# Comparison of global and local equations for bias within a structured data set

**Sam W. Coleman[a] and Jim Halgerson[b]**

[a]*USDA, ARS, Subtropical Agricultural Research Station, 22271 Chinsegut Hill Rd., Brooksville, FL 34601, USA*

[b]*Agronomy and Plant Genetics, Univ. Minnesota,St. Paul, MN 55108, USA*

## Introduction

Plant breeders and other researchers create large data sets of similar samples that are suited for analysis by NIRS. Often, however, one wishes to use archived databases or those from the same experiment in earlier years to predict new samples or samples somewhat different from those in the calibration data set. As the technology has matured, equations developed and used by agricultural scientists in both research and advisory capacities have become broader and more diverse. While broad and diverse data sets have been successfully used[1], precision may be sacrificed for robustness. Buxton and Mertens[2] reported internal bias within a data set caused by classification variables such as plant species, harvest date, year, etc., that could be identified for the experiment. One objective of this study was to examine the structure within a data set (plant genetic differences and plant maturity) for bias using two calibration methods, global equations and locally weighted regressions. Another objective was to investigate whether internal structure could be used to evaluate the robustness of equations.

## Methods

The data set consisted 240 total samples from a small grain forage variety test with 10 cultivars clipped over six harvest dates (plant maturity), and four replicates of each. To examine the data set for internal bias among cultivars and plant maturity, all samples were used for a global calibration with modified partial least squares[3] using software supplied by InfraSoft International, Port Matilda, PA[+]. Two math treatments, 1,4,4,1 and 2, 12, 8, 1, both without scatter correction were used. Due to similarities in results, only data from 1,4,4,1 will be reported. All samples were used to calculate an equation for each analyte, and the equations were then used to predict all samples (ALLPLS).

To evaluate the robustness of prediction methods an alternate method was used. Two data sets were generated for each level of the structural attribute (cultivar), one with only the level of the target cultivar ($C_i$) and one with all levels except the target cultivar (N-$C_i$), resulting in 20 different data sets (Table 1). Global equations for each attribute were developed for each of the ten N-$C_i$ data sets (CULGBL), and were used to predict the unknown samples from the eliminated cultivar ($C_i$). This has the advantage of testing the equation on samples that are unrelated to the calibration samples in some attribute and should provide a more robust evaluation than software routine that randomly eliminates one sample at a time. For comparison, locally weighted regression[4] was

[+] Mention of a trade name is solely to inform the reader and does not imply endorsement or recommendation by the USDA.

performed on samples in each of the 10 different cultivar ($C_i$) data sets using the N-Ci samples from the remaining nine cultivars (CULLOC) as a reference library, thus maintaining the same unrelatedness as with global equations above. For each NIRS method, differences were calculated between the laboratory reference values and the predicted values.

**Table 1. Method for using internal structure to examine robustness and bias within a data set.**

| Calibration samples (N-$C_i$) | Evaluation samples ($C_i$) |
|---|---|
| Cultivar 1-9 | Cultivar 10 |
| Cultivar 1-8, 10 | Cultivar 9 |
| Cultivar 1-7, 9-10 | Cultivar 8 |
| … | … |
| Cultivar 2-10 | Cultivar 1 |

Three analytes were evaluated representing different levels of chemical specificity. From previous experience, they also provided different levels of relationship to NIR spectra. Crude protein (CP) is a specific chemical element determined by the kjeldahl[5] method as N*6.25. Neutral detergent fiber[6] (NDF) is a gravimetric assay that estimates the proportion of plant cell walls. In vitro dry matter digestibility[7] (IVDMD) is a routine bioassay to estimate digestibility and depends on ruminal microorganisms.

Bias was investigated in two ways using SAS[8]. Analysis of variance was used to determine if the differences between reference values and NIRS predicted values were significantly different among cultivars and harvest dates. The second test sought to determine if the cultivars ranked the same way using the different methods for determination

## Results

Residual standard errors (SE) were calculated for each method and analyte by fitting the statistical model including rep, cultivar, harvest date, and all 2x interactions (Table 2). Standard errors for NIRS methods were similar, or in some cases smaller, than those for the reference method. There were also smaller SE for locally weighted regression than for global equations for every analyte.

**Table 2. Standard errors among field replicates and bias in prediction with NIRS based on different methods.**

| | Crude protein | | Neutral-detergent fiber | | In vitro digestibility | |
|---|---|---|---|---|---|---|
| Method | SE[a] | Bias[b] | SE | Bias | SE | Bias |
| Reference | 0.61 | -- | 1.5 | -- | 1.86 | -- |
| ALLPLS[c] | 0.63 | 0 | 1.45 | -0.03 | 1.42 | 0.04 |
| CULGBL[d] | 0.61 | 0.01 | 1.33 | -0.07 | 1.81 | -0.23 |
| CULLOC[e] | 0.55 | -0.07 | 1.11 | 0.06 | 1.18 | -0.14 |

[a] Standard error of field replicates. Includes variation in sampling and method.
[b] Difference in reference analysis and that predicted by NIRS method.
[c] ALLPLS - prediction of all samples from an equation calculated from the same samples.
[d] CULGBL - prediction of a specific cultivar from a global PLS equation calculated using the remaining cultivars.
[e] CULLOC - prediction of samples of a specific cultivar using locally weighted regression based on a database of all samples except those within the predicted cultivar.

While average bias (Table 2) should theoretically be zero for the ALLPLS method since all samples were used for calibration and prediction, bias was different among harvest dates for CP and among cultivars and harvest dates for the less defined attributes NDF and IVDMD. This is similar to the method employed by Buxton and Mertens (1991) when internal bias was first reported with predictions from NIRS, and suggests that even within the data set, spectral data are not consistently related to the analytical data. It is possible that the reference method may contribute to some of the internal bias, as evidenced by increased number of significant sources for less chemically defined analytes (e.g., NDF and IVDMD).

When cultivars were left out of the calibration and equations used in an extrapolation mode (CULGBL and CULLOC), average bias was similar for the two prediction methods, and was nominal. All treatment effects (replicate cultivar, harvest date and interactions) contributed significant bias. Increased bias could be expected, especially for those cultivars that contain the least and most CP, NDF and IVDMD. When these extreme cultivars were not represented in the calibration samples, extrapolations beyond the scope of the data were evidenced. The structural (i.e., cultivar) elimination technique is a good measure of the robustness of the data set and the may provide evidence toward its ability to predict other samples. In almost every case, the minimum and maximum bias (Table 3) among cultivars was within the standard errors for the experiment (field reps; Table 1). Therefore, we might conclude that because of the very small error among samples within a cultivar, small internal bias due to analysis method is statistically significant. However, it is likely not biologically significant.

**Table 3. Range in bias among cultivars as predicted by different methods.**

| | Crude protein | | Neutral-detergent fiber | | In vitro digestibility | |
|---|---|---|---|---|---|---|
| Method | Low | High | Low | High | Low | High |
| ALLPLS[c] | -0.08 | 0.11 | -0.55 | 0.65 | -0.93 | 1.26 |
| CULGBL[d] | -0.44 | 0.30 | -1.14 | 0.94 | -1.94 | 2.57 |
| CULLOC[e] | -0.31 | 0.27 | -0.32 | 0.88 | -1.84 | 2.10 |

[a]ALLPLS - prediction of all samples from an equation calculated from the same samples.
[b] CULGBL - prediction of a specific cultivar from a global PLS equation calculated using the remaining cultivars.
[c] CULLOC - prediction of samples of a specific cultivar using locally weighted regression based on a database of all samples except those within the predicted cultivar.

Another way to assess the biological significance of internal bias involves ranking the cultivars and see if method influences the decision a plant breeder might make concerning which samples to keep or discard. Cultivar rankings and groupings by Duncan's multiple range test (Table 4) demonstrate that once again, the specificity of the analyte has caused different degrees of discrepancy. For CP, ALLPLS prediction with NIRS ranked cultivars almost exactly as the reference kjeldahl method. However, as the reference method became less chemically defined, then some ranking differences were observed. This implies the relationship between spectral and nutritional attributes were different among the internal structural components of the data set. Specifically, the four cultivars with the highest NDF belonged to the same statistical group (superscript a) when predicted with the ALLPLS equation, but cultivars 8 and 5 were different from 7 and 4 by the reference method. All four were statistically different from one another for both reference and ALLPLS CP concentration. The cultivars with low NDF were better separated using ALLPLS than with the reference method. Cultivar 9 had greater IVDMD ($P < .05$) than cultivar 10

by the reference method, but with the ALLPLS prediction, but they were reversed and not statistically different.

Greater deviations from reference methods were observed when cultivars were eliminated from the calibration sample set. The highest ranked cultivar for crude protein was the same for all methods, but the CULGBL procedure reversed the cultivars with the most NDF. Both CULxxx methods ranked cultivar 10 with the greatest concentration of IVDMD, similar to ALLPLS, but different from the reference method. The reason is that minimum and maximum internal bias among cultivars involved these two cultivars (Figure 1).

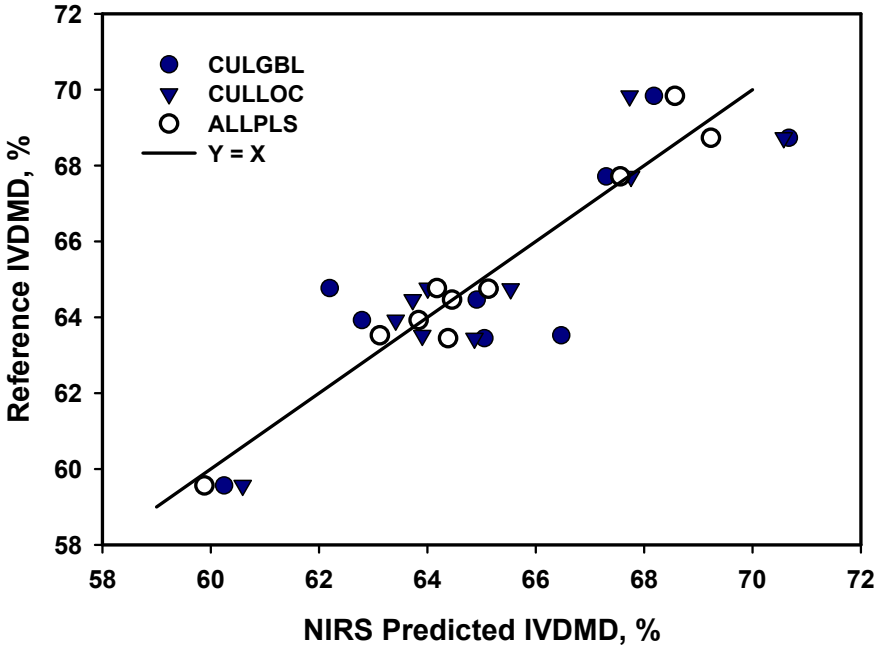


**Figure 1. Relationship of IVDMD determined by reference and three NIRS methods. Cultivar 9 had highest IVDMD by reference method whereas cultivar 10 had highest value by all three NIRS methods.**

Cultivar 9 had the greatest amount of IVDMD by the reference method, and had the greatest negative bias (-1.94 g $kg^{-1}$ for CULGBL and –1.84 g $kg^{-1}$ for CULLOC, respectively). In contrast, cultivar 10 had the greatest positive bias (2.57 for CULGBL and 2.10 g $kg^{-1}$ for CULLOC, respectively). While these levels of bias were within the SE, they were enough to result in a reversal of the cultivars. The CULLOC method also placed cultivar 9 below cultivar 3, the third ranked cultivar for other methods. While these differences could be troubling, depending on the level of specificity and precision of the reference analysis, most plant breeders would probably take the best 3 or 4 cultivars for further evaluation, rather than only one, because selection is seldom based on a single trait. While the three attributes investigated in this research ranked differently by the reference method, the three NIRS methods ranked cultivar 10 as the best according all analytes.

**Table 4. Ranking of cultivars for crude protein concentration by different methods of analysis.**

| Crude Protein | | | | Neutral Detergent Fiber | | | | In vitro digestibility | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reference | ALLPLS | CULGBL | CULLOC | Reference | ALLPLS | CULGBL | CULLOC | Reference | ALLPLS | CULGBL | CULLOC |
| $10^{a}$ | $10^{a}$ | $10^{a}$ | $10^{a}$ | $8^{a}$ | $8^{a}$ | $7^{a}$ | $8^{a}$ | $9^{a}$ | $10^{a}$ | $10^{a}$ | $10^{a}$ |
| $9^{b}$ | $9^{b}$ | $3^{a}$ | $9^{a}$ | $5^{abc}$ | $5^{a}$ | $8^{a}$ | $7^{a}$ | $10^{b}$ | $9^{a}$ | $9^{b}$ | $3^{b}$ |
| $2^{b}$ | $3^{b}$ | $2^{b}$ | $3^{a}$ | $7^{bcd}$ | $7^{a}$ | $4^{a}$ | $4^{ab}$ | $3^{b}$ | $3^{b}$ | $3^{c}$ | $9^{b}$ |
| $3^{b}$ | $2^{b}$ | $9^{b}$ | $2^{b}$ | $4^{bcde}$ | $4^{a}$ | $5^{a}$ | $5^{ab}$ | $4^{c}$ | $6^{c}$ | $1^{c}$ | $6^{c}$ |
| $8^{c}$ | $8^{c}$ | $1^{c}$ | $8^{c}$ | $2^{de}$ | $6^{b}$ | $6^{b}$ | $2^{bc}$ | $6^{c}$ | $7^{cd}$ | $6^{d}$ | $2^{c}$ |
| $1^{d}$ | $1^{d}$ | $8^{c}$ | $1^{d}$ | $6^{de}$ | $3^{b}$ | $1^{bc}$ | $1^{c}$ | $7^{cd}$ | $2^{cd}$ | $2^{d}$ | $4^{d}$ |
| $7^{d}$ | $7^{d}$ | $7^{d}$ | $7^{e}$ | $1^{ef}$ | $1^{b}$ | $3^{c}$ | $6^{c}$ | $8^{cd}$ | $4^{d}$ | $7^{d}$ | $1^{d}$ |
| $6^{e}$ | $6^{e}$ | $6^{d}$ | $6^{e}$ | $3^{ef}$ | $2^{bc}$ | $2^{c}$ | $3^{c}$ | $1^{d}$ | $8^{de}$ | $8^{e}$ | $7^{d}$ |
| $4^{f}$ | $4^{f}$ | $4^{e}$ | $4^{f}$ | $9^{fg}$ | $9^{cd}$ | $9^{d}$ | $9^{d}$ | $2^{d}$ | $1^{e}$ | $4^{e}$ | $8^{d}$ |
| $5^{g}$ | $5^{g}$ | $5^{f}$ | $5^{g}$ | $10^{g}$ | $10^{d}$ | $10^{d}$ | $10^{d}$ | $5^{e}$ | $5^{f}$ | $5^{f}$ | $5^{e}$ |

$^{abcdefg}$ Numbers in the same column with similar superscript are not different ($P < .05$).

## Conclusion

Nutritive value attributes of forage predicted by NIRS methods had equal or lower residual error than determinations by the reference method. While internal bias existed, it was within the overall SE for that analyte. When CULGBL and CULLOC predictions were compared to reference values, some predicted cultivar means were ranked differently from the laboratory values, but most reversals were within a statistical grouping. The bias torque was applied by the extreme (best and worst) cultivars. However, one would essentially draw the same conclusions concerning the best cultivars regardless of analytical method, including the reference method.

Local predictions produced slightly lower range in bias among cultivars and among harvest dates than global equations. Local predictions produced lower residual mean square errors than global equations and predictions based on the same calibration and evaluation data sets. Therefore locally weighted regression may provide some advantage over global equations when extrapolating predictions outside of the structural parameters of the calibration data set.

## References

1. S.M. Abrams, J.S. Shenk, M.O. Westerhaus, and F.E. Barton II. J. Dairy Sci. 70, 80 (1987)
2. D. R. Buxton and D. R. Mertens, Crop Sci.,31, 212 (1991)
3. H.A. Martens and S.A. Jensen, in Proc. 7th World Cereal Congr. J. Holas and J. Kratchovil (Eds)., Elsevier, Amsterdam p. 607, (1982).
4. T. Isaksson and T. Naes, in Making Light Work: Advances in Near Infrared Spectroscopy. Ian Murray and Ian A. Cowe (Eds)., VCH, Weinheim, FRG., p. 140, (1991).
5. AOAC 15th Edn. Association of Official Analytical Chemists, Arlington, VA, USA (1990)
6. H.K. Goering, and P.J. Van Soest, Agric. Handbook No. 379. USDA, ARS, Washington, DC, USA (1970)
7. Marten and R.F. Barnes, in Standardization of Analytical Methodology for Feeds, W.J. Pigden, C.C. Balch and M. Graham (Eds). p. 49, (1979).
8. SAS User's Guide: Statistics. (1990). SAS Inst., Inc. Cary, NC.