Crop Yield Relationship to Remote Sensing Data Using Intensified Weighted Nonlinear Regression Models

Honggang Bu, David Franzen, and Lakesh Sharma,
Department of Soil Science, North Dakota State University, Fargo, ND, 58105

Introduction

Remote sensing technologies have been widely applied to early-season crop yield prediction and fertilizer management. In predicting the crop yield based on ground-based active optical sensing data, the ordinary statistical unweighted regression analyses are the most popular choices, which assume that each data point provides equally precise information about the deterministic part of the total process variation. Obviously, this ideal situation is extremely difficult to arrive at in reality. In situations where 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 study 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.

Experimental Setup and Data Collection

Crops: 2012 spring wheat and corn
Spring wheat sites: Gardner and Valley City
Corn sites: Durbin and Valley City
Experimental design: RCBD with 4 replications and 6 nitrogen rate treatments (0, 45, 90, 135, 180, 225 Kg/ha)
Plot size: 9.14m×9.14m
Sensors: GreenSeeker® (NTech Industries, Inc., Ukiah, CA, USA) and Crop Circle® (Holland Scientific Inc., Lincoln, Nebraska, USA)
Sensing stages: spring wheat Feekes 4; Corn V6 and V12
Index used: In Season Estimate of Yield (INSEY) (Stone, et al. 1996)
INSEY=NDVI/accumulated GDD

Intensified Weighted Nonlinear Regression

Constructing appropriate weights is the most important step in establishing weighted nonlinear regression. The initial weights ω_i(i=1, 2, .., n) for constructing weighted regression models are defined below:

$$\omega_i = {}^{T_i}\!/_{\sum_i T_i}$$

$$T_i = 1/CV_i,\ \ i = 1, 2, \ldots, n$$

where $n \in \mathrm{N}$ 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:

$$v_i = {}^{\omega_i^k}\!/_{\sum_i \omega_i^k},\ i = 1, 2, \ldots, n,$$

where $k \in \mathrm{N}$ and $\mathrm{k} \geq 2$ 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 use W1 to indicate models with initial weight, W2 to indicate models with weight set of *k*=2, W3 to indicate models with weight set of *k*=3, and so on. Specially, we use W0 to indicate unweighted regression.

Analyzing Methods

Basic models: $y = a \cdot e^{b \cdot x}$ and $y = a \cdot x^2 + b \cdot x + c$
Objective: Comparison between proposed weighted models and unweighted regression models
Statistical software used: Matlab 8.0 (The MathWorks Inc., 2012)
Indicators of model performance: Model statistical significance and R^2

Meaning of the abbreviations used

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

Results

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

Note 1: Only the polynomial quadratic regression results in terms of R^2 are listed because they outperformed the exponential regression results
Note 2: NS in the following tables means the model is not significant at the 0.05 level of confidence

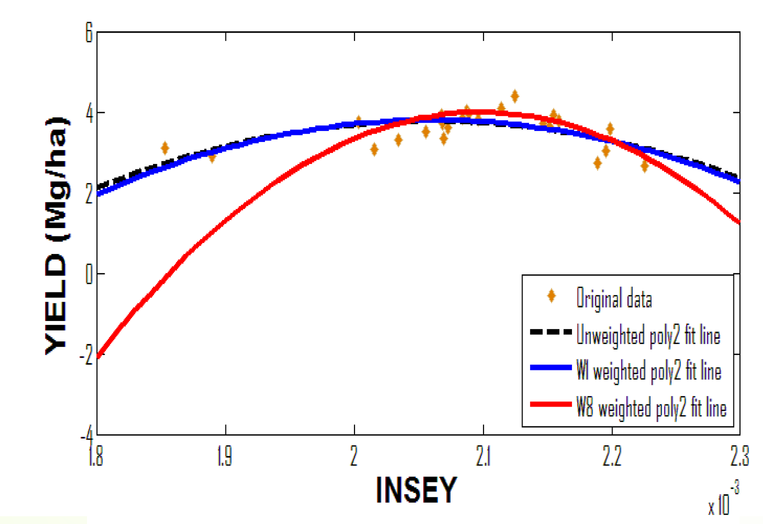

Figure 1. Valleycity wheat GreenSeeker Red INSEY regression models comparison

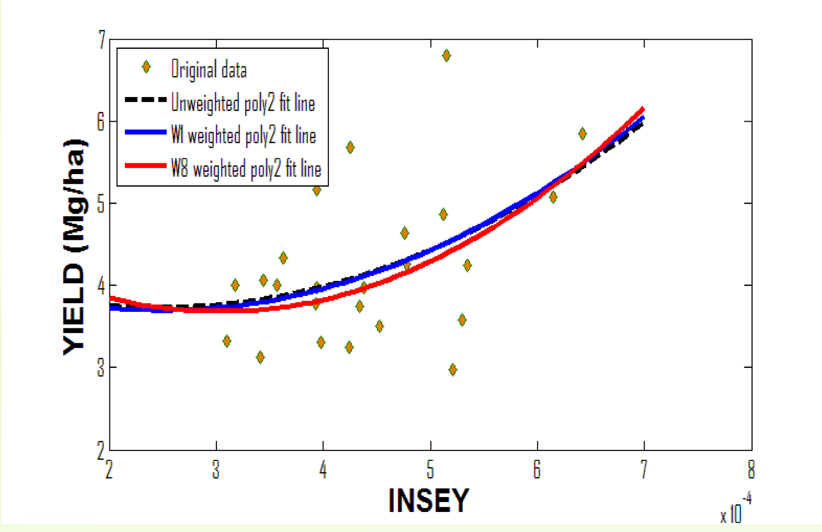

Figure 2. Durbin Corn Crop Circle Red edge INSEY regression models comparison

http://www.trimble.com/Agriculture/greenseeker.aspx

http://hollandscientific.com/crop-circle-acs-470-multi-spectral-crop-canopy-sensor/

Discussions

◇ Compared to the unweighted regression in terms of the R^2 performance and model significance of the relationships between crops yield and sensing data

Based on single site,
W1-weighted quadratic polynomial regression didn't help much;
W8-weighted quadratic polynomial regression were significantly superior.

Based on pooled two sites,
All weighted quadratic polynomial regressions were either inconsistent or showing a decreasing trend in performance.

◇ Reasons for the poor performance of the proposed method on pooled data:

1. pooled data consisted of greater sample number, enabling unweighted regression models to yield a more stable and significant relationship.

2. The other may be that the differences in crop growth and in turn sensing differences between each two sites were too great, confounding the effect of weighting.

Conclusions

Using single site data, the proposed intensified weighted nonlinear regression models significantly outperformed their corresponding unweighted regression models in terms of R^2 and model significance.

Using pooled site data, the propose methodology did not improve the performance of prediction models.

This study 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

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.