Assessment of Grain Protein Deviation In Hard Winter Wheat Germplasm

Sue Latshaw[a], Scott Haley[a], Merle Vigil[b], and Eric Storlie[a]

[a]Department of Soil & Crop Sciences, Colorado State University, Fort Collins, CO

[b]USDA-ARS, NPA Central Great Plains Research Station, 40335 County Rd. GG, Akron, CO

Abstract

Hard winter wheat producers maximize profits by growing varieties with high grain yield (YLD) and with adequate grain protein concentration (GPC) to satisfy marketing and end-use standards. Simultaneous selection for both traits is confounded by the commonly observed negative relationship between them. Additionally, substantial genotype by environment (GxE) effects reduce the accuracy of selection. Selection for both traits, in the face of large GxE interactions, may miss valuable genotypes.

We report on an assessment for deviations from the negative relationship between these traits, termed 'Grain Protein Deviation' (GPD) among a set of adapted winter wheat varieties and breeding lines (n=399) developed by the Colorado State University Wheat Breeding Program. We calculated the residuals of the regression of GPC on YLD across four environments in the 2011-12 growing season as the first step towards evaluating multi-environment trial data for these lines for the 2006-2013 growing seasons.

Estimates of GPD for entries from individual trial data were not stable across environments. Combined analysis of GPD and GPY calculated from entry means across environments may improve selection accuracy for high GPC and high YLD lines.

Objectives

- Calculate grain protein deviation among 399 adapted winter wheat varieties and advanced generation Colorado State University breeding lines in four environments.
- Compare grain protein yield (GPY) and GPD values calculated from a subset of 26 released varieties.
- Develop a method to accurately assess GPD and GPY to apply towards evaluation of their genetic architecture and for development of genomic selection models.

Methods

FIELD TRIAL EXPERIMENTAL DESIGN

2011-2012 (2 locations) Mean values for grain yield (kg ha^{-1}) and grain protein concentration (g kg^{-1}) were determined for adapted varieties and advanced generation CSU breeding lines developed from 2002-2009 (n=399). The experimental design was a latinized augmented row-column design within a split-block arrangement with nitrogen treatments as main plots and genotypes as subplots. Nitrogen (urea 46-0-0) was surface broadcast before planting at high and low rates for each location (Table 1). Harvested area was 2.8 m^2. The trial at Fort Collins, CO was under full irrigation, while the trial at Akron, CO was under dryland management. Precipitation totaled 11.1 cm at Akron from the planting to harvest dates.

GRAIN PROTEIN DEVIATION ESTIMATION

GPDs were initially estimated as the standardized residuals of the least squares regression of GPC on YLD (Monaghan et al. 2001). High positive values for GPDs were defined as those above 1.96, or the quantile of the Normal Standard for P=0.025.

GPD = $r_i/(s\text{-}hat*\sqrt{1-h_i})$ where, s-hat = $\sqrt{\{1/(n-p)\}*\Sigma\, r_i^2}$, r_i = residual for observation i, n = number of observations, p = number of estimated parameters, h_i = ith value of the diagonal of the hat matrix

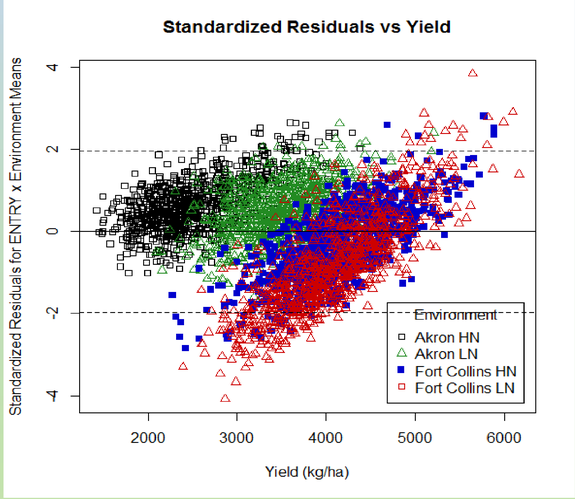
To improve trait estimates, the effects of outliers on the regression equation used for calculating GPDs were minimized to improve the accuracy of GPDs (Oury and Goudin 2007). Grain protein yield (GPY) was calculated as GPC*YLD. All procedures were performed in the R Programming language (R Development Core Team 2012).

Results

Table 1
Locations, nitrogen sources, and mean YLD and GPC for the four environments (environment = location + nitrogen rate).

Location	Residual Soil NO_3 (kg ha^{-1})	%Organic matter	Applied Nitrogen (urea)		Mean YLD (kg ha^{-1})	Mean GPC (g kg^{-1} DW)
			HN (kg ha^{-1})	LN (kg ha^{-1})		
Fort Collins	7.8	1.73	112	56	HN 4132	105
					LN 4069	100
Akron	25.7	1.5	112	22	HN 2541	151
					LN 3510	129

Figure 2
Scatterplot of standardized residuals (GPD) from the linear regression of GPC on YLD for 399 lines in four environments, plotted against YLD.



GRAIN PROTEIN DEVIATION FOR ENTRY MEANS IN EACH OF FOUR ENVIRONMENTS

GPDs were estimated as the standardized residuals of the least squares regression of GPC on YLD for entry means in the four environments. There were 186 entries with GPDs above 1.96 in at least one environment (n=147, 1 env; n=32, 2 env; n=6, 3 env; n=1, 4 env).

GPD ESTIMATION AFTER REMOVAL OF OUTLIERS FROM THE REGRESSION

GPDs were calculated for genotype means across all environments. The regression equation of GPC on YLD was iteratively recalculated after removing those standardized residuals with an absolute value that exceed a threshold of 2.5% (GPD>|1.96|). The iterations stopped when all points fell below the threshold (Figure 2). Predicted GPDs for entry means over the four environments were then calculated for the whole dataset from the new regression equation for the final trimmed data set (Figure 3).

Figure 2
Boxplots showing the quantiles of the standardized residuals for the original data and the six steps for trimming the data (threshold=|1.96|). Numbers of entries included and the R-squared values for the models are indicated.

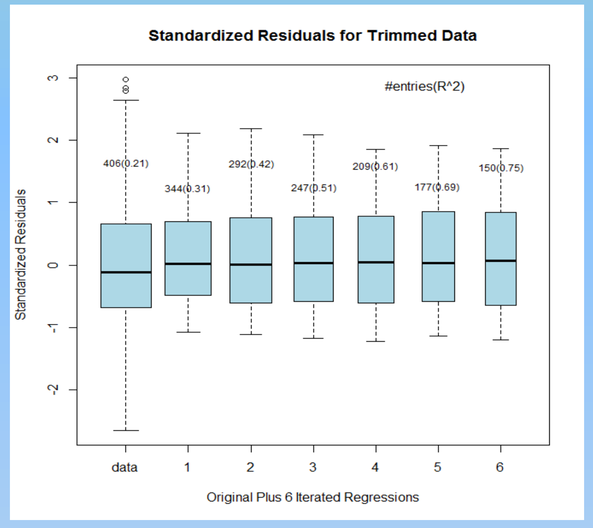
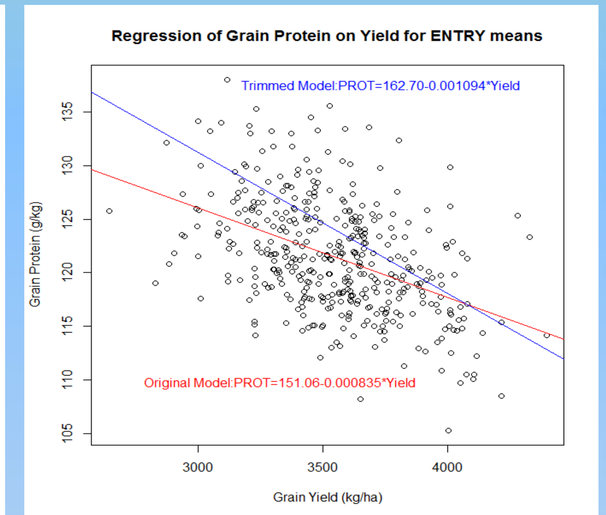


Figure 3
Scatterplot of entry means for YLD and GPC. Linear regression lines are plotted for the complete data (red, N=399) and for the final trimmed data (blue, N=150).

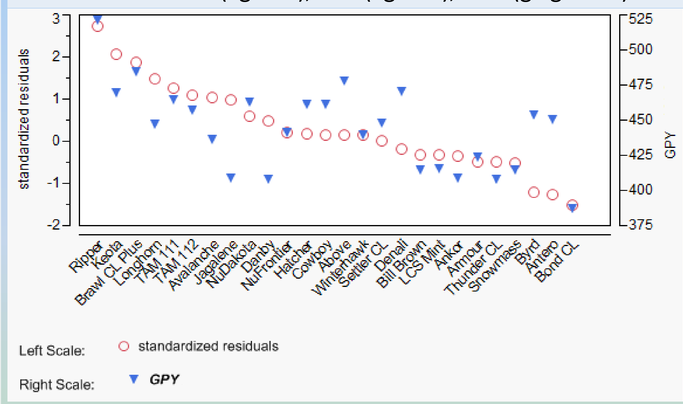


GPDs & GPY FOR RELEASED VARIETIES

GPY and predicted GPD for the 26 released varieties in the trial were plotted (Figure 4). The correlation of these traits was calculated (r=0.60, p=0.001, α=0.05). Genotype values for GPY averaged 443 kg protein ha^{-1} for these varieties.

- Ripper had the highest GPD and GPY, having both high YLD and GPC.
- Byrd and Antero had low GPDs and high GPY with high YLD, but very low GPC in the Fort Collins trials.
- Below average values of GPY were observed for Jagalene and Danby, although both had positive values for GPD, with low YLD and high GPC.
- Above and Denali each had above average YLD and below average GPC, resulting in relatively high GPY, but near zero GPD values.

Figure 4: Standardized residuals and GPY for 26 Varieties
GPY=YLD*GPC (kg ha^{-1}), YLD (kg ha^{-1}), GPC (g kg^{-1} DW)



Summary

A set of 399 genotypes (released varieties and breeding lines) were evaluated for GPD using data from four environments in this preliminary analysis. GPD evaluation may facilitate identification of lines with high values for both GPC and YLD. Entry rankings for GPD for were not stable across environments. Combined analysis of GPDs and GPY calculated from entry means across environments may enable simultaneous selection for high GPC and high YLD. The analysis will be expanded to a large set of multi-year, multi-site data for these lines to evaluate accuracy of trait estimation.

References

1. Monaghan, J. M., et al. (2001). The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. Euphytica 122(2): 309-317.
2. Oury, F.-X. and C. Godin (2007). Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes? Euphytica 157(1-2): 45-57.
3. R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.