Mirosław CZECHLOWSKI, Tomasz WOJCIECHOWSKI
Uniwersytet Przyrodniczy w Poznaniu, Instytut Inżynierii Biosystemów
ul. Wojska Polskiego 28, 60-637 Poznań, Poland
e-mail: mczech@up.poznan.pl

THE UTILIZATION OF INFORMATION ABOUT LOCAL VARIABLE ENVIRONMENTAL CONDITIONS TO PREDICT THE QUALITY OF WHEAT GRAIN DURING THE HARVEST

Summary

The study presents the correlation between the quality of winter wheat grain, understood as the content of protein, and the parameters of harvested grain (moisture and yield) and locally variable environmental conditions (relative altitude, the content of total Kjeldahl nitrogen, exchangeable phosphorus and potassium, magnesium, the pH coefficient, the content of organic matter in soil). The results obtained on the basis of the data collected from 5 production fields of the total area of 112 ha give grounds for the conclusion that by application of multiple regression it is possible to construct a relatively precise model for prediction of the content of protein in wheat grain even on the basis of the measurement of easily obtainable information about the relative altitude and yield. However, the effectiveness of the model will be limited to a small field area. The construction of universal model using information about locally variable environmental conditions is difficult due to the strong variability of the correlation between the analysed traits describing environmental conditions and the content of protein in wheat grain.
The study was a part of the development project No. R12 0073 06 entitled "Development and validation of the technology for separation grain stream during cereals selective harvesting ", financed by the Polish Ministry of Science and Higher Education.
***Key words**: grain quality, multiple regression, environmental conditions, selective grain harvest, VIS-NIR spectroscopy*

WYKORZYSTANIE INFORMACJI O LOKALNIE ZMIENNYCH WARUNKACH ŚRODOWISKOWYCH W CELU PRZEWIDYWANIA JAKOŚCI ZIARNA PSZENICY PODCZAS ZBIORU

Streszczenie

W pracy przedstawiono zależności pomiędzy jakością ziarna pszenicy ozimej, rozumianą jako zawartość białka, a parametrami zbieranego ziarna (wilgotność i wielkości plonu) oraz lokalnie zmiennymi warunkami środowiskowymi (względna wysokość n.p.m., zawartość azotu ogólnego, wymiennego fosforu i potasu, oraz magnezu, współczynnik pH, zawartość materii organicznej w glebie). Wyniki uzyskane w oparciu o dane zgromadzone na 5 produkcyjnych polach o łącznej powierzchni 112,78 ha pozwalają stwierdzić, że stosując regresję wieloraką można zbudować stosunkowo dokładny model do predykcji zawartości białka w ziarnie pszenicy nawet w oparciu o pomiar łatwych do pozyskania informacji o względnej wysokości n.p.m. i wielkości polonu, jednak jego skuteczność będzie ograniczona do niewielkiego obszaru powierzchni. Budowa uniwersalnego modelu wykorzystującego informacje o lokalnie zmiennych warunkach środowiskowych jest utrudniona ze względu na silną zmienność zależności pomiędzy analizowanymi cechami opisującymi warunki środowiskowe, a zawartością białka w ziarnie pszenicy.
Pracę zrealizowano w ramach projektu rozwojowego nr R12 0073 06 pt: „Opracowanie i walidacja technologii rozdziału strumienia ziarna podczas selektywnego zbioru zbóż" finansowanego przez MNiSW
***Słowa kluczowe**: jakość ziarna, regresja wieloraka, warunki środowiskowe, selektywny zbiór zbóż, spektroskopia VIS-NIR*

1. Introduction

The process of combine harvesting of crops is a basic and widely applied method of harvesting cereals in large-area farmlands [2]. During harvest all threshed grain, regardless of its quality, comes into one grain container. However, as is widely known, the farmland area may be considerably diversified in terms of soil abundance in nutrients or moisture [3, 4, 5, 10, 14] and it may be characterised by individual landform, which influences the quality and quantity of crops. Usually the highest content of protein can be observed in terrains located at a high level, whereas the yield is higher in terrains located at lower levels [3, 9].

So far research projects have usually been limited to monitoring and recording the content of protein in harvested grain [7, 8, 15]. Until recently attempts to divide the stream of grain were only made in stationary conditions in granary [16]. However, in this solution the considerable distance between the place of measurement and the place of grain harvesting does not guarantee effective division due to the multiple mixing of grain during harvest, reloading and transport. As a result, the quality variance observed in the field is lost and the grain represents averaged traits describing its quality. Therefore attempts to divide grain during combine harvesting are justified [12, 13].

However, the authors of this study are of the opinion that the decision algorithm used for controlling the process of grain stream division, which is based only on the data obtained from the spectrometer assessing the quality of harvested grain, may be unreliable. As Maertens [8] proved, this fact may be particularly evident in the case of very dynamic variations in the parameters describing grain quality and simultaneous considerable delays of the signal due to the time of the flow of grain through the threshing and cleaning mechanisms of the combine harvester. At the same time the authors think that the probability of making the right decision to send grain into one of the two chambers of the grain tank in the harvester may be increased by using the information about variable environmental conditions in the direct neighbourhood of the harvester at work.

Therefore, due to the fact that the authors had a database on grain parameter variability (moisture and winter wheat grain yield) and the variability of soil parameters enhancing yield (the content of total nitrogen, exchangeable phosphorus and potassium, magnesium, pH coefficient, the content of organic matter in soil) they made an attempt to check how much the parameters influence the content of protein in winter wheat grain. However, the main goal of the study was to select from the aforementioned data those which enable prediction of changes in the distribution of the content of protein in wheat grain, on the one hand, and which will be relatively easy to obtain during the work of a combine harvester, on the other hand.

2. Material and methods

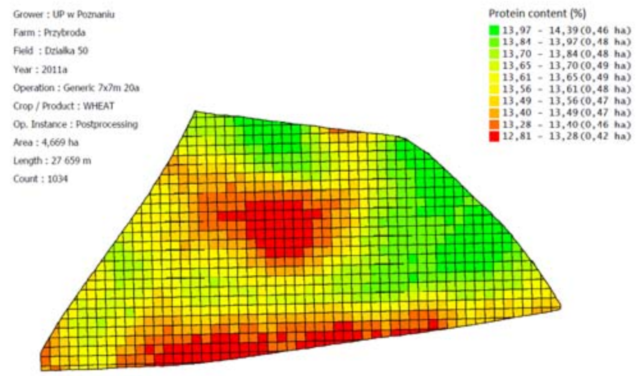
The research used the data obtained in 2011 from five winter wheat production fields. The fields belong to three experimental farms of Poznań University of Life Sciences and they are situated in three different locations in the western part of the Wielkopolska region (Poland). Soil and grain samples for further analyses, grain spectrums and basic data about the location of spectrums and samples were obtained from the fields of the total area of 112 ha.

Soil samples were collected from the 0-0.25 m layer eight weeks before harvesting. Mixed samples of the weight of about 1000 g were collected in a regular square network, where the area of one cluster was 1 ha for the fields larger than 20 ha and 0.5 ha for the fields smaller than 20 ha. An individual mixed sample was made up of 16-18 primary samples. A precise grid of sample collection was made by means of a GNSS Novatel Smart V1 receiver with a TDS Recon recorder and 3R Area Pro field mapping software. The laboratory analysis of the soil samples was made at the Laboratory of National Chemical-Agricultural Station in Poznań, accredited by the State Accreditation Centre for the measurements which are the research subject. For the soil samples the content of total nitrogen was labelled with the Kjeldahl method, the content of absorbed phosphorus (P_2O_5) and potassium (K_2O) with the Egner-Riehm method, magnesium (MgO) – with the Schachtschabel method, the organic matter – with the Tiurin method and the pH – with the potentiometric method in 1nKCl.

In order to obtain the spectrums of wheat grain and to collect grain samples during harvest a ClaasLexion 480 combine harvester was used. It was equipped with an AgroSpec spectrometer (tec5), a GNSS NovAtel PROPAK V3 RT2 receiver with the RTK correction (ASG-EUPOS system), a standard Claas system for yield measurement – Quantimeter and an automatic grain sample collection system constructed at the Institute of Biosystems Engineering in Poznań. The absorption spectrums of radiation were recorded by means of diffuse reflection at the wavelength ranging from 400 to 2170 nm, with interpolated resolution up to 2 nm. A contact measurement probe installed in the measurement channel accumulating the grain sample collected from the grain conveyor of the combine harvester was used for this purpose. While the combine harvester was working, 21.2 thousand spectrums were recorded and 599 grain samples were collected, for which the geographical position and altitude AMSL were also recorded with sub centimetre accuracy. The obtained values of altitude AMSL were converted to relative altitude, where the lowest point in the field was assigned the value of zero.

A Foss Infratec 1241 grain analyzer was used to measure the content of protein in the dry weight of the grain. The grain moisture was measured according to PN-EN ISO in all of the collected samples. On the basis of the results calibration models were built by means of PLS method [1, 6, 11] ($R2=0.75$; $RMSECV=0.59$ for protein content in dry mass and $R2=0.85$; $RMSECV=0.58$ for grain moisture), which was available from Unscrambler X software (CAMO Software AS). The prediction of the content of protein and grain moisture was made in the post-processing mode on the basis of the collected spectrums.

a)



b)

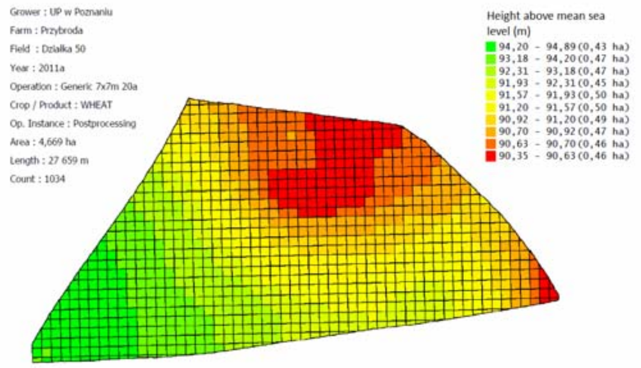


Fig. 1. Examples of interpolated maps showing the variability of: a) protein content distribution, b) altitude AMSL, for the field No. P.50 at the Experimental Agricultural Farm Przybroda

The data which were obtained with the aforementioned methods and the data about the yield (dry), obtained from the board computer of the combine harvester, were initially prepared by means of a spread sheet (MS Office Excel). Then, on their basis digital maps (Fig. 1) of spatial variability of the aforementioned parameters were made with SMS Advance Demo software (AgLeader). Next, the maps were interpolated into a network sized 7 x 7 m and the data in this form were exported to Statistica 10 package, where the essential part of data analysis took place.

Multiple regression was applied to analyse of the influence of selected traits of locally variable environmental conditions on the quality of winter wheat grain. Its aim was to prove the significance $\alpha = 0.05$ of the traits influencing the quality of harvested grain, where the content of protein was determined as quality (independent trait). The Pearson linear correlation analysis was also used to determine the influence of individual independent traits on the dependent trait. The aforementioned analyses were made independently for each of the five fields under investigation and for the entire set of data.

3. Results

Although for the entire set of data the multiple regression analysis confirmed the statistical significance of $\alpha = 0.05$ for all the independent traits in relation to the dependent trait (the content of protein), for individual fields, which were analysed separately, the presence of traits without significance to the grain quality represented as the content of protein was found. The following values were usually indicated as statistically insignificant: relative altitude (B.21, P.50), potassium content (P.2, P.50) and yield (B.21, P.2).

The comparison of the results of multiple regressions (Table 1) shows that the adjustment of the regression model to the set of data deteriorates along with the increase in the area under investigation. This fact may indicate considerable diversification of the character of the influence of individual traits on the grain quality in different areas of the fields. This observation is somewhat surprising in view of the fact that the fields under investigation had relatively undiversified landform (usually slightly sloping in one direction) and small elevations (B.2 Δh=8.14m; B.21 Δh=7.42 m; P.2 Δh=3.78 m; P.50 Δh= 4.45 m; S.69 Δh=2.70 m).

The values of Pearson correlation coefficients in Table 2 also indicate considerable diversification of the strength of the influence of individual traits on the quality of grain between individual fields. However, the obtained results point to the fact that the content of protein in grain is most influenced by the content of total nitrogen and organic matter in soil. Simultaneously, the analysis of correlations between individual traits for the entire set of data indicates a relatively strong relationship between the traits (r = 0.80). At the same time the content of total nitrogen and organic matter in soil are definitely negatively correlated with the relative altitude (r = 0.48 and r = 0.60 respectively). This relationship results in the observed negative correlation between the relative altitude and the content of protein.

The small influence of such traits as the content of phosphorus and magnesium in soil on the content of protein in wheat grain is confirmed by low values of partial and semi-partial correlations (about 0.05) obtained for the entire set of data. Similar low values of partial and semi-partial correlations obtained for the dry yield are contrasted by the high toleration value (1-R2=0.92) obtained for this variable. It suggests the fact that in spite of the observed low correlation between the content of protein and yield the trait contributes some variability to the regression model.

On the basis of the aforementioned results the authors made a decision to determine the variability of the accuracy of the multiple regression model with a smaller number of independent variables and a change in the cardinality of the model set. On the basis of proved correlations between the variables three sets of data were selected. The first contained data on the content of total nitrogen and organic matter in soil, which was supplemented with the values of relative altitude and yield. The content of total nitrogen in soil was excluded from the second set. The third set contained only the data about the relative altitude and yield. Table 3 shows the obtained results.

4. Conclusions

The obtained results give grounds for the conclusion that in the case of small sets of data (approximate field area - 5 ha) it is also possible to obtain satisfying results of prediction of protein content if the input data of the regression model are limited to relative altitude and yield. This observation is valuable due to the fact that this information is

Table 1. A comparison of the results of multiple regressions for individual fields and for the entire set of data

Farm	Field number	Field area (ha)	Number of records	Multiple R	Multiple R^2	Corrected R^2	Standard error of estimate
RGD Brody	B.2	37.54	1270	0.78	0.61	0.61	0.73
RGD Brody	B.21	53.69	9942	0.68	0.46	0.46	0.77
RGD Przybroda	P.2	4.44	6715	0.42	0.17	0.17	0.38
RGD Przybroda	P.50	4.98	1034	0.72	0.52	0.51	0.18
RGD Swadzim	S.69	12.13	2256	0.67	0.45	0.45	0.78
	EDS*	112.78	21218	0.45	0.20	0.20	0.98

* the entire data set

Table 2. A comparison of Pearson correlation coefficients between the content of protein and individual independent traits

Field number	Yield mass dry	Grain moisture	Relative altitude	Soil N_{total}	Soil P_2O_5	Soil K_2O	Soil MgO	Soil OM	Soil pH_{KCl}
B.2	0.12	0.15	-0.36	0.00	-0.51	0.29	0.51	0.03	-0.59
B.21	0.00	0.52	-0.31	0.44	0.34	0.35	0.35	0.42	0.11
P.2	-0.03	-0.16	0.31	-0.18	-0.09	-0.14	-0.25	-0.22	0.15
P.50	0.13	-0.28	-0.29	0.19	0.12	0.23	-0.23	-0.43	0.36
S.69	0.01	-0.38	-0.14	0.33	-0.06	0.19	-0.01	0.37	0.22
EDS*	-0.03	0.14	-0.18	0.31	-0.07	0.12	0.10	0.34	0.00

* the entire data set

Table 3. A comparison of the results of multiple regressions for different data sets of independent variables and cardinality of model sets

Field number	Applied variables	Number of records	Multiple R	Multiple R^2	Corrected R^2	Standard error of estimate
P.50	A N OM Y*	1035	0.68	0.47	0.46	0.19
	A OM Y**		0.67	0.45	0.45	0.19
	A Y***		0.63	0.40	0.40	0.20
B.21	A N OM Y	9942	0.46	0.21	0.21	0.93
	A OM Y		0.42	0.18	0.18	0.95
	A Y		0.31	0.09	0.09	0.99
EDS*	A N OM Y	21218	0.34	0.12	0.12	1.03
	A OM Y		0.34	0.12	0.11	1.03
	A Y		0.18	0.03	0.03	1.08

*ANOMY – relative altitude, total nitrogen, organic matter, yield, **AOMY – relative altitude, organic matter, yield, ***AY – relative altitude, yield

available in most of currently manufactured high-efficiency combine harvesters, so it will not be necessary to bear high costs to equip the harvester with additional sensors to obtain the information.

As the field area increases, the accuracy of the regression model decreases rapidly. It can be at least partially prevented by entering information about the variability of the content of total nitrogen and organic matter in soil into the model. Our calculations confirm the fact that the high correlation between the data gives a possibility to omit one of them without a significant loss to the accuracy of the regression model. The authors think that it would be much easier to equip a combine harvester with a sensor measuring the content of organic matter in soil. It could use the NIRS technology (Near Infrared Spectroscopy).

However, in the case of very large fields even the application of this solution does not always guarantee appropriately accurate prediction of the protein content so that it can be used as supplementary information in controlling the process of separation of the grain stream. For example, it is noticeable at the moments when too dynamic variations of the values returned by the spectrometer used for evaluation of the grain quality make it difficult to determine the trend of variations in the content of protein.

Therefore, further research will check the impact of locally variable soil moisture on grain quality, because the lack of water, even with sufficient availability of minerals compounds, may also result in a decrease in wheat grain protein content.

5. References

[1] Dayal B. S., MacGregor J. F. Improved PLS Algorithms, J. Chemom., 1997, 11, 73-85.

[2] Dreszer, K. A., Gieroba, J. & Roszkowski, A. Kombajnowy zbiór zbóż. Warszawa, IBMER, 1998, (Chapter 2).

[3] Fiez, T.E., Miller, B.C. and Pan, W.L. Winter wheat yield and grain protein across varied landscape positions. Agronomy Journal, 1994, 86: 1026-1032.

[4] Jadczyszyn, T. Zmienność gleb, plonów i potrzeb nawozowych w granicach pola produkcyjnego. Inżynieria Rolnicza, 2001, 13 (33), 168-173.

[5] Kollárová, K., Krajčo, J., Plačko M. & Rutkowski K. Ocena zmienności przestrzennej wilgotności gleby na podstawie map konduktywności elektrycznej. Inżynieria Rolnicza, 2007, 6(94), 73-80.

[6] Lindgren F., Geladi P., Wold S. The kernel algorithm for PLS, J. Chemom., 1993, 7, 45-59.

[7] Long, D. S., Engel, R. E., & Carpenter, F. M. On-Combine Sensing and Mapping of Wheat Protein Concentration, Crop Management (On-line), 2005, Published 27 May 2005.

[8] Maertens, K., Reyns, P., De Baerdemaeker, J. On-line measurement of grain quality with NIR technology. Transactions of the ASAE, 2004, 47(4), 1135-1140.

[9] Moore, I.D., Gessler, P.E., Nielsen, G.A. and Peterson, G.A. Soil attribute prediction using terrain analysis. Soil Science of America Journal, 1993, 57:443-452.

[10] Nolan, S.C., Goddard, T.W., Penney, D.C. & Green, F.M. Yield response to nitrogen within landscape classes. Proceedings of the 4th International Conference on Precision Agriculture, St. Paul, MN, July 19-22, 1998, 479-485.

[11] Rannar S., Lindgren F., Geladi P., Wold S. A PLS kernel algorithm for data sets with many variables and fewer objects, Part 1: Theory and Algorithm, J. Chemom., 1994, 8, 111-125.

[12] Risius, H., Hahn, J., Huth, M., Korte, H., Luetke Harmann, T. Near Infrared Spectroscopy for Sorting Grain according to Specified Quality Parameters on a Combine Harvester. In B.: 67th International Conference on Agricultural Engineering LAND.TECHNIK AgEng (pp.187-192), Stuttgart-Hohenheim: VDI, Germany, 2008,

[13] Risius, H., Hahn, J., Korte, H., Monitoring of grain quality and segregation of grain according to protein concentration threshold on an operating combine harvester. In B.: Book of Abstracts XVII.th World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR | SCGAB) (pp. 28), Québec City: QC, Canada. 2010.

[14] Stewart. C.M., Mc Bratney, A.B. &Skerritt, J.H. Site-specific Durum wheat quality and its relationship to soil properties in a single field in Northern New South Wales. Precision Agriculture, 2002, 3, 155-168.

[15] Taylor, J.; Whelan, B.; Thylén, L.; Gilbertsson, M.; Hassall, J. Monitoring wheat protein content on-harvester - Australian experiences. In B. J. V. Stafford (Eds.) Precision agriculture '05, 5th European Conference on Precision Agriculture, Conference paper (pp. 369-375), Uppsala, Sweden, 2005.

[16] Thylen, L., Gilbbertsson, M., Rosenthal, T., Wrenn, S. Sorting of Grain on the Farm - Experiences with an Online Protein Sensor. In B: D.E. Maier (Eds.) International Quality Grains Conference (pp. 1-8), Indianapolis: Purdue University, USA, 2004.

The paper was published in the Papers Book of International Conference of Agricultural Engineering. CIGR-AgEng2012. July 8-12, 2012.Valencia. Spain.