CROP YIELD RELATIONSHIP TO REMOTE SENSING DATA USING INTENSIFIED WEIGHTED NONLINEAR REGRESSION MODELS

H. Bu, D. Franzen, and L. Sharma
North Dakota State University, Fargo, ND

Abstract

Yield prediction is important for making in-season agronomic input decisions as well as for greater logistical decisions. In predicting the crop yield based on ground-based active optical sensing data, the ordinary statistical unweighted linear or nonlinear regression models are the most popular choices. However, these unweighted models may not be accurate enough for practical use because they are based on the assumption that each data point for regression is obtained with equal precision and that each data point contributes equally to the model construction. Using unweighted models relying on the average sensing information alone, some important sensing information such as the variation of the sensor readings within each subplot, is left unmined. To improve the performance of the prediction models, feasibility of developing and applying weighted nonlinear regression models was explored. Novel intensified weights were developed based on the coefficient of variation of sensing data within each subplot. The experiments involved two crops, spring wheat and corn; two sites for each crop, using two ground-based active optical sensors, GreenSeeker™ and Holland Crop Circle™, two NDVI-based crop indices, red NDVI and red edge NDVI, and two general types of regression models, exponential and quadratic (weighted or unweighted) models. Results indicated that the proposed intensified weighted nonlinear regression models outperform their corresponding unweighted regression models in terms of R^2. Results also showed that this methodology did not improve predictions using pooled multiple site-year data. The reason may be that pooled data consisted of greater sample number enabling unweighted regression models to yield a more stable and significant relationship.

Introduction

Remote sensing technologies, such as ground-based active optical sensing and satellite imaging, enable people to collect and analyze data in a non-destructive, efficient, non-costly, and effective way, and therefore have been widely applied to early-season crop yield prediction and fertilizer management practices. In predicting the crop yield based on ground-based active optical sensing data, the ordinary statistical unweighted linear or nonlinear regression analyses are the most popular choices. One of the common assumptions underlying most unweighted regression modeling methods, including linear and nonlinear least squares regression, is that each data point provides equally precise information about the deterministic part of the total process variation. In other words, the standard deviation of the error term is constant over all values of the predictor or explanatory variables (NIST/SEMATECH, 2012). Obviously, this ideal situation is extremely difficult to arrive at in reality. Extensive soil sampling, optical sensor measurements of plants, and geostatistical analyses, indicated that statistically significant differences in nitrogen availability existed at a 1 m2 spatial field (Raun et al., 1998; Solie et al., 1999). In situations like this, when it may not be reasonable to assume that every observation should be treated equally, information of coefficient of variation (CV) of sensor readings may be a useful aid to improving regression models performance. CV information has been used to improve the crop in-season N recommendation algorithm (Raun et al., 2005). This paper explores an alternative way of using sensor readings' CV-based information in early-season crop yield prediction by incorporating it into weighted nonlinear regression to maximize the efficiency of parameter estimation. This is done by attempting to give each data point its proper amount of influence over the parameter estimates. A procedure that treats all of the data equally would give less precisely measured points more influence than they should have and would give highly precise points too little influence.

Materials and Methods

Experimental setup and data collection

Spring wheat and corn data generated from N rate trials during 2012 were used in the analyses. The two spring wheat and two corn experimental sites were Gardner and Valley City, and Durbin and Valley City, in North Dakota respectively. The experimental design of each N-rate study as a random complete block design with four replications and six N treatments. The experimental unit within each experimental site was 30 feet by 30 feet. Two ground-based active optical sensors, GreenSeeker® (NTech Industries, Inc., Ukiah, CA, USA) and Crop Circle® (Holland Scientific Inc., Lincoln, Nebraska, USA), were used to collect crop canopy NDVI ((near infrared-red) / (near infrared + red)) data at the 4-leaf stage (Feekes 4) for spring wheat yield prediction and V6 and V12 stages for corn yield prediction. The in-season estimate of yield (INSEY), essentially an estimate of the rate of accumulated biomass, was calculated by dividing NDVI by positive accumulated GDD from planting date (Stone, et al. 1996) and then used as the independent variable in statistical regression. The GreenSeeker provided one red NDVI and the Crop Circle provided one red NDVI and one red-edge NDVI. Hence we have one GreenSeeker red INSEY (GSINSEY), one Crop Circle red INSEY (CCINSEY), and one Crop Circle red-edge INSEY (CCREINSEY). Dry grain yield was used as the dependent variable in each regression model.

Intensified weighted nonlinear regression

In performing statistical regression, the best-fit curve is often assumed to be that which minimizes the sum of squared residuals. This is the ordinary or unweighted least squares approach. However, in cases where the dependent variable does not have constant variance a sum of weighted squared residuals may be minimized: $S = \sum_{i=1}^{n} r_i^2$ and there are many options to solve this optimization problem (Bukac, 2008). Since CV is a kind of normalized standard deviation reflecting the variability of sensor readings, it is a good candidate to be used as the weight in regression. The initial weights ω_i(i=1, 2, .., n) for constructing weighted regression models are defined below:

$$\omega_i = {T_i}/{\sum_i T_i}$$

$$T_i = 1/CV_i, \quad i = 1,2,\dots,n$$

where $n \in \mathbb{N}$ is the number of data points involved in the regression. To further strengthen the impact of those subplots each with smaller sensor reading variations and weaken the influence of those subplots each with larger sensor reading variations, a series of intensified weights based on the initial weights were defined:

$$v_i = {\omega_i^k}/{\sum_i \omega_i^k}, \quad i = 1,2,\dots,n,$$

where $k \in \mathbb{N}$ and $k \geq 2$ is the power of the initial weights. For each determined k we have a corresponding set of weights. We call these new weights the Intensified Weights. For simplicity, we call the initial weight set W1, the intensified weight set with $k = 2$ W2, the intensified weight set with $k = 3$ W3, and so on.

Analyzing Methods

Exponential function $y = a \cdot e^{b \cdot x}$ and polynomial quadratic function $y = a \cdot x^2 + b \cdot x + c$ were adopted in this study to compare the performance of the proposed method with that of the corresponding unweighted regression models. All these statistical weighted or unweighted regression models were built and analyzed using Matlab 8.0 (The MathWorks Inc., 2012) based on weighted or nonweighted least squares method. Model statistical significance and R^2 were used as the indicators of the performance of each regression model.

Results, Discussion, and Summary

Since our experimental results indicated that in most cases the polynomial quadratic regression models outperform the corresponding exponential regression models, only the polynomial quadratic regression results in terms of R^2 are listed in Table 1 through Table 3, with Table 1 being the regression results for spring wheat and the other two tables being the results for V6 corn and V12 corn, respectively. Weights of W1 and W8 were selected to incorporate into the regression models. W0 in the tables represents unweighted regression, and NS means the model is not significant at the 0.05 level of confidence. The meanings of the other abbreviations in these tables are listed below:

GGR: Gardner GreenSeeker red INSEY
GCR: Gardner Crop Circle red INSEY
GCRE: Gardner Crop Circle red edge INSEY
VGR: Valley City GreenSeeker red INSEY
VCR: Valley City Crop Circle red INSEY
VCRE: Valley City Crop Circle red edge INSEY
GVGR and DVGR: two-site pooled GreenSeeker red INSEY
GVCR and DVCR: two-site pooled Crop Circle red INSEY
GVCRE and DVCRE: two-site pooled Crop Circle red edge INSEY

Two figures each comparing the regression effects among unweighted, W1-weighted, and W8-weighted are also given below, with Figure 1 concerning the relationships between Valley City wheat GreenSeeker red INSEY and wheat dry grain yield and Figure 2 being the relationships between Durbin V6 corn Crop Circle red edge INSEY and corn dry grain yield.

Table 1. Regression results for spring wheat wheat vs INSEY in terms of R^2.

weight	GGR	GCR	GCRE	VGR	VCR	VCRE	GVGR	GVCR	GVCRE
W0	0.5175	0.3236	0.3526	0.3874	0.2949	0.3030	0.5226	0.3362	0.2439
W1	0.5543	0.3845	0.4086	0.3966	0.3010	0.3116	0.5181	0.3278	0.2262
W8	0.6847	0.6026	0.7107	0.6590	0.3811	0.4046	0.5346	0.3687	0.1757

Table 2. Regression results for 6-leaf corn yield vs INSEY in terms of R^2.

weight	DGR	DCR	DCRE	VGR	VCR	VCRE	DVGR	DVCR	DVCRE
W0	0.1204 (NS)	0.1825 (NS)	0.2133	0.0527 (NS)	0.0648 (NS)	0.0984 (NS)	0.2642	0.6961	0.5826
W1	0.1554 (NS)	0.2495	0.2801	0.0641 (NS)	0.0568 (NS)	0.0822 (NS)	0.2441	0.6754	0.5813
W8	0.3807	0.6132	0.6827	0.2620	0.0206 (NS)	0.0094 (NS)	0.1111 (NS)	0.6332	0.5493

Table 3. Regression results for 12-leaf corn in terms of R^2

weight	DGR	DCR	DCRE	VGR	VCR	VCRE	DVGR	DVCR	DVCRE
W0	0.0966 (NS)	0.1315 (NS)	0.1502 (NS)	0.0173 (NS)	0.0935 (NS)	0.1249 (NS)	0.2305	0.2117	0.3110
W1	0.1008 (NS)	0.1142 (NS)	0.1308 (NS)	0.0121 (NS)	0.1190 (NS)	0.1450 (NS)	0.1775	0.1426	0.2281
W8	0.0854 (NS)	0.6487	0.4413	0.3043	0.3690	0.3733	0.0337 (NS)	0.1022	0.1787

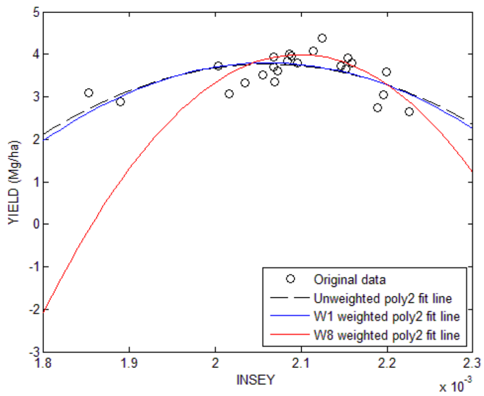

Figure 1. Valleycity wheat GreenSeeker Red INSEY regression models comparison

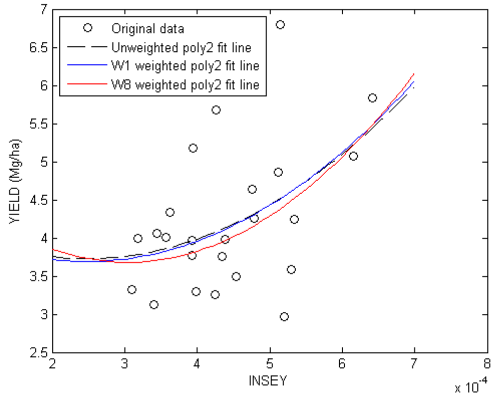

Figure 2. Durbin Corn Crop Circle Red edge INSEY regression models comparison

From Table 1 it can be seen that for each single site regression, the intensified weight W8 greatly helped improve the spring wheat regression models' R^2 performance compared to the unweighted regression, while W1 didn't help much. The performance of the w8-weighted regressions for pooled two wheat sites, however, displayed inconsistency. Table 2 and Table 3 revealed that for single-site corn yield prediction using either V6 or V12 sensing data, in most cases the unweighted and W1-weighted regression models were not statistically significant with very small R^2 value. A main reason for this is that there was the extreme drought weather in the year of 2012. But in many cases, the W8-weighted regression models significantly outperformed their corresponding unweighted or W1-weighted regression models in that the models became significant and the R^2 values greatly increased. Again, the W8-weighted regression models didn't improve the pooled two corn sites regression performance; rather, the W8 weight decreased the R^2 value.

The poor performance of intensified weighted nonlinear regression for pooled data in this study probably owed to two facts. One may be that pooled data consisted of greater sample number enabling unweighted regression models to yield a more stable and significant relationship, and the other may be that there were that the differences in crop growth and in turn sensing differences between each two sites were too great, confounding the effect of weighting. Therefore developing a more versatile series of weights for use in weighted regression models for crop yield prediction is a still a challenge. However, the exercise strengthens our confidence that an unweighted approach to relating yield and INSEY is a valid approach to establishing yield prediction in spring wheat and corn at an early growth stage.

References

Bukac, J., 2008. Weighted Nonlinear Regression. Analysis in Theory and Applications, 24: 330–335.

NIST/SEMATECH, 2012. e-Handbook of Statistical Methods: Engineering statistics handbook. http://www.itl.nist.gov/div898/handbook/, Date created: 6/01/2003, Last updated: 4/01/2012.

Raun, W.R., J.B. Solie, G.V. Johnson, M.L. Stone, R.W. Whitney, H.L. Lees, H. Sembiring, and S.B. Phillips, 1998. Micro-variability in soil test, plant nutrient, and yield parameters in bermudagrass. Soil Sci. Soc. Am. J. 62: 683–690.

Raun, W.R., J.B. Solie, M.L. Stone, K.L. Martin, K.W. Freeman, R.W. Mullen, H. Zhang, J.S. Scheppers, and G.V. Johnson, 2005. Optical sensor-based algorithm for crop nitrogen fertilization. Commun. Soil Sci. Plant Anal. 36: 2759–2781.

Solie, J.B., W.R. Raun, M.L. Stone, 1999. Submeter spatial variability of selected soil and bermudagrass production variables. Soil Sci. Soc. Am. J. 63: 1724–1733.

The MathWorks, Inc. (2012). Matlab R2012b, MA, USA.

PROCEEDINGS OF THE

43^{rd}

NORTH CENTRAL EXTENSION-INDUSTRY SOIL FERTILITY CONFERENCE

Volume 29

November 20-21, 2013
Holiday Inn Airport
Des Moines, IA

PROGRAM CHAIR:

Carrie Laboski
University of Wisconsin
1525 Observatory Dr.
Madison, WI 53706-1207
(608) 263-2795
laboski@wisc.edu

PUBLISHED BY:

International Plant Nutrition Institute
2301 Research Park Way, Suite 126
Brookings, SD 57006
(605) 692-6280
Web page**: www.IPNI.net**

ON-LINE PROCEEDINGS:

http://extension.agron.iastate.edu/NCE/