, the quantity of disease and pest infestation or classification of the pathogens (Huang et al., 2015). Near infrared reflectance spectroscopy (NIRS) which is one of the spectroscopic technique offers advantages in determining the qualitative and quantitative properties of organic materials (Youngentob et al., 2012). It is sensitive enough to detect the changes in the free and structure-bounded components within a sample (Salgó and Gergely, 2012). The measurements are based on the correlation among the molecular vibration of an organic molecule and their interactions with the amount of absorbed and reflected near infrared radiation by the chemical structure (Jin et al., 2017) containing OH, NH, CH and SH bounds within a sample (Qiu et al., 2018). The absorption bands are

characterized by very broad and overlapped peaks (Kosmowski and Worku, 2018), which inhibit the visual assessment and direct interpretation of the spectra (Büchi, 2013). The chemometric software is used to find a statistical correlation between the spectra and the known biochemical values of the relevant sample to evaluate the entwined spectral data. Thus, a large number of biochemical parameters of a very small amount of a sample can be predicted in a very short time without any sample preparation (Burns, 2008). The potential of NIR spectroscopy to predict the qualitative and quantitative properties in food, pharmaceutical, and agricultural industries has been investigated (Fagan et al., 2011).

Corn is the second most plentiful cereal grown for human consumption in worldwide. Knowledge of seed quality and germination rate is a prerequisite to ensure a successful crop establishment. At this stage, near-infrared spectroscopy is being a suitable technique for getting fast and cost-effective information. However, very few studies conducted with different species have tried to predict the seed viability and germination rate by using NIRS. Therefore, the main objective of this study was to exhibit the potential of using NIR spectroscopy combined with chemometrics to predict the seed germination rate of corn genotypes to reduce the need for wet chemical analysis and intensive labor and time for consumers and the seed industry.

Materials and Methods

Plant Material

Commercial and local corn (*Zea mays* L.) seed samples (14 of 18 samples belong to the same genotype of various ages) were used as the research material (Table 1).

Maize genotypes were sampled from a genetic collection stored for different periods from 3 months to 2 years under uncontrolled room conditions.

Table 1. Plant material and their origin

N	Corn varieties	Origin
1	Syperformer 1	Syngenta Turkey
2	Syperformer 3	""
3	Syperformer 4	""
4	Syperformer 5	""
5	Syperformer 6	""
6	Syperformer 7	""
7	Syperformer 11	""
8	Syperformer 12	""
9	Syperformer 13	""
10	Syperformer 14	""
11	Syperformer 15	""
12	Syperformer 17	""
13	Syperformer 29	""
14	Syperformer 30	""
15	Colored popcorn	Local (Çanakkale)
16	Popcorn breed line	Breed line
17	Karadeniz Hard Seed	Local (Trabzon)
18	Karadeniz Hard Seed (non-viable)	Local (Trabzon)

NIRS Analyses

To obtain the reflectance spectra, the unprocessed seed samples were scanned by using a fourier-transform near infrared (FT-NIR) spectrophotometer (NIRFlex N-500, BÜCHI Labortechnik AG, Switzerland) before germination test. The spectral measurements were implemented in glass Petri dishes (100×20 mm) containing about 90 g of corn seed. Each spectrum was obtained with the electromagnetic scan on the reflectance mode from 1000 to 2500 nm wavelength (4000-10000.cm^{-1}) with a spectral resolution of 4 cm^{-1}. All spectra were gathered after 32 scans with threefold repetition for each sample. Partial least squares regression (PLSR) which establishes a linear relationship between the spectral data and reference data (germination rate) obtained through a germination test was used to construct the calibration models. Principal components (PC) were used to create a multivariate discriminant model with the help of chemometric software NIRCal 5.5 (BÜCHI Labortechnik AG, Switzerland). Pre-treatments were employed to correct and reconstruct the spectrums during the data analysis. In the development of the NIRS models, 66% of the 54 spectra and the remaining 33% were used for the calibration and validation, respectively.

Reference Data

The germination rate (percentage) of the spectroscopically identified corn seed samples were measured according to ISTA (2008). Fifty seeds were placed on 3 layers of moist, non-toxic, germination paper (Anchor Paper Co., St. Paul, Minn.). The papers were rolled, placed in a plastic container (21.5 × 32.5 × 5.5cm), and incubated in an incubator at 25°C for 7 days in dark conditions. When the emerging radicle was at least 2 mm long that seed was considered as germinated. The seedling abnormalities at the final assessment were checked also.

Statistics

The performance of the calibration was expressed in terms of the coefficient of calibration and validation ($R^2{}_{CAL}$, $R^2{}_{VAL}$) and the standard errors of calibration (SEC) and prediction (SEP). In the validation step, the overall error between the predicted and reference values was assessed by using the root mean square error of prediction (RMSEP). Additionally, residual predictive deviation (RPD) which is the expression of the division of SEP to the standard deviation (Bellon-Maurel et al., 2010) was also calculated.

Result and Discussion

A significant variation was detected between corn genotypes in terms of germination rates after seven days (Figure 1).

The effect of the seed storage period, in other words, seed age on the germination rate was clear. The germination rate was detected between 18% and 100 % depending on the genotypes.

The genotype-specific spectrums obtained by the electromagnetic scanning of corn seed samples on the reflectance mode from 1000 to 2500 nm were differentiating from each other at different wavelength regions (Figure 2).
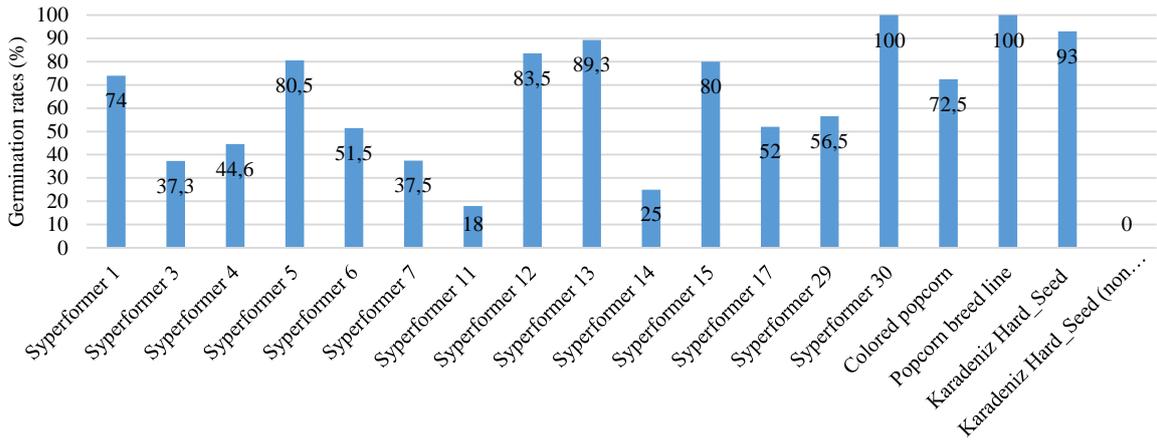
Figure 1. Germination rates (*reference values*) of corn genotypes after 7 days of incubation
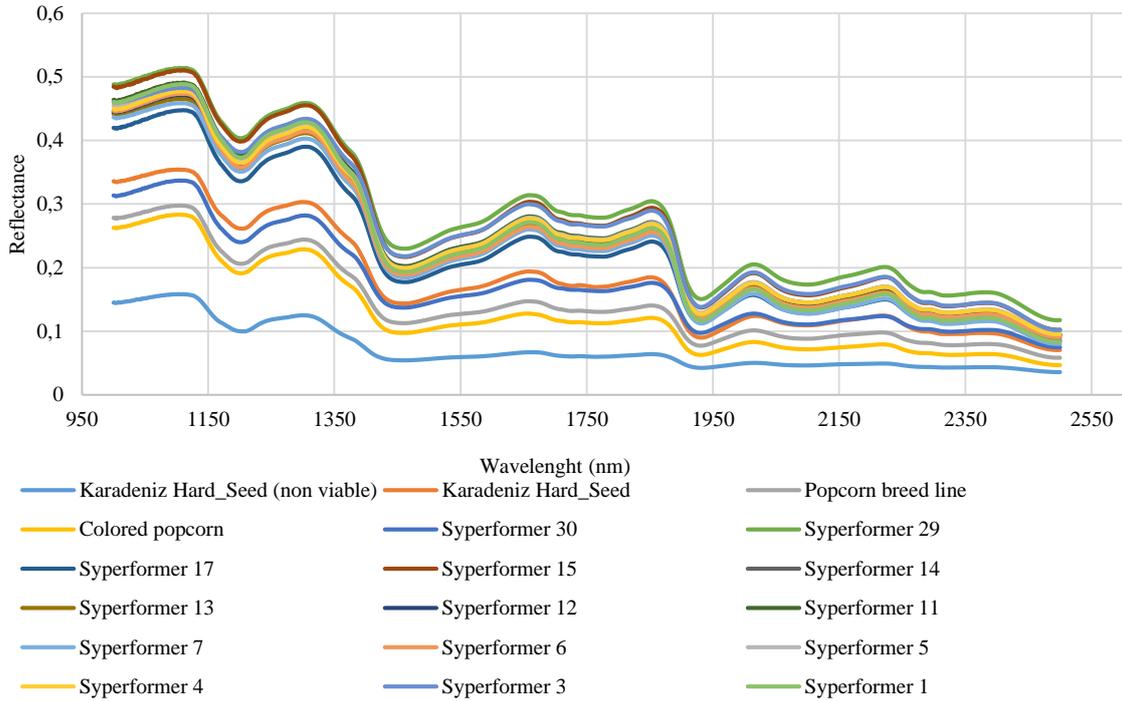

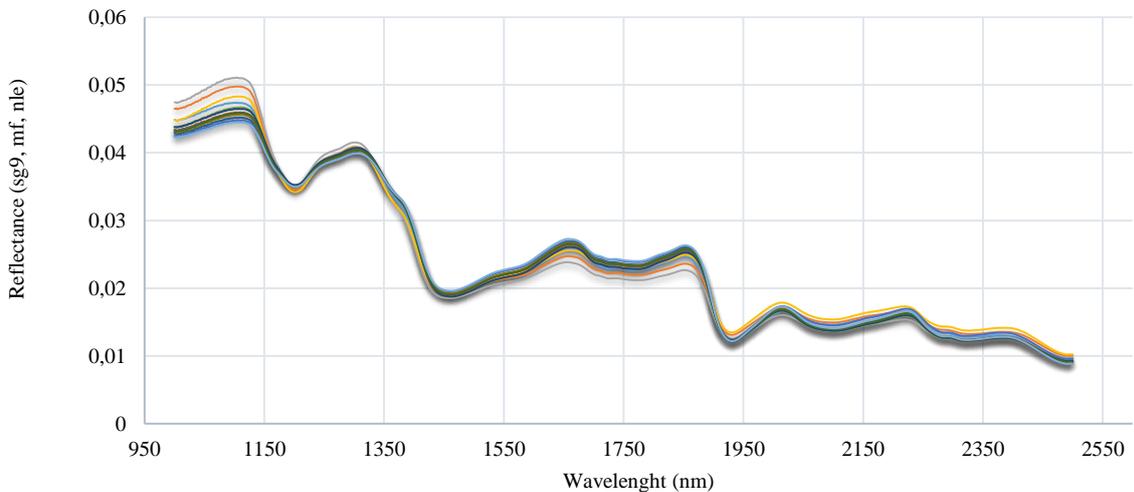Figure 2. The sample specific original spectrums

Figure 3. Statistically pretreated (sg9, mf and nle) spectrums

Non-viable seeds (0%) were obtained artificially by boiling and drying of the seeds to observe the differences of the non-viable seed spectrum. It is clearly visible on the bottom of Figure 2 with a differentiated spectrum.

Differences were detected between the repetitions of the spectral scanning performed for each genotype. For this reason, the original spectra graph is presented in Figure 2 as the average reflection of the three replicated spectral scans. Although not all wavelength regions in the spectrums contain information about seed germination, these regions were not excluded from the evaluation even being a noisy part and all wavelength regions were used for calibration. The non-viable sample spectra were eliminated from both calibration and validation sets during the NIRS model development process since it was determined as an outlier.

As a result of displacement, regulation and the improvements of the original spectrums, 61.5% of the total spectrum was used for calibration and 28.5% for validation, unlike the initial programming. The rest of the 10% spectrums were specified as outliers also. The spectral measurement can be affected by many factors leading to interference. So, the chemometric pretreatments are needed to optimize the performance of the model (Dupuy, 2010). The pretreatments never change the wavelength of the original spectrums or the property-related spectra regions. Different statistical pretreatments or their combinations were applied in the current study to the original spectrums. In the final NIRS model the best accuracy was obtained from the pretreatment combinations of smooth savitzky-Golay 9 Points (sg9) + MSC full and (mf) + normalization to unit length (nle). As a result of the statistical improvements applied to the raw spectra, the reflection graph converted as in Figure 3.

Figure 4 shows the relationship between the original versus predicted germination rates of corn genotypes in the final NIRS model.

The best results in terms of higher R^2 and lower prediction error of SEP were compared for the cross-validation model of PLSR. The descriptive statistical data associated with the NIR-based calibration and validation are given in Table 2.

In terms of the reliability of the NIRS model, the coefficient R^2 is expected to as possible as close to one. The average bias between calibration and validation data sets for the property of the germination rate was quite low, consequently resulting in relatively normal linear distribution and rather high coefficients were observed on the plots (Figure 4). On the other hand, the parameters of SEC and SEP should be almost overlapped with each other for a good prediction performance. In these aspects, the NIRS model developed for corn seed germination with this study may be considered as reliable. Al-Amery et al. (2018) reported that the performance statistics of R^2 and SEC of their NIRS model for the prediction of soybean seed germination rate as 0.57 and 11.42%. The authors separated the low and high germinating seed groups easily but had difficulty in distinguishing the genotypes from each other with their model. On the contrary, our results showed that the developed calibration could provide a realistic germination rate at a genotype level, with high accuracy. However, RMSEP is the most important and robust figure to judge the quality of a PLSR model since it

shows the random and systematic errors in the prediction as different from SEP (Ferreira et al., 2013). The RMSEP is actualized as 8.88% in this study. According to this, the model needs to be improved. Because this percentage can lead to a slight deviation on the planting density in the field conditions. RPD is a non-dimensional statistic for the evaluation of an NIR spectroscopy calibration model (Esbensen et al., 2014). Williams (2001) emphasizes that the usage of RPD with R^2 will be more meaningful for evaluating the efficiency of NIRS models. The RPD value in the PLS model was calculated as 4.18 for the germination rate. According to the classification of Chang et al. (2001) when the RPD>2 and measured vs predicted R^2 was between 0.80 and 1.0 it means that the NIRS prediction as in the best category. Therefore, the model provided with this study is quite reasonable for the prediction of corn seed germination rate.

Table 2. The performance statistics of the final NIRS model developed for the germination rates (%) of corn genotypes

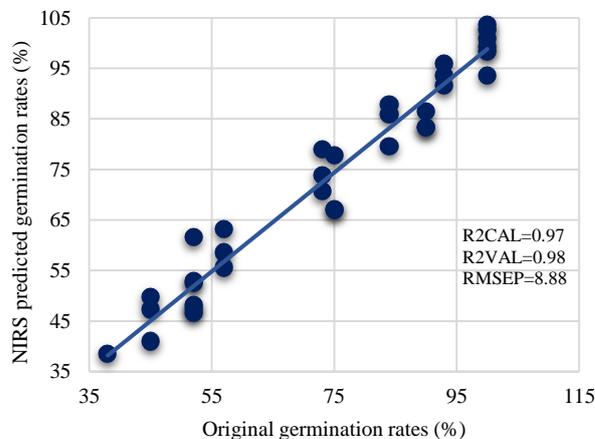| Property	Germination rate (%)
R^2_{CAL}	0.97
R^2_{VAL}	0.98
SEC (%)	4.17
SEP (%)	4.61
RMSEP (%)	8.88
RPD	4.18
Pred. Bias	-1.2560

Figure 4. Original vs NIRS predicted seed germination rates (%) of corn genotypes

Tigabu and Odén (2004) stated that they could separate the senescence in the seeds of the species *Pinus patula* by the PLS regression model. The researchers have been emphasized that the wavelength regions that belong to lipid and protein in the old and young seeds come into prominence with respect to that classification. Therefore, it is possible to say that the spectral difference in terms of seed germination rates is actually because of the result of the biochemical difference caused by senescence. Actually many studies conducted on seed senescence mentioned that the chromosomal disorders, DNA degradation, abnormalities on RNA and so protein synthesis, and the anomalies on the enzymatic activity as a result of seed

aging (Walters, 1998; McDonald, 1999). Significant reductions in total protein (Kalpana and Madhava, 1995), unsaturated fatty acids, total lipid, and phospholipid content were detected with seed senescence (Thapliyal and Conner, 1997). As a result of these biochemical abnormalities, seed germination performance, speed and uniformity decrease (Kalpana and Madhava, 1995). Therefore, the NIRS chemometry generates a discrimination model by using these biochemical

differentiation. The question is what the most effective parameter of this discrimination is. In general, the longer wavelength region had a larger weight and hence a larger influence on the discriminant model for any property (Tigabu and Odén, 2004). Esbensen (2009) stated that important reflectance bands have higher regression coefficients in a multivariate model. Therefore, the highest regression coefficients were found to be around the wavelengths of 1112 nm and 1231 nm (Figure 5).

Table 3. The important reflectance bands and the related chemical products on the prediction of corn seed germination rates

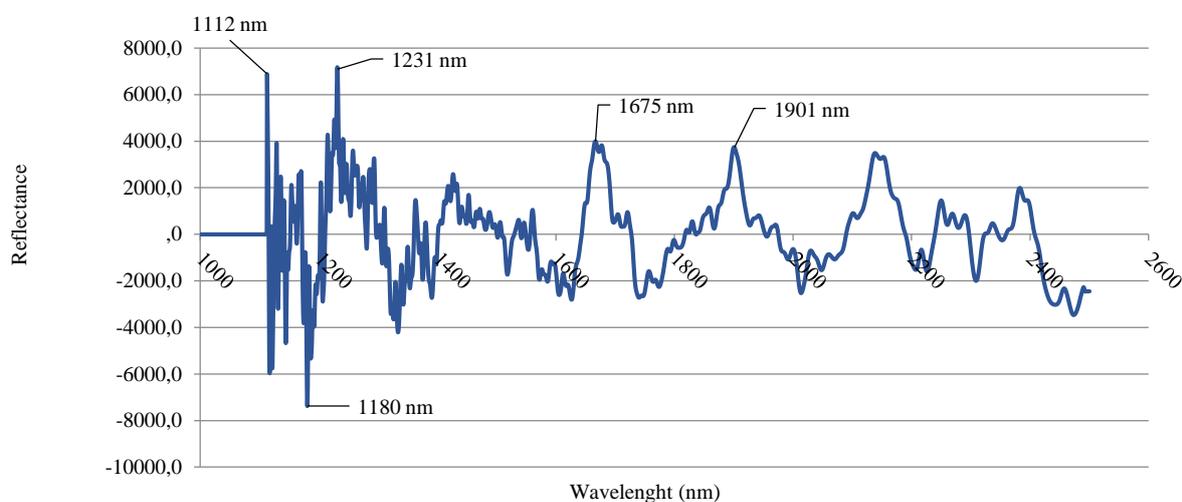| Wavelength (nm)	Chemical bonds	Product Example
1112 nm	2x C-H str.+2x C-C str.	Benzene, cyclopropane,
1180 nm	C-H str. second overtone	CH3, HC=CH
1231 nm	C-H str. second overtone	CH3, HC=CH
1675 nm	C-H str. first overtone	cis-RCH=CHR, aromatic, CH3
1901 nm	O-H str.+2x C-O str. C=O str. second overtone	Cellulose, starch , -CO2H

Figure 5. The important wavebands for corn seed germination rates
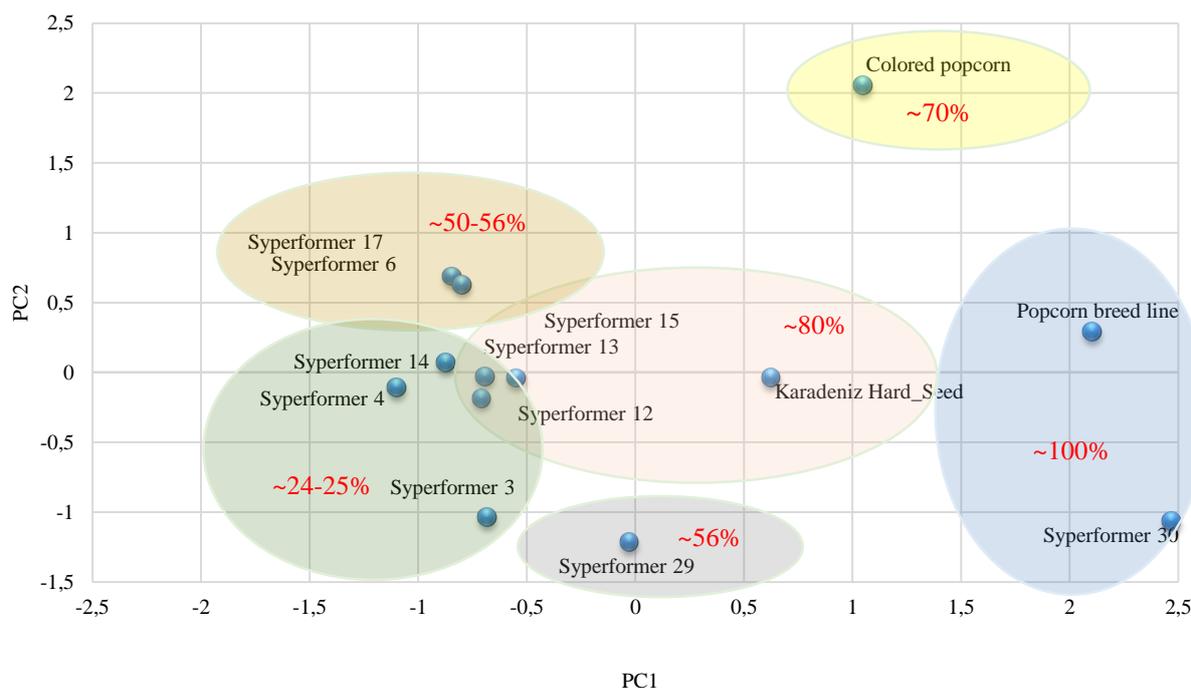

Figure 6. Score plots of germination rates using the FT-NIRS data

The important wavelengths on the prediction of germination rates, their chemical bonds, and the related products were explained in Table 3.

Principal component analyses (PCA) transforms the large data sets into relatively few meaningful dimensions that represent the majority of the information present in the original data. The primary principal component (PC1) is used to reconstruct the spectra, and the second primary component (PC2) is used to decompose the models formed from the data set. According to the score plot of the PC1 and PC2 (Figure 6), the PLSR model was able to differentiate corn genotypes very well by the first two-component in respect to their germination rates.

As shown in Figure 6 the PC1 and PC2 successfully differentiated the low and high germinating corn genotypes but failed to decompose the genotypes that are medium germinating (50-60%).

Although the performance indicators of the developed NIRS model provided sufficient information about the calibration accuracy, the model should be re-tested with non-calibration samples (external) and the statistical relationship between the actual values and the NIRS predictions should be demonstrated. The lack of clear distribution in PCA also led to some deviations in the external validation test with a developed NIRS model (Figure 7). Despite these deviations, the regression coefficient (R^2) between the laboratory and the predicted germination rates actualized as 0.79. This is a statistically acceptable coefficient and it shows that the accuracy of the developed calibration is quite high.
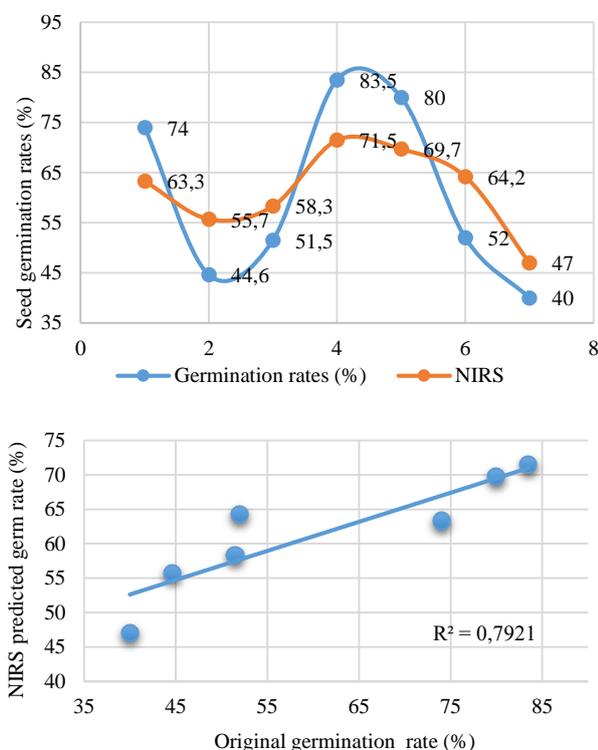

Figure 7. Functionality of NIRS calibration developed for corn seed germination rate

Conclusion

As a conclusion, the developed NIRS model for the germination rates of corn genotypes is rather promising for very fast screening of a huge sample set without using any reagents. It is a preferable technique, especially for a very small amount of seed material since the spectral prediction method does not destroy the seed structural integrity. However, in order to increase the reliability of the NIRS model for different species and for commercial purposes, it is necessary to generate the original spectrum and reference values from very different species.

Acknowledgement

References

Al-Amery M, Geneve RL, Sanches MF, Armstrong PR, Maghirang EB, Lee C, Hildebrand DF. 2018. Near-infrared spectroscopy used to predict soybean seed germination and vigour. Seed Science Research, 1-8. doi:10.1017/s0960258518000119.

Bellon-Maurel V, Fernandez-Ahumada E, Palagos P, Roger JM, McBratney AB. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends in Analytical Chemistry, 29 (9): 1073-1081. doi:10.1016/j.trac.2010.05.006

Burns D, Ciurczak E. 2008. Handbook of Near-Infrared Analysis. 3rd Ed. CRC Press. Boca Raton, FL. ISBN 9780849373930

Büchi 2013. Operation manual, NIRCal 5.5, Version A. BÜCHI Labortechnik AG, Flawil, Switzerland.

Chang CW, Laird DA, Mausbach MJ, Hurburgh CR. 2001. Near-infrared Reflectance Spectroscopy - Principal Components Regression Analyses of Soil Properties. Soil Science Society of America Journal, 65: 480-490. doi:10.2136/sssaj2001.652480x.

Dupuy N, Galtier O, Dréau Y Le, Pinatel C, Kister J, Artaud J. 2010. Chemometric Analysis of Combined NIR and MIR Spectra to Characterize French Olives. European Journal of Lipid Science and Technology, 112 (4): 463-475. doi:10.1002/ejlt.200900198

Esbensen KH. 2009. Multivariate Data Analysis – In Practice. 5th Edition. Camo Software AS, Oslo, Norway. ISBN 82-993330-3-2.

Esbensen KH, Geladi P, Larsen A. 2014. The RPD Myth. NIR News, 25(5): 24–28. doi:10.1255/nirn.1462

Fagan CC, Everard CD, McDonnell K. 2011. Prediction of moisture, calorific value, ash and carbon content of two dedicated bioenergy crops using near-infrared spectroscopy. Bioresource Technology 102: 5200–5206. doi: 10.1016/j.biortech.2011.01.087

Ferreira MH, Braga Jez WB, Sena MM. 2013. Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. Microchemical Journal, 109: 158–164. doi:10.1016/j.microc.2012.03.008

Huang M, Wang QG, Zhu QB, Qin JW, Huang G. 2015. Review of seed quality and safety tests using optical sensing technologies. Seed Science & Technology, 43: 337-366. doi:10.15258/sst.2015.43.3.16

ISTA, 2008, International Rules for Seed Testing. International Seed Testing Association, Bassersdorf. Switzerland. ISSN: ISBN 0251-0952.

Jin X, Chen X, Xiao L, Shi C, Chen L, Yu B, Yi Z, Yoo JH, Heo K, Yu CY, Yamada T, SacksEJ, Peng J. 2017. Application of visible and near-infrared spectroscopy to classification of *Miscanthus* species. PLoS ONE, 12(4): e0171360. doi:10.1371/ journal.pone.0171360

Kalpana R, Madhava Roa KV. 1995. On the ageing mechanism in pigeonpea (*Cajanus cajan* (L) Millsp.) seeds. Seed Science & Technology, 23: 1–9.

Kosmowski F, Worku T. 2018. Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. PLoS ONE, 13(3): e0193620. doi:10.1371/journal.pone.0193620

McDonald MB. 1999. Seed deterioration: physiological, repair and assessment. Seed Science & Technology, 27: 177–237.

Qiu G, Lü E, Lu H, Xu S, Zeng F, Shui Q. 2018. Single-Kernel FT-NIR Spectroscopy for Detecting Supersweet Corn (*Zea mays* L. Saccharata Sturt) Seed Viability with Multivariate Data Analysis. *Sensors,* 18(4): 1010. doi:10.3390/s18041010

Rahman A, Cho B. 2016. Assessment of seed quality using non-destructive measurement techniques: A review. Seed Science Research, 26(4): 285-305. doi:10.1017/S0960258516000234

Salgó A, Gergely S. 2012. Analysis of wheat grain development using NIR spectroscopy. Journal of Cereal Science, 56: 31-38. doi:10.1016/j.jcs.2012.04.011

Thapliyal RC, Connor KF. 1997. Effect of accelerated ageing on viability, leachate exudation and fatty acid content of *Dalbergia sissoo* Roxb. Seeds. Seed Science & Technology, 25: 31–319.

Tigabu M, Odén PC. 2004. Rapid and non-destructive analysis of vigour of *Pinus patula* seeds using single seed near infrared transmittance spectra and multivariate analysis. Seed Science and Technology, 32(2): 593-606. doi:10.15258/ sst.2004. 32.2.28.

Walters C. 1998. Understanding the mechanisms and kinetics of seed ageing. Seed Science Research, 8: 223-244. doi:10.1017/ S096025850000413X

Williams PC. 2001. Implementation of Near-Infrared Technology. In: Williams P.C., Norris K., editors. Near-Infrared Technology in the Agricultural and Food Industries. 2nd ed. American Association of Cereal Chemists; St. Paul, MN, USA. ISBN-10: 1891127241

Youngentob KN, Renzullo LJ, Held AA, Jia XP, Lindenmayer DB, Foley WJ. 2012. Using imaging spectroscopy to estimate integrated measures of foliage nutritional quality. Methods Ecol. Evol., 3: 416-426. doi:10.1111/j.2041-210X.2011. 00149.x

Zhang T, W Wei, B Zhao, R Wang, M Li, L Yang, J Wang, Q Sun. 2018. A Reliable Methodology for Determining Seed Viability by Using Hyperspectral Data from Two Sides of Wheat Seeds. *Sensors*, 18 (3): 813. doi:10.3390/s18030813.