Korean J. Crop Sci 50(4) 268~275(2005)

Identifying Factors for Corn Yield Prediction Models and Evaluating Model Selection Methods

Jiyul Chang*† and David E. Clay**

*Center for Biocomplexity Studies, South Dakota State University, Brookings, SD, 57007, USA
**Plant Science Department, South Dakota State University, Brookings, SD, 57007, USA

ABSTRACT : Early predictions of crop yields can provide information to producers to take advantages of opportunities into market places, to assess national food security, and to provide early food shortage warning. The objectives of this study were to identify the most useful parameters for estimating yields and to compare two model selection methods for finding the 'best' model developed by multiple linear regression. This research was conducted in two 65-ha corn/soybean rotation fields located in east central South Dakota. Data used to develop models were small temporal variability information (STVI: elevation, apparent electrical conductivity (EC_a), slope), large temporal variability information (LTVI: inorganic N, Olsen P, soil moisture), and remote sensing information (green, red, and NIR bands and normalized difference vegetation index (NDVI), green normalized difference vegetation index (GDVI)). Second order Akaike's Information Criterion (AICc) and Stepwise multiple regression were used to develop the best-fitting equations in each system (information groups). The models with $\Delta_i \leq 2$ were selected and 22 and 37 models were selected at Moody and Brookings, respectively. Based on the results, the most useful variables to estimate corn yield were different in each field. Elevation and EC_a were consistently the most useful variables in both fields and most of the systems. Model selection was different in each field. Different number of variables were selected in different fields. These results might be contributed to different landscapes and management histories of the study fields. The most common variables selected by AICc and Stepwise were different. In validation, Stepwise was slightly better than AICc at Moody and at Brookings AICc was slightly better than Stepwise. Results suggest that the AICc approach can be used to identify the most useful information and select the 'best' yield models for production fields.

Keywords: model selection, AICc, stepwise multiple regression, yield prediction

Yield predictions can assist in harvest planning, modeling, assuming food security, and providing early food shortage warnings. Researchers have used remote sensing information to predict crop yields, because remote sensing information is relatively easy to collect and shows large area at once.

Staggenborg & Taylor (2000) reported that GDVI explained 40 % of corn yield variability observed in 13 Kansas fields. Weigand *et al.* (1999) showed that NDVI was correlated with corn yield in Texas fields and the predicted yield from an equation using NIR, red, and yellow-green bands accounted for 85 % of actual yield. Chang *et al.* (2003) used remote sensing data taken in three different times (beginning, middle, and end of growing season) to predict corn yields. This study showed that reflectance measured early in the season provided information about soil water and color, while reflectance measured in late summer provided information about plant conditions.

To develop yield prediction models, simple or multiple regression analysis has been used. When multiple regression is used, collinearity may be a problem. Collinearity is the property that at least one predictor is a 'near' linear combination of other predictors (Chatterjee *et al.*, 2000). If collinearity occurs the coefficient values are sensitive to slight changes in the data set and to the addition or deletion of variables in the equation. Principal Component Analysis (PCA) is one of the common methods in multivariate analysis to avoid collinearity problem (Johnson, 1998), but the number of information (variables) to calculate PCs is not reduced (very much data driven) and standardization may cause information to be lost.

Model selection (e.g. variable selection in regression) is a balance between bias and variance (Burnhaam & Anderson, 2001, 2002). Models with too few parameters (variables) have bias, whereas models with too many parameters (variables) may have poor precision. In the analysis of empirical data, one must face the question 'What model is the best one to explain variability of data at hand?'. In linear regression model, goodness-of-fit index or coefficient of determination (R^2) based on least square errors has been used widely to evaluate the 'best' model. The problem of using R^2 is that R^2 increases with increasing the number of explanatory variables of equation in same data set always. The Akaike's Information Criterion (AIC) is another method to find the

†Corresponding author: (Phone) 605-688-6666 (E-mail) jiyul.chang@sdstate.edu
<Received August 2, 2005>

'best' model. AIC is based on Kullback-Leibler (K-L) information and statistical maximum likelihood and has been used to select the best model. K-L information is a measure (a 'distance' in a heuristic sense) between conceptual reality, f, and approximating model, g, and the equation is,

$$I(f,g) = \int f(x)\ln\left(\frac{f(x)}{g(x|\theta)}\right)dx \quad [1]$$

where f and g are n-dimensional probability distributions. K-L information, denoted $I(f, g)$, is the 'information' lost when model g is used to approximate reality, f. The analyst seeks an approximating model that loses as little information as possible; this is equivalent to minimizing $I(f, g)$, over the set of models of interest. Akaike's derivation for large samples relied on the K-L information and the equation of AIC is,

$$\mathrm{AIC} = -2\ln(L(\theta|data)) + 2K \quad [2]$$

where $\ln(L(\theta|data))$ is the value of the maximized log-likelihood over the unknown parameters (θ), given the data and the model, and K is the number of estimable parameters in that approximating model. AIC is easy to compute from the results of least square estimation in the case of linear models or from the results of a likelihood-based analysis. The model which has a minimum AIC is selected as 'best' for the empirical data. AIC is not a test of models, so there is no any level of 'significance'. Instead, there are concepts of evidence and a 'best' inference, given the data and the set of a priori models.

AIC-based model selection is equivalent to certain cross-validation methods. The difference between model AIC and the minimum AIC is very useful to rescale AIC values such that the model with the minimum information criterion has a value of 0,

$$\Delta_i = \mathrm{AIC}_i - \mathrm{minAIC} \quad [3]$$

the Δ_i values easy to interpret, and allow a 'strength of evidence' comparison and ranking of candidate models. The large the $_i$, the smaller the likelihood of that model being the best model in the set of candidate models considered. It is important to know which model is second best as well as some measure of its standing with respect to the best model. There is a rule to assess the relative merits of models in the set: models having $\Delta_i \leq 2$ have substantial support (evidence), those where $4 \leq \Delta_i \leq 7$ have considerably less support, while models having $\Delta_i > 10$ have essentially no support.

Alternatively, Akaike weighs, w_i, can be used as indicators of the 'weight of evidence' for the model i,

$$w_i = \frac{\exp(-\Delta i/2)}{\sum_{r=1}^{R}\exp(-\Delta r/2)} \quad [4]$$

It is convenient to normalize these likelihoods such that they sum to 1. The w_i can be interpreted as the probability that model i is the best K-L model in the set of R models being considered.

When K is large relative to sample size n (which includes when n is small, for any K), second-order AIC (AICc) should be used, and the equation is

$$AICc = -2\ln(L(\theta)) + 2K + \frac{2K(K+1)}{(n-K-1)} \quad [5]$$

This should be used unless $n/K > \sim 40$. AICc values replace AIC in equation [3] and [4] to calculate Δ_i and w_i.

Westphal *et al.* (2003) used AIC to select set of candidate logistic regression models (important variables) to estimate distribution of 31 bird species at four different landscape characteristics in the Mt. Lofty Ranges, South Australia. They reported that AIC can be important technique in the toolkits of landscape ecologists. When developing model, it is important to identify useful information. In this study, three categories of data information were used for developing yield prediction models. They are STVI, LTVI, and RS. The STVI includes elevation, EC_a, and slope which do not change much for a long time. The LTVI includes inorganic N, Olsen P, and soil moisture which are changing and should be measured periodically. The RS includes aerial images taken in end of growing season. The advantage of RS information is that large area can be shown in one image. Two different model selection methods (AIC and Stepwise) were compared to find the best model equation to estimate corn yields. The objectives of this study were to identify the most useful parameters for estimating yields and to compare two model selection methods for finding the 'best' model developed by multiple linear regression.

MATERIALS AND METHODS

Ground Scouting Data

This research was conducted in two 65-ha corn/soybean rotation fields located in east central South Dakota. The latitude and longitude values for Moody were 44.17 °N and 96.62 °W, respectively, and for Brookings were 44.23 °N and 96.65 °W, respectively. Elevation ranged from 518 to 534 m and slope ranged from 0 to 7.2% in Moody. In Brookings, elevation ranged from 505 to 518 m and slope ranged from 0 to 10.6%. Soil information for these fields was previously reported in Clay *et al.* (2001).

In each field, soil and corn yield information were collected from 50 different points in 1999 and 2000 for Moody and Brookings, respectively. At 50 sampling points in each field soil samples were collected. Each soil sample consisted of 15 individual cores. Each sampling point was located with a carrier phase frequency Global Positioning System (DGPS). The differential correction was obtained from Omnistar (Omnistar, Inc., Houston, TX). Elevation was measured with DGPS.

Soil samples were air-dried (35 °C) and ground (2-mm) to prepare them for analysis. Inorganic N was extracted from soil with 1.0 *M* KCl using a 10:1 solution to soil ratio and analyzed on an Astoria Analyzer 300 (Astoria-Pacific Inc., Clackamas, OR). Olsen P was extracted from 2 g of soil with 40 ml of 0.5 *M* $NaHCO_3$ at a pH value of 8.5. The soil extract was filtered, a color reagent containing ascorbic acid and molybdate was added, and color development was measured on a colorimeter set at 882 nm (Olsen & Sommers, 1982). EC_a was measured with a Veris 3100 (Veris Technology, Salinas, KS). Gravimetric soil moisture to a depth of 60-cm was measured at least every month. Slope (%) was calculated based on elevation using ArcView (ESRI, Inc. Redland, CA). Grain was harvested by a combine equipped with a calibrated Ag Leader 2000 (AgLeader Technology, Inc., Ames, IA) yield monitor in 1999 and 2001 for Moody, and in 2000 and 2002 for Brookings (Lems, 2001). The soil and grain data were transformed to natural log.

Aerial Images

The remote sensing images in 1999 and 2000 were collected with a digital camera mounted on a plane flying at 1500 m above MSL (mean sea level) between 10 AM and 2 PM at local time on cloud free days. Spatial resolution of the images was approximately 1-m. The wavelengths collected were: green (557-582 nm), red (647-672 nm), and NIR (720-920 nm). Images were collected on 21 September and on 29 August, in 1999 and 2000, respectively.

At least four control points in each field were used by IMAGINE (ERDAS, Inc. Atlanta, GA) for geo-registration. All data that were prepared for point coverage (point layers) were overlaid on the aerial images. ArcView and ArcView Spatial Analyst (ESRI) were used for mapping and spatial analysis. The aerial images were transformed into grid files, which were used to calculate NDVI and GDVI. The equations for NDVI and GDVI are:

$$\mathrm{NDVI} = (\mathrm{NIR-Red}) / (\mathrm{NIR + Red}) \qquad [6]$$

$$\mathrm{GDVI} = (\mathrm{NIR-Green})/(\mathrm{NIR+Green}) \qquad [7]$$

where NIR is pixel value (digital number) of between 700 and 900 nm, red is pixel value between 600 and 700 nm, and green is pixel value between 500 and 600 nm.

The pixel values collected from sampling points were divided by reference pixel values of each band. Based on the assumption that the adjacent gravel road was an invariant target, it was used as a pixel reference value. After normalization, the remote sensing data were transformed to natural log.

Model Development and Selection

Yield prediction models were based on six different systems which were STVI, LTVI, STVI+LTVI, STVI+RS, LTVI+RS, and STVI+LTVI+RS as explanatory variables. The independent variables were corn yields in 1999 and 2000 at Moody and Brookings, respectively.

AICc method was used instead of AIC to find the best models in each system, because number of samples over number of variables is less than 40. PROC REG in SAS (SAS Institute, 1995) was used to calculate AIC values for each model. Equation [5] was used to calculate AICc. Δ_i and w_i of AICc were calculated with equation [3] and [4] instead of AIC to see how models were supportable. For choosing the best-fitting equations from all possible multiple regressions, models with $\Delta_i \leq 2$ were selected.

Stepwise method in SAS was used to develop the best-fitting equations in each system. The stepwise method is a modification of the forward-selection. After a variable is added, the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an *F* statistic significant at the specific entry level. The stepwise process ends when none of the variables outside the model has an *F* statistic significant at the specific entry level and every variable in the model is significant at the specific entry level. The significant level for entry (SLE) was 0.2 and the significant level for stay (SLS) was 0.05 to avoid the collinearity problem. The yield models generated from data collected in 1999 and 2000 were used to predict corn yields in 2001 and 2002 at Moody and Brookings, respectively. Variance inflation factor (VIF) was calculated to see collinearity between variables. If VIF is bigger than 10, there is enough collinearity present to state causing sever problem (Chatterjee *et al.* 2000). These statistical analyses were conducted using PROC REG in SAS.

RESULTS AND DISCUSSION

Identifying the Most Useful Factors

Corn yield was negatively correlated with elevation, P, and red band and positively correlated with EC_a, soil moisture before planting and in summer, NIR, NDVI, and GDVI at Moody (Table 1). At Brookings, corn yield was negatively correlated with elevation, green, and red band and positively correlated with EC_a, N, soil moisture before planting and in summer, NIR, NDVI, and GDVI.

Based on the criterion ($\Delta_i \leq 2$), 22 and 37 models were selected in Moody and Brookings, respectively. The best models in each system were selected from all possible models. At Moody, the number of selected models were 2 out of 7 models in STVI system, 3 out of 15 models in LTVI system, 2 out of 127 models in STVI+LTVI system, 4 out of 255 models in STVI+RS system, 7 out of 511 models in LTVI+RS system, and 4 out of 4095 models in STVI+LTVI+RS system by AICc (Table 2). The range of R^2 values generally increased with increasing number of variables. The most common variables, which were selected in more than 50% of selected models, in STVI system were elevation. In LTVI models, the most common variables were P and soil moisture before planting. In STVI+LTVI models, the most common variables were elevation, EC_a, P, and soil moisture before planting. In STVI+RS models, the most common variables were elevation, slope, EC_a, and NDVI. In LTVI+RS models, the most common variables were P, soil moisture in summer, green and red bands, and NDVI. In STVI+LTVI+RS models, the most common variables were elevation slope, EC_a, P, soil moisture before planting, and NDVI.

Elevation, EC_a, P, soil moisture before planting, and NDVI were the most common variables in systems. Eleva-

Table 1. Correlation coefficient matrix between variables in three systems (STVI. small temporal variability information, LTVI: large temporal variability information, and RS: Remote Sensing) and corn yield in 1999 and 2000 at Moody and Brookings, respectively.

	STVI			LTVI				RS				
	Elev. (m)	Slope (%)	EC_a (mS m^{-1})	N (---- mg kg^{-1} ----)	P	(a) Soil moisture (-------- % ---------)	(b) Soil moisture	Green	Red	NIR	NDVI	GDVI
	---------- r ----------											
Moody												
Slope	0.01											
EC_a	-0.89**	0.06										
N	-0.02	-0.05	-0.01									
P	0.80**	0.13	-0.76**	0.05								
(a) Soil moisture	-0.82**	-0.07	0.76**	-0.13	-0.66**							
(b) Soil moisture	-0.82**	-0.12	0.73**	-0.06	-0.71**	0.97**						
Green	-0.42**	-0.21	0.12	-0.07	-0.46**	0.10	0.18					
Red	0.43**	-0.07	-0.51**	0.06	0.25	-0.67**	-0.63**	0.50**				
NIR	-0.84**	0.14	0.82**	-0.15	-0.64**	0.79**	0.77**	0.09	-0.70**			
NDVI	-0.72**	0.10	0.70**	-0.14	-0.49**	0.81**	0.78**	-0.12	-0.89**	0.92**		
GDVI	-0.32*	0.32*	0.47**	-0.10	-0.12	0.50**	0.42**	-0.60**	-0.88**	0.69**	0.81**	
Corn (Mg ha^{-1})	-0.78**	-0.06	0.61**	-0.10	-0.45**	0.84**	0.82**	0.12	-0.70**	0.84**	0.89**	0.58**
Brookings												
Slope	0.43**											
EC_a	0.12	0.06										
N	0.02	-0.05	0.26									
P	-0.25	0.08	0.12	0.39**								
(a) Soil moisture	-0.26	-0.17	0.31*	0.43**	0.04							
(b) Soil moisture	-0.51**	-0.16	0.47**	0.37*	0.32*	0.78**						
Green	0.07	-0.00	-0.59**	-0.41**	-0.20	-0.63**	-0.73**					
Red	0.37*	0.09	-0.54**	-0.28	-0.25	-0.59**	-0.80**	0.92**				
NIR	-0.79**	-0.42**	0.24	-0.01	0.11	0.25	0.58**	-0.16	-0.45**			
NDVI	-0.59**	-0.28	0.47**	0.17	0.17	0.51**	0.79**	-0.71**	-0.90**	0.77**		
GDVI	-0.50**	-0.24	0.57**	0.27	0.20	0.58**	0.84**	-0.81**	-0.94**	0.70**	0.97**	
Corn (Mg ha^{-1})	-0.38*	-0.19	0.69**	0.34*	0.27	0.54**	0.81**	-0.84**	-0.90**	0.52**	0.84**	0.91**

(a) soil moisture: soil moisture before planting, (b)soil moisture· soil moisture in summer * and ** are significant at 95 and 99% level, respectively.

Table 2. The models selected by AICc in each system and variables selected in more than 50% of selected models. The dependent variables were corn yield collected in 1999 and 2000 at Moody and Brookings, respectively (number of all variables, number of all possible models).

Systems	Num. of selected models	Num of variables	Range of R^2 values	Ranges of w_i	The most selected variables in system
STVI (3, 7)					
Moody	2	1 - 2	0.606-0.633	0.33-0.41	elevation, EC_a
Brookings	1	2	0.703	0.75	elevation, EC_a
LTVI (4, 15)					
Moody	3	1 - 3	0.713-0.732	0 09-0.26	P, (a) soil moisture
Brookings	3	1 - 3	0.653-0.683	0.12-0.28	(a) soil moisture, (b) soil moisture
STVI+LTVI (7, 127)					
Moody	2	4 - 5	0.841-0.844	0.13-0.33	elevation, EC_a, P, (a) soil moisture
Brookings	7	2 - 4	0.780-0.809	0.06-0.16	EC_a (b) soil moisture
STVI+RS (8, 255)					
Moody	4	3 - 6	0 895-0 916	0.08-0 15	elevation, slope, EC_a, green, NDVI
Brookings	5	3 - 4	0.912-0 915	0.05-0.14	elevation, EC_a, green
LTVI+RS (9, 511)					
Moody	7	4 - 5	0.896-0.906	0 03-0.08	P, (a)soil moisture, (b)soil moisture, green, red, NDVI
Brookings	11	2 - 4	0 865-0.881	0.01-0.04	NDVI, GDVI
STVI+LTVI+RS (12, 4095)					
Moody	4	6 - 8	0 939-0.947	0.05-0.10	elevation, slope, EC_a, P, (a)soil moisture, NDVI
Brookings	10	3 - 6	0.912-0.926	0 01-0.04	elevation, EC_a, green

(a) soil moisture: soil moisture before planting, (b)soil moisture: soil moisture in summer.
STVI: small temporal variability information, LTVI: large temporal variability information, and RS: Remote Sensing

Table 3. The intercept and estimated parameters of the best models in each system selected by AICc and stepwise methods and amount of yield variability explained by the equations with corn yield in 1999 at Moody.

Systems	Intercept	STVI			LTVI				RS					R^2
		Elev.	Slope	EC_a	N	P	(a) Soil moisture	(b) Soil moisture	Green	Red	NIR	NDVI	GDVI	
STVI														
AICc	1707	-269		-3.07										0.633
Stepwise	1212	-192												0.606
LTVI														
AICc	-25.47						10.64							0.713
Stepwise	-25 47						10.64							0.713
STVI+LTVI														
AICc	1360	-219		-3.42		1.54	8.59							0.841
Stepwise	-25.47						10.64							0.713
STVI+RS														
AICc	1121	-174	-3 27	-4 79								3.50		0.905
Stepwise	12.18											5.58	-1 40	0.849
LTVI+RS														
AICc	1.54					1.30		3 66		3.63		5.82	-0.85	0.906
Stepwise	12.18											5.58	-1 40	0.849
STVI+LTVI+RS														
AICc	1202	-189	-3 19	-4.25		1.03	3 53					2.65		0.936
Stepwise	12.18											5 58	-1.40	0.849

(a) soil moisture: soil moisture before planting, (b)soil moisture: soil moisture in summer
STVI small temporal variability information, LTVI: large temporal variability information, and RS: Remote Sensing

tion and EC_a in STVI system were selected in most these models. These results were contributed to areas with high elevation having low soil water than area with low elevation and EC_a being positively correlated with soil water, which was a major factor influencing crop yield (Chang *et al.* 2003). Soil moisture before planting and P in LTVI system were the most common in models and correlated positively and negatively with corn yield, respectively. Inorganic N was selected in 2 models out of 16 models (models of STVI, STVI+LTVI, STVI+RS, and STVI+LTVI+RS). NDVI was selected in most of models and correlated positively with corn yield. NIR band and GDVI was selected in 1 and 3 models out of 15 models (models of LTVI+RS, STVI+RS, and STVI+LTVI+RS), respectively. There were severe collinearity problems between soil water before planting and in summer (VIF: 17 and 19, respectively), between red band and NDVI (VIF: 22 and 20, respectively), and between elevation and EC_a (VIF: 16 and 8, respectively) (data not shown).

At Brookings, the number of selected models were 1 model in STVI system, 3 models in LTVI system, 7 models in STVI+LTVI system, 5 models in STVI+RS system, 11 models in LTVI+RS system, and 10 models in STVI+LTVI+RS system by AICc (Table 2). Similar results with Moody were observed in Brookings, but in STVI+RS system, even though the highest number of variables (4) was same with STVI+LTVI and LTVI+RS systems, the range of R^2 values was higher than the two systems. The most common variables in each system were elevation and EC_a in STVI system, soil moisture before planting and in summer in LTVI system, EC_a and soil moisture in summer in STVI+LTVI system, elevation, EC_a, and green band in STVI+RS system, NDVI and GDVI in LTVI+RS system, and elevation, EC_a, and green band in STVI+LTVI +RS system.

Elevation, EC_a, and green band were the most common in systems. Elevation and EC_a in STVI variables were the most common in models. Soil moisture before planting and in summer in LTVI system were the most common in models. P was selected in 1 model out of 31 models (models of LTVI+RS, STVI+RS, and STVI+LTVI+RS). Green band was the most common in models. NIR band was not selected in any model. Variables between green, red, NDVI, and GDVI (VIF: 40, 64, 100, and 81, respectively) and between red, NIR, NDVI, and GDVI (VIF: 41, 14, 46, and

Table 4. The intercept and estimated parameters of the best models in each system selected by AICc and stepwise methods and amount of yield variability explained by the equations with corn yield in 2000 at Brookings

Systems	Intercept	STVI			LTVI				RS					R^2
		Elev.	Slope	EC	N	P	(a) Soil moisture	(b) Soil moisture	Green	Red	NIR	NDVI	GDVI	
STVI														
AICc	165	-26.4		0.810										0.703
Stepwise	165	-26.4		0.810										0.703
LTVI														
AICc	-0.327						-0.445	1.107						0.677
Stepwise	-0.983							1.139						0.653
STVI+LTVI														
AICc	69.9	-11.4		0.535				0.663						0.804
Stepwise	-1.45			0.432				0.873						0.779
STVI+RS														
AICc	136	-21.7		0.427					-1.08					0.912
Stepwise	2.19			0.279									1.05	0.872
LTVI+RS														
AICc	3.43											-0.73	2.48	0.869
Stepwise	3.43											-0.73	2.48	0.869
STVI+LTVI+RS														
AICc	136	-21.7		0.427					-1.08					0.912
Stepwise	2.19			0.279									1.05	0.872

(a)soil moisture soil moisture before planting, (b)soil moisture: soil moisture in summer.
STVI: small temporal variability information, LTVI: large temporal variability information, and RS: Remote Sensing

Table 5. The intercept, slope, and R^2 of sımple lınear regressıon between the estımated corn yıelds usıng the best models selected by AICc and Stepwise methods and corn yıelds ın 2001 and 2002 at Moody and Brookıngs, respectıvely.

Systems	Moody			Brookıngs		
	Intercept	Slope	R^2	Intercept	Slope	R^2
AICc						
STVI	1.71 (±2.58)	0.87 (±0.28)	0 515**	0.46 (±0 56)	0.60 (±0.26)	0.410**
LTVI	1.26 (±2.07)	0.92 (±0.23)	0.650**	0 17 (±0.47)	0.74 (±0.22)	0.592**
STVI+LTVI	1.13 (±1.48)	0.93 (±0 16)	0.790**	0.34 (±0.45)	0.66 (±0.21)	0.558**
STVI+RS	1.60 (±1.53)	0.88 (±0.17)	0.758**	0.40 (±0.41)	0 63 (±0.19)	0.578**
LTVI+RS	0 10 (±1.55)	1 01 (±0.17)	0 811**	0.46 (±0 46)	0.60 (±0.21)	0.507**
STVI+LTVI+RS	1.38 (±1.26)	0.90 (±0.14)	0.831**	0.40 (±0.41)	0.63 (±0.19)	0.578**
Stepwise						
STVI	1.72 (±2.70)	0.87 (±0.30)	0.492**	0.46 (±0.56)	0.60 (±0.26)	0.410**
LTVI	1.26 (±2.07)	0.92 (±0.23)	0.650**	0.21 (±0.51)	0.72 (±0.24)	0.538**
STVI+LTVI	1.26 (±2.07)	0.92 (±0.23)	0 650**	0.29 (±0 45)	0.68 (±0 21)	0.574**
STVI+RS	1 62 (±1.74)	0.88 (±0.19)	0.708**	0.51 (±0.46)	0.59 (±0.22)	0.484**
LTVI+RS	1.62 (±1.74)	0.88 (±0.19)	0.708**	0 46 (±0.46)	0.60 (±0.21)	0.507**
STVI+LTVI+RS	1.62 (±1.74)	0.88 (±0.19)	0.708**	0.49 (±0 47)	0.59 (±0.22)	0.484**

Parentheses are confidence ınterval at 95% level. ** ıs sıgnıficant at 95% level.
STVI. small temporal varıability information, LTVI: large temporal varıability ınformation, and RS· Remote Sensıng

43, respectively) had severe collinearity problems (data not shown).

Based on the results, the most useful variables to estimate corn yield were different in each field. Elevation and EC_a were consistently the most useful variables in both fields and most of the systems. These results might be contributed to different landscapes and management histories of the study fields.

Evaluation of Model Selection Methods

Table 3 and 4 show the intercept and estimated parameters of the best models in each system. Care should be used in considering the coefficients. For example, at Moody, EC_a and P had negative and positive coefficients, while these were positively and negatıvely correlated with yield, respectively. The equations developed by Stepwise method had fewer explanatory variables and had lower R^2 values than equations selected by AICc method in all systems. In Stepwise method, NDVI, GDVI, and soil moisture before planting were the most common in models. GDVI correlated positively with corn yield, but the coefficient was negative in the equations. At Brookings, similar results were observed with Moody. Soil moisture before planting had negative coefficient in equation. EC_a, and GDVI were the most common in models developed by Stepwise method.

Table 5 show the intercept, slope, and R^2 of simple linear regression between the estimated corn yields using the best models selected by AICc and Stepwise methods and corn yields in 2001 and 2002 at Moody and Brookings, respectively. The relationship between estimated value and measured values should be 1 : 1 (intercept=0 and slope=1). Based on 95 % Confidence Interval (CI), STVI+RS and STVI+LTVI+RS system models selected by AICc were slightly different from 0 in intercept and other models were not different from 0 at Moody. All models were not different from 1 in slope. All models selected by AICc had higher R^2 values than models developed by Stepwise. At Brookings, STVI+RS and STVI+LTVI+RS system models developed by Stepwise were slightly different from 0 in intercept and other models were not different from 0. All models were slightly different from 1 in slope. All models, except STVI+LTVI system model, selected by AICc had higher R^2 values than models developed by Stepwise.

These results showed that model selection was different in each field. Different number of variables were selected in different fields. The most common variables selected by AICc and Stepwise were different. In validation, Stepwise was slightly better than AICc at Moody and at Brookings AICc was slightly better than Stepwise. Results suggest that the AICc approach can be used to identify the most useful information and select the 'best' yield models for production fields.

REFERENCES

Burnham, K. P. and D. R Anderson. 2001. Kullback-Leibler ınformation as a basis for strong inference in ecological studies. Wildlife Research. 28 · 111-119.

Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference: A practical pnformation-theoretic approach. Springer-Verlag, New York, New York, USA, 488p.

Chang, J., D. E Clay, K. Dalsted, S A. Clay, and M. O'Neill. 2003. Corn (Zea mays L.) Yield prediction using multispectral and multidate reflectance. Agron. J. 95 : 1447-1453.

Chatterjee, S, A. S Hadi, and B. Price, 2000 Regression analysis by example. (3rd ed). John Wiley & Sons, Inc. New York, New York, USA, 359p.

Clay, D. E , J. Chang, D. D. Malo, C. G Carlson, C. L. Reese, S. A Clay, M Ellsbury, and B. Berg 2001. Factors influencing spatial variability of soil apparent electrical conductivity. Commun. Soil Sci. Plant Anal 32 : 2993-3008.

Johnson, D 1998. Applied multivariate methods for data analysts. Duxbury Press.

International Thomson Publishing Inc. Pacific Grove, CA, USA, 567p.

Lems, J., D. E. Clay, D Humburg, T. A. Doerge, S. Christopherson, and C. L. Reese. 1999 Yield monitors-Basic steps to ensure system accuracy and performance. SSMG-31. *In* D.E. Clay *et al.* (ed.) Site specific management guidelines Potash and Phosphate Institute, Norcross, GA.

Olsen, S. R and L. E. Sommers. 1982. Phosphorus. p 403-430. *In* A. Klute (ed.) Methods of soil analysis. Part 2, 2nd ed. ASA. Madison, WI.

SAS Institute. 1995. SAS/INSIGHT User's Guide (Version 6, 3rd ed.) SAS Inst., Cary, NC.

Staggenborg, S. A. and R. K. Taylor. 2000. Predicting grain yield variability with infrared images. [CD-ROM computer file] *In* P.C. Robert *et al* (ed.). Proceedings of the 5th International Conference on Precision Agriculture. July 16-19, 2000. Bloomington, MN. ASA-CSAA-SSSA, Madison WI.

Weigand, C. L., C. Yang, and J. M. Bradford. 1999. Mapping soil and yield variations using aerial videography, GIS, ground observations, image analysis, and yield monitors p 151-162 *In* Proc. 17th Biannual Workshop on Color Photography and Videography in Resource Assessment. May 5-7, 1999.

Westphal, M. I., S. A. Field, A. J. Tyre, D Paton, and H. P. Possingham. 2003. Effects of landscape pattern on bird species distribution in the Mt. Lofty Ranges, South Australia. Landscape Ecology. 18 . 413-426.