Biosystems Engineering (2002) **83** (3), 291–298
doi:10.1016/S1537-5110(02)00195-2, available online at http://www.idealibrary.com on IDEAL®
PA—Precision Agriculture

Use of Hyperspectral Imagery for Identification of Different Fertilisation Methods with Decision-tree Technology

Chun-Chieh Yang[1]; Shiv O. Prasher[1]; Joann Whalen[2]; Pradeep K. Goel[1]

[1] Department of Agricultural and Biosystems Engineering, Macdonald Campus of McGill University, Box 950/21, 111 Lakeshore Road, Ste-Anne-de-Bellevue, Qué., Canada H9X 3V9; e-mail of corresponding author: prasher@macdonald.mcgill.ca
[2] Department of Natural Resource Sciences, Macdonald Campus of McGill University, Ste-Anne-de-Bellevue, Qué., Canada H9X 3V9

(Received 7 November 2001; accepted in revised form 9 August 2002)

This paper introduces data mining technology designed to classify agricultural fields under different manure/fertiliser application strategies. During the summer of 2000, airborne hyperspectral data were collected three times at two field sites in southwestern Quebec, Canada. One field site contained 24 plots (20 m by 24 m) that were amended with manure treatments and planted with maize and soya beans. The second field site contained 18 plots (18·5 m by 75 m) that received chemical fertilisers and were planted with maize. Reflectances of 71 wave bands of hyperspectral data (400 nm for violet to 940 nm for near infrared) were collected from 5 subplots within each of the 42 plots. The decision-tree algorithm of data mining technology was used to distinguish between manure and chemical fertiliser treatments. The decision-tree algorithm divides the data to reduce the deviance, and classifies them into the pre-defined categories as many tree branches. The success of the classification rate was as high as 91% for the early planting season, 99% for the mid-planting season, and 95% for the late planting season. The accuracy of the results demonstrates that data mining technology could be used for remote-sensing imagery classification of fertiliser applications.

1. Introduction

The application of animal manure to agricultural land has been viewed as an excellent way to recycle nutrients and organic matter that can support crop production and maintain, or improve, soil quality. Generally, soil organic matter and biological activity increase, and soil physical properties such as aggregation and tilth improve following manure applications (Hafez, 1974; Sommerfeldt *et al.*, 1988; Haynes & Francis, 1993). Storing more carbon in soils (carbon sequestration) has been proposed as one way of mitigating atmospheric CO_2 increases. Manure could have a role in carbon sequestration due to its positive effects on crop production (organic residues from crops are the source of new soil carbon) and by improving soil aggregation. Organic matter is distributed in aggregates, and its resistance to decomposition depends on its physical location in aggregates and aggregate stability (Adu & Oades, 1978; Six *et al.*, 1998). Numerous studies have found the proportion of water-stable macroaggregates increases soon after soils begin receiving manure, and it has been proposed that long-term manure applications will increase the amount of organic matter associated with soil minerals and improve microaggregate stability (Aoyama *et al.*, 1999).

Despite the many positive consequences of applying manure to agricultural soils, a major concern in areas of high animal density is the potential for increased atmospheric and water pollution resulting from improper storage or application of manure. In areas where many intensive livestock operations exist, manure applications are often limited to land in the vicinity of the operations, with the result that manure is often applied with high frequency and high rates. As a result, the nearest land may be amended with large quantities of manure on a frequent basis. The export of N from manure-amended soils through greenhouse gas emissions and transport processes, such as leaching, surface runoff and erosion have been well documented (Adams *et al.*, 1994; Chang & Janzen, 1996; Goss & Goorahoo, 1995; Zebarth *et al.*, 1999). Migration of P from manure-amended soils to ground and surface waters has been linked to eutrophication of aquatic systems (Sharpley *et al.*, 1994; Heathwaite, 1997).

Clearly, manure applications must be managed carefully to minimise nutrient export from agricultural systems into the atmosphere and waterways. Policy-makers and regional planners require information on the amount of agricultural land that receives manure to identify areas at risk for water pollution and develop new guidelines for the siting of intensive livestock operations, which will improve manure utilisation on agricultural land. At present, this type of information is collected by surveying livestock producers, a time-consuming and costly practice that generally provides information for a small percentage of the total producers operating in a region. New technologies to rapidly collect such information are urgently needed.

Remote sensing by aircraft or satellite can provide information on agronomic practices at the scale of an individual producer's fields. Hyperspectral images, which are based on reflectances from the visible and near-infrared regions of the electromagnetic spectrum, have been used successfully to monitor crop cover, crop health and soil moisture in agricultural fields (Barnes & Baker, 2000; Chang *et al.*, 2001; Chen *et al.*, 2000; Lillesand & Kiefer, 1994). Thenkabail *et al.* (2000) recommended optimal hyperspectral wave bands, wave centres, and wave widths in the visible and near-infrared spectral ranges to identify crop characteristics and vegetation indices for cotton, potato, soya beans, maize, and sunflower. Crop residues could be differentiated from soils using spectral reflectance (Daughtry, 2001; Nagler *et al.*, 2000). However, hyperspectral images have not been used previously to detect impacts of manure application to agricultural fields. This technology, if applicable, could be used to rapidly map the spatial distribution of manure applied across many agricultural fields. Such knowledge could lead to significant environmental and economic benefits.

However, accurate interpretation of hyperspectral imagery from agricultural fields is difficult due to spectral mixing (Barnes & Baker, 2000; Lillesand & Kiefer, 1994). Lately, machine learning algorithms, such as data mining, are being used for classification problems (Teorey, 1999; Witten & Frank, 2000). These technologies, i.e. decision trees, neural networks, association analysis, and instance-based learning, search and discover knowledge from the data. Decision trees recursively split input data into branches to reduce the deviance within the data in each branch until all of the data are assigned to proper categories (Han & Kamber, 2001). Neural networks generate the implicit relationship between the input data and the outputs through a group of parallel computation units and their inter-connections (Gurney, 1997; Yang *et al.*, 2000). Association analysis determines the association rules between the input data attributes and the output values, and then selects the most frequently applied rules from other alternatives (Han & Kamber, 2001). Instance-based learning stores instance data with specific output attributes, and groups the input data with the most proper and closest instance data in the same class (Witten & Frank, 2000).

Among the above data mining methods, decision-tree algorithms have become popular (Flamig, 2000; Witten & Frank, 2000). They not only provide an efficient classification method, but have the additional advantage of providing ease of interpretation of the rules used to filter data sets to their appropriate categories, while simultaneously bringing to light the relative importance of different variables in the system studied.

The primary goal of this study was to explore the use of hyperspectral imagery in differentiating fields receiving organic manure from those with chemical fertilisers. Additionally, the imagery was used to identify maize and soya bean crops as a function of fertiliser treatments. The hyperspectral data were collected with a Compact Airborne Spectrographic Imager (Borstad Associates Ltd., Sidney, Canada) sensor on an airborne platform from two experimental fields, one for maize and soya bean applied with organic manure and another for maize with chemical fertilisers. Several models utilising the classification and regression trees (C&RT) method, one of the decision-tree algorithms, were used to classify these hyperspectral data. The models were trained to recognise the fertiliser-application strategies and the crop type for the experimental plots using different dates of data collection. The study evaluated the capability of the decision-tree models for the classification of fertiliser-application strategy on different dates of a growing season.

2. Materials and methods

2.1. *Field experiments*

The hyperspectral data were obtained over two adjacent fields, receiving organic and chemical fertilisers, at the Macdonald Campus Experimental Farm, in southwestern Quebec, Canada. The soil was a Typic Endoaquent (St. Amable and Courval series) with sandy loam or loamy sand surface textures. Field A (organic manure) contained 24 plots (20 m by 24 m): 16 were assigned randomly to maize (*Zea mays* L.) and eight to soya beans (*Glycine max* (L.) Merr.). Field B (chemical fertiliser) contained 18 plots (18·5 m by 75 m), all planted with maize. The experimental plots were not specifically established for this study, so the number of maize and soya bean plots were not equal.

In 2000, Field A was seeded on 30 May, whereas Field B was seeded on 8 May. However, it was observed that the relatively cold and wet spring significantly delayed the germination of maize by 2–3 weeks in Field B. The maize in Field A had grown to the same developmental stage as maize in Field B by mid-June. Fertilisers were applied at the seeding dates in both fields. Organic fertilisers, composted cattle manure, were applied to Field A. Cattle manure (Les Composts du Québec) was applied prior to seeding at a rate of 22 mg ha^{-1} (wet weight), composed of, on average, 15·4 g[N] kg^{-1} (dry weight basis), 3·3 g[P] kg^{-1}, 18·7 g[K] kg^{-1} with a moisture content of 0·68 kg[H_2O] kg^{-1}. Chemical fertilisers, 100 kg[P_2O_5] ha^{-1} on all plots and 39 kg[N] ha^{-1} as diammonium phosphate (18-46-0) banded at seeding, were applied to Field B.

Hyperspectral images with a resolution of 2 m by 2 m were taken on 30 June, 5 August, and 25 August, 2000 using a Compact Airborne Spectrographic Imager (Borstad Associates Ltd., Sidney, Canada). Proper radiometric, geometric, and atmospheric corrections were applied on the collected hyperspectral data. Seventy-one wave bands were measured: six in the violet range (408·73–445·79 nm), seven in the blue range (453·21–497·90 nm), ten in the green (505·37–572·82 nm), two in the yellow (580·34–587·86 nm), four in the orange (595·39–618·02 nm), ten in the red (625·57–693·76 nm) and 32 in the near infrared (701·36–939·33 nm). Band widths slightly varied from 4·27 nm in the violet to 4·40 nm in the near infrared. From the hyperspectral images, each plot was visually divided into five subplots of equal size to not only obtain more data points per treatment but also properly represent the detailed attributes within each plot. The reflectances at each wave band for a given subplot were averaged among all pixels clearly associated with the subplot (i.e. pixels not overlapping other subplots). A set of 120 data points (5 subplots by 24 plots) were collected for Field A and 90 data points (5 subplots by 18 plots) for Field B, each of which contained the average reflectance values of the 71 wave bands. The average reflectances of subplots were used in the development of the decision-tree models.

2.2. *Description of the decision-tree algorithm*

AnswerTree version 2.1 was used to generate decision trees (SPSS Inc., Chicago, IL, USA). This data mining software package was run on a personal computer (PC) with a Pentium II-600 microprocessor, 28 GB hard disk space, 128 MB of random access memory (RAM), under the Windows 2000 operating system. The C&RT (Breiman *et al*., 1984) was chosen from several decision-tree algorithms offered in the package for this study because the input variables, the reflectance of wave bands, were continuous, as opposed to categorical.

The C&RT is a recursive algorithm that splits the entire data set from the root node into smaller subsets to reduce the deviance and correct for the total sum of the squares. In the tree structure, each subset is termed a node. A mother node contains data that can be split into another subset, called the child node. When node data cannot be split into additional subsets, it is called a terminal node. Once the first subset has been created, the algorithm repeats the procedure for each subset until all data are categorised as terminal nodes. *Figure 1* shows the general architecture of a C&RT model.

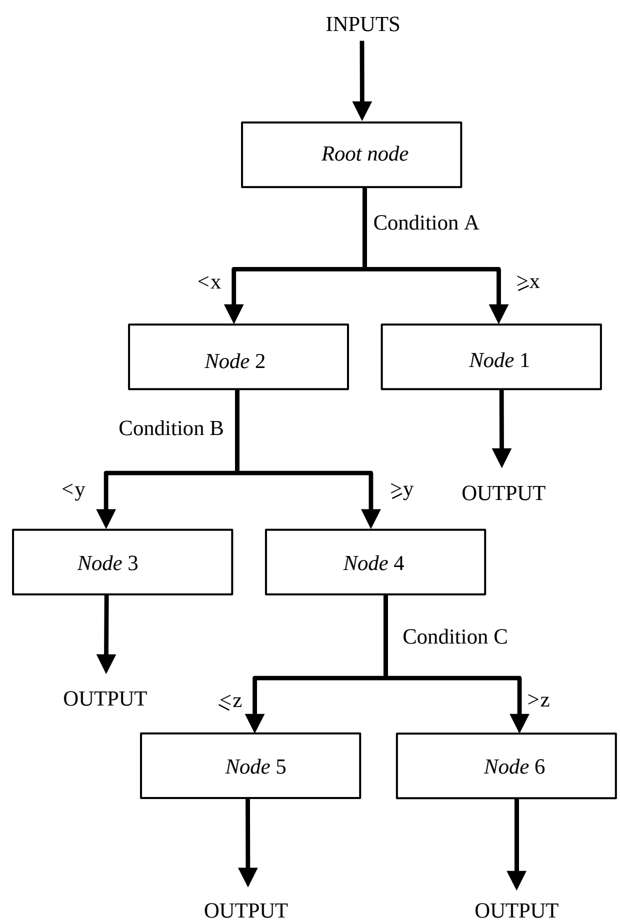

Fig. 1. General architecture of a decision-tree model; x, y, and z are the threshold values of the split points for the corresponding split conditions A, B, and C to split the data from the mother nodes into child nodes

A search for a split point is conducted over all input variables, and the reduction in deviance $D(total)$ is maximised as follows:

$$D(total) = \sum (Y_i - Y_{bar})^2 \quad (1)$$

where: Y_i is the value of the target variable and Y_{bar} is the mean for the node. At each node where splitting of data into two mutually exclusive subsets takes place, the reduction in deviance Δ is as follows:

$$\Delta_{j,total} = D(total) - (D(L) + D(R)) \quad (2)$$

where, for split j at this node, $D(L)$ and $D(R)$ are the deviancies of the left and right subsets. Thus, the C&RT algorithm first searches for Δ_{max} over all input variables and possible split points, where the number of data in the left and right subsets is larger than a pre-defined minimum threshold. Moreover, for the second and subsequent splits, $D(N)$ should be larger than $0{\cdot}01 \times D(total)$. Thus, if the deviance of a node is not greater than 1% of the deviance of the original data set, the node is not split any further and becomes a terminal node. Furthermore, a maximum number of tree levels, a minimum number of data set for the mother nodes and a minimum number of data set for the child nodes need to be pre-defined. When the tree level reaches the pre-defined maximum level or the data, either before splitting in the mother node or after splitting in the child node, reaches the pre-defined minimum amount, the nodes are forced to be terminal nodes even if the data in these nodes can potentially be split again. The above conditions are termed stopping rules.

Except for obtaining the final decision-tree model, stopping rules are set to prevent the model from overfitting the training data. Moreover, the decision tree can be pruned to reduce the tree level if the tree is judged to already overfit the training data. Overfitting of the training data can make the decision trees fit the training data perfectly, but poorly fit other independent data.

For the categorical output variables, the suggested splitting criterion is the Gini impurity g (Breiman *et al.*, 1984; SPSS, 1998), defined as follows:

$$g(m) = 1 - \sum_i p^2(i|m) \quad (3)$$

where: g is the Gini impurity and i is the number of output categories at the node m, p is the proportion of data sent from the mother node m to the child node. When all of the data in the node m can be classified as one category, $g(m)$ is equal to zero, since the summation is of only one term whose value is one. Based on this criterion, the decision tree continues to build more tree levels and to classify the data into more categories, until one of the stopping rules is satisfied.

2.3. *Decision-tree model development*

In this study, three types of models were developed to classify data from the experimental plots. Since the atmospheric environment, weather, and lighting condition varied from date to date, all of the decision-tree models were developed for a single date respectively. Each date of data collection would represent different planting seasons (30 June for the early planting season, 5 August for the middle planting season, and 25 August for the late planting season). In Model (1), the decision trees were developed to distinguish plots receiving organic manure (Field A) from those amended with chemical fertilisers (Field B). In Model (2), the decision trees were developed to differentiate maize plots (Fields A and B) from the soya bean plots (Field A). In Model (3), the decision trees were developed to identify three types of plots: organic-manure-amended plots under maize production (Field A), organic-manure-amended plots under soya bean production (Field A), and chemical-fertiliser-amended plots under maize production (Field B). All 210 data from Fields A and B were used for the development of Models (1) and (3). For the development of Model (2), two submodels were generated with a different training data set. Model (2a) was developed with 120 data from Field A, while Model (2b) was generated with all 210 data from Fields A and B. Separate models were generated from the hyperspectral data collected at each sampling date, for a total of 12 models that were trained and analysed.

Preliminary model runs showed that one of the stopping rules was usually applied before the model needed to grow more than five tree levels. It indicated that no more than five tree levels were required to classify the data, so the boundary for the maximum tree level was set at five. The minimum number of data was set at five for the mother nodes and two for the child nodes. To precisely evaluate the performance of decision trees, the decision-tree models were cross-validated using the n-fold approach (SPSS, 1998; Weiss and Kulikowski, 1991), where n is the number of data sets into which the training data is going to split, with a value for the sample size n of 10, as recommended by Weiss & Kulikowski (1991). For the development of each model, the training data were randomly split into ten sets, with a random seed of 2 000 000 as the default number, set by the software. A decision tree was then generated from nine of the ten sets and validated by the remaining data set. This process was repeated ten times so that each data set was used to validate the model once. The average recognition rate and standard error for the cross-validation were calculated by the software as conservative estimates of model accuracy (SPSS, 1998). After the cross-validation, a final decision tree was generated and tested with all the training data. The success recognition rate and standard error for the final decision tree was also calculated and reported as the optimal estimation for model accuracy. The misclassification matrix for the final decision tree with all training data was also presented and analysed.

3. Results and discussion

From generally high success classification rates for cross-validation, as shown in Table 1, decision trees successfully recognised the hyperspectral images for different fertiliser application strategies and crop types when the hyperspectral measurements of reflectance were used as the inputs for the decision trees for image classification. A model with a single classification purpose could have a success recognition rate as high as 99% for Model (1) and 100% for Model (2) (Table 1). Even models with a multi-classification purpose such as Model (3) could have a success classification rate as high as 97% (Table 1). The results for cross-validation also indicate that the success classification rate could be as high as 86% for Model (1), 99% for Model (2) and 87% for Model (3) (Table 1). However, the significant variation in the success classification rates between cross-validation and the final model, Model (1) had a 13% difference and Model (3) had a 10% difference, indicated that the model performance for the classification of fertiliser use would vary from plot to plot. This type of classification could be difficult for some plots.

Model (2a) was developed with fewer data than Model (2b), and the results indicate that more training data improved crop classification, without overfitting the data during the early stages of crop growth. The success classification rate for cross-validation increased from 64% in Model (2a) to 75% in Model (2b) for the June 30 collection date presumably because more training data was included in Model (2b) (Table 1). Such an improvement indicates that, for the early planting season, the model training might still be insufficient and more training data should be collected for better performance of the model.

Hyperspectral reflectance from crops early in the season (*i.e.* 30 June) before canopy closes may have had interference from weeds and soil. The classification performances of decision trees in all models were highest for the data collected on 5 August (middle planting season), followed by data collected on 25 August (late planting season) and 30 June (early planting season). The optimal date for hyperspectral data collection appeared to be in the middle planting season, when plants were large enough that different crop species were easily identified and hyperspectral reflectance from the soil surface was still visible (*i.e.* not obscured by the plants). Ideally, more measurements should have been made to evaluate plant physiological developments, such as leaf area index, leaf nitrogen, and photosynthesis rate, and support the results given in Table 1. In the absence of these measurements, it can still be observed that the decision-tree approach is successful in identifying fertiliser treatments using the imagery collected in the early planting season when the plants were quite small.

There was more uncertainty associated with decision trees generated from data that was classified by fertiliser source than crop type at all sampling dates. Therefore, the classification and misclassification matrix for data collected on 5 August was generated to better understand the variation associated with different fertilisers and crops.

Table 2 indicated that the decision tree misclassified only three of the120 subplots applied with both chemical fertilisers and organic manure. For image classification, the decision tree for Model (1) required nine wave bands as split points. However, the split points might vary according to different data-collecting dates and locations. The results showed that it was possible to

Table 1
Results of success classification rates and standard errors for decision-tree models

Date	*Classification rate, % (standard error)*							
	Model (1)		*Model (2a)*		*Model (2b)*		*Model (3)*	
	Cross-validation	*Final model*	*Cross-validation*	*Final model*	*Cross-validation*	*Final model*	*Cross-validation*	*Final model*
30 June	71 (0·03)	91 (0·02)	64 (0·04)	94 (0·02)	75 (0·03)	93 (0·02)	59 (0·03)	83 (0·03)
5 August	86 (0·02)	99 (0·01)	99 (0·01)	100 (0·00)	99 (0·01)	100 (0·00)	87 (0·02)	97 (0·01)
25 August	77 (0·03)	95 (0·02)	96 (0·02)	98 (0·01)	98 (0·01)	99 (0·01)	78 (0·03)	95 (0·01)
Number of subplots for model development	210		120		210		210	

Table 2
Classification and misclassification matrix for Model (1) with hyperspectral data from 5 August 2000

Model outputs	*Actual classification outputs*	
	Organic manure	*Chemical fertiliser*
Organic manure	90	3
Chemical fertiliser	0	117

Wave bands selected by the model to split data: 423·53 nm (violet), 430·95 nm (violet), 468·09 nm (blue), 490·44 nm (blue), 497·90 nm (blue), 580·34 nm (yellow), 655·83 nm (red), 693·76 nm (red), 724·20 nm (near infrared).

accurately recognize the fertiliser-application strategy from the remote-sensing images. It is hypothesized that the hyperspectral response of a manure field is different from that of an inorganically fertilised field due to higher carbon contents and organic matters in the former. Accurate results required hyperspectral information from both plants and soils, as well as more training data for fields where chemical fertiliser was applied. In this study, the number of training data for chemical fertiliser application (120 subplots) was higher than for organic manure application (90 subplots). Furthermore, a greater number of subplots applied with chemical fertilisers were misclassified than were ones applied with organic manure.

The decision trees generated to differentiate crop type accurately classify hyperspectral images as maizes or soya beans (Table 3). Regardless of the number of training data, 120 for Model (2a) and 210 for Model (2b), the number of misclassified plots was always zero. Cross-validation results also indicated a very high success classification rate of 99% (Table 1). Furthermore, there was only one wave band selected for the split point, 724·20 nm of near infrared, for Models (2a) and (2b). When the decision tree was trained for the data collected on 25 August, the wave band selected for the split point was 739·45 nm of near infrared, close to the split point of 724·20 nm for 5 August. The less successful classification results for 30 June could be caused by the plant size being too small (i.e., insufficient hyperspectral reflectance by the crops). The hyperspectral response of the two crop types is different due to different leaf properties and canopy structures.

The results show that the decision trees can simultaneously classify hyperspectral images from 210 subplots into two different output categories, fertiliser source and crop type (Table 4). One of the 90 maize subplots receiving chemical fertiliser was misclassified as maize with organic manure, and five of the 80 maize subplots amended organic manure were misclassified as maize with chemical fertiliser. Soya bean subplots receiving organic manure were all classified correctly. The results indicate that it was easier to differentiate maize plots from the soya bean plots than to distinguish soils that received chemical fertilisers from those amended with organic manure. The split point selected to distinguish crops in Model (2) was also selected for Model (3), but the split points selected by the decision tree to distinguish fertilisers in Model (1) were not identical to those selected for Model (3) (Tables 2–4). It was observed that wave bands at 423·53 and 490·44 nm were selected for both Models (1) and (3). Overall, the performance of the decision tree for Model (3) in classifying data collected on 5 August into three categories had a cross-validation success rate of 87% (Table 1).

The results indicated that no more than 12 wave bands of the 71 reflectance inputs were required for the decision-tree model development. However, the wave bands selected as critical inputs for the classification rules varied from model to model, and from date to date. More training data and further investigations are required before the critical wave bands could be applied universally in the classification rules. Future studies will include the development of decision-tree models based on more site-specific hyperspectral data from different dates, years, and locations. This site-specific modelling has the potential to efficiently recognize the type of

Table 3
Classification and misclassification matrix for Model (2) with hyperspectral data from 5 August, 2000

Model outputs	*Actual classification outputs*			
	Model (2a)		*Model (2b)*	
	Maize	*Soya bean*	*Maize*	*Soya bean*
Maize	80	0	170	0
Soya bean	0	40	0	40

Wave bands selected by the model to split data: 724·20 nm (near infrared).

Table 4
Classification and misclassification matrix for Model (3) with hyperspectral data from 5 August, 2000

Model outputs	*Actual classification outputs*		
	Chemical/ Maize	*Manure/ Maize*	*Manure/ Soya bean*
Chemical/Maize	89	5	0
Manure/Maize	1	75	0
Manure/Soya bean	0	0	40

Wave bands selected by the model to split data: 423·53 nm (violet), 460·65 nm (blue), 490·44 nm (blue), 686·17 nm (red), 724·20 nm (near infrared), 739·45 nm (near infrared).

fertiliser uses in different locations. Such a recognition can help to improve the precision and efficiency of fertiliser use in agriculture, which will contribute to the long-term economic and environmental sustainability of this troubled industry.

4. Conclusions

Several decision tree models were developed to classify hyperspectral data, collected on 30 June, 5 August and 25 August, 2000, from a maize/soya bean field amended with organic manure and a maize field amended with chemical fertilisers. The decision trees had a 71–86% success rate in distinguishing fertiliser sources, and a 64–99% success rate in classifying different crop types correctly. A single decision tree was developed to analyse output categories, fertiliser source and crop type, simultaneously. The success classification rate ranged from 59 to 87%. It was determined that the best date for collecting hyperspectral data for these types of classification is during the middle growing season, such as early August, with a planting date in May. During these dates, the hyperspectral data would generally represent the conditions of both plants and soils.

The results show that the data mining technology could be applied to remote-sensing image classification to differentiate agricultural soils that receive organic manure or chemical fertilisers, especially on the basis of the results obtained with the first flight. Had measurements on plant physiological developments been made concurrently, more concrete conclusions about the utility of the decision-tree approach could have been established. Nevertheless, it appears that this technology has potential to rapidly classify the agricultural land base which receives organic and chemical fertilisers. With accurate and fast classification results from the decision trees, watersheds and regions at risk of pollution from fertilisers can be easily located. At the present time, this information can only be obtained by detailed surveys and censuses of producers in areas of interest. Data mining technology could greatly improve the speed at which information on fertiliser use is transmitted to extension agents, agricultural and environmental agencies, as well as local and regional policy-makers. This technology has the potential to improve fertiliser use and reduce pollution from the agricultural sector.

References

Adams P L; Daniel T C; Edwards D R; Nichols D J; Pote D H; Scott H D (1994). Poultry litter and manure contributions to nitrate leaching through the vadose zone. Soil Science Society of America Journal, **58**(4), 1206 –1211

Adu J K; Oades J M (1978). Physical factors influencing decomposition of organic matter in soil aggregates. Soil Biology and Biochemistry, **10**(2), 109–115

Aoyama M; Angers D A; N'Dayegamiye A (1999). Particulate and mineral-associated organic matter in water-stable aggregates as affected by mineral fertiliser and manure applications. Canadian Journal of Soil Science, **79**(2), 295–302

Barnes E M; Baker M G (2000). Multispectral data for mapping soil texture: possibilities and limitations. Applied Engineering in Agriculture, **16**(6), 731–741

Breiman L; Friedman J H; Olshen R A; Stone C J (1984). Classification and Regression Trees. Wadsworth, Belmont, CA, USA

Chang C; Janzen H H (1996). Long-term fate of nitrogen from annual feedlot manure applications. Journal of Environmental Quality, **25**(4), 785–790

Chang C-W; Laird D A; Mausbach M J; Hurburgh Jr C R (2001). Near-infrared reflectance spectroscopy — principal components regression analyses of soil properties. Soil Science Society of America Journal, **65**(2), 480–490

Chen F; Kissel D E; West L T; Adkins W (2000). Field-scale mapping of surface soil organic carbon using remotely sensed imagery. Soil Science Society of America Journal, **64**(2), 746–753

Daughtry C S T (2001). Discriminating crop residues from soil by shortwave infrared reflectance. Agronomy Journal, **93**(1), 125–131

Flamig B (2000). Discover the value of data mining: turning piles of stagnant information into golden opportunities. Reference Series: How Computers Work, Part II **4**(3), 120–125

Goss M J; Goorahoo D (1995). Nitrate contamination of groundwater: measurement and prediction. fertiliser Research, **42**(1–3), 331–338

Gurney, K (1997). An Introduction to Neural Networks. University College London Press, London, UK

Hafez A A R (1974). Comparative changes in soil physical properties induced by admixtures of manures from various domestic animals. Soil Science, **118**(1), 53–59

Han J; Kamber M (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco, USA

Haynes F J; Francis G S (1993). Changes in microbial biomass C, soil carbohydrate composition and aggregate stability induced by growth of selected crop and forage species under field conditions. Journal of Soil Science, **44**(4), 665–675

Heathwaite A L (1997). Sources and pathways of P loss. In: Phosphorus Loss from Soil to Water (Tunney H; Carton O T; Brookes P C; Johnston A E, eds.), pp. 205–223. CAB International, New York, NY, USA

Lillesand T M; Kiefer R W (1994). Remote Sensing and Image Interpretation. John Wiley & Sons, Inc., New York, NY, USA

Nagler P L; Daughtry C S T; Goward S N (2000). Plant litter and soil reflectance. Remote Sensing of Environment, **71**(2), 207–215

Sharpley A N; Chapra S C; Wedepohl R; Sims J T; Daniel T C; Reddy K R (1994). Managing agricultural phosphorus for protection of surface waters: issues and options. Journal of Environmental Quality, **23**(3), 437–451

Six J; Elliott E T; Paustian K; Doran J W (1998). Aggregation and soil organic matter accumulation in cultivated and native grasslands. Soil Science Society of America Journal, **62**(5), 1367–1377

Sommerfeldt T G; Chang C; Entz T (1988). Long-term annual manure applications increase soil organic matter and nitrogen, and decrease carbon to nitrogen ratio. Soil Science Society of America Journal, **52**(6), 1668–1672

SPSS (1998). AnswerTree 2.0 User's Guide. SPSS Inc., Chicago, IL, USA

Teorey T J (1999). Database Modeling & Design. Morgan Kaufmann Publishers, San Francisco, CA, USA

Thenkabail P S; Smith R B; Pauw E D (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. Remote Sensing of Environment, **71**(2), 158–182

Weiss S M; Kulikowski C A (1991). Computer Systems That Learn. Kaufmann Publishers, San Mateo, CA, USA

Witten I H; Frank E (2000). Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations. Morgan Kaufmann Publishers, San Francisco, CA, USA

Yang C-C; Prasher S O; Landry J-A; DiTommaso A (2000). Applications of artificial neural networks in image recognition and classification of crop and weeds. Canadian Agricultural Engineering, **42**(3), 147–152

Zebarth B J; Younie M F; Hall J W; Telford G A (1999). fertiliser banding influence on spatial and temporal distribution of soil inorganic nitrogen in a corn field. Soil Science Society of America Journal, **63**(6), 1924–1933