

## FACTORS AFFECTING FARINOGRAPH AND BAKING ABSORPTION II. RELATIVE INFLUENCE OF FLOUR COMPONENTS<sup>1</sup>

K. H. TIPPLES, J. O. MEREDITH,<sup>2</sup> and J. HOLAS,<sup>3</sup> Canadian Grain Commission, Grain Research Laboratory, Winnipeg, Man. R3C 3G9

### ABSTRACT

Cereal Chem. 55(5): 652-660

Interrelationships between six independent variables (protein, wet gluten, wet gluten/protein, damaged starch, pentosans, 100/amylograph viscosity) and six water absorption values (farinograph and baking absorption in two baking methods with variable malt level) were examined statistically for a series of flours and flour streams milled from Canadian hard red spring wheat. The strongest single predictor variable for farinograph absorption was starch damage. Addition of protein content as a second variable gave an excellent prediction of absorption; adding further variables to the prediction equation

was not advantageous. By contrast, baking absorption in both long and short processes was largely a function of gluten/protein content, and the best explanatory combinations were any two of protein, wet gluten, and wet gluten/protein. Damaged starch not only had little influence on baking absorption but also appeared to have a negative effect, particularly when malt was used in the formula in a method involving bulk fermentation. Differences in pentosan content and amylograph viscosity between flour streams did not appear to have a major influence on fluctuations in either farinograph or baking absorption.

In the first part of this article (1) we emphasized the large differences that can exist in analytic, rheologic, and baking properties between individual mill fractions. In this article, we will examine the relative contribution of various flour components (protein, wet gluten, damaged starch, pentosans, and  $\alpha$ -amylase activity) to absorption for the same series of hard red spring wheat flour streams. Using multiple linear regression analysis, we attempted to find equations or sets of equations that best predicted the relationships between the flour components and the absorption data. In particular, we examined differences between farinograph water absorption and baking absorption.

<sup>1</sup>Paper No. 396 of the Canadian Grain Commission, Grain Research Laboratory, Winnipeg, Man. R3C 3G9.

<sup>2</sup>Research statistician, Canadian Grain Commission, Economics & Statistics Division.

<sup>3</sup>Recipient of a postdoctoral fellowship from the National Research Council of Canada for the year 1973-1974. Present address: Research Institute of Milling and Baking, Na Pankraci 30, 140 04 Praha, Czechoslovakia.

## METHODS

Often the best way to summarize multifactorial data is by a simple equation or set of equations. We wished to find an equation, or set of equations that best described the data, both from the viewpoint of a good fit in the sense of least squares and from the viewpoint of providing some insight into the relative contribution of different flour components to absorption. The dependent variables, or responses, are the baking absorption and farinograph absorption data. The independent variables, to be used in predicting or explaining the absorption data, were selected from the analytic data for the flour streams.

### Analytic Data Considered for Prediction of Absorption

The analytic data considered relevant for prediction of absorption were protein (P) ( $N \times 5.7$ ), wet gluten (G), damaged starch (D), amylograph peak viscosity (V), and total pentosans (Pn) (soluble and insoluble). Since V and  $\alpha$ -amylase activity are inversely related, we used  $100/V$  as an indicator of  $\alpha$ -amylase activity. The apparent influence of D on V described earlier (1) was not corrected.

Preliminary regression analyses indicated that considering soluble and insoluble pentosans separately was not advantageous. The values for Pn therefore were used for subsequent statistical analyses. Preliminary analyses also showed flour protein content alone to be a surprisingly poor predictor of absorption. A protein quality factor (G/P) was obtained by dividing wet gluten by flour protein.

Table I lists the variables that were considered for statistical analysis. The values of these variables for the 20 flours and flour streams examined are shown in Table II.

### Statistical Methods

A number of different approaches to the multiple regression problem have been considered (2). A straightforward method is to choose on physicochemical grounds those variables that should be included as independent predictor

TABLE I  
Variables Considered for Statistical Analysis

Variable	Abbreviation
Independent	
Protein content ( $N \times 5.7$ ), %	P
Wet gluten, %	G
Protein quality factor	G/P
Damaged starch, Farrand units	D
100/Amylograph peak viscosity	100/V
Total pentosans, %	Pn
Dependent	
Farinograph absorption, %	FA
Remix baking absorption, 0.3% malt	RBA.3
Remix baking absorption, 0.1% malt	RBA.1
Remix baking absorption, no malt	RBA.0
Chorleywood baking absorption, 0.1% malt	CBA.1
Chorleywood baking absorption, no malt	CBA.0

**TABLE II**  
**Values of Independent and Dependent Variables**

Flour Stream	Independent Variables						Dependent (Absorption) Variables					
	P (%)	G (%)	D (FU) <sup>b</sup>	100/V	Pn (%)	G/P	FA (%)	RBA.3 (%)	RBA.1 (%)	RBA.0 (%)	CBA.1 (%)	CBA.0 (%)
1 & 2B	15.2	47.8	17	.130	1.45	3.14	61.9	64	67	69	69	71
3B	16.1	50.6	16	.124	1.54	3.14	63.2	64	67	70	69	72
4B	19.2	62.7	15	.143	1.57	3.27	66.4	68	71	75	73	77
1M	12.5	37.0	29	.116	1.41	2.96	62.6	60	61	64	63	66
2M	12.1	35.4	22	.108	1.53	2.93	60.5	62	63	65	65	67
3M	11.6	33.8	40	.119	1.59	2.91	64.6	60	61	64	63	66
4M	11.5	33.6	63	.137	1.62	2.92	69.9	60	61	68	63	70
5M	13.4	33.8	89	.244	1.97	2.52	79.2	56	58	62	60	64
T	14.0	37.9	50	.213	1.95	2.47	68.6	60	61	63	63	65
2Q	12.3	34.0	56	.185	1.91	2.76	69.1	59	60	64	62	67
LG	13.5	33.3	54	.196	1.99	2.71	69.5	56	57	59	59	61
S	13.0	37.9	33	.133	1.73	2.92	64.0	61	62	65	64	67
C1-1M	12.2	36.3	37	.145	1.71	2.95	64.5	60	62	68	64	70
C1-S	13.3	39.2	42	.154	1.81	2.89	66.8	61	62	66	63	69
C2-2M	11.9	34.8	43	.154	1.78	2.84	65.6	60	61	67	63	69
C2-S	13.6	39.3	45	.116	1.49	2.98	68.0	60	61	67	63	69
C1,2-2M	11.6	33.8	78	.122	1.58	2.92	73.8	60	61	69	63	69
C1,2-S	13.4	38.1	52	.141	1.61	2.91	69.5	61	62	67	64	70
F	12.9	38.3	26	.127	1.61	3.02	62.9	62	62	65	66	68
G	12.9	38.9	18	.132	1.42	3.02	61.3	63	63	66	65	69

<sup>a</sup>See Table I for abbreviations.

<sup>b</sup>FU = Farrand units.

variables, and to estimate the coefficients and their errors. Promising subsets of the predictor variables may be selected for further multiple regression analysis. This method, however, is largely intuitive, and more systematic methods have been developed.

Stepwise multiple regression has become a standard technique and can be used in two different forms: the forward, or step-up, method and the backward, or step-down, method. The two methods will not necessarily select the same variables, and neither is guaranteed to find the same set of variables as found by an exhaustive search of all possible subsets of variables in which, for example, the regression of Y is computed on each variable (X) singly and all possible pairs on all possible triples. The best pair does not necessarily include the best single variable, and the best triplet does not necessarily include the best pair, although this may occur.

The large number of regression computations needed for an exhaustive search led to research on optimal methods of selecting best subsets according to some definition of "best." Daniel and Wood (3) used Mallows' (4) total error squared, and recommended an exhaustive search if possible. They developed a systematic, directed search that first involved fitting the full equation. Boyce et al (5) and Beale et al (6) have developed another method based on general properties of functions defined on subsets.

The method chosen for this work was a complete computer-assisted examination of all  $(2^k - 1) = 63$  regressions using RBA.3 as the dependent variable. With the experience gained, a directed search similar to that of Daniel and Wood (3) was made to establish the regression models for the remaining five different dependent variables. This procedure allowed a full examination of the independent variables, identification of their interrelationships, and exploration of all of the most promising combinations of explanatory variables. The general goodness of fit was measured by the sample correlation coefficient (r) or its square ( $r^2$ ). These quantities must improve, or at worst, stay the same, as each new explanatory variable is added to a prediction equation.

Other measures used, which allow for loss in degrees of freedom as each additional variable is used, were standard error of estimate (SEE) and total error squared (Cp) (5), which is calculated from the formula

$$C_p = (RSS/s^2) - (n-2p)$$

where RSS is the residual sum of squares;  $s^2$ , the SEE for the equation using all six variables; and p, one plus the number of variables used in the regression. Cp is a measure of both random errors about the fitted line and biases or specification error due to using a simplified model with fewer explanatory variables.

## RESULTS AND DISCUSSION

### Correlation Matrix

The first step in the analysis was the calculation of the correlation matrix for all 12 variables. The matrix is presented in Table III along with the means and standard deviations. The most striking feature of the matrix was that whereas D, Pn, and 100/V were all positively and relatively highly correlated with FA, they were all negatively correlated with baking absorption. D appeared to be the predominant factor. Hence, while increased values of D result in higher FA, the

**TABLE III**  
Means, Standard Deviations, and Correlations Between Variables<sup>a</sup>

Variables	Correlation Coefficients										
	G	G/P	D	Pn	100/V	FA	RBA.3	RBA.1	RBA.0	CBA.1	CBA.0
P	0.755** <sup>b</sup>	0.426*	-0.444*	-0.016	0.114	-0.084	0.662	0.762**	0.545**	0.741**	0.531*
G		0.471*	-0.283	-0.102	0.026	-0.001	0.529*	0.600**	0.571**	0.593**	0.576**
G/P			-0.747**	-0.796**	-0.795**	-0.644**	0.916**	0.880**	0.846**	0.886**	0.853**
D				0.651**	0.663**	0.927**	-0.772**	-0.716**	-0.393	-0.733**	-0.404
Pn					0.875**	0.697**	-0.600**	-0.551**	-0.537**	-0.558**	-0.530*
100/V						0.778**	-0.582**	-0.490*	-0.513*	-0.502*	-0.511*
FA							-0.578**	-0.480*	-0.203	-0.506*	-0.215
RBA.3								0.968**	0.821**	0.969**	0.834**
RBA.1									0.852**	0.987**	0.847**
RBA.0										0.831**	0.991**
CBA.1											0.829**
Mean	37.3	2.91	41.2	1.66	0.147	66.6	60.8	62.2	66.2	64.2	68.4
Standard Deviation	10.4	0.19	20.5	0.18	0.034	4.6	2.7	3.1	3.4	3.2	3.4

<sup>a</sup>See Table I for abbreviations.

<sup>b</sup>Significance at the 5, 1, and 0.1% levels is indicated by one, two, and three asterisks, respectively.

reverse may be true for baking absorption, particularly in longer methods and when a source of added  $\alpha$ -amylase is used.

By contrast, P and G were positively correlated with baking absorption but were not significantly correlated with FA, presumably due to the overriding influence of D. The relatively high negative correlation of G/P with FA was undoubtedly related to the fact that the flour streams with high D were generally those of lower quality having a lower gluten/total protein ratio.

Considering these relationships, we were not surprised that FA was negatively correlated with baking absorption.

### Investigation of Models for Predicting RBA.3

All possible regressions of RBA.3 were completely enumerated on combinations of the six explanatory variables. Baking absorption in the RBA.3 method was chosen for this detailed initial examination, since it was most likely to be different from FA. This baking method involves the longest time from initial mixing to absorption assessment (more than 3 hr) and the highest level of added malt  $\alpha$ -amylase, making the potential for dough softening quite high. Most of the "best" equations involved combinations of any two of P, G, and G/P; little was to be gained by considering other variables.

A summary of the most promising equations for predicting RBA.3 is presented in Table IV, along with their  $r^2$ , SEE, and Cp values. The three different criteria do not necessarily select the same equation as best. Using Cp as the criterion, one would choose the equation with P + G/P, while consideration of SEE would suggest the four-variable equation with P, G/P, D, and Pn. Because of the high correlation between P and G when both are used together, the value of the coefficient for P is generally negative—certainly a nonintuitive result. When either G/P + P or G/P + G are used, however, the coefficients are intuitively

TABLE IV  
Summary of Most Promising Equations for Predicting RBA.3<sup>a</sup>

No. of Variables	Regression Coefficient							$r^2$	SEE <sup>b</sup>	Cp <sup>c</sup>
	Constant	G/P <sup>d</sup>	G <sup>a</sup>	P <sup>a</sup>	D <sup>a</sup>	Pn <sup>a</sup>	100/V <sup>a</sup>			
1	23.5	12.9**** <sup>d</sup>	0.31***					0.839	1.10	20.6
	48.6							0.734	1.41	42.8
2	22.8	10.9***	0.16***	0.49***				0.929	0.752	2.21
	29.5	8.7***		-1.91***				0.928	0.753	2.26
	56.3	0.77***		0.924				0.777	3.30	
3	57.4	7.7***	0.67***	-1.64***	-0.020			0.934	0.742	2.88
	33.4		0.15***	-0.015	0.934			0.744	2.93	
	26.7		0.46***	-0.014	0.934			0.748	3.08	
4	55.3	9.9***	0.81***	-2.18***	-0.027	2.6		0.941	0.727	3.39
	24.4		0.12*	-0.020	2.3	0.940		0.735	3.68	
	19.2		11.6**	0.36	-0.019	2.3		0.939	0.740	3.83

<sup>a</sup>See Table I for abbreviations.

<sup>b</sup>SEE = standard error of estimate.

<sup>c</sup>Cp = total error squared.

<sup>d</sup>Significance at 5, 1, and 0.1% levels indicated by one, two, and three asterisks, respectively.

appealing. That is, each unit increase in P content contributes about 0.5 units of increase in absorption, and each change of one unit in G/P increases the absorption by 10 units. If G + G/P are considered, each unit increase in G contributes about 0.2-unit increase in absorption, or about one-third as much as a unit increase in P. This is consistent, because G is about three times the value of P. Equations involving G, P, and G/P exhibit erratic behavior of the coefficients, because these three variables are not independent.

#### Models for FA

The strongest single predictor variable for FA was D (Table V). Addition of P to D as a second variable produced a dramatic drop in SEE—from 1.8 to 0.42. Further addition of variables produced only a small further reduction in SEE, and the value of Cp was increased. The increase in FA amounted to about 1 percentage unit for every 4 Farrand units of increase in D, while each 1 percentage unit increase in P increased absorption by 1 percentage unit. By contrast, when the same two variables were used for predicting RBA.3, each 13 units of increase in D caused baking absorption to decrease by about 1 percentage unit.

TABLE V  
Summary of Most Promising Equations for Predicting Farinograph Absorption

No. of Variables	Regression Coefficients									
	Constant	G/P <sup>a</sup>	G <sup>a</sup>	P <sup>a</sup>	D <sup>a</sup>	Pn <sup>a</sup>	100/V <sup>a</sup>	r <sup>2</sup>	SEE <sup>b</sup>	Cp <sup>b</sup>
1	58.0				0.207***			0.897	1.77	290
2	42.7			1.03***	0.248***			0.993	0.416	2.07
	44.5		0.283*** <sup>b</sup>		0.268**			0.982	0.642	24.1
3	43.5			1.06***	0.254***	-0.88		0.993	0.411	2.70
	44.1		0.252***		0.240***		18.8**	0.993	0.414	2.92
4	44.6			1.02	0.249***	-1.72	6.8	0.994	0.410	3.72
	45.3		0.249***		0.240***	-1.00	23.2**	0.994	0.416	4.16

<sup>a</sup>See Table I for abbreviations.

<sup>b</sup>See footnotes, Table IV.

TABLE VI  
Influence of D as an Absorption Predictor When Used in Conjunction With P and G<sup>a</sup>

Absorption	Regression Coefficients						
	Constant	G <sup>a</sup>	P <sup>a</sup>	D <sup>a</sup>	r <sup>2</sup>	SEE <sup>b</sup>	Cp <sup>b</sup>
RBA.1	53.8	0.77*** <sup>b</sup>	-1.59***	-0.006	0.960	0.675	1.17
CBA.1	56.7	0.77***	-1.63***	-0.012	0.949	0.779	1.34
RBA.0	58.4	1.43***	-3.88***	0.092***	0.928	0.981	3.48
CBA.0	61.4	1.4***	-3.90***	0.088***	0.920	1.034	2.09

<sup>a</sup>See Table I for abbreviations.

<sup>b</sup>See footnotes, Table IV.

### Investigation of Models for RBA.1, RBA.0, CBA.1, and CBA.0

The general pattern was similar to that observed for RBA.3. Again, the best explanatory combinations were any two of P, G, and G/P. For each of the five baking absorptions, the third variable to enter the equation was D (Table VI). As the malt level decreased for both the Remix and Chorleywood methods, however, the influence of D changed from negative (although insignificant), through zero, to positive, and the regression coefficient became highly significant. Also, the tests performed with no malt were the most difficult to predict. They required three or four explanatory variables rather than only two, and the SEE was higher. For most of the samples, some malt was required in both baking systems to achieve adequate gas production and maximum loaf volume (1).

A summary of the best absorption prediction equations is presented in Table VII for all six absorption variables. This table emphasizes how the influence of D on absorption depends largely on the presence or absence of  $\alpha$ -amylase in the formula.

### CONCLUSIONS

The equations derived above for baking absorption would not necessarily be useful in predicting absorption for flours milled from wheats showing a wide range in quality. Individual absorption factors for damaged and undamaged starch, total protein, and gluten may vary. Our experience is that different wheats can produce flours with absorptions quite different from predicted values. This is particularly evident for varieties of inferior or superior quality, but may also apply to the effects of year, location, and weather damage on a given wheat variety.

Although the actual numerical values of the equation coefficients would likely be somewhat different for a different set of samples, we believe that the general observations concerning the interrelationships between farinograph water absorption, baking absorption, and major flour components are valid.

The results presented here confirm the findings of other workers (7,8) that as far as FA is concerned, both P and D have a direct positive effect on absorption and that use of values for both these factors can produce a good prediction of

TABLE VII  
Summary of Best Absorption Prediction Equations

Dependent Variable	Regression Coefficients					
	Constant	P <sup>a</sup>	G/P <sup>a</sup>	D <sup>a</sup>	SEE <sup>b</sup>	Cp <sup>b</sup>
RBA.3	22.8	0.49*** <sup>b</sup>	10.9***	...	0.752	2.21
RBA.1	18.9	0.81***	11.2***	...	0.676	0.02
RBA.0	- 7.0	0.60***	20.9***	0.104***	0.998	4.03
CBA.1	20.2	0.77***	11.6***	...	0.775	0.02
CBA.0	- 3.8	0.55**	20.9***	0.100***	1.03	1.97
FA	42.7	1.03***	...	0.248***	0.416	2.07

<sup>a</sup>See Table I for abbreviations.

<sup>b</sup>See footnotes, Table IV.



water absorption. When we used Farrand's (7) prediction equation ( $FA = 1.4P + 0.38D - [1.6M + 0.004D (M+P)] + 57.3$ ) with our set of data, we obtained predicted absorption values that were consistently lower than experimental values by about 1.5 percentage units. This constant difference was largely accounted for by the difference in dough consistency (600 versus 500 BU) in the two studies, and possibly also by a difference in farinograph bowls. The close agreement suggests that the same relation between P, D, and FA existed for our set of flour streams as for Farrand's set of commercial flours, streams, and air-classified fractions.

In contrast to the marked influence of D on FA, baking absorption in both long and short processes is largely a function of G content. D not only has little influence but also appears to have a negative effect on baking absorption, particularly when malt is used in the formula in a method involving bulk fermentation. This is consistent with conclusions from our earlier work with pin-milled flours (9). Differences in pentosan content did not appear to have a major influence on fluctuations in either farinograph or baking absorption.

Differences in amylograph viscosity for the samples in this study reflected largely the influence of D, and differences in  $\alpha$ -amylase activity between streams did not appear to have a major influence on absorption. The decrease in baking absorption with increasing addition of malt in both baking methods, however, illustrated the marked negative effect of  $\alpha$ -amylase activity. The additional initial water uptake of D is counteracted as the D is attacked by  $\alpha$ -amylase with time. The resultant dough softening forces the baker to reduce the amount of water added so that he can successfully handle the dough at makeup. Whereas British millers have successfully maximized CBA by judicious increases in the level of D combined with a carefully controlled optimum addition of malt (7), problems could arise if the ramifications of the relationships between flour properties, baking process, and malt addition are not fully appreciated.

#### Literature Cited

1. HOLAS, J., and TIPPLES, K. H. Factors affecting farinograph and baking absorption. I. Quality characteristics of flour streams. *Cereal Chem.* (in press).
2. SNEDECOR, G. W., and COCHRANE, W. G. *Statistical Methods*. Ed. 6. p. 412. Iowa State University Press: Ames (1967).
3. DANIEL, C., and WOOD, F. S. *Fitting Equations to Data*. Wiley: New York (1971).
4. MALLOWS, C. L. Choosing a subset regression. Presented at the Central Regional Meeting of the Institute of Mathematical Statisticians, Manhattan, KS (1964).
5. BOYCE, D. E., FARHI, A., and WEISCHEDEL, R. Optimal subset selection. In: BECKMAN, H. E., and KUZU, H. D. (eds.). *Economics and Mathematical Systems*. Springer-Verlag: New York (1974).
6. BEALE, E. M. L., KENDALL, M. G., and MANN, D. W. The discarding of variables in multivariate analysis. *Biometrika* 54: 357 (1967).
7. FARRAND, E. A. Starch damage and alpha amylase as bases for mathematical models relating to flour water absorption. *Cereal Chem.* 46: 103 (1969).
8. MEREDITH, P. Dependence of water absorption of wheat flour on protein content and degree of starch granule damage. *N.Z. J. Sci.* 9: 324 (1966).
9. TIPPLES, K. H., and KILBORN, R. H. Effect of pin-milling on the baking quality of flour in various breadmaking methods. *Cereal Sci. Today* 13: 331 (1968).

[Received November 4, 1976. Accepted January 3, 1978]