Use of Regression and Discriminant Analyses to Develop a Quality Classification System for Hard Red Winter Wheat[1]

Scott Baker,[2] Timothy J. Herrman,[2,3] and Thomas Loughin[4]

ABSTRACT

Cereal Chem. 76(6):890–893

An attempt to create a segregation system that uses rapid quality detection instrumentation and represents value to millers and bakers led to the development of a single value called "dough factor." The dough factor value represents the amount of flour-water dough that can be produced by a given unit of wheat. Samples of hard red winter wheat (≈100/location) collected from five Kansas country elevators during the 1995 and 1996 harvests were evaluated for dough factor. Single kernel properties, sample protein content, and test weight measurements were subjected to regression and discriminant analyses for the purpose of developing a dough factor classification system. Regression analysis identified kernel weight, kernel weight standard deviation, and protein as important characteristics for predicting dough factor, however, the resulting model possessed poor predictive ability (adjusted R^2 = 0.39). Classifying wheat into dough factor groups of <107, 107–112.9, and ≥113 using discriminant analysis resulted in an accuracy of 56%, while discriminant analysis correctly placed wheat into two dough factor groups (<113 and ≥113) with an 80% accuracy. Creation of a dough factor classification system using single kernel measures, kernel protein, and wheat cultivar correctly classified 86.3 and 68.8% of the wheat samples into dough factor groups <113 and ≥113, respectively. In the dough factor group ≥113, cost savings associated with higher flour yields and water absorption were $0.15/cwt of flour and $0.65/1,000 lb of dough, respectively. Increases in processing efficiency for both the miller and the baker would be expected to further differentiate the value between the two dough factor groups.

The Grain Quality Acts of 1986 and 1990 contain congressional mandates for developing tools to facilitate adoption of a quality-based marketing system. As an outgrowth of this legislation, the introduction of the single kernel characterization system (SKCS) and near-infrared transmittance (NIR-T) whole grain analyzers have improved the capability of participants in the marketplace to rapidly predict and segregate wheat based on end-use properties.

In the Southern Plains, country elevators receive >90% of the wheat during harvest. Thus, a quality-based marketing system will likely involve the incorporation of rapid quality identification technology at the country elevator. Baker et al (1997) concluded that receiving systems at Kansas country elevators possess the engineering capability to segregate wheat during harvest. However, a need exists to develop a segregation method that represents end-use value to millers and bakers.

Previous research pertaining to segregating wheat at country elevators focused on test weight, dockage, grade, and protein (Dunne and Anderson 1975, Kenkel 1996). Kenkel (1996) found that segregation by test weight and dockage at Oklahoma country elevators could increase grain-handling profits by reducing cleaning costs. In addition, ≈25% of the loads sampled in that study (representing ≈500,000 bu) could be marketed directly to flour mills for a protein premium of $0.10 to $0.15/bu. Dunne and Anderson (1975) found that protein content between grades may not differ within a crop year but could vary by as much as 2% between crop years. Variation in wheat quality can interrupt production schedules, increase processing costs, require additional storage capacity, or reduce product quality (Wilson and Preszler 1992).

In a companion article, Baker et al (1999) introduces the conceptual framework for segregating wheat using the SKCS and NIR-T whole grain analyzer to classify wheat based on flour extraction and water absorption. Combining these two properties into a single value called "dough factor" enables one to describe the amount of flour-water dough (w/w) that can be produced from a given quantity of wheat. Dough factor increases in response to increased test weight, kernel weight, and kernel size. More importantly, Baker et al (1999) found that the relationship between dough factor groups and single kernel properties remained consistent across all locations. This indicates that wheat quality characteristics within dough factor groups are similar regardless of location. These consistent differences indicated that dough factor can be segregated into three groups: <107, 107–112.9, and ≥113.

The next step in developing a segregation system, using the concept of dough factor, entails identification of parsimonious variables measured by SKCS and NIR-T that best predict end-use performance. To that end, this article contains the results of regression and discriminant analyses used to develop a dough factor classification system.

MATERIALS AND METHODS

Samples of hard red winter (HRW) wheat (≈100/location) collected from five Kansas elevators in 1995 and 1996 were evaluated as whole grain, milled, and subjected to mixogram analysis of flour water absorption as described by Baker et al (1999). HRW wheat cultivars included Karl 92, Ike, 2163, 2180, Hickock, Tomahawk, Arapahoe, and Victory.

Regression analysis was performed using SAS statistical software (SAS Institute, Cary, NC). PROC REG was used to identify which linear, quadratic, and cross-product terms were significant. The most significant regression models were selected based on these variables using PROC REG with adjusted R^2 and mean square error serving as selection criteria. Discriminant analysis was performed in a Microsoft Excel program using Mahalanobis distances as described by Ragsdale (1998).

RESULTS AND DISCUSSION

Regression Analysis

Regression analysis used single kernel characteristics and protein to predict dough factor. Results indicated that the most significant variables were kernel weight, kernel weight standard deviation, protein, kernel weight squared, protein squared, and the cross-product of kernel weight and kernel weight standard deviation. Each of these variables was significant. However, a low adjusted R^2 of 0.39 prevents the use of regression to precisely predict dough factor. Adding test weight to the model provided only a marginal increase in the adjusted R^2 (0.42) and a small decrease in the mean square

[1] Contribution No 99-184-J, Kansas Agricultural Experiment Station.
[2] Graduate research assistant and associate professor, respectively, Department of Grain Science and Industry, Kansas State University, Manhattan, KS 66506-2201.
[3] Corresponding author. Phone: 785-532-4082. Fax: 785-532-4017. E-mail: tjh@wheat.ksu.edu
[4] Associate professor, Department of Statistics, Kansas State University, Manhattan, KS 66506-2201.

Publication no. C-1999-1005-05R.

error. Even though precise prediction of dough factor is not possible using regression, the regression analysis can be used to generate contour plots to identify regions of optimal response.

Contour plots (Figs. 1–3) illustrate how the greatest frequency of samples possessing a dough factor value 113 possessed kernel weights >28 mg. However, when kernel weight standard deviation was low and protein content high, wheat samples with average kernel weights as low as 24 mg had dough factor values >113. As kernel weight increased >30 mg, kernel weight standard deviation became less important for protein values >12.5%. Dough factor values <107 were associated with kernel weights <26 mg and possessed a relatively high kernel weight standard deviation. These contour plots also demonstrate the relationship between single kernel properties and protein content on dough factor.

Discriminant Analysis

Tables I and II show classification matrices for dough factor groups <107, 107–112.9, and ≥113 for different combinations of kernel weight, kernel weight standard deviation, kernel size, and protein. The total percentage of samples correctly placed in their respective dough factor group was <60%. However, samples belonging to dough factor groups <107 and ≥113 were correctly placed 70 and 60%, respectively. Differentiation between groups 107–112.9 and <107 was poor, with at least 34% of group 107–112.9 being improperly placed in dough factor group <107 in both models, and 20% of the 107–112.9 samples improperly placed in dough factor group ≥113. Reducing the number of dough factor groups to <113 and ≥113 improved the placement accuracy to ≈79%. Samples in the lower dough factor group were placed with greater accuracy (85.1%) than the highest dough factor group (63.6%). The addition of test weight to the discriminant analysis did not improve the correct placement of wheat into the respective dough factor groups.

Kernel Weight Standard Deviation							
10.0	< 107	< 107	< 107	< 107	< 107	107	107
9.5	< 107	< 107	< 107	< 107	107	107	107
9.0	< 107	< 107	< 107	< 107	107	107	107
8.5	< 107	< 107	< 107	< 107	107	107	107
8.0	< 107	< 107	< 107	107	107	107	107
7.5	< 107	< 107	< 107	107	107	107	107
7.0	< 107	< 107	107	107	107	107	113
6.5	< 107	< 107	107	107	107	113	113
6.0	< 107	107	107	107	113	113	113
5.5	< 107	107	107	113	113	113	113
5.0	107	107	107	113	113	113	113
Kernel Weight (mg)	20.0	22.0	24.0	26.0	28.0	30.0	32.0

Fig. 1. Contour plot for dough factor group using kernel weight, kernel weight standard deviation, and protein (11.5%).

Kernel Weight Standard Deviation							
10.0	< 107	< 107	< 107	< 107	107	107	107
9.5	< 107	< 107	< 107	< 107	107	107	107
9.0	< 107	< 107	< 107	< 107	107	107	107
8.5	< 107	< 107	< 107	107	107	107	113
8.0	< 107	< 107	< 107	107	107	107	113
7.5	< 107	< 107	107	107	107	113	113
7.0	< 107	< 107	107	107	107	113	113
6.5	< 107	< 107	107	107	113	113	113
6.0	< 107	107	107	113	113	113	113
5.5	< 107	107	107	113	113	113	113
5.0	107	107	113	113	113	113	113
Kernel Weight (mg)	20.0	22.0	24.0	26.0	28.0	30.0	32.0

Fig. 2. Contour plot for dough factor group using kernel weight, kernel weight standard deviation, and protein (12.5%).

Dough Factor Classification System

A system for classifying wheat into two dough factor groups (<113 and ≥113) is presented in Table IV. Table V shows how the dough factor classification system correctly classified 76% of the wheat samples. The primary differences between the results from discriminant analysis and the dough factor classification system were that the latter correctly identified a higher percentage of dough factor group ≥113 and improperly classified a larger percentage of dough factor group <113.

Wheat quality characteristics for the samples correctly and incorrectly classified by the dough factor classification system are shown in Table VI. The classification system placed wheat that possessed similar single kernel characteristics into dough factor group <113. However, the incorrectly identified samples (≥113) possessed higher test weight and protein values. Similarly, the samples improperly classified as dough factor group ≥113 had single kernel characteristics similar to those of samples correctly identified, but the incorrectly identified samples had lower test weight and protein values.

The distribution of the dough factor group by cultivar (Table VII) indicates that separate classification systems by cultivar would improve classification accuracy. For example, 12% of the samples

TABLE I
Classification Matrix for Three Dough Factor Groups (<107, 107–112.9, and ≥113) Using Kernel Weight, Kernel Weight Standard Deviation, and Protein

	Predicted				
Actual	**<107**	**107–112.9**	**≥113**	**Total**	**% Correct**
<107	108	26	6	140	77.1
107–112.9	206	208	89	503	41.4
≥113	27	61	162	250	64.8
Total	341	295	257	893	53.5

TABLE II
Classification Matrix for Three Dough Factor Groups (<107, 107–112.9, and ≥113) Using Kernel Weight, Kernel Weight Standard Deviation, Kernel Diameter, and Protein

	Predicted				
Actual	**<107**	**107–112.9**	**≥113**	**Total**	**% Correct**
<107	103	32	5	140	73.6
107–112.9	171	246	86	503	48.9
≥113	19	75	156	250	62.4
Total	293	353	247	893	56.6

TABLE III
Classification Matrix for Two Dough Factor Groups (<113 and ≥113) Using Kernel Weight, Kernel Weight Standard Deviation, Kernel Diameter, and Protein

	Predicted			
Actual	**<113**	**≥113**	**Total**	**% Correct**
<113	547	96	643	85.1
≥113	91	159	250	63.6
Total	638	255	893	79.1

for cultivar 2163 had a dough factor of ≥113, whereas >40% of the Karl and Ike samples had a dough factor of ≥113. Cultivars 2163, Karl, and Ike accounted for 38, 32, and 14% of the samples, respectively. The use of separate classifications systems for Karl/Ike and 2163/ other cultivars resulted in a 4.8% improvement in overall accuracy (Table VIII). The primary source of improvement resulted from a decreased number of samples with dough factor <113 that were incorrectly identified as dough factor ≥113. The classification system for wheat cultivars possessing superior quality placed Karl/Ike into dough factor group ≥113 if single kernel weight is >22 mg, single kernel diameter is >1.9 mm, kernel weight standard deviation is <6.5, and protein content is 12.5%. The classification system for wheat cultivars possessing less desirable quality placed 2163/other cultivars into the dough factor group ≥113 whenever kernel diameter is >2.3 mm.

Varietal and environmental influences on kernel weight and kernel weight standard deviation also have been reported by Watson and Heyne (1977) and Puppala et al (1998). The practice of categorizing wheat cultivars by milling and baking properties for HRW wheat appears in a Kansas State University Research and Extension bulletin prepared by McCluskey et al (1998). While the planted acres of particular cultivar change annually as new wheat cultivars are released, a dough factor classification system would likely benefit from differentiating superior and less desirable quality wheats within the selection matrices.

Despite the improper classification associated with using the same dough factor classification system for all cultivars, important differences were observed between the two groups. Based on the average dough factor values for each group, the dough factor group 113 would produce 3.4% more dough from the same amount of wheat. The increased flour yield resulted in a savings in wheat costs of $0.15/cwt of flour (based on a wheat price of $2.86/bu) for the miller and the increased water absorption resulted in a reduction in flour costs of $0.65/1,000 lb of dough (based on flour price of $9/cwt) for the baker. Lower milling costs/cwt of flour produced also would be expected due to the influence of higher kernel weight and kernel size on mill adjustments and daily milling capacity (Kremer 1993, Gwirtz et al 1996). Increases in processing efficiency also would be expected to result from better dough-handling properties (Herrman et al 1995) associated with higher protein flour.

TABLE IV
Minimum Kernel Size, Kernel Weight, Kernel Weight Standard Deviation (SD), and Protein Requirements for the Dough Factor Classification System

Kernel Size (mm)	Kernel Weight (mg)	Kernel Weight SD	Protein[a] (%)	Dough Factor Group
1.7	. . .	. . .	. . .	<113
1.9	20	. . .	. . .	<113
	22	. . .	. . .	<113
	24	. . .	. . .	<113
2.1	22	. . .	. . .	<113
	24	<7.0	12.5	≥113
	26	<8.0	12.5	≥113
	28	<9.0	12.5	≥113
2.3	26	<8.5	12.0	≥113
	28	<9.5	12.0	≥113
	30	<10.0	12.0	≥113
2.5	30	<10.0	12.0	≥113

[a] 12% mb.

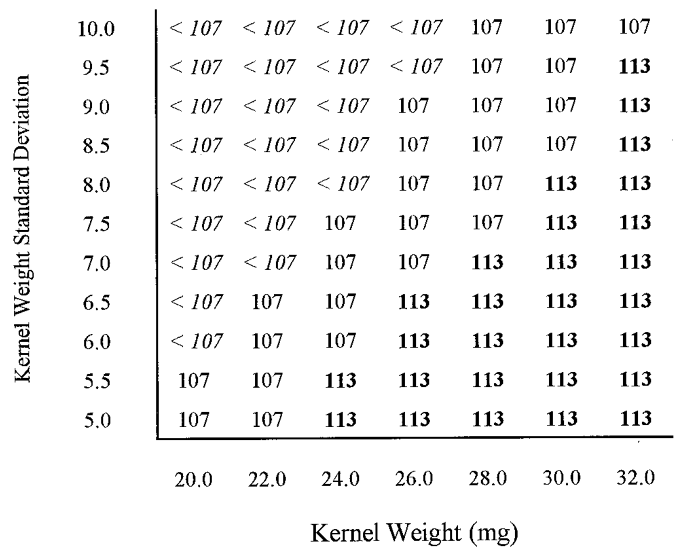

Fig. 3. Contour plot for dough factor group using kernel weight, kernel weight standard deviation, and protein (13.5%).

TABLE V
Classification Matrix for the Dough Factor Classification System

	Predicted			
Actual	**<113**	**≥113**	**Total**	**% Correct**
<113	511	132	643	79.5
≥113	77	173	250	69.2
Total	588	305	893	76.6

TABLE VII
Distribution of Two Dough Factor Groups (<113 and ≥113) by Cultivar

Cultivar	<113	≥113
2163	88.1	11.9
Ike	58.1	41.9
Karl	56.2	43.8
Other[a]	77.4	22.6

[a] 2180, Tomahawk, Hickock, Arapahoe, Victory.

TABLE VI
Quality Characteristics of Correctly and Incorrectly Classified Wheat Samples Using the Dough Factor Classification System

Dough Factor Group		Kernel Weight (mg)	Kernel Weight SD	Kernel Diameter (mm)	Test Weight (lb/bu)	Protein[a] (%)	Flour Yield[b] (%)	Water Absorption[b] (%)	Dough Factor
Predicted	**Actual**								
<113	<113	25.2	7.2	2.1	56.6	12.5	67.3	61.2	108.5
	≥113	25.3	6.9	2.1	57.9	13.0	70.1	63.4	114.6
	Avg	25.2	7.2	2.1	56.8	12.5	67.7	61.5	109.3
≥113	<113	28.5	7.7	2.3	58.9	13.2	67.8	62.5	110.1
	≥113	29.1	7.5	2.3	59.8	13.6	70.3	64.0	115.3
	Avg	28.8	7.6	2.3	59.4	13.5	69.2	63.4	113.0
Total	Avg	26.4	7.3	2.2	57.7	12.8	68.2	62.1	110.6

[a] 12% mb.
[b] 14% mb.

TABLE VIII
Classification Matrix for Separate Dough Factor Classification Systems for Karl/Ike and 2163/Other Cultivars

	Predicted			
Actual	**<113**	**≥113**	**Total**	**% Correct**
<113	555	88	643	86.3
≥113	78	172	250	68.8
Total	633	260	893	81.4

ACKNOWLEDGMENTS

We thank the Grain Inspection, Packers, and Stockyards Administration Federal Grain Inspection Service and the Agricultural Research Service for project funding, and Patrick McCluskey for use of laboratory milling and dough testing equipment.

LITERATURE CITED

Baker, S., Herrman, T., and Fairchild, F. 1997. Capability of Kansas grain elevators to segregate wheat during harvest. SRP781, Agric. Exper. Stn. and Cooperative Extension Serv. Kansas State University: Manhattan, KS.

Baker, S., Herrman, T. J., and Loughin, T. 1999. Segregating hard red winter wheat into dough factor groups using single kernel measurements and whole grain protein analysis. Cereal Chem. 76:884-889.

Dunne, W., and Anderson, J. A. 1975. Traditional systems for segregating Canadian wheat by protein content. Can. J. Plant Sci. 55:1-18.

Finney, K. A., and Shogren, M. D. 1972. A ten-gram mixograph for determining and predicting functional properties of wheat flours. Baker's Dig. 46(2):32-35, 38-42, 77.

Gwirtz, J. A., Eustace, W. D., and Curran, S. P. 1996. Effect of kernel size and test weight on the Kansas State pilot mill. Assoc. Oper. Millers Tech. Bull. June: 6731-6741.

Herrman, T., Bequette, R. K., McCluskey, P. J., Pedersen, J. R., Ponte, J. G., and Shroyer, J. P. 1995. Quality oriented marketing of hard red winter wheat. MF-2119, Cooperative Extension Service. Kansas State University: Manhattan, KS.

Kenkel, P. 1996. Grain segregation can improve profits. Feed & Grain June/July: 20-23.

Kremer, H. 1993. The effect of test weight on milling results. Am. Assoc. Oper. Millers Tech. Bull. March: 6189-6190.

McCluskey, P. J., Chung, O. K., and Herrman, T. J. 1998. Milling and bread-baking qualities of hard winter wheat varieties 1998 Kansas Update. MF-1077. Kansas State University Agricultural Experiment Station and Cooperative Extension Service: Manhattan, KS.

Puppala, V., Herrman, T. J., Bockus, W. W., and Loughin, T. M. 1998. Quality response of twelve hard red winter wheat cultivars to foliar disease across four locations in Central Kansas. Cereal Chem. 75:94-99.

Ragsdale, C. T. 1998. Discriminant analysis. In: Spreadsheet Modeling and Decision Analysis: A Practical Introduction to Management Science. South-Western College Publishing: Cincinnati, OH.

Watson, C. A., and Heyne, E. G. 1977. Individual kernel weight distribution of 12 varieties of hard red winter wheat. Cereal Chem. 54:161-166.

Wilson, W. W., and Preszler, T. 1992. End-use performance uncertainty and competition in international wheat markets. Am. J. Agric. Econ. 74:556-563.

[Received November 18, 1998. Accepted June 7, 1999.]