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

THE USE OF ARTIFICIAL NEURAL NETWORKS TO PREDICT THE SPATIAL VARIABILITY OF GRAIN QUALITY DURING COMBINE HARVEST OF WHEAT

Summary

The aim of the study was to attempt to build and validate the neural model controlling the qualitative selection of the stream of grain mass as early as the stage of combine harvesting of winter wheat. The model uses the highest possible number of data describing locally changeable environmental conditions such as: protein content, moisture and yield of wheat grain, soil abundance in basic nutrients (total Kjeldahl nitrogen, exchangeable phosphorus and potassium, magnesium) and additionally - the pH coefficient, content of organic matter in soil and the relative altitude. The construction of the neural model was preceded with a multiple regression analysis. The results of the analysis (α = 0.05) indicated statistical significance of all of the traits under analysis, which influence grain quality and are defined as the content of protein. The MLP neural network (9-30-1) consisted of one hidden layer containing 30 neurons, one output and nine inputs. The network learning was done with the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm in a single phase during 827 epochs with the SOS error function. 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 National Centre for Research and Development.
***Key words**: artificial neural network, MLP, neural prediction, selective grain harvest, VIS-NIR spectroscopy*

WYKORZYSTANIE SZTUCZNYCH SIECI NEURONOWYCH DO PROGNOZOWANIA ZMIENNOŚCI PRZESTRZENNEJ JAKOŚCI ZIARNA PODCZAS ZBIORU KOMBAJNOWEGO PSZENICY

Streszczenie

Celem pracy było podjęcie próby budowy i walidacji modelu neuronowego sterującego selekcją jakościową strumienia masy ziarna już na etapie kombajnowego zbioru pszenicy ozimej. Model wykorzystuje jak najwięcej danych opisujących lokalnie zmienne warunki środowiskowe takie jak: zawartości białka, wilgotność i wielkość plonu ziarna pszenicy, zasobność gleby w podstawowe składniki pokarmowe (azot ogólny, fosfor i potas wymienny, magnez) oraz dodatkowo współczynnik pH, zawartość materii organicznej w glebie oraz wysokość względną NPM. Budowę modelu neuronowego poprzedzono analizą regresji wielorakiej. Wyniki tej analizy na poziomie α = 0,05 wskazały istotność statystyczną wszystkich badanych cech wpływających na jakość ziarna zdefiniowaną jako zawartość białka. Zbudowana sieć neuronowa typu MLP (9-30-1) składała się jednej warstwy ukrytej zawierającej 30 neuronów, jednego wyjścia i dziewięciu wejść. Uczenie sieci z wykorzystaniem algorytmu BFGS wykonano jednofazowo w trakcie 827 epok z funkcją błędu SOS. Pracę zrealizowano w ramach projektu rozwojowego nr R12 0073 06 pt: „Opracowanie i walidacja technologii rozdziału strumienia ziarna podczas selektywnego zbioru zbóż" finansowanego przez NCBIR.
***Słowa kluczowe**: sztuczne sieci neuronowe, MLP, predykcja neuronowa, 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, 5, 6, 11, 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 grain elevators [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 likelihood of making the right decision to channel grain into one of the two chambers of the grain container 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 topsoil yield parameters (the content of total Kjeldahl 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 make an attempt to build and validate the neural model [1, 4] controlling the qualitative selection of the grain mass stream at the stage of combine harvesting of winter wheat. The model uses the highest possible amount of data describing locally changeable environmental conditions.

2. Material and methods

In order to build the necessary database to generate the neural network data obtained in 2011 from five winter wheat production fields were used. The fields belong to three experimental farms of Poznań University of Life Sciences and they are located 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.

8 weeks before harvesting soil samples were collected from the layer placed at 0-0.25 m. 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 network of the places of sample collection was made by means of a GNSS Novatel Smart V1 kit with a TDS Recon recorder and 3R Area Pro software for field mapping. The laboratory analysis of the soil samples was made at the Laboratory of the District Agri-Chemical 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 Claas Lexion 480 combine harvester was used. It was equipped with an AgroSpec spectrometer (Tec5), a GNSS NovAtel PROPAK V3 RT2 receiver with the RTK update of the ASG-EUPOS system, a standard system for yield measurement – Quantimeter and an automatic grain sample collection system constructed at the Institute of Agricultural 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 subcentimetre 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 analyser was used to measure grain moisture (according to PN-EN ISO 712) and the content of protein in the dry weight of the grain in all of the collected samples. On the basis of the results calibration models were built by means of PLS (R2=0.75; RMSECV=0.59 for protein content in SM 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.

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 spreadsheet (MS Office Excel). Then, on their basis digital maps 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.

The construction of the neural model was preceded with multiple regression analysis, whose 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.

This methodology was also used in an earlier project [10]. The effect of that project was a Kohonen self-organising neural network. Unfortunately, the results of the analysis of the network were not satisfying, so the aim of this project was to create a new model based on a multilayer perceptron neural network.

The neural model was prepared on the basis of the MLP neural network based on the same data as those used in multiple regression. The network structure with one hidden layer was assumed and the BFGS learning algorithm (Broyden-Fletcher-Goldfarb-Shanno) with a maximum value of 1000 epochs was used.

3. Results

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) used for construction of the model. The multiple correlation coefficient of the model was R = 0.44 and the determination coefficient was R2=0.19. The standard error of estimate was 0.99.

Table 1 shows a Pearson correlation analysis of all the independent traits under analysis.

The MLP (9-30-1) neural network (Fig. 1) consisted of three layers: the input layer containing 9 inputs, the hidden layer with 30 neurons and the output layer with one dependent feature – the content of protein. The network learning was carried out in one step with the BFGS algorithm. During the 827 epochs (Fig. 2), the SOS (sum of squares) error function and the method of activation were used - the logistic method for the hidden layer and the exponential method for the output layer. Figure 2 shows a graph of the network training error for the training set (highlighted in blue) and the validation set (highlighted in red).

The learning file accounted for 50% of all the cases, and the test set and the validation set accounted for 25% of the cases. Figure 3 shows a fragment of the learning file (Fig. 3).

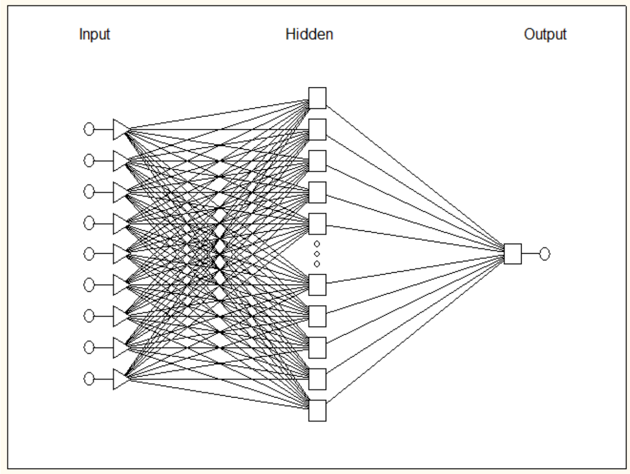


Fig. 1. A view of the MLP network

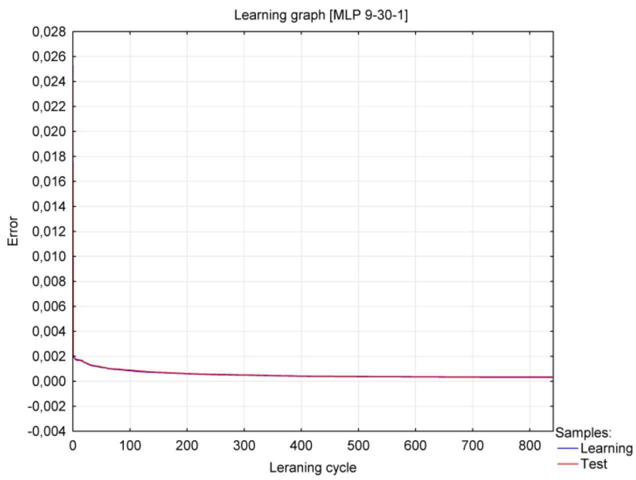
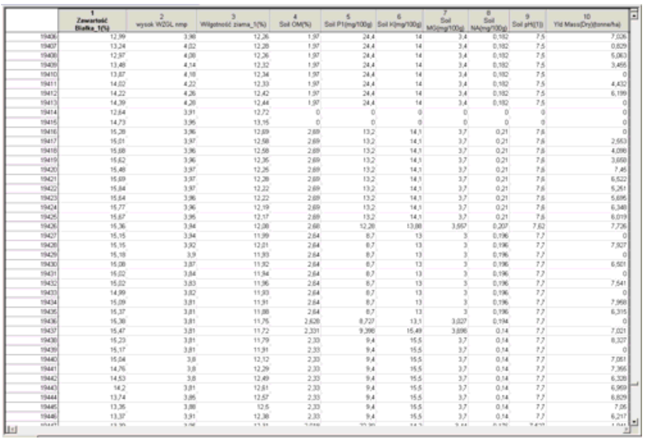


Fig. 2. A learning diagram of the MLP neural network with 827 epochs

The errors achieved different values: 0.092 - for the learning set, 0.100 - for the test set and 0.098 – for the validation set. The qualitative coefficients achieved the following values: 0.920 - for the learning set, 0.916 - for the test set, 0.915 - for the learning set.

Table 1 shows an analysis of the network sensitivity. It indicates the highest rank for the relative altitude - 30.68. Another trait affecting the content of protein under analysis is the organic matter in soil, which achieved the value of 29.52. Similarly to the multiple regression, all the traits under analysis were important during creation of the neural model (error ratio > 1).



	1 Zawartość Białka_1(%)	2 wysok WZGL nmp	3 Wilgotność ziarna_1(%)	4 Soil OM(%)	5 Soil P1(mg/100g)	6 Soil K(mg/100g)	7 Soil MG(mg/100g)	8 Soil NA(mg/100g)	9 Soil pH(1)	10 Yld Mass(Dry)(tonne/ha)
19406	12,99	3,98	12,26	1,97	24,4	14	3,4	0,182	7,5	7,036
19407	13,24	4,02	12,28	1,97	24,4	14	3,4	0,182	7,5	0,829
19408	12,97	4,08	12,26	1,97	24,4	14	3,4	0,182	7,5	5,063
19409	13,48	4,14	12,32	1,97	24,4	14	3,4	0,182	7,5	3,466
19410	13,87	4,18	12,34	1,97	24,4	14	3,4	0,182	7,5	0
19411	14,02	4,22	12,33	1,97	24,4	14	3,4	0,182	7,5	4,432
19412	14,22	4,26	12,42	1,97	24,4	14	3,4	0,182	7,5	6,199
19413	14,39	4,28	12,44	1,97	24,4	14	3,4	0,182	7,5	0
19414	12,64	3,91	12,72	0	0	0	0	0	0	0
19415	14,73	3,95	13,15	0	0	0	0	0	0	0
19416	15,28	3,96	12,69	2,69	13,2	14,1	3,7	0,21	7,6	0
19417	15,01	3,97	12,58	2,69	13,2	14,1	3,7	0,21	7,6	2,653
19418	15,68	3,96	12,58	2,69	13,2	14,1	3,7	0,21	7,6	4,098
19419	15,62	3,96	12,36	2,69	13,2	14,1	3,7	0,21	7,6	3,658
19420	15,48	3,97	12,26	2,69	13,2	14,1	3,7	0,21	7,6	7,46
19421	15,69	3,97	12,28	2,69	13,2	14,1	3,7	0,21	7,6	6,522
19422	15,84	3,97	12,22	2,69	13,2	14,1	3,7	0,21	7,6	5,251
19423	15,64	3,96	12,22	2,69	13,2	14,1	3,7	0,21	7,6	5,696
19424	15,77	3,96	12,19	2,69	13,2	14,1	3,7	0,21	7,6	6,348
19425	15,67	3,95	12,17	2,69	13,2	14,1	3,7	0,21	7,6	6,019
19426	15,36	3,94	12,08	2,68	12,28	13,88	3,657	0,207	7,62	7,726
19427	15,15	3,94	11,99	2,64	8,7	13	3	0,196	7,7	0
19428	15,15	3,92	12,01	2,64	8,7	13	3	0,196	7,7	7,927
19429	15,18	3,9	11,93	2,64	8,7	13	3	0,196	7,7	0
19430	15,08	3,87	11,92	2,64	8,7	13	3	0,196	7,7	6,501
19431	16,02	3,84	11,94	2,64	8,7	13	3	0,196	7,7	0
19432	15,02	3,83	11,96	2,64	8,7	13	3	0,196	7,7	7,541
19433	14,99	3,82	11,93	2,64	8,7	13	3	0,196	7,7	0
19434	15,09	3,81	11,91	2,64	8,7	13	3	0,196	7,7	7,968
19435	15,37	3,81	11,88	2,64	8,7	13	3	0,196	7,7	6,315
19436	15,38	3,81	11,75	2,628	8,727	13,1	3,027	0,194	7,7	0
19437	15,47	3,81	11,72	2,331	9,398	15,49	3,698	0,14	7,7	7,021
19438	15,23	3,81	11,79	2,33	9,4	15,5	3,7	0,14	7,7	8,327
19439	15,17	3,81	11,91	2,33	9,4	15,5	3,7	0,14	7,7	0
19440	15,04	3,8	12,12	2,33	9,4	15,5	3,7	0,14	7,7	7,061
19441	14,76	3,8	12,29	2,33	9,4	15,5	3,7	0,14	7,7	7,366
19442	14,53	3,8	12,49	2,33	9,4	15,5	3,7	0,14	7,7	6,328
19443	14,2	3,81	12,61	2,33	9,4	15,5	3,7	0,14	7,7	6,969
19444	13,74	3,85	12,57	2,33	9,4	15,5	3,7	0,14	7,7	6,829
19445	13,36	3,88	12,6	2,33	9,4	15,5	3,7	0,14	7,7	7,06
19446	13,37	3,91	12,38	2,33	9,4	15,5	3,7	0,14	7,7	6,217

Fig. 3. A part of the learning file

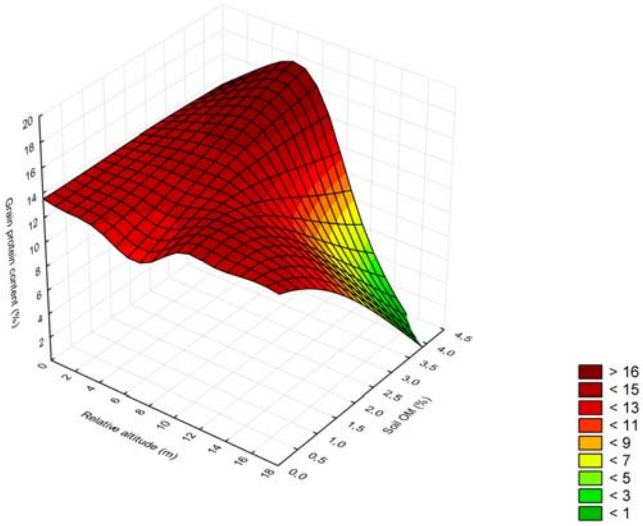


Fig. 4. Network response surface

The surface of response was based on the results of the analysis of the neural network sensitivity. One predicted feature (protein content) and two dependent traits were used: relative altitude and organic matter in soil.

4. Conclusions

The first stage of the study involved a multiple regression analysis. The significance level ($\alpha = 0.05$) gave information about the significance of all independent traits in the model. The information was the basis for further analyses on the neural model, including all variables (the dependent trait and independent traits). Unfortunately, the low value of the determination coefficient of the regression model, R2 = 0.19, does not guarantee the quality of the model. The standard error of estimate was 0.99, which results in the deviation value of 7.1%.

The neural model built on the MLP network proved that the highest rank in the first position belongs to the independent variable – the relative altitude. The analysis of the network sensitivity showed another important trait – the organic matter in soil.

The correlation analysis proves the influence of the independent traits (relative altitude and organic matter in soil) on the content of protein in wheat. The first trait (organic matter in soil) achieved the maximum level R = 0.33, but the other one amounted to R = -0.178.

The results obtained from the correlation analysis correspond to the results obtained from the neural model, although for the relative altitude the correlation is negative. It may indicate a non-linear nature of the problem studied.

Further research should include a more detailed analysis of the data for individual fields. It should include the correlation and the error ratio between the multiple regression and the neural model.

Table 1. A comparison of the multiple regression model parameters and network sensitivity analysis

Parameter	Pearson correlation Grain protein content (%)	Network sensivity analysis - error ratio	Network sensivity analysis - rank
Relative altitude (m)	-0.178	**30.68**	**1**
Yield dry weight (tonne/ha)	-0.031	1.56	9
Grain moisture (%)	0.131	10.51	5
Soil N_{total} (mg/100g)	0.306	9.92	6
Soil P_2O_5 (mg/100g)	-0.064	12.45	3
Soil K_2O (mg/100g)	0.120	12.14	4
Soil MgO (mg/100g)	0.098	8.83	7
Soil pH (1)	0.004	7.97	8
Soil OM (%)	0.335	**29.52**	**2**

5. References

[1] Alvarez R.: Predicting average regional yield and production of wheat in the Argentinian Pampas with an artificial neural network approach. European Journal of Agronomy, 2009, 30(2): 70-77.

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

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

[4] Huang Y., Lan Y., Thomson S. J., Fang A., Hoffmann W. C., Lacey R. E.: Development of soft computing and applications in agricultural and biological engineering. Computers and Electronics in Agriculture, 2010, 71(2), 107-127.

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

[6] 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.

[7] Long D.S., Engel R.E., Carpenter F.M.: On-Combine Sensing and Mapping of Wheat Protein Concentration, Crop Management (On-line), 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., Peterson G.A.: Soil attribute prediction using terrain analysis. Soil Science of America Journal, 1993, 57:443-452.

[10] Niedbała G., Czechlowski M., Wojciechowski T.: The use of artificial neural networks to predict the spatial variability of grain quality during combine harvest of wheat. Proceedings of the International Conference of Agricultural Engineering CIGR-AgEng2012, Valencia, Spain, July 8-12, 2012, P2112.

[11] 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.

[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., McBratney 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.