THE SPECTRAL-TEMPORAL RESPONSE SURFACE AND ITS USE IN THE MULTI-SENSOR, MULTI-TEMPORAL CLASSIFICATION OF AGRICULTURAL CROPS

Carlos VIEIRA, Paul MATHER, and Michael McCULLAGH
School of Geography,
The University of Nottingham, Nottingham, NG7 2RD, UK
vieira@geography.nottingham.ac.uk
Paul.Mather@nottingham.ac.uk
Michael.McCullagh@nottingham.ac.uk

KEY WORDS: Multitemporal, Classification, Multisensor/multispectral fusion, Land use, Integration.

ABSTRACT

A method for classifying agricultural crops using multi-temporal, multi-spectral and multi-source remotely-sensed data is described. The procedure characterises all the pixels in a scene by considering their intensity values as a function of time of imaging and spectral waveband. An analytical surface is interpolated through these data points, which may be irregularly spaced. Two fitted function interpolation methods were used to generate and parameterise the analytical surfaces. Then, the surface coefficients were input to three different supervised classifiers (Maximum Likelihood, Artificial Neural Network and Minimum Distance Rule algorithms). Results show that classification accuracy is significantly improved in comparison with the use of any single-date image. The advantages of the methodology described in this paper are that it takes account of the reflectance spectra at different points in the growing season, and that the time periods between images, as well as the wavebands, need not be the same at each date. Thus, the procedure can handle data from sensors such as SPOT HRV and Landsat TM. In addition, the use of coefficients to represent the analytical surfaces significantly reduces the amount of data processing, whilst maintaining information reliability.

RÉSUMÉ

Les résultats d'une méthode innovatrice pour classer les cultures agricoles en utilisant des données obtenues par télédétection, multitemporelles, de spectres multiples et de sources multiples sont décrits. Le procédé d'abord les donnés par classe de formation individuelle en considérant les valeurs de ces donnés comme une fonction du temps de prise des images et de la bande de fréquences. Une surface analytique est ajustée sur ces points de données, qui peuvent être irrégulièrement espacés. Initialement une surface polynomiale à deux variables a été utilisée, et les coefficients de cette surface étaient les entrées d'un réseau neural artificiel, un algorithme à probabilité maximale et un processus de distance euclidienne minimale, respectivement. Les résultats montrent que la précision de la classification des cultures agricoles est nettement améliorée par rapport à l'utilisation d'image à date unique. Les avantages de la méthodologie décrite dans cet article sont qu'on prend en considération les spectes de réflectance en différents points dans la saison de croissance, et que les intervalles de temps entre les images n'ont pas besoin d'être les mêmes. Le procédé peut aussi traiter de données de détecteurs différents, tels que SPOT HRV et Landsat TM. En outre, l'utilisation de coefficients pour représenter les surfaces analytiques réduit de façon importante la quantité de traitement de données, tout en maintenant la fiabilité de l'information.

1 INTRODUCTION

The use of single-date images to classify agricultural crops has a number of significant drawbacks. Firstly, the different crop types represented in the area under study may be at different stages of growth. Secondly, the temporal 'profile' of the spectral reflectance curve of each crop is not taken into account. Such profiles may be of considerable value in discriminating between crop types, which may be difficult to distinguish at certain points in the growth cycle. Thirdly, results derived from data obtained by different sensors may not be comparable due to differences in spectral and spatial characteristics.

Haralick *et al.* (1980), Badhwar *et al.* (1982), Badhwar (1984), Lambin and Strahler (1994a,b) and Ortiz (1997) consider the problem of characterising the temporal dimension but none utilises the method proposed here, which explores the use of the spectral temporal response surface (STRS) to characterise a pixel's behaviour over time for each waveband. The method described in this paper represents each training data sample as a set of points in a three-dimensional space, the axes of which are time, wavelength and reflectance. Hence, a set of training samples

representing the same crop type measured at *n* time points over the growing season can be visualised as points lying on a surface, each point having coordinates (time, waveband, reflectance). An analytical surface fitted to this distribution can be parameterised and its coefficients, rather than the pixel values in each spectral band, used as features in image classification procedures.

The advantages of this methodology are that:

- It provides a description of the pixel's spectral response over a growing season – the STRS representation is a 2D analogue of the spectral signature;
- Allows the use of data from different sensors with differing numbers of wavebands. The temporal spacing of observations can be irregular, and so the procedure is not adversely affected if images cannot be obtained on specific dates, or if cloud obscures some training data collection sites;
- It is economical, as the number of coefficients is smaller than the total number of bands.

2 DATA

The study area is located near the town of Littleport in Cambridgeshire, eastern England. This area is approximately at mean sea level with gently undulating topography. The agriculture of the region is characterised by the use of rotational crop plantation techniques, which were started early in this century. Imagery in the optical bands of the TM sensor for June 11 and 27; July 20; and August 14 (1994) are available plus SPOT HRV images for May 13; June 28; July 30 and August 14 (1994). In addition, Field Data Printouts for 1994 were collected from local farmers to generate the ground truth data set. Image processing operations were performed using ERDAS Imagine (version 8.3) and IDRISI systems. The ANN applications used the SNNS software. In addition, in-house programs were written to carry out some specific procedures. All images were registered to the Ordnance Survey of Great Britain National Grid projection at a 1:25.000 scale. This geometric correction of the satellite data was performed using 17 Ground Control Points (GCPs) and performing a nearest neighbour resampling method (resolution of 30 metres), since this sampling technique does not alter the pixel brightness during resampling (Jensen, 1986). The root-mean-square (RMS) errors for all the images were less than 0.5 pixels.
The reflectance values (without consider atmospheric correction) for the images ρ^*, can be derived based on the equation:

$$\rho^* = \frac{L}{E_s d \cos(\theta_s)} \tag{1}$$

where L denotes the radiance which is calculated on the basis of the digital number (*DN*) of the image pixels, E_s denotes the band averaged equivalent solar distance, d is a multiplicative factor to account for the variation in the solar constant as a function of the changing Earth-sun distance, and θ_s is the solar zenith angle. It is important to mention that in case of Landsat TM imagery, the radiance L is calculated by $L = (DN)A + B$, where A is known as gain and B denotes offset (these values are presented by Olsson (1995)), while for SPOT HRV imagery L is calculated using the formula $L = (DN) / A$, where A is called absolute calibration gain and is provided by SPOT image parameter sheet. Tso (1997) gives comprehensive reviews of these formulations.

3 THE SPECTRAL TEMPORAL RESPONSE SURFACE (STRS)

The type and sequence of procedures used in the generation and potential use of the STRS representations are outlined in Figure 1. As can be seen, at each of the major stages in the analysis, different procedures were implemented and their results were compared. Details are present in the following sub-sections.

3.1 Temporal-Spectral Model

The concept of the temporal-spectral matrix, in which a pixel's reflectance is represented as a function of its temporal and spectral distribution, is introduced. This matrix is represented in 3D space with the temporal, spectral and radiometric properties represented by the *x*, *y* and *z* axes respectively.

The image acquisition dates are expressed in the form of Julian days (*x* axis) and the spectral dimension (*y* axis) are characterised by their medial waveband value computed in the form of wavelengths. Thus, the spectral bands were labelled using the medial wavelength values of 0.458, 0.56, 0.66, 0.83, 1.645, 2.215 - to the six available TM channels (except the thermal infrared TM band 6) - and 0.545, 0.645, 0.84 - to the three HRV channels respectively. The radiometric properties are expressed in form of reflectance's values along the z-axis. Furthermore, for each pixel, 36 three-dimensional control points were generated (4 TM images with 6 bands plus 4 SPOT HRV images with 3 bands). It is important mention that the

values along the x, y and z axes are scaled into the interval between 0 and 1, sometimes referred as normalisation, before the interpolation phase.

3.2 Analytical Surface Interpolations

The next stage involves the interpolation of an analytical surface through these control points, which may be irregularly spaced. Interpolation methods involve one of two contrasting general approaches, which use either a *fitted function* or a *weighted average* methods to calculate the interpolated z value at a given (x, y) point. Fitted function methods determine the coefficients of an analytic function that shows a smoothed representation of reflectance variation (z axis). The alternative approach, to obtain a representative surface height at a given location, is by directly summing the data influences that are within a range; these are weighted average methods. Weighted average surfaces emphasise local detail and require more computer time per interpolation point than do the fitted function surface, while fitted function interpolation provide a generalised representation of the surface, and thus override aberrant, anomalous, or noisy data (Lam, 1983, Watson, 1996). As we are interested in the general shape of the STRS, we limited our choice of interpolator to the fitted function types. Two methods were tested - polynomial trend surface analysis and collocation. These methods are briefly described in the following sections.

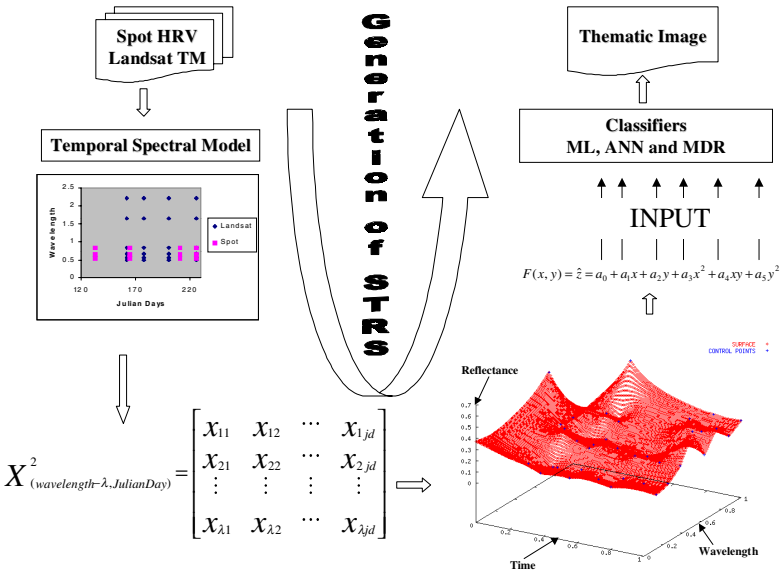

Figure 1. An outline of the methodology followed in this study to generate the STRS representations

3.2.1 Polynomial Trend Surface Analysis: Initially these control points are used to fit a surface using the Polynomial Trend Surface (PTS) interpolator. A bivariate polynomial surface is a linear combination of elements known as basis functions that are expressed in terms of powers and cross-products of the x and y coordinates ($z = 1$, $z = x$, $z = y$, $z = xy$, $z = x^2$, $z = y^2$, and so on). For any polynomial degree, g, the interpolated surface value ($\hat{z}$) at any location (x, y) is a combination of power basis functions of that location. For example, the interpolated value F ($= \hat{z}$) at the point (x,y) using a second-order bivariate polynomial function is given by:

$$F(x,y) = \hat{z} = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \quad (2)$$

The coefficients a_i are calculated using the method of least squares (Burrough, 1986; Davis, 1973). Mather (1976) presents a method of computing the coefficients a_i using the Gram-Schmidt orthogonalisation process, which is relatively simple to program and use, and - even in badly-conditioned problems - is surprisingly accurate, though the degree of accuracy is affected by the size of the residuals (Mather, 1975). A bivariate polynomial surface can be calculated for increasing orders of the polynomial. The first three terms of the right-hand side of equation (2) describe a planar surface, while the addition of terms 3 to 5 inclusive results in a second-order or quadratic surface. The complexity of the surface increases with order but the surface does not pass through all of the data points (representing training data) unless the number of coefficients in the polynomial equation equals n - 1, where n is the number of data points. Although a surface order of 7 (36 coefficients) explained over 99% of the sum of squares, using a surface order of 3 (10 coefficients) experimentally proved to be enough to characterise the analytical surfaces.

3.2.2 Collocation. In addition, the same control points are used to fit a surface using the Collocation Interpolator. This method, which produces a surface that passes through all the control points, is described by Watson (1996). Given observations $z(P_i)$ at a set of spatially-distributed locations P_i, a system of simultaneous linear equations is set up and solved for the coefficients a_i. This is illustrated in Figure 2 for four points, but any number may be used. The term $C(P_i\ P_j)$ is a function of the distance on the x - y plane between points P_i and P_j, modified by the arbitrary nonnegative constant e_j, which forces all distances less than this arbitrary limit to be positive.

$$\begin{bmatrix} e_1 & C(P_1P_2) & C(P_1P_3) & C(P_1P_4) \\ C(P_2P_1) & e_2 & C(P_2P_3) & C(P_2P_4) \\ C(P_3P_1) & C(P_3P_2) & e_3 & C(P_3P_4) \\ C(P_4P_1) & C(P_4P_2) & C(P_4P_3) & e_4 \end{bmatrix} \bullet \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} z(P_1) \\ z(P_2) \\ z(P_3) \\ z(P_4) \end{bmatrix}$$

where $C(P_iP_j) = \sqrt{(P_{ix}-P_{jx})^2 + (P_{iy}-P_{jy})^2 + e_i^2}$

Figure 2. Linear equations for the Collocation interpolator.

Hardy (1971) suggests e = 0.815 times the average distance between the data points, when large scale variations of the distances are predicted. Since the reference system (*xyz*), specifically in this research, ranges between 0 and 1, it follows that relatively small-scale variations on the distances are expected, and the value e can be assigned to 0. Other possible kernel functions are described by Schut (1976). The surface coefficients, a_i (or vector **a**), whose sum is zero or near zero, are obtained by solving the system of linear equations according to the following model: **Da = z,** where **D** is the matrix of distances between the points P_i; **z** is the vector of reflectance's variations $z(P_i)$; and **a** is a vector of unknown coefficients which are to be estimated from the data. The method of least squares is used to find the vector coefficients **a** using the standard formula: $\mathbf{a} = \mathbf{(D'D)^{-1}D'z}$. These estimated values a_i become the coefficients for an arbitrary location. If P is the interpolation point with Cartesian coordinates (x, y), and the coefficients, a_i, have been determined, then the interpolated value, $F(x, y)$, is given by:

$$F(x,y) = a_1 C(P-P_1) + a_2 C(P-P_2) + a_3 C(P-P_3) + a_4 C(P-P_4) \quad (3)$$

One pixel example of the PTS and Collocation analytical surfaces is shown in the Figure 3 for the crop wheat.

3.3 Classification

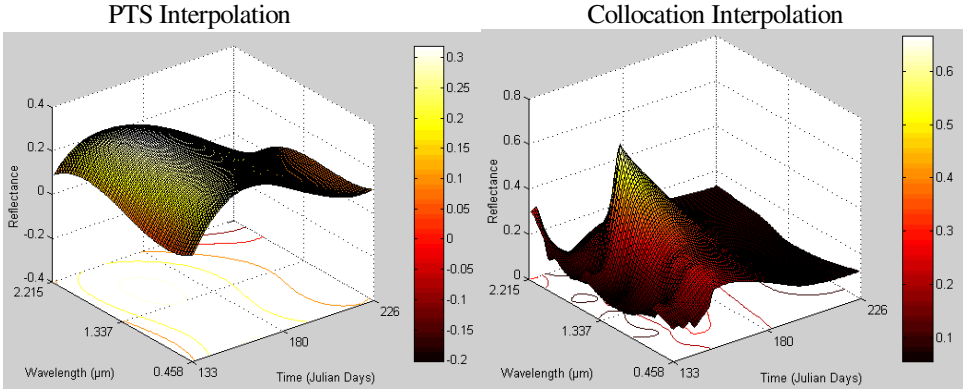

Figure 3. Analytical surfaces and contours for the crop wheat.

The training data for each class consist of measurements on n pixels at k different points in time, and on p wavebands. Each pixel has an intensity or reflectance value v. In terms of the procedures for surface fitting described in the section 3.2, the location (x, y) coordinates are time and waveband, which are scaled into the interval between 0 and 1. The reflectance values constitute the set of points to which an analytical surface is fitted. There is one surface for each pixel in the training data and it is also expected that each individual class has approximately the same surface shape generated by the interpolation process. The training data set to be used in subsequent classification is now the set of coefficients describing the temporal-spectral surface that is fitted to these data points. As the interpolated coefficients show different magnitudes on their values, they were again scaled collectively to the interval between 0 and 1 before the training and test phases. In addition, a technique called Logical Channel Approach (LCA) is implemented in order to compare the results of this new method against a classification based on raw data. For each pixel, the thirty-six reflectance values are considered as individual logic channels, therefore generating a thirty-six dimensional vector, to be also used as input into the classification process.

Once normalised, the coefficients of the surface are input to each of three supervised classifiers: Maximum Likelihood (ML), Artificial Neural Network (ANN) - Backpropagation model, and Minimum Distance (MD) rule.

The Gaussian maximum likelihood method is a well known supervised classification algorithm that is based on the assumption that the probability density functions $p(x/w_i)$ for each class w_i, are multivariate normal (Gaussian). (Tou and Gonzalez, 1974; Mather, 1999). Because of its analytical tractability, the multivariate normal density function has received considerable attention, and many examples of its successful application can be found in the literature. In this approach, each individual class distribution is characterised by its mean vector and covariance matrix, which are determined from the training samples. The ML classification is performed using algorithms developed by Mather (1999), adapted to classify 3D surface coefficients and LCA vectors (based on the raw data set).

The neural network architecture chosen is the multi-layer perceptron using the back-propagation algorithm. This method is perhaps the best known in the context of remote sensing (Lippman, 1987, Benediktsson *et al.*, 1990, Bischof *et al.*, 1992, Civco, 1993, Paola, and Schowengerdt, 1995, Kanelloupoulos *et al.*, 1997). The definition of the structure and parameters of the ANN is crucial for its successfully performance. The input layer of the ANN used one node for each of the surface coefficients. The ordering of the input data is not important in the standard back-propagation neural network model (where the input neurones are connected to all hidden layer nodes) beyond maintaining consistency of input data from one training sample to the next. In the scope of this research, the input layer have 10, 36 and 36 input nodes respectively, corresponding to the 10 PTS coefficients, 36 Collocation coefficients and 36 vector elements in LCA (based on the raw data set), used as discriminant variables. All the neural network configurations tested have 6 output nodes, corresponding to 6 general

crop classes. The number of hidden layers and the number of hidden nodes were found using the *growing procedure* as presented by Hirose *et al.* (1991). The optimal number of hidden layers and number of nodes per hidden layer was found to be two hidden layers and 20 nodes per layer for the PTS coefficients, and one hidden layers and 20 nodes per layer for the Collocation interpolator and LCA vectors. The learning rate and momentum are kept constant at 0.2 and 0.9, respectively, during the comparison of architectures. In this research, the simplest way to assign a class to the input data was to select the class from the output node with the highest value, providing that values was more than an arbitrary threshold value of 0.7.

The minimum distance classifier involves the calculation of the Euclidean distance in feature space between a pixel to be classified and a class centroid, the latter being calculated from training data (Mather, 1999). The pixel is given the label of the closest centre. The same data (normalised surface coefficients and raw data) as described above are used to perform the MDR classification.

3.4 Accuracy Assessment

After the classification phase, standard accuracy measures derived from a confusion matrix were computed, using a test data set based on the Field Data Printouts. The measures based on the confusion matrix were overall accuracy, individual class accuracy, producer's accuracy and user's accuracy. The calculations associated with these measures are described in standard textbooks (e.g., Mather, 1999). The Kappa coefficient, conditional Kappa for each class, and test Z statistics, all of them widely used statistic derived from the contingency matrix, were also computed (Congalton and Green, 1998).

In addition, a pairwise test statistics for testing the significance of the classifiers (represented here by their respective confusion matrices), were performed utilising the Kappa coefficients. These results were summarised in form of a *significance matrix*, in which the major diagonal elements indicate if the respective classification result is meaningful. In this single confusion matrix case, the *Z* value can be computed using the formula:

$$Z = Ka \Big/ \sqrt{\mathrm{var}(Ka)} \tag{4}$$

where *Z* is standardised and normally distributed and *var* is the large sample variance of the Kappa coefficient *K*. If $Z \geq Z_{\alpha/2}$, the classification is significant better than a random classification, where $\alpha/2$ is the confidence level of the two-tailed *Z* test and the degree of freedom are assumed to be infinity. On the other hand, the off diagonal elements give an indication, again if $Z \geq Z_{\alpha/2}$, that the two independent classifiers are significantly different. The formula used to test for significance between the two independent Kappa coefficients is:

$$Z = \left| Ka_1 - Ka_2 \right| \Big/ \sqrt{\mathrm{var}(Ka_1) - \mathrm{var}(Ka_2)} \tag{5}$$

where the Ka_1 and Ka_2 are the two Kappa coefficients in comparison. Congalton and Green (1998) give a comprehensive review of these measures.

3.5 Feature Selection

Finally, a forward selection method, as described by Mather (1999), is applied to the coefficients in order to estimate the subset of *m* coefficients (features) that best combine classification accuracy and computational economy. The forward selection method starts with the best subset of size $m = 1$ and call this feature f_1. Then, the next step is to find the best subset of size $m = 2$ including f_1, that is f_1 and one other coefficient (feature), in which the classification accuracy (using ML algorithm) is maximised. The best subset at the end of the second cycle will be f_1, f_2. The procedure continues with subset f_1, f_2, f_3, and so on.

4 Results and Discussions

A stratified random sample of 6000 temporal-spectral pixels, with 1000 pixels per class, was selected, representing the six most common cover types in the study area: Potatoes, Sugar Beet, Wheat, Fallow, Onions and Peas. This data set was divided systematically into training (3600 pixels) and test (2400 pixels) samples, in which each class has the same number of pixels. All the classifiers were trained with the same data set (3600 patterns). Classification accuracy was expressed as the percentage correct allocation to the independent test data set (2400 patterns). With the exception of the ANN (threshold = 0.7), the pixels received the label of the output class having the highest probability or minimum distance. Classification accuracy achieved by the three selected classifiers, using the three different sets of discriminant variables, for instance: spectral-temporal coefficients derived from the two spatial interpolation methods and LCA (based on the raw data), are summarised in Table 1.

Based on the results it is noteworthy that all of the classifiers performed well on this data set. The performance of the ANN and ML classifiers is superior to the MDR model, and the conditional Kappa coefficients (* 100) of these two classifiers maintain high levels for the six classes. Observing also the consistence in the performance of the three classifiers when using the three different set of discriminant variables a general conclusion is that these three features set have in themselves a discriminatory capability and they are able to produce very similar results. If this is the case, the findings lend support to the hypothesis that the use of any one of these feature sets has equally a promising potential for classifying croplands.

INTERPO.	LCA			PTS			COLLOCATION		
CLAS/MET	MDR	ANN	ML	MDR	ANN	ML	MDR	ANN	ML
Potatoes	58.9	95.9	96.3	58.1	92.8	97.0	61.6	92.5	96.6
Sugar Beet	55.7	94.8	86.8	54.4	91.0	89.9	54.0	91.2	87.4
Wheat	78.3	94.5	87.5	69.8	88.5	84.4	80.6	91.0	87.5
Fallow	84.7	99.1	82.2	82.2	95.9	77.3	86.0	96.8	81.1
Onions	94.3	97.6	98.2	84.9	95.4	96.6	95.3	97.3	97.9
Peas	62.3	98.2	99.0	68.8	94.4	96.9	66.0	97.0	99.0
OVERALL(%)	76.3	93.2	92.7	74.9	89.0	91.3	78.1	92.0	92.5
Total Pixels	2400	2301	2400	2400	2269	2400	2400	2318	2400
Unrecognised	-	99	-	-	131	-	-	82	-
Kappa	0.716	0.920	0.912	0.699	0.870	0.896	0.738	0.905	0.910

Table 1. Classification accuracies for six agricultural crops using Logical Channel Approach (LCA), Polynomial Trend Surface (PTS) and Collocation; and three classification algorithms - maximum likelihood (ML), minimum distance rule (MDR) and artificial neural network (ANN). The table shows individual classification accuracy for each crop (Conditional Kappa * 100), overall accuracies, the number of unrecognised pixels and the value of the Kappa coefficients. These accuracies were calculated from an independent dataset test (2400 patterns).

The individual class accuracies determined by the three classification algorithms (MDR, ANN, ML) show consistency across the three approaches (LCA, PTS, Collocation). The highest overall classification accuracy is produced by the ANN using the LCA approach (93.2%). It should be noted that this does not imply that the ANN is best for each class. For example, the highest accuracy for potatoes, onion and pea classes was achieved by the ML classifier (Table 1). The MDR procedures was not able to discriminate well between the potato, sugar beet and peas classes due to the fact that they seem not to be linearly separable, but fallow and onions showed high accuracy for all experiments. MDR was consistently the worst performing algorithm.

Results of the pairwise test of significance utilizing the Kappa coefficients are summarized in the Table 2. At the 95 percent confidence level, the critical value ($Z_{\alpha/2}$) is 1.96. Therefore, the major diagonal elements (representing the single error matrices) show that all the classifications are significant better than a random assignment of labels ($Z >$ 1.96). Moreover, with exception of the eleventh classifier pairs (off diagonal elements), there was significant difference between the overall accuracies of these classifications. It could be concluded that these classifier pairs (bright blue) produce equal classifications between them, therefore, it would be best to use the cheaper, quicker, or more efficient approach. However, on a classifier-by-classifier comparison, more interesting feature is exhibited. The majority of the experiments suggest that the ML and ANN classifiers are able to produce statistically equivalent accuracy. Consequently, one could argue that the use of the ML procedure instead of an ANN is cheaper as decision.

Although Table 1 (confirmed by Table 2) shows an acceptable level of overall accuracy, a forward selection method was applied to the coefficients in order to test the hypothesis that a subset of coefficients (features) might produces results that are acceptable, yet at level computational cost. The forward selection method is performed using the ML classifier. Table 3 presents the results of the Kappa analysis that compares the error matrices, two at a time, to determine if they are significantly different. The results of this pairwise test for significance between a subset of coefficients and using the whole set of original coefficients (i.e., 10 PTS and 36 Collocation), reveals that using the feature sequence {1, 9, 10, 6, 8, 7, 3, 5} PTS and {1, 22, 21, 15, 16, 26, 27, 23, 17, 34, 35, 8, 4} Collocation coefficients the results are not significantly different at the 95% confidence level.

CLASSIFIERS	MDRL	ANNL	MLL	MDRP	ANNP	MLP	MDRC	ANNC	MLC
Kappa	*0.716*	*0.92*	*0.912*	*0.699*	*0.87*	*0.896*	*0.738*	*0.905*	*0.91*
Variance	*0.000107*	*0.000036*	*0.000041*	*0.000112*	*0.000054*	*0.000047*	*0.000102*	*0.000042*	*0.000041*
MDRL	69.13								
ANNL	29.85	154.05							
MLL	16.11	0.91	142.94						
MDRP	1.15	18.16	17.22	65.92					
ANNP	12.14	5.27	4.3	13.27	118.16				
MLP	14.51	2.63	1.7	15.62	2.59	130.13			
MDRC	1.52	15.49	14.55	2.66	10.57	12.94	72.96		
ANNC	15.48	1.69	0.77	16.6	3.57	0.95	13.92	139.47	
MLC	15.95	1.14	0.22	17.06	4.1	1.49	14.38	0.55	141.61

Table 2. Results of Kappa Analysis for comparison among the classifiers. In all the classifiers' names were appended the character: P (Polynomial) or C (Collocation) to indicating the method of interpolation and L (Logical) to refer to the raw data set. The table also presents the Kappa coefficients and variance for each classifier. The Z values (in major diagonal) were computed using formula (4). The Z values (off diagonal elements) were computed using formula (5).

It can be concluded that using more than these feature sequences of coefficients will produce approximately equal classification results. In order to reduce the computational requirements we can reduce the number of features to these dimensions without losing too much information in the data and also maintaining the classification accuracy.

On balance we can use the surface coefficients as a set of representative features to summarise the spectral-temporal information within the images and also reduce the amount of time processing during the classification procedures. Specifically in this application, 8 PTS coefficients or 13 Collocation coefficients can be used directly as input to the process of classification without losing accuracy. Moreover, their use seems to overcome the problems resulting from presence of clouds in the multitemporal images, since the interpolated coefficients express information about general behaviour on the surfaces, and therefore override aberrant or noisy data. Furthermore, the approach outline in this study has proven to be effective in identifying these cropland classes. This conclusion is further supported by comparison of the results listed in Table 1 with those achieved by single-date classification. Using the ML procedure, an overall accuracy of 57% was achieved using Landsat TM data, and for the SPOT HRV data set, the overall accuracy fell to 42%. The worst performing classifier (i.e., MDR) using our methodology achieved an overall accuracy of 74.9% and the best was 93.2%.

The classified image generated by the ML classifier, using PTS coefficients, Collocation coefficients, LCA vectors and their respective ground truth image are shown in Figure 4. These thematic images show isolated "noise" pixels problems, which could be easily smoothed out by employing mode filter (contextual information), under the assumption that pixels of a given class are likely to be surrounded by pixels of the same class.

5 CONCLUSIONS

A methodology for classifying agricultural crops combining multi-temporal, multi-spectral and multi-source remotely-sensed data has been shown to be effective in identifying general agricultural crop classes over an area in East Anglia. Classification accuracies in excess of 90% were achieved, even though parts of some of the images are covered by clouds. The basic assumption of the method, that different crops have different spectral-temporal trajectories, has been used in

Coeffic.	KAPPAtsa	VARtsa	Ztsa	Coeffic.	KAPPAcol	VARcol	Zcol
1				1			
9	0.627	0.000127	20.39	22	0.627	0.000126	21.9
10	0.74	0.000101	12.82	21	0.684	0.000116	18.04
6	0.81	0.00008	7.63	15	0.738	0.000102	14.38
8	0.848	0.000066	4.52	16	0.762	0.00095	12.69
7	0.863	0.000061	3.18	26	0.777	0.000091	11.58
3	0.874	0.000056	2.17	27	0.814	0.000078	8.8
5	0.883	0.000053	1.3	23	0.83	0.000073	7.49
2	0.882	0.000053	1.4	17	0.844	0.000068	6.32
4	0.896	0.000047	0	34	0.862	0.000061	4.75
				35	0.88	0.000054	3.08
				8	0.891	0.000049	2
				4	0.9	0.000046	1.07
				19	0.904	0.000044	0.65
				18	0.907	0.000043	0.32
				7	0.909	0.000042	0.11
				12	0.909	0.000042	0.11
				25	0.911	0.000041	0.11
				28	0.911	0.000041	0.11
				33	0.914	0.00004	0.11
				20	0.914	0.00004	0.44
				9	0.915	0.000039	0.56
				32	0.914	0.00004	0.44
				3	0.915	0.000039	0.56
				2	0.914	0.00004	0.44
				24	0.914	0.00004	0.41
				36	0.915	0.000039	0.56
				29	0.915	0.000039	0.56
				30	0.916	0.000039	0.67
				31	0.915	0.000039	0.56
				10	0.913	0.00004	0.33
				11	0.911	0.000041	0.11
				14	0.911	0.000041	0.11
				13	0.914	0.00004	0.44
				5	0.91	0.000041	0
				6	0.91	0.000041	0

Table 3. The Z statistic for the pairwise comparison of the error matrices. The matrices are compared, two at time, with whole set of original coefficients (i.e., 36 coefficients) to determine if they are significant different. If the absolute value of the test Z statistic is grater than 1.96, the result is significant different at the 95% confidence level.

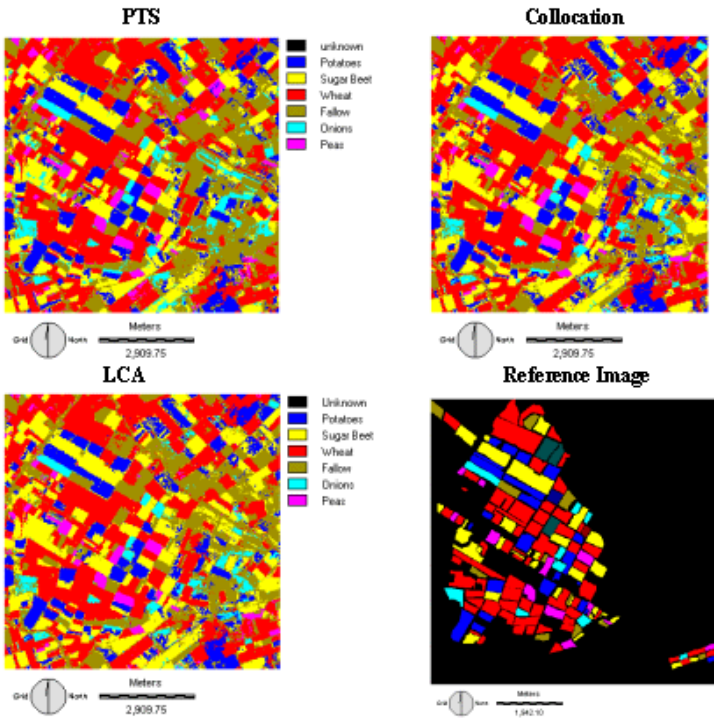

Figure 4. Classified image by ML classifiers using PTS coefficients, Collocation coefficients and LCA vectors (based on raw data).

earlier studies. However, the methods used to characterise the spectral reflectance changes over a growing season using a spectral-temporal surface represents a promising new approach, for several reasons. First, the method can deal with multi-sensor data, as the spectral bands measured at each date do not need to be the same. Second, data points obscured by clouds can be filtered out throughout the interpolation and parameterisation procedures of the analytical surfaces. Third, the overall spectral variation of a given crop class over the growing season is captured by a set of coefficients, which are fewer in number than the training data pixels and hence produce computationally more efficient classifiers. Finally, the ML classifier is the algorithm that best combine classification accuracy and computational economy.

ACKNOWLEDGEMENTS

Mr. Vieira's research is supported by CAPES (Brazilian Research Council). We are grateful to Logica PLC and SPOT Image for permission to use their images. We are also grateful to Mr. T. Kavzoglu for providing software to allow us to display the analytical surfaces using the MATLAB System.

REFERENCES

Badhwar, G. D., 1984. Classification of corn-soybean using multitemporal Thematic Mapper data. Remote Sensing of Environment, 16, pp. 175-182.

Badhwar, G. D., Autin, W. W., and Carnes, J. G., 1982. A semi-automatic for multitemporal classification of a given crop within a Landsat scene. Pattern Recognition, 15, pp. 217-230.

Benediktsson, J. A., Swain, P. H., and Ersoy, O. K., 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 28, pp. 540-552.

Bischof, H., Schneider, W., and Pinz, A. J., 1992. Multispectral classification of Landsat-images using neural networks. IEEE Transactions on Geoscience and Remote Sensing, 30, pp. 482-490.

Burrough, P. A., 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Clarendon Press.

Civco, D. L., 1993**.** Artificial neural networks for land-cover classification and mapping. International Journal of Geographic Information Systems, 7, pp. 173-183.

Congalton, R. G., and Green, K., 1998. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. New York: Lewis Publishers.

Davis, J. C., 1973. Statistics and Data Analysis in Geology. New York: John Wiley and Sons.

Haralick, R. M., Hlavka, C. A, Yokoyama, R., and Carlyle, S.M., 1980. Spectral-temporal classification using vegetation phenology. IEEE Transactions on Geoscience and Remote Sensing, 18, pp. 167-174.

Hardy, R. L., 1971. Multiquadric equations of topography and other irregular surfaces. Journal of Geophysical Research, 43, pp. 475-492.

Hirose, Y., Yamashita, K., and Hijiya, S., 1991. Back-propagation algorithm which varies the number of hidden units. Neural Networks, 4, pp. 61-66.

Kanellopoulos, I., Wilkinson, G. G., Roli, F., and Austin, J. (eds.), 1997. Neurocomputation in Remote Sensing Data Analysis. Berlin: Springer-Verlag.

Lam, N., 1983. Spatial Interpolation Methods: a Review. The American Cartographer, 10, pp. 129-149.

Lambin, E. F. and Strahler, A. H., 1994(a). Change-vector analysis in multitemporal space: a tool to detect and categorise land-cover change processes using high temporal-resolution satellite data. Remote Sensing of Environment, 48, pp. 231-244.

Lambin, E. F. and Strahler, A. H., 1994(b). Indicators of land-cover change for change-vector analysis in multitemporal space at coarse spatial scales. International Journal of Remote Sensing, 15, pp. 2099-2119.

Lippmann, R. P., 1987. An introduction to computing with neural nets. IEEE ASSP Magazine, April, pp. 4-22.

Jensen, J. R., 1986. Introductory Digital Image Processing – A Remote Sensing Perspective. Englewood Cliffs, NJ: Prentice-Hall.

Mather, P. M., 1975. The use of orthogonal polynomials in least square problems. Computer Applications in the Natural and Social Sciences, New Series, 2(10), Department of Geography, University of Nottingham.

Mather, P. M., 1976. Computational Methods of Multivariate Analysis in Physical Geography. Chichester: John Wiley and Sons.

Mather, P. M., 1999. Computer Processing of Remotely-Sensed Images: An Introduction. Chichester: John-Wiley and Sons, Second edition.

Olsson, H., 1995. Radiometric Calibration of Thematic Mapper data for forest change detection. International Journal of Remote Sensing, 24, pp. 81-96.

Ortiz, M. J., Formaggio, A. R., and Epiphano, J. C. N., 1997. Classification of croplands through integration of remote sensing, GIS and historical database. International Journal of Remote Sensing, 18, pp. 95-105.

Paola, J. D., and Schowengerdt, R. A., 1995. A review and analysis of backpropagation neural networks for classification of remotely-sensed multispectral imagery. International Journal of Remote Sensing, 16, pp. 3033-3058.

Schut, G. H., 1976. Review of interpolation methods for digital terrain models. The Canadian Surveyor, 30, pp. 389-412.

Tou, J. T. and Gonzalez, R.C., 1974. Pattern Recognition Principles. Reading, Massachusetts: Addison - Wesley.

Tso, B.C.K., 1997. An Investigation of Alternative Strategies for Incorporating Spectral, Textural, and Contextual Information in Remote Sensing Image Classification. Unpublished Ph.D. thesis, Department of Geography, The University of Nottingham, pp. 44-53.

Watson, D. F., 1996. Contouring A Guide to the Analysis and Display of Spatial Data. Oxford: Pergamon Press.