

2006

Novel Analysis of Hyperspectral Reflectance Data for Detecting Onset of Pollen Shed in Maize

Amy L. Kaleita

Iowa State University, kaleita@iastate.edu

Brian L. Steward

Iowa State University, bsteward@iastate.edu

Robert P. Ewing

Iowa State University

Daniel A. Ashlock

University of Guelph

Mark E. Westgate

Iowa State University

See next page for additional authors

Follow this and additional works at: http://lib.dr.iastate.edu/abe_eng_pubs



Part of the [Agriculture Commons](#), and the [Bioresource and Agricultural Engineering Commons](#)

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/abe_eng_pubs/16. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

Novel Analysis of Hyperspectral Reflectance Data for Detecting Onset of Pollen Shed in Maize

Abstract

Knowledge of pollen shed dynamics in and around seed production fields is critical for ensuring a high yield of genetically pure corn seed. Recently, changes in canopy reflectance using hyperspectral reflectance have been associated with tassel emergence, which is known to precede pollen shed in a predictable manner. Practical application of this remote sensing technology, however, requires a simple and reliable method to evaluate changes in spectral images associated with the onset of tassel emergence and pollen shed. In this study, several numerical methods were investigated for estimating percentage of plants with visible tassels (VT) and percentage of plants that initiated pollen shed (IPS) from remotely sensed hyperspectral reflectance data (397 to 902 nm). Correlation analysis identified regions of the spectra that were associated with tassel emergence and anthesis (i.e., 50% of plants shedding pollen). No single band, however, generated correlations greater than 0.40 for either VT or IPS. Classification using an artificial neural network (ANN) was predictive, correctly classifying 83.5% and 88.3% of the VT and IPS data, respectively. The extensive preprocessing necessary and the "black box" nature of ANNs, however, rendered analysis of spectral regions difficult using this method. Partial least squares (PLS) analysis yielded models with high predictive capability (R^2 of 0.80 for VT and 0.79 for IPS). The PLS coefficients, however, did not exhibit a spectrally consistent pattern. A novel range operator-enabled genetic algorithm (ROE-GA), designed to consider the shape of the spectra, had similar predictive capabilities to the ANN and PLS, but provided the added advantage of allowing information transfer for increased domain knowledge. The ROE-GA analysis is the preferred method to evaluate hyperspectral reflectance data and associate spectral changes to tassel emergence and the onset of pollen shed in corn on a field scale.

Keywords

Genetic programming, Partial least squares, Remote sensing, Spectral analysis, Tassel emergence

Disciplines

Agriculture | Bioresource and Agricultural Engineering

Comments

This article is from *Transactions of the ASABE*, 49, no. 6 (2006): 1947–1954.

Authors

Amy L. Kaleita, Brian L. Steward, Robert P. Ewing, Daniel A. Ashlock, Mark E. Westgate, and Jerry L. Hatfield

NOVEL ANALYSIS OF HYPERSPECTRAL REFLECTANCE DATA FOR DETECTING ONSET OF POLLEN SHED IN MAIZE

A. L. Kaleita, B. L. Steward, R. P. Ewing, D. A. Ashlock, M. E. Westgate, J. L. Hatfield

ABSTRACT. Knowledge of pollen shed dynamics in and around seed production fields is critical for ensuring a high yield of genetically pure corn seed. Recently, changes in canopy reflectance using hyperspectral reflectance have been associated with tassel emergence, which is known to precede pollen shed in a predictable manner. Practical application of this remote sensing technology, however, requires a simple and reliable method to evaluate changes in spectral images associated with the onset of tassel emergence and pollen shed. In this study, several numerical methods were investigated for estimating percentage of plants with visible tassels (VT) and percentage of plants that initiated pollen shed (IPS) from remotely sensed hyperspectral reflectance data (397 to 902 nm). Correlation analysis identified regions of the spectra that were associated with tassel emergence and anthesis (i.e., 50% of plants shedding pollen). No single band, however, generated correlations greater than 0.40 for either VT or IPS. Classification using an artificial neural network (ANN) was predictive, correctly classifying 83.5% and 88.3% of the VT and IPS data, respectively. The extensive preprocessing necessary and the “black box” nature of ANNs, however, rendered analysis of spectral regions difficult using this method. Partial least squares (PLS) analysis yielded models with high predictive capability (R^2 of 0.80 for VT and 0.79 for IPS). The PLS coefficients, however, did not exhibit a spectrally consistent pattern. A novel range operator-enabled genetic algorithm (ROE-GA), designed to consider the shape of the spectra, had similar predictive capabilities to the ANN and PLS, but provided the added advantage of allowing information transfer for increased domain knowledge. The ROE-GA analysis is the preferred method to evaluate hyperspectral reflectance data and associate spectral changes to tassel emergence and the onset of pollen shed in corn on a field scale.

Keywords. Genetic programming, Partial least squares, Remote sensing, Spectral analysis, Tassel emergence.

In open-pollinated crops, such as corn (*Zea mays* L.), production of genetically pure or transgene-free seed depends on effective reproductive isolation from adventitious (foreign) sources of pollen. Reproductive isolation has two interrelated components: biological and physical (Ireland et al., 2006). Effective biological isolation requires knowledge of flowering dynamics, density of pollen shed, and competition for female silks by both desirable and adventitious pollen sources. Physical isolation simply limits the capacity of adventitious pollen to enter the field. Distance between fields, orientation relative to prevailing winds, and field size all have been shown to impact genetic purity of harvested seed to some degree (Ireland et al., 2006). Of these two components, biological isolation is far more difficult to manage because of environmental effects on plant development

and spatial variation between and across fields. Technologies that rapidly and reliably assess corn flowering dynamics on a field scale, however, would enable producers to assess the effectiveness of their isolation strategy in season. This has direct economic consequence for seed producers, and is paramount for ensuring gene confinement for corn engineered to produce industrial or pharmaceutical products (Wolt et al., 2005).

Research on corn pollen biology has shown that the process of pollen shed can be reliably predicted from the progress of tassel development (Westgate et al., 2003). Tassels advance through a series of stages from emergence through completion of pollen shed in a well-defined pattern. Therefore, once the initiation of tassel emergence within a plant canopy is detected, the subsequent timing and intensity of pollen shed in the field can be readily estimated. If the amount and timing of pollen shed are known, this information can be combined with particulate transport models to assess risk of pollen entry into nearby fields (Westgate et al., 2000; Goggi et al., 2006). As such, the effectiveness of reproductive isolation between fields can be assessed by quantifying the timing of tassel emergence (commonly referred to as tasseling) within the fields. Currently, however, there is no simple or cost-effective means to acquire this critical information on a field scale.

Because corn tassels have a different cellular makeup and color than the vegetative parts of the corn plant, it is ostensibly possible to detect the tassels as they emerge from the whorl of leaves at the top of the canopy from the spectral reflectance of the whole canopy. Recent research confirms

Submitted for review in July 2005 as manuscript number IET 5976; approved for publication by the Information & Electrical Technologies Division of ASABE in September 2006.

The authors are **Amy L. Kaleita**, ASABE Member Engineer, Assistant Professor, and **Brian L. Steward**, ASABE Member Engineer, Associate Professor, Department of Agricultural and Biosystems Engineering, and **Robert P. Ewing**, Assistant Scientist, Department of Agronomy, Iowa State University, Ames, Iowa; **Daniel A. Ashlock**, Associate Professor, Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada; **Mark E. Westgate**, Professor, Department of Agronomy, Iowa State University, Ames, Iowa; and **Jerry L. Hatfield**, Plant Physiologist, USDA-ARS National Soil Tilth Lab., Ames, Iowa. **Corresponding author:** Amy L. Kaleita, Iowa State University, 211 Davidson Hall, Ames, IA 50011; phone: 515-294-5167; fax: 515-294-2552; e-mail: kaleita@iastate.edu.

that the presence of tassels alters canopy reflectance between 400 and 900 nm. For example, Goel et al. (2003) noted differences in remotely derived estimates of leaf area index (LAI) from one tassel stage to another. Viña et al. (2004) reported a consistent increase in canopy reflectance in the 600-700 nm region when tassels were present. It may therefore be possible to utilize changes in spectral reflectance in this range as a means to identify tassel emergence and the onset of pollen shed.

With the advent of hyperspectral sensing technology, more detailed data related to tasseling stage are potentially available. With that availability, however, comes the challenge of extracting meaningful relationships from the overwhelming quantity of data. Conventional linear methods such as principal component analysis (PCA) or partial least squares (PLS) regression have been used to relate linear combinations of spectral bands to independent physical variables of interest. These "soft modelling" approaches (Wold, 1982) are powerful but limited for several reasons. They apply only linear relationships between spectral bands and the variables of interest, they fail to incorporate domain knowledge or physical first principles into the analysis (Martens and Martens, 2001), and they treat each channel as independent, thereby ignoring additional information implicitly embedded in the structure of the spectrum.

An alternate approach is to use domain knowledge to develop indices based on simple mathematical operations such as ratios, differences, and sums of spectral intensities at specific wavelengths, e.g., the normalized difference vegetation index (NDVI; Rouse et al., 1973). Several studies have addressed the utility of derivative approaches to analyze remote sensing data (Demetriades-Shah et al., 1990; Clevers et al., 2002). Other studies have used parameters based on spectral shape (Whiting et al., 2004). With increasing numbers of spectral bands, identifying or searching for the best combinations of bands and operations can become intractable because of the overwhelming number of possible combinations. Furthermore, interpreting physical meaning from the results of these numerous combinations is not straightforward.

Evolutionary computation, and specifically the subset called genetic algorithms, is an adaptive search method that deals powerfully with difficult search problems, often avoiding local extremes in the search space (Goldberg, 1989). Evolutionary computation has only rarely been used in spectral analysis. Tang et al. (2000) used a genetic algorithm (GA) to identify a region in hue-saturation-intensity color space for vegetation segmentation in field images. Rauss et al. (2000) used genetic programming (GP) to develop a classification algorithm for an image with 28 spectral bands. Burman (1997, 1999) used evolutionary computation methods to perform automatic object detection and recognition in hyperspectral images. Leardi (2000) used a GA to select spectral features for use in PLS-based analysis. Yao and Tian (2003) used a GA to select hyperspectral bands to include in a principal components transformation for information extraction.

The above studies focused on linear combinations of individual reflectance or absorbance bands. Incorporating local domain knowledge, however, can provide a substantial amount of additional information for spectral analysis. Recently, Steward et al. (2004) developed a genetic algorithm that builds a predictive model from range-based

operations such as median and slope, making use of information contained in the spectral shape. The use of these range operators, which are non-linear, expands the search space beyond purely linear approaches. Furthermore, building models from range operators rather than particular wavebands makes the resulting models more robust to differences between sensors with different bands and bandwidths.

The objectives of this study were: (1) to determine whether it is possible to classify tasseling and pollination stages of corn canopies based on analysis of canopy hyperspectral reflectance spectra, and (2) to establish a reliable analytical technique to estimate the onset of these reproductive stages from hyperspectral data.

DATA COLLECTION

The corn canopies observed in this study were grown at the Iowa State University Bruner Farm (42° 0.6' N, 93° 44.2' W) during the 2003 growing season. Half of the field was planted with the cultivar Asgrow 740 (Monsanto, St. Louis, Mo.), the rest with Dekalb 611 (Monsanto, St. Louis, Mo.). Plots measured 6.1 × 12.2 m and were planted at 8 plants per m² replicated in three blocks. Each cultivar had two treatments: control and 50% detasseled; detasseling was done manually when tassels were first visible prior to pollen shed. There were thus a total of 12 plots used. Five measurement areas of approximately 1 m² were established within each plot. This experiment was a continuation of the third experiment presented in Fonseca et al. (2003).

During the tasseling and pollen shed period in late July and early August, the measurement areas in each plot were scouted daily to determine the fraction of plants that had reached the visible tassel (VT) or initial pollen shed (IPS) growth stage (Fonseca et al., 2003). Figure 1 shows the temporal dynamics of tassel development for a typical Asgrow 740 canopy. The population passed from 0% to 100% of plants at each stage in a sigmoid fashion. Initial pollen shed follows tassel emergence by approximately 2 days, and pollen shed ends about 5 to 6 days later. At the population level, the canopy began to shed pollen when about 20% of the tassels were visible. Anthesis (50% plants shedding pollen) occurred when about 80% of the tassels were visible. Maximum pollen shed for the crop occurs between 50% of plants shedding and 50% of plants completed shedding.

Visible and near-infrared (NIR) reflectance spectra of the measurement areas were recorded periodically on clear days beginning prior to tassel emergence (VT) and continuing through 100% shedding. An ASD FieldSpec spectrophotometer (Advanced Spectral Devices, Boulder, Colo.) with a 25° conical field of view was mounted on a boom at 2 m above the top of canopy and facing downwards to measure a canopy area of approximately 0.8 m². The FieldSpec collected 751 spectral variables at 1 nm intervals across a spectral range from 325 nm to 1075 nm. To reduce sampling noise from the spectrophotometer, each recorded spectrum was an average of 25 instantaneous scans. Spectra were collected from each of the five measurement areas within each plot. Preliminary inspection of the spectra indicated that reflectance at wavebands below 397 nm and above 902 nm had a low signal-to-noise ratio, so spectra were truncated to 397 nm to 902 nm (505 variables). Reflectance data were trans-

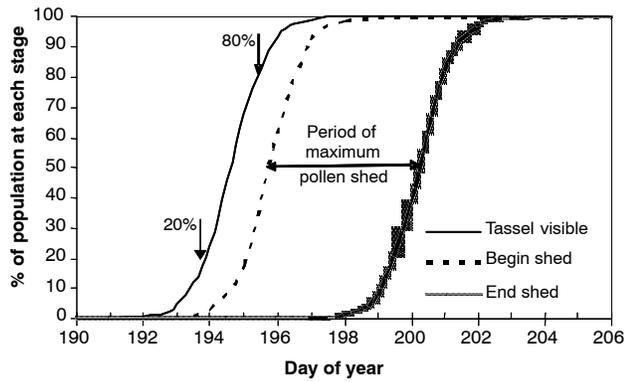


Figure 1. Dynamics of tassel development for Asgrow 740. The population passes through each stage in a sigmoid fashion. Note that plants start shedding pollen when about 20% of tassels are visible. Anthesis is defined as the date 50% of the plants have begun to shed pollen. This occurs when about 80% of the population has tassels visible. Maximum pollen shed for the crop occurs between 50% of plants shedding and 50% of plants completed shedding.

formed to absorbance using the $1/\log_{10}$ transform. To conserve computation time, the spectra were downsampled to include only every fourth band, for a total of 127 bands.

A total of 412 reflectance spectra for the VT and IPS stages were used. Preliminary data analysis revealed little or no difference in reflectance spectra from the two hybrids, so the analysis presented here uses data aggregated from both hybrids. A potential limitation of this data set is the relatively small number of plants in each measurement area, as it typically contained three to four plants within the field of view of the sensor. Additionally, the developmental reference variables (percent of plants at the VT or IPS stage) have a limited number of discrete values: 15 unique values of VT from 0 to 1, and 17 unique values of IPS from 0 to 1 (fig. 2).

Based on the typical reproductive patterns described above, three population classes were established for spectral analysis. Spectral data for canopies at less than 20% VT and

IPS were placed in class 1; spectra for canopies between 20% and 80% VT and IPS were placed in class 2; and spectra for canopies with greater than 80% VT and IPS were placed in class 3. Normalized mean reflectance spectra for each of these classes are shown in figure 3.

ANALYSIS

Four analytical techniques were investigated: a correlation analysis, an artificial neural network, a partial least squares approach, and a genetic algorithm based on operations over spectral ranges. A correlation analysis is the simplest approach: each band is considered individually, and its linear relationship to the reference variable is tabulated. In both the neural network and partial least squares approaches, band combinations are used. In contrast, the genetic algorithm was developed to use combinations not of individual bands but rather of operators, such as median and slope, performed on spectral ranges. In this way, the shapes of the spectra are analyzed for their relationship to the reference variable.

CORRELATION ANALYSIS

A correlation analysis was performed to determine how individual spectral bands related the reference development data (VT and IPS) to absorbance (fig. 4). This analysis assumed that there were no interrelationships among spectral bands. Correlation magnitudes for both reference variables were less than 0.40 for all wavelengths, indicating that no single band's absorbance can be used as a standalone indicator of visible tassels or initial pollen shed.

ARTIFICIAL NEURAL NETWORK (ANN)

The spectra were sorted by class, and then each class was divided into training, testing, and validation sets with proportions of one quarter, one quarter, and one half, respectively. Spectra were then normalized to a zero mean

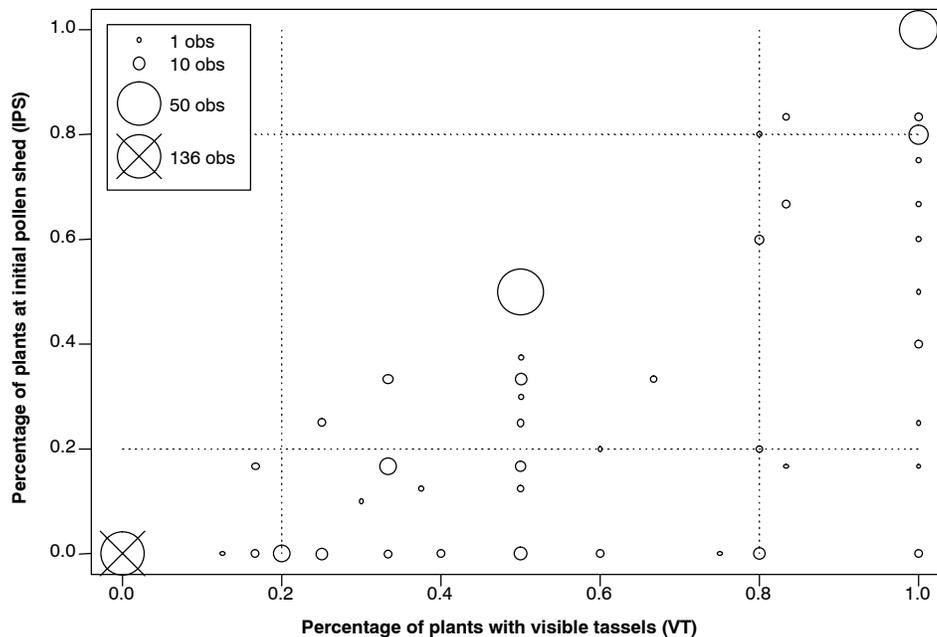


Figure 2. Values of VT and IPS plotted, with the size of the symbol proportional to the number of data points with that combination of VT and IPS (412 data points total). Class thresholds at 20% and 80% are also shown.

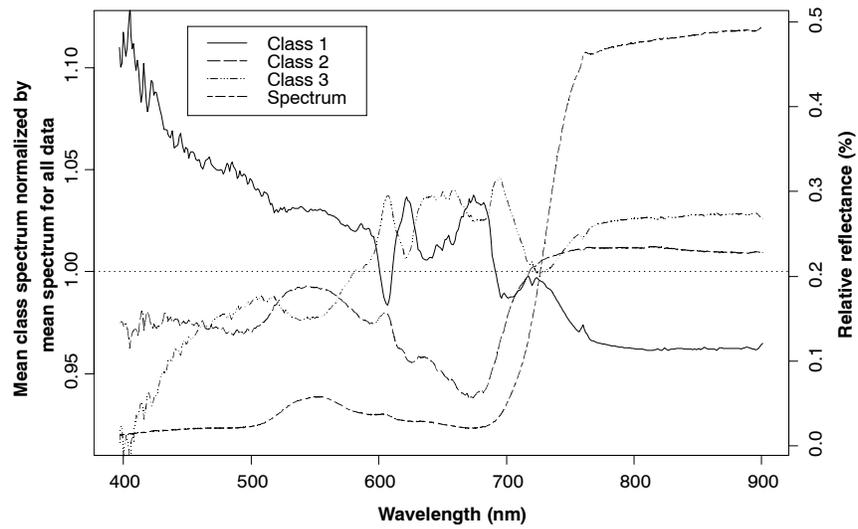


Figure 3a. Mean reflectance spectra for each of the three classes of VT: VT less than 20% (class 1), VT between 20% and 80% (class 2), and VT greater than 80% (class 3). Spectra were normalized by the mean spectrum of the three classes, which is plotted for reference.

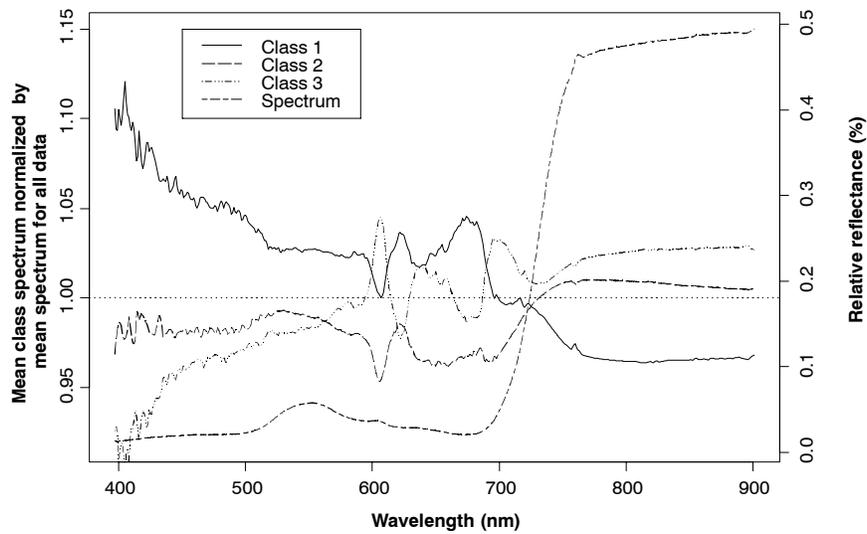


Figure 3b. Mean reflectance spectra for each of the three classes of IPS: IPS less than 20% (class 1), IPS between 20% and 80% (class 2), and IPS greater than 80% (class 3). Spectra were normalized by the mean spectrum of the three classes, which is plotted for reference.

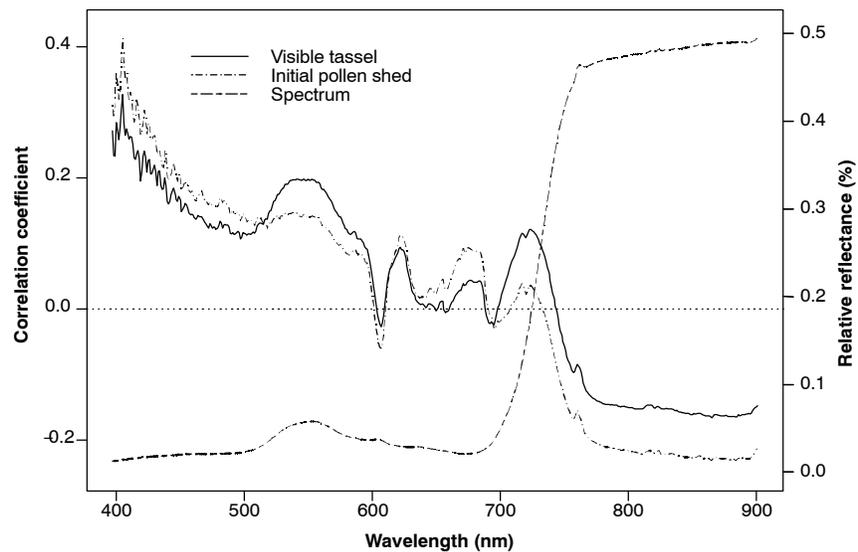


Figure 4. Correlation by wavelength between spectra and reference variables. The mean reflectance spectrum is shown for reference.

Table 1. Summary of classification results for the artificial neural network (ANN), partial least squares (PLS), and range operator-enabled genetic algorithm (ROE-GA).

Method	VT Classification Accuracy (%)			Overall
	Class 1	Class 2	Class 3	
ANN	88.9	82.1	78.6	83.5
PLS	88.4	76.8	51.3	78.8
ROE-GA	81.0	84.2	75.0	80.6

Method	IPS Classification Accuracy (%)			Overall
	Class 1	Class 2	Class 3	
ANN	93.2	76.9	88.8	88.3
PLS	73.7	83.6	61.2	71.7
ROE-GA	89.9	92.3	44.4	82.5

and unit standard deviation. Principal components were calculated to reduce the dimension of the spectral data further, and only those seven components that individually captured more than 0.1% of the variance in the data set were retained. An artificial neural network was used as the classifier. The network consisted of seven input neurons corresponding to the number of selected principal components, a single hidden layer consisting of 20 neurons, and an output layer consisting of three neurons, one for each class. The neural network was trained using back-propagation with the stopping condition that validation performance did not improve within five epochs. After training, the VT and IPS test sets were applied to the neural network, and the correct classification ratio (number classified correctly to total number classified) was calculated for each. The correct classification ratios were 83.5% and 88.3% for the VT and IPS data, respectively (table 1).

The performance of the neural network indicates that, contrary to the results of the correlation analysis, there are signals embedded in the spectra that are related to VT and IPS; however, extensive processing of the data may be necessary to draw out these relationships. A disadvantage of the neural network is that, because of the complex structure of the network operations, it is not possible to use the final

results from the network to identify specific spectral regions that are influenced by VT or IPS.

PARTIAL LEAST SQUARES (PLS)

A PLS analysis was performed on the absorbance data for each of the two reference variables, with 20 latent variables used to develop the final model. PLS identifies the most predictive orthogonal components in the original dataset for use in predicting the reference variable. In this way, redundancies in the data are eliminated, and composite signals related to the reference variables are identified. In each case, PLS analysis gave models with a good fit to the reference data (R^2 of 0.80 and 0.79 for VT and IPS, respectively). For comparison to other methods, we also applied the 20-80-100 classification scheme defined above to the PLS results, which gave 78.8% and 71.7% accuracy for VT and IPS, respectively (table 1). Class 3 (values of VT or IPS greater than 80%) was the most frequently misclassified, with 48.7% (VT) and 38.8% (IPS) being incorrectly identified as class 2. Coefficients of the PLS models did not show a consistent pattern across the spectrum (fig. 5). Thus, while it is possible with this method to analyze pieces of the spectrum influenced by VT or IPS by looking at the coefficients, it is difficult to identify diagnostic regions of the spectrum from these coefficients.

RANGE OPERATOR-ENABLED GENETIC ALGORITHM (ROE-GA)

Using MATLAB (The Mathworks, Natick, Mass.) script, we developed a range operator-enabled genetic algorithm (ROE-GA) to search the space of possible spectral ranges in combination with several basic mathematical range operators. Unlike other analytical techniques, the ROE-GA approach exploits features of the shape of the spectra in addition to individual bands. Features extracted in the form of the range operator outputs were used as inputs into PLS, which was used to calibrate a model for predicting the reference variables, VT and IPS. Cross-validation, discussed

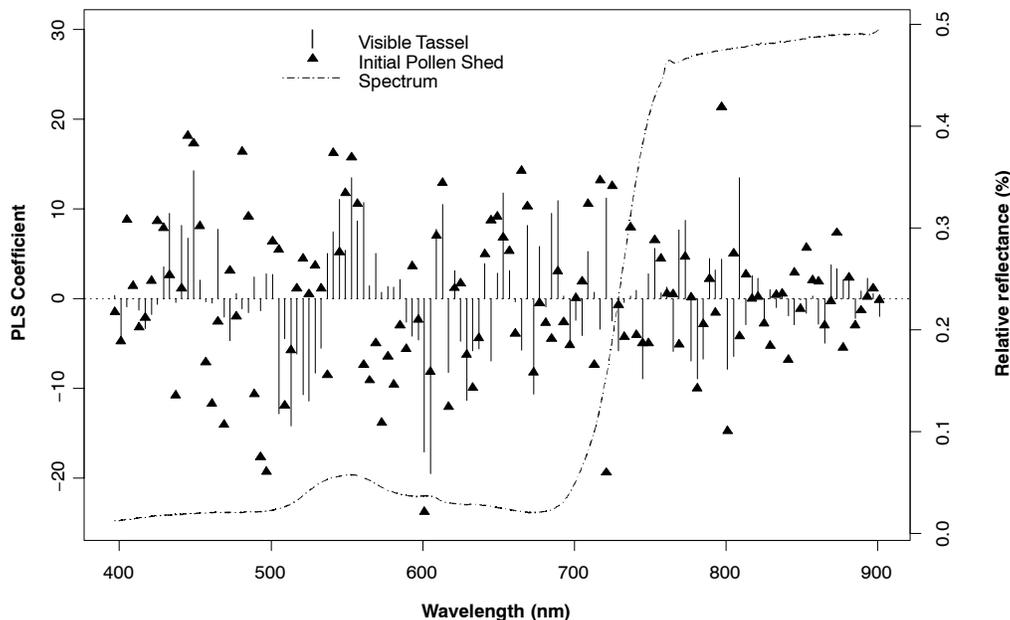


Figure 5. PLS coefficients for models of VT and IPS. The mean reflectance spectrum is shown for reference.

below, was performed using multivariate analysis algorithms in PLS Toolbox (Eigenvector Research, Manson, Wash.), yielding the standard error of cross-validation (SECV) as the fitness metric for the GA.

A population of 256 individuals was randomly initialized. Every individual in the population contained 25 range operation combinations, referred to as chromosomes in GA terminology (Goldberg, 1989). Each chromosome consisted of three numbers representing: (1) the spectral starting point (*S*); (2) the length (*L*) of the range, with a maximum length of 100 bands; and (3) a code (*O*) for the operator. For convenience, we refer to an individual chromosome as an *SLO*. In this way, each individual in the 256-member population consisted of 25 range operation combinations.

Five simple range operators were used: maximum, minimum, median, slope (determined from linear regression of fit to the absorbance data versus waveband, over the range defined by *S* and *L*), and curvature (determined from a second-order polynomial fit to the absorbance data versus waveband, over the range defined by *S* and *L*). A sixth operator was also used, which returned a value of zero for that *SLO*, effectively turning it off. We initially experimented with other operators, including mean and variance, but discarded them because they were infrequently selected. Furthermore, in the case of the mean operator, there was likely some content duplication with the median operator.

Cross-validation was accomplished using random segment selection in which the complete spectral dataset was divided into five segments. Each segment was predicted using a PLS calibration model calculated from the remaining four segments. This five-segment cross-validation was repeated ten times. *SLO*-based PLS model performance was measured using SECV. The minimum SECV across spectral models with 1 to 20 latent variables was used as a measure of model performance and thus the fitness for each individual member of the population.

Absolute fitness replacement was implemented by sorting the population by fitness and discarding the less fit half of the population. A new lower half was then generated from the upper half using two-point crossover. Each new individual produced by crossover was selected for mutation. The mutation rate (the number of chromosomes per individual that can mutate) was variable, starting with a very high mutation rate, 25 chromosomes, for 75 generations, and then linearly decreasing to a lower rate, four chromosomes, at the 225th generation. When a chromosome was mutated, only a single component was affected: the starting point *S*, length *L*, or operator *O*. Mutations for *S* and *L* were limited to changes of 10-band intervals. Mutation of *O* was conversion to another operator selected at random. The ROE-GA stopped when either 300 generations had passed or when more than half of the population contained duplicate individuals. More details on the ROE-GA are available in Steward et al. (2004).

This process was repeated five times, each realization beginning with a different randomly initialized population. The best fit model from the five ending populations had a root mean square error of prediction (RMSEP) of 0.149 for VT and 0.141 for IPS. Classification accuracy with the 20-80-100 scheme using the best fit model was 80.6% and 82.5% for VT and IPS, respectively (table 1). As with the PLS, class 3 was the most frequently misclassified. For both VT and IPS, ROE-GA classification was slightly better than PLS classification, but not as good as the neural network classification.

Use of the ROE-GA approach also allowed us to identify portions of the spectrum that contain relevant information for predicting each reference variable. To illustrate, we examined the five fittest individuals from each of the five realizations. Those *SLOs* that were used by 15 or more of these top 25 individuals were mapped onto the mean reflectance spectrum to determine which spectral features were selected by the ROE-GA method (tables 2 and 3; fig. 6).

Within the VT models, 10 spectral range operations were found in 15 or more models out of the top 25 (table 2). Within the IPS models, 11 popular *SLOs* were identified (table 3). The proportion of operator types was different for the two reference data sets: for VT, 61% of the popular *SLOs* used slope or curvature operators, whereas for IPS, only 46% of the popular *SLOs* were slope or curvature operators, and 53% of the *SLOs* used the median operator. This illustrates that the ROE-GA was able to identify different operator sets that are more appropriate for different reference variables. It further indicates that some spectral shape features contain predictive information.

Table 2. Regions identified as influential for VT prediction by ROE-GA. Only those *SLOs* that were used by 15 or more of the top five models from five populations are presented. Operator (*O*), spectral range determined by *S* and *L*, and feature of the vegetation spectrum in that region are listed.

Operator	Range (nm)	No. of Uses	Spectral Feature (letter in fig. 6)
Maximum	433-673	25	Visible region (a)
Median	593-793	20	Yellow through red edge (i)
	713-833	20	Red edge through NIR plateau (l)
	873-901	25	Far edge of spectra (n)
Slope	753-901	31 ^[a]	Entire NIR plateau (m)
Curvature	513-593	15	Green peak (c)
	553-673	15	Middle of green peak through red (e)
	593-673	15	Yellow through red (f)
	593-713	25	Yellow through beginning of red edge (g)
	873-901	25	Far edge of spectra (o)

^[a] Because the algorithm accepted the entire spectrum first for the *SLO* identification, and then truncated the wings of the data before any computation, these represent some *SLOs* that were longer than the accepted spectral range but functionally identical to shorter *SLOs*.

Table 3. Regions identified as influential for IPS prediction by ROE-GA. Only those *SLOs* that were used by 15 or more of the top five models from five populations are presented. Operator (*O*), spectral range determined by *S* and *L*, and feature of the vegetation spectrum in that region are listed.

Operator	Range (nm)	No. of Uses	Spectral Feature (letter in fig. 6)
Median	473-713	25	Green peak through beginning of red edge (b)
	593-713	15	Yellow through beginning of red edge (g)
	593-753	20	Yellow through beginning of NIR plateau (h)
	673-873	20	Red through NIR plateau (j)
	793-873	15	Within NIR plateau (n)
	873-901	25	Far edge of spectra (o)
Slope	553-593	25	Downslope of green peak before small yellow peak (d)
	753-901	25	Entire NIR plateau (m)
	873-901	20	Far edge of spectra (o)
Curvature	713-753	15	Red edge (k)
	873-901	20	Far edge of spectra (o)

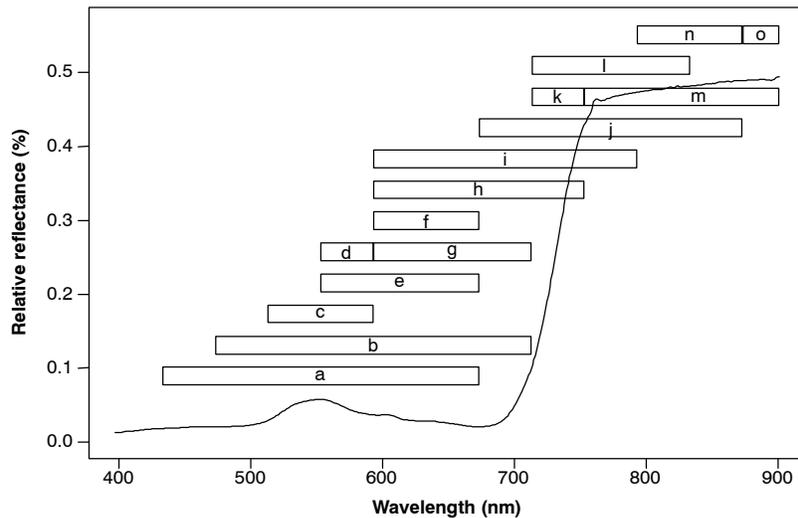


Figure 6. Ranges identified as influential by ROE-GA. Tables 2 and 3 give the explanation of the ranges shown in this figure by letter. The mean reflectance spectrum is shown for reference.

Many of the successful *SLOs* occurred in regions of the reflectance spectrum where one would expect changes to occur due to reproductive stage. For example, several of the *SLOs* were close to or included the green peak of the reflectance curve. Similarly, several *SLOs* included changes associated with the red edge or the near-infrared plateau. However, several other ranges were identified as influential, indicating the need for further spectral analysis using additional, more specific analytical techniques or domain expertise. The ROE-GA results indicate features of the spectrum that are useful in predicting phenological stage itself.

CONCLUSIONS

Advances in remote sensing technology offer opportunities for improved spectral analysis techniques that take advantage of the high spectral resolution possible from modern sensors. Several analysis methods were compared in order to analyze their utility for estimating corn pollen shed stage from hyperspectral data. We conclude that corn canopies undergo detectable and diagnostic changes in reflectance in portions of the visible and near-infrared spectrum during tasseling and initial pollen shed. These changes in reflectance can be used to develop models to accurately classify plants according to their development stage.

Simple correlation analysis did not give useful results. While this type of analysis could potentially indicate regions of the spectra that are influenced by reproductive changes, no single band, when considered alone, was predictive. Classification using an artificial neural network (ANN) was also predictive, but the extensive preprocessing necessary and the "black box" nature of the ANN mean that analysis of different regions of the spectra is difficult using this method. Similarly, partial least squares (PLS) analysis yielded models with high predictive capability, but because the PLS coefficients did not exhibit spectrally consistent patterns, PLS could not be used to improve our understanding of the relationship between corn plant development and spectral reflectance or absorbance.

A range operator-enabled genetic algorithm (ROE-GA) had similar predictive capabilities to the ANN and the PLS, but had the added advantage of allowing information transfer for increased domain knowledge. It also demonstrated the utility of using operators like slope and curvature, which consider the shape of the spectra. We conclude that this is a valuable tool for hyperspectral data analysis, and show how it provides a reliable means to estimate the onset of pollen shed in corn canopies.

ACKNOWLEDGEMENTS

This research of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 3612, was supported by Hatch Act and State of Iowa funds. This research was also supported in part by the Iowa State University Special Research Initiation Grant Program and the Iowa Space Grant Consortium. Special thanks are given to Wolfgang Oesterreich for his efforts in collecting the field data with the spectroradiometers.

REFERENCES

- Burman, J. A. 1997. Non-literal pattern recognition method for hyperspectral imagery exploitation using evolutionary computing methods. In *Imaging Spectrometry III: Proc. SPIE* 3118: 250-261.
- Burman, J. A. 1999. Hybrid pattern recognition method using evolutionary computing techniques applied to the exploitation of hyperspectral imagery and medical spectral data. In *Image and Signal Processing for Remote Sensing V: Proc. SPIE* 3781: 348-357.
- Clevers, J. G., S. M. De Jong, G. F. Epema, F. D. Van Der Meer, W. H. Bakker, A. K. Skidmore, and K. H. Scholte. 2002. Derivation of the red edge index using the MERIS standard band setting. *Int. J. Remote Sensing* 23(16): 3169-3184.
- Demetriades-Shah, T. H., M. D. Steven, and J. A. Clark. 1990. High-resolution derivative spectra in remote sensing. *Remote Sensing Environ.* 33(1): 55-64.
- Fonseca, A. E., M. E. Westgate, L. Grass, and D. L. Dornbos, Jr. 2003. Tassel morphology as an indicator of potential pollen production in maize. Online. *Crop Management* doi:10.1094/CM-2003-0804-01-RS.

- Goel, P. K., S. O. Prasher, J. A. Landry, R. M. Patel, A. A. Viau, and J. R. Miller. 2003. Estimation of crop biophysical parameters through airborne and field hyperspectral remote sensing. *Trans. ASAE* 46(4): 1235-1246.
- Goggi, A. S., P. Caragea, H. Lopez-Sanchez, M. Westgate, R. Arritt, and C. Clark. 2006. Statistical analysis of outcrossing between adjacent maize grain production fields. *Field Crops Research* 99(2-3): 147-157.
- Goldberg, D. E. 1989. Sizing populations for serial and parallel genetic algorithms. In *Proc. 3rd Int. Conference on Genetic Algorithms*, 70-79. San Francisco, Cal.: Morgan Kaufman.
- Ireland, D. S., D. O. Wilson, Jr., M. E. Westgate, J. S. Burris, and M. J. Lauer. 2006. Managing reproductive isolation in hybrid seed corn production. *Crop Sci.* 46(4): 1445-1455.
- Leardi, R. 2000. Application of genetic algorithm-PLS for feature selection in spectral datasets. *J. Chemometrics* 14(5-6): 643-655.
- Martens, H., and M. Martens. 2001. *Multivariate Analysis of Quality*. New York, N.Y.: Wiley.
- Rauss, P. J., J. M. Daida, and S. Chaudhary. 2000. Classification of spectral imagery using genetic programming. In *Proc. Genetic and Evolutionary Computation Conference*. San Francisco, Cal.: Morgan Kaufmann.
- Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. In *Proc. 3rd ERTS Symposium*, 1: 48-62. Greenbelt, Md.: NASA SP-351.
- Steward, B. L., R. P. Ewing, D. A. Ashlock, A. L. Kaleita, and S. M. Shaner. 2004. Range operator enabled genetic algorithms for hyperspectral analysis. In *Intelligent Engineering Systems through Artificial Neural Networks: Smart Engineering Systems Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Complex Systems, and Artificial Life* 14: 295-300. C. H. Dagli, A. L. Buczak, D. L. Enke, M. J. Embrechts, and O. Ersoy, eds. New York, N.Y.: ASME Press.
- Tang, L., L. Tian, and B. L. Steward. 2000. Color image segmentation with genetic algorithm for in-field weed sensing. *Trans. ASAE* 43(4): 1019-1027.
- Viña, A., A. A. Gitelson, D. C. Rundquist, G. Keydan, B. Leavitt, and J. Schepers. 2004. Monitoring maize (*Zea mays* L.) phenology with remote sensing. *Agron. J.* 96(4): 1139-1147.
- Westgate M. E., D. Ireland, B. Ashton, D. P. Todey, G. S. Takle, and C. Daniel. 2000. Predicting maize pollen travel using particulate dispersion models. In *ASA-CSSA-SSSA Annual Meetings Abstracts*. Madison, Wisc.: ASA-CSSA-SSSA.
- Westgate, M. E., J. Lizaso, and W. D. Batchelor. 2003. Quantitative relationships between pollen shed density and grain yield in maize. *Crop Sci.* 43(3): 934-942.
- Whiting, M. L., L. Li, and S. L. Ustin. 2004. Predicting water content using Gaussian model on soil spectra. *Remote Sensing Environ.* 89(4): 535-552.
- Wold, H. 1982. Soft modelling: The basic design and some extensions. In *Systems under Indirect Observations. Causality-Structure-Prediction*, 1: 263-271. Amsterdam, The Netherlands: North-Holland.
- Wolt, J. D., Y.-Y. Shyy, P. Christensen, K. S. Dormin, and M. Misra. 2005. Quantitative exposure assessment for confinement of maize biogenic systems. *Environ. Biosafety Res.* 3(4): 183-196.
- Yao, H., and L. Tian. 2003. A genetic algorithm-based selective principal component analysis (GA-SPCA) method for high-dimensional data feature extraction. *IEEE Trans. Geosci. Remote Sensing* 41(6, pt. 1): 1469-1478.