Evaluation of Critical Agro-Meteorological Parameters Affecting Wheat Quality on the Canadian Prairies using Partial Least Squares (PLS) Regression

Manasah Mkhabela[1]*, Paul Bullock[1] and Harry Sapirstein[2]
[1]Department of Soil Science, [2]Department of Food Science
University of Manitoba, Winnipeg, MB, R3T 2N2

ABSTRACT

The Canadian Prairies experience a wide range of growing season weather conditions, which consequently have a major impact on wheat quality. The objective of this study was to identify agro-meteorological factors that impact spring wheat quality using partial least squares (PLS) regression. Crop and weather data collected from several wheat experiments conducted in Manitoba and Saskatchewan from 2003 through 2006 were used in the study. Fifty three (53) agro-meteorological indices (predictor variables) were derived from the weather data and used in the PLS regression. Wheat quality characteristics (i.e., Grain Protein Content, Farinograph Absorption, Dough Development Time and Loaf-Volume) for two varieties (AC Barrie and Superb) were analysed and used as the response variables. Wheat quality data for AC Barrie were used to develop the PLS models, which were in-turn used to predict quality characteristics for Superb.

Results showed that 3 separate three-**variable** PLS models explained 83% of variability in Grain Protein Content, 80% of variability in Dough Development Time and 69% of variability in Loaf-Volume, respectively. Meanwhile, a **four-variable** PLS model explained 79% of the variability in Farinograph Absorption. When the developed PLS models were used to predict quality characteristics for the variety Superb, the overall average (mean) of the predicted values were not significantly different ($p>0.05$) from the overall average of the observed values for all quality characteristics except Farinograph Absorption. Predicted values correlated well with the observed values with R^2 values ranging from 0.69 for Loaf-Volume to 0.96 for Grain Protein Content, indicating that the PLS models accounted for 69 to 96% of the variability in the various wheat quality characteristics.

*Corresponding author: Manasah Mkhabela
Email: Manasah.Mkhabela@umanitoba.ca

INTRODUCTION

Spring wheat is one of the major crops grown on the Canadian Prairies, of which, about 80% is exported. However, wheat quality fluctuation in the Prairies is a major issue for both local and international wheat customers/buyers with expectations of consistent quality to meet the needs of the baking industry. Wheat dough quality is very important in baking industries for producing quality products according to consumers demand. Besides genetics and management, the environment, especially the growing season weather, has a significant impact on wheat quality. The Canadian Prairies experience a wide range of weather conditions within and between growing seasons. Thus, the overall objective of this study was to identify agro-meteorological factors that impact spring wheat quality in western Canada using partial least squares (PLS) regression. PLS regression is particularly suited when the matrix of predictors has more variables than observations and there is multicollinearity among the predictor variables (which is usually the case with weather variables).

MATERIALS AND METHODS

Crop and weather data collected from small-plot experiments (Figure 1) during the 2003 through 2006 growing seasons from Carman and Winnipeg in Manitoba and Regina, Swift Current and Melfort in Saskatchewan were used in the study. The experimental layout at all sites was a randomised complete block (RCB) design with three replicates. Fifty three (53) agro-meteorological indices (predictor variables) categorised into (i) Water Supply, (ii) Water Demand, (iii) Water Balance and (iv) Water Use were derived from the weather data and used in the partial least squares (PLS) regression. Wheat quality characteristics (response variables) including (i) Grain Protein Content (GPC), (ii) Farinograph Absorption (FarAB), (iii) Dough Development Time (DDT) and (iv) Loaf-Volume (LVol) for two varieties (AC Barrie and Superb) were analysed using standard laboratory methods and used in the PLS regression.

The number of PLS latent variables (components) was selected using a cross-validation method. Quality data for AC Barrie were used to develop the PLS models (training or calibration set), which were in-turn used to predict wheat quality characteristics for Superb (test or validation set). A prior ***t-test*** analysis had shown that there was no significant difference ($p>0.05$) between the quality characteristics of the two cultivars, except for Farinograph Absorption (Table 1). Variable importance of projection (VIP) was used to select the most influential predictor variables for each quality characteristic. VIP values (typically set at >0.8) describe the relative contribution of the predictor variables to the PLS latent variables. To avoid model over-fitting problems, predictor variables with VIP scores ≥1.3 were selected as the most influential in this study. Statistical Analysis and Model fitting were performed using R statistical packages.

Figure 1: Wheat small-plot experiments at different growth stages and weather recording instruments.

RESULTS AND DISCUSSION

Table 1: Summary statistics of quality characteristics for cultivars AC Barrie and Superb.

	Protein Content (%)	Farinograph Absorption (%)	Dough Development Time (minutes)	Loaf Volume (cc)
Average (mean):				
AC Barrie	14.54	60.69	5.13	961.44
Superb	14.06	62.48	5.29	1013.72
p-value	*0.44*	*0.03*	*0.84*	*0.18*
Standard Deviation:				
AC Barrie	1.67	2.04	2.07	123.67
Superb	1.71	2.34	2.23	76.13
Coefficient of Variation:				
AC Barrie	11.46	3.37	40.44	12.86
Superb	12.14	3.75	42.11	7.51
Range (Min & Max):				
AC Barrie	10.80-16.70	57.70-65.13	1.65-9.33	666.67-1145.83
Superb	10.47-16.83	59.20-67.80	2.20-9.33	833.33-1116.67

Three separate **3-latent-variable** PLS models explained 83% of the variability in wheat GPC and 86% of the variability in the predictor variables; 80% of the variability in DDT and 85% of the variability in the predictor variables; 69% of variability in LVol and 85% of the variability in the predictor variables. Meanwhile, a **4-latent-variable** PLS model explained 79% of the variability in FarAB and 89% of the variability in the predictor variables. Six, 4, 6 and 7 predictor variables were identified as the most influential factors (VIP≥1.3) affecting PC, FarAB, DDT and LVol variability, respectively.

When the developed models were used to predict quality characteristics for the variety Superb, the predicted values correlated very well with the observed values with R^2 values of 0.96 ($p<0.001$) for GPC, 0.75 ($p<0.001$) for DDT, 0.69 ($p<0.001$) for LVol and 0.84 ($p<0.001$) for FarAB. This indicates that the models explained from 69% to 96% of the variability in the four quality characteristics. Figs. 2 and 3 show the linear relationship between observed and predicted grain protein content and the linear relationship between observed and predicted farinograph absorption, respectively. A ***t-test*** showed that the overall average (mean) of the predicted values were not significantly different ($p>0.05$) from the overall average (mean) of the observed values for all the wheat quality characteristics except for FarAB, which was generally under-predicted (Fig. 3). The root mean square error (RMSE) values were 0.32 for GPC, 0.85 for DDT, 59.15 LVol, and 0.71 for FarAB.

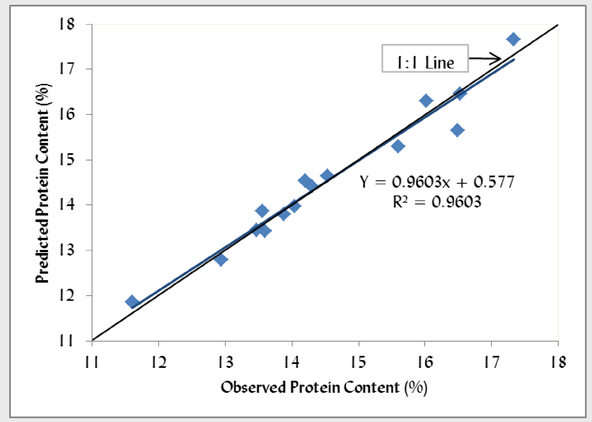

Figure 2: Linear relationship between observed grain protein content and predicted grain protein content for the cultivar Superb.

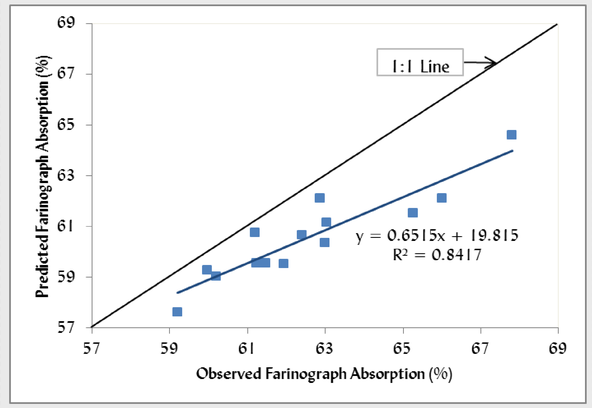

Figure 3: Linear relationship between observed farinograph absorption and predicted farinograph absorption for the cultivar Superb.

CONCLUSION

Sixty nine to 83% of the variability in the four wheat quality characteristics studied was explained by either a **3- or 4-latent variable** PLS model. The developed models were able to predict all quality characteristics for the variety Superb with reasonable accuracy, except for FarAB, which was generally under predicted. Overall predicted quality characteristics values were not significantly different ($p>0.05$) from observed values. The models will be further tested and validated using data that is currently being collected across the Canadian Prairies in a project that started during the 2015 cropping season.

ACKNOWLEDGEMENTS

Funding was provided by NSERC Strategic Grant and a CWB graduate fellowship. Assistance in conducting the field trials and grain quality analysis was provided by staff from AAFC, CGC and the University of Manitoba.