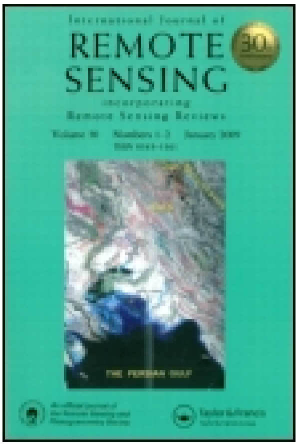

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Statistical learning algorithms for identifying contrasting tillage practices with Landsat Thematic Mapper data

Pijush Samui [a], Prasanna H. Gowda [b], Thomas Oommen [c], Terry A. Howell [b], Thomas H. Marek [d] & Dana O. Porter [e]

[a] Centre for Disaster Mitigation and Management, VIT University, Vellore, Tamil Nadu, 632014, India

[b] Conservation and Production Research Laboratory, USDA-ARS, Bushland, TX, 79012, USA

[c] Department of Geological and Mining Engineering and Sciences, Michigan Technological University, Houghton, MI, 49931, USA

[d] Texas AgriLife Research, Texas A&M University, Amarillo, TX, 79106, USA

[e] Texas AgriLife Extension, Texas A&M University, Lubbock, TX, 79403, USA

Available online: 27 Mar 2012

International Journal of Remote Sensing
Vol. 33, No. 18, 20 September 2012, 5732–5745

Statistical learning algorithms for identifying contrasting tillage practices with Landsat Thematic Mapper data

PIJUSH SAMUI†, PRASANNA H. GOWDA*‡, THOMAS OOMMEN§, TERRY A. HOWELL‡, THOMAS H. MAREK¶ and DANA O. PORTER|

†Centre for Disaster Mitigation and Management, VIT University, Vellore, Tamil Nadu 632014, India

‡Conservation and Production Research Laboratory, USDA-ARS, Bushland, TX 79012, USA

§Department of Geological and Mining Engineering and Sciences, Michigan Technological University, Houghton, MI 49931, USA

¶Texas AgriLife Research, Texas A&M University, Amarillo, TX 79106, USA

|Texas AgriLife Extension, Texas A&M University, Lubbock, TX 79403, USA

(*Received 5 April 2011; in final form 27 December 2011*)

Tillage management practices have a direct impact on water-holding capacity, evaporation, carbon sequestration and water quality. This study examines the feasibility of two statistical learning algorithms, namely the least square support vector machine (LSSVM) and relevance vector machine (RVM), for identifying two contrasting tillage management practices using remote-sensing data. LSSVM is firmly based on statistical learning theory, whereas RVM is a probabilistic model where the training takes place in a Bayesian framework. Input to the LSSVM and RVM algorithms were reflectance values at different bandwidths and indices derived from Landsat Thematic Mapper (TM) data. Ground-truth data for this study were collected from 72 commercial production fields in two counties located in the Texas High Plains of the south-central USA. Numerous LSSVM- and RVM-based tillage models were developed and evaluated for tillage classification accuracy. The percentage correct and kappa statistic were used for the evaluation. The results showed that the best LSSVM and RVM models included the use of TM band 5 or vegetation indices that involved TM band 5, indicating sensitivity of near-infrared reflectance of crop residue cover on the surface. This is consistent with other remote-sensing models reported in the literature. Overall classification accuracies of the best LSSVM and RVM models were 87.8 and 90.2%, respectively. The corresponding kappa statistics for those models were 0.75 and 0.80, respectively. Furthermore, comparison of the best LSSVM and RVM models with the published logistic regression-based tillage models developed with the same data indicated the superiority of the RVM model over LSSVM and logistic regression models in determining contrasting tillage practices with Landsat TM data.

1. Introduction

Tillage practices affect evaporation (Schwartz *et al.* 2010), infiltration (Vervoort *et al.* 2001), run-off (Takken *et al.* 2001), carbon sequestration (West and Marland 2002)

*Corresponding author. Email: Prasanna.Gowda@ars.usda.gov

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2012 Taylor & Francis
http://www.tandfonline.com
http://dx.doi.org/10.1080/01431161.2012.671555

and soil erosion (Takken *et al.* 2001) from agricultural fields due to wind and water erosion. Consequently, models that simulate agricultural systems require tillage as input (Gowda *et al.* 2003). Therefore, identifying and mapping contrasting tillage practices over large areas is an imperative task in environmental modelling. However, collecting tillage data over large areas is a time-consuming and costly task (Sudheer *et al.* 2010). In the past, researchers have developed and used different methods for mapping tillage practices (DeGloria *et al.* 1986, Motsch *et al.* 1990). However, the success of these methods depends on the interpreter's ability (Sudheer *et al.* 2010). Recently, numerous regression-based spectral models have been adopted to determine tillage practices (Sullivan *et al.* 2004, Thoma *et al.* 2004, Daughtry *et al.* 2005, Sullivan *et al.* 2006). Logistic regression-based tillage models are widely used statistical tools for deriving tillage information from Landsat Thematic Mapper (TM) data (van Deventer *et al*. 1997; Vina *et al.* 2003, Bricklemyer *et al.* 2006, Gowda *et al.* 2008). However, logistic regression models have some limitations: (1) they should be thoroughly evaluated before using them in different geographic regions to adjust the cut-point probability values (Gowda *et al.* 2008) and (2) available data are forced to conform to a predefined model form, which may not be correct for every location (Sudheer *et al.* 2010). Sudheer *et al.* (2010) successfully adopted artificial neural network (ANN) to identify contrasting tillage practices in the Texas High Plains. But the ANN models have limitations such as lower convergence speed, a black box approach, less generalized performance and absence of probabilistic output (Park and Rilett 1999, Kecman 2001).

In this study, the applicability of two new-generation statistical learning algorithms, namely the least square support vector machine (LSSVM) and relevance vector machine (RVM), is verified for identifying two contrasting tillage practices using satellite remote-sensing data. LSSVM is similar to the support vector machine (SVM) and is based on statistical learning theory. However, LSSVM adopts a least squares linear system as a loss function instead of using quadratic programming as in SVM (Suykens *et al.* 1999). Researchers have successfully used LSSVM for solving different problems such as estimation of carbon content of agricultural land, site characterization, hyperspectral image classification and crop identification (Tang *et al.* 2006, Mathur and Foody 2008, Samui and Sitharam 2008). RVM is a statistical learning algorithm developed by Tipping (2000) that uses Bayesian inference to obtain parsimonious solutions for classification and regression problems compared to LSSVM. RVM has an identical functional form to SVM, but provides probabilistic classification and uses 'automatic relevance determination' to choose sparse basis sets (Bishop 1995), which pushes non-essential weights to 0. Recently, researchers have demonstrated the robustness of RVM for applications such as land-cover classification, well-log acoustic velocity prediction, settlement of shallow foundations and seismic attenuation prediction (Foody 2008, Ghosh and Majumdar 2008, Samui and Sitharam 2008). The main objective of this study was to evaluate the feasibility of adopting LSSVM and RVM for determining contrasting tillage practices in the Texas High Plains using Landsat TM data. The goals of this study were to

- develop and examine the feasibility of LSSVM- and RVM-based models for collection of tillage information and
- conduct a comparison of the developed LSSVM and RVM models with logistic regression models reported by Gowda *et al.* (2008).

2. Study area

This study was conducted with tillage data collected from 72 commercially operated production farms (31 in Moore County and 41 in Ochiltree County) in the Texas High Plains underlain by the Ogallala Aquifer (figure 1), which is being depleted by extensive pumping with minimal recharge. Moore County is located in the north-central part of the Texas High Plains and has a total land area of 236 826 ha. Two-thirds of the land in this county is in the nearly level, smooth uplands of the High Plains (USDA-SCS 1975) with most in row crop and cereal grain production. Corn, sorghum and wheat are the major crops in the county. In 2004, Moore County ranked 5th in corn production and accounted for about 5.7% of the total corn produced in the state (NASS 2005). Ochiltree County has an area of 234 911 ha, with more than 70% of the land in row crop production. Sorghum, wheat and corn are the major crops in this county. In 2004, Ochiltree County ranked 8th in sorghum production and accounted for about 2.4% of the total sorghum produced in the state (NASS 2005). Typical planting dates for major crops in the study area vary from the 2nd week of April to the 3rd week of May. Annual average precipitation is about 481 and 562 mm for Moore and Ochiltree counties, respectively. Crop water needs are supplemented with groundwater from the underlying Ogallala Aquifer. Nearly level to gently sloping fields with silty clay soils of the Sherm series occupy nearly all of the crop land in both Moore and Ochiltree counties. Conventional tillage practices within the study area typically consist of offset disc operations in the fall. Common conservation tillage practices are no ploughing in the fall and sweep or disc ploughing operations at planting, which leaves at least 30% of the surface covered with crop residue after planting.

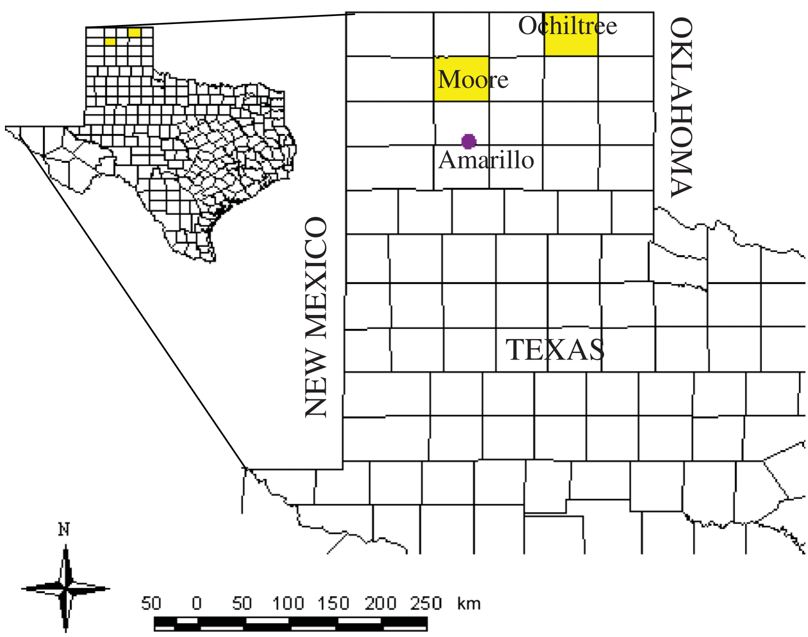

Figure 1. Location of Moore and Ochiltree counties in the Texas High Plains, USA.

3. Materials and methods

Development and evaluation of tillage models in this study mainly consisted of two steps: (1) development of models using the LSSVM and RVM techniques and (2) evaluation of tillage models with statistical measures of classification accuracy (i.e. percentage correct or overall classification accuracy and kappa (k) values). Two level-1 processed, precision-corrected Landsat TM scenes acquired, one on 10 May 2005 for Ochiltree County (Path 30/Row 35) and the other on 17 May 2005 for Moore County (Path 31/Row 35), were used for developing and evaluating LSSVM- and RVM-based tillage models. On the day of the Landsat 5 satellite overpass, ground-truth data were collected from 31 and 41 randomly selected commercial production fields planted with major crops in Moore and Ochiltree counties, respectively. Ground-truth data included geographic coordinates obtained using a handheld GPS, infrared images taken at a 2 m height using the Agricultural Digital Camera (ADC, Dycam Inc., Chatsworth, CA, USA) and digital pictures of the residue cover taken with a 5 megapixel digital camera.

Crop residue cover was estimated by classifying the infrared images using Multispec© image processing software developed by the Purdue Research Foundation (West Lafayette, IN, USA). Tillage classification was based on the percentage of the soil surface covered with crop residue. Conservation tillage systems were defined in this study as those that retained at least 30% of the soil surface covered with crop residue after a crop was planted. Ground-truth pixel locations on each image were identified using the GPS coordinates for extracting spectral reflectance data for each TM band image. In Landsat TM data, reflectance values are stored as brightness values (or digital numbers) in an 8 bit format. The raw brightness values for ground-truth pixels were extracted and analysed using image processing software. For model development and evaluation, the mean reflectance data from nine pixels (the ground-truth pixel and the surrounding eight pixels) were used. The Moore County data set was used for model development/calibration and the Ochiltree County data set was used for evaluating the models. For LSSVM and RVM model development, indices were developed with all possible combinations of two bands from all seven Landsat 5 TM bands. The TM indices included difference indices, sum indices, product indices, ratio indices and normalized difference indices. A comparative study has also been conducted between the developed LSSVM- and RVM-based tillage models used in this study and the logistic regression models developed by Gowda *et al.* (2008). Comparison has been made for both training and testing data sets. The following subsections briefly describe the LSSVM and RVM methods and the model evaluation criteria used for evaluation.

3.1 *Least square support vector machine*

Vapnik (1995) introduced SVMs for solving pattern recognition problems. SVM maps the low-dimensional data to a higher dimensional space and constructs an optimal separating hyperplane in the transformed space. This involves solving a quadratic programming problem. The dominant feature of SVM that makes it attractive is that classes that are non-linearly separable in the original feature space can be linearly separated in the higher dimensional feature space. This makes SVM capable of solving complex non-linear classification problems. The LSSVM uses the general concepts of SVM but utilizes the formulation in least squares and, as a result, the solution follows directly from solving a set of linear equations, instead of quadratic programming

(Suykens and Vandewalle 1998). Further details on LSSVM can be found in Suykens *et al.* (2002). In this study, the collection of tillage information has been considered as a binary classification problem. A binary classification problem is considered as having a set of training vectors ($\boldsymbol{D}$) belonging to two separate classes:

$$\boldsymbol{D} = \{(x_1, y_1), \ldots, (x_l, y_l)\}, \quad \mathbf{x} \in \mathrm{R}^n,\ y \in \{-1, +1\}, \tag{1}$$

where $\mathbf{x} \in \mathrm{R}^n$ and is an n-dimensional data vector with each sample belonging to either of two classes labelled as $y \in \{-1, +1\}$ and l is the number of training data. In this study, we use different input parameters for seven different models as shown in table 1. In the current context of classifying the tillage information, the two classes labelled as +1 and –1 are conservation tillage and conventional tillage, respectively.

For the case of two classes, one assumes the following:

$$\begin{aligned} w^{\mathrm{T}}\varphi(x_k) + b \geq 1, &\quad \text{if } y_k = +1 \ \text{(conservation tillage)}, \\ w^{\mathrm{T}}\varphi(x_k) + b \leq 1, &\quad \text{if } y_k = -1 \ \text{(conventional tillage)}, \end{aligned} \tag{2}$$

which is equivalent to

$$y_k\left[w^{\mathrm{T}}\phi(x_k) + b\right] \geq 1, \quad k = 1, \ldots, l, \tag{3}$$

where $\phi(\cdot)$ is a non-linear function that maps the input space into a higher dimensional space, b is the bias, T is the transpose and w is the weight. According to the structural risk minimization principle, the risk margin is minimized by formulating the following optimization problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}w^{\mathrm{T}}w + \frac{\gamma}{2}\sum_{k=1}^{l} e_k^2, \\ &\text{subject to } y_k\left[w^{\mathrm{T}}\phi(x_k) + b\right] = 1 - e_k,\ k = 1, \ldots, l, \end{aligned} \tag{4}$$

where γ is the regularization parameter, determining the trade-off between the fitting error minimization and smoothness, and e_k is an error variable. This optimization problem (4) is solved by Lagrangian multipliers (Suykens *et al.* 1999), and its solution is given by

$$y(x) = \operatorname{sign}\left[\sum_{k=1}^{l} \alpha_k y_k K(x, x_k) + b\right], \tag{5}$$

where α_{k} is the Lagrange multiplier, $K(x, x_k)$ is the kernel function and sign(•) is the signum function. Its resultant is +1 (conservation tillage) if the element is greater than or equal to 0 and –1 (conventional tillage) if it is less than 0.

This study adopts the above methodology for classifying contrasting tillage practices. In the LSSVM modelling, the data were divided into two subsets: a training data set for constructing the model and a testing data set for estimating the model performance. Thus, in this study, a total of 31 data points in Moore County were considered for training, and the other 41 data points in Ochiltree County were considered for testing. The training and testing data sets used in this study were also used by

Table 1. Performance of different LSSVM models.

Model	Input parameter	Design value of σ	Design value of γ	Training performance (%)	Testing performance (%)	Equation
I	TM5	1	200	91.4	75.6	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{2}\right\} - 0.0024\right]$
II	TM5, TM6	2	10	91.4	85.4	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{8}\right\} + 0.0566\right]$
III	D15, D16	2	60	88.6	87.8	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{8}\right\} + 0.1859\right]$
IV	R35, R36	4	120	91.4	85.4	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{32}\right\} + 0.1108\right]$
V	R45, R46	2	150	91.4	87.8	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{8}\right\} + 0.5333\right]$
VI	NDTI45, NDTI46	3	50	88.6	87.8	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{18}\right\} - 0.0194\right]$
VII	NDTI15, NDTI56	3	90	91.4	87.8	$y = \mathrm{sign}\left[\sum_{i=1}^{35} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^{\mathrm{T}}}{18}\right\} + 0.4451\right]$

Note: D15 = difference between TM bands 1 and 5; D16 = difference between TM bands 1 and 6; R35, R36, R45 and R46 = ratio of TM bands 3 and 5, 3 and 6, 4 and 5 and 4 and 6, respectively; NDTI45, NDTI46, NDTI15 and NDTI56 = normalized difference between TM bands 4 and 5, 4 and 6, 1 and 5 and 5 and 6, respectively.

Gowda *et al.* (2008) to develop logistic regression-based tillage models. The radial basis function $\left(K(x, x_k) = \exp\left\{-\frac{(x_k - x)(x_k - x)^{\mathrm{T}}}{2\sigma^2}\right\}\right)$ was adopted as a kernel function for LSSVM. The design values of the regularization parameter, γ and width (σ) of the radial basis function were determined by a trial-and-error approach during the training process.

3.2 *Relevance vector machine*

RVM, introduced by Tipping (2000), is a sparse linear model. The key feature of RVM is that its target function attempts to minimize the number of errors made in the training data set while simultaneously minimizing the margin between feature spaces. In this section, a brief introduction on how RVM is used for classification is presented. Consider a set of example input vectors $\{\boldsymbol{x_i}\}_{i=1}^{N}$ given along with a corresponding set of targets, $\mathrm{t} = \{\mathrm{t}_i\}_{i=1}^{N}$. For the classification problem, $\boldsymbol{x_i}$ should belong to either of the two classes (−1, +1). In the current context of determination of tillage practice, the two classes are labelled as −1 for conventional tillage and +1 for conservation tillage. Table 2 shows different input parameters for different RVM models. RVM constructs a logistic regression model based on a set of sequence features derived from the input patterns, i.e.

$$\mathrm{P}\left(\mathrm{C}_1/\mathrm{x}\right) \approx \sigma\left\{\mathrm{y}(\mathrm{x};\mathrm{w})\right\}, \quad \text{with } \mathrm{y}(\mathrm{x};\mathrm{w}) = \sum_{i=1}^{N} \boldsymbol{w_i}\Phi_i(\mathrm{x}) + \boldsymbol{w_0}, \tag{6}$$

where Φ_i is the ith component of the basis vector function:

$$\boldsymbol{\Phi}(\mathrm{x}) = \left(\Phi_1(\mathrm{x}), \Phi_2(\mathrm{x}), \ldots, \Phi_N(\mathrm{X})\right)^{\mathrm{T}} = \left[1, \mathrm{K}(\boldsymbol{x_i}, \boldsymbol{x_1}), \mathrm{K}(\boldsymbol{x_i}, \boldsymbol{x_2}), \ldots, \mathrm{K}(\boldsymbol{x_i}, \boldsymbol{x_N})\right]^{\mathrm{T}}, \tag{7}$$

$\mathrm{W} = (\boldsymbol{W_0}, \ldots, \boldsymbol{W_N})^{\mathrm{T}}$ is a vector of weights, $\sigma\{y\} = (1 + \exp\{-y\})^{-1}$ is the logistic sigmoid link function and $K\left(x_i, x_j\right)_{j=1}^{N}$ is the kernel term. Assuming a Bernoulli distribution for P (t/x), the likelihood can be written as

$$\mathrm{P}(\mathrm{t}/w) = \prod_{i=1}^{N} \sigma\left\{\mathrm{y}(\boldsymbol{x_i};\mathrm{w})\right\}^{\mathrm{t}_i}\left[1 - \sigma\left\{\mathrm{y}(\boldsymbol{x_i};\mathrm{w})\right\}\right]^{1-\mathrm{t}_i}. \tag{8}$$

To form a Bayesian training criterion, we must also impose a prior distribution over the vector of model parameters or weights, p(w). The RVM adopts a separable Gaussian prior, with a distinct hyperparameter, α_i, for each weight:

$$\mathrm{p}(\mathrm{w}/\alpha) = \prod_{i=1}^{N} N\left(\boldsymbol{w_i}/0, \alpha_i^{-1}\right). \tag{9}$$

The optimal parameters of the model are then derived by minimizing the penalized negative log-likelihood:

$$\log\left\{\mathrm{P}(\mathrm{t}/\mathrm{w})\,\mathrm{p}(\mathrm{w}/\alpha)\right\} = \sum_{i=1}^{N}\left[\mathrm{t}_i \log \mathrm{y}_i + (1 - \mathrm{t}_i)\log(1 - \mathrm{y}_i)\right] - \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{A}\mathrm{w}, \tag{10}$$

Table 2. Performance of different RVM models.

Model	Input parameter	Design value of σ	Number of relevance vector	Training performance (%)	Testing performance (%)	Equation
I	TM5	0.05	7	94.3	85.4	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.005}\right\}$
II	TM5, TM6	0.04	15	94.3	87.8	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.0032}\right\}$
III	D15, D16	0.03	19	88.6	87.8	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.0018}\right\}$
IV	R35, R36	0.04	7	91.4	85.4	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.0032}\right\}$
V	R45, R46	0.03	6	91.4	90.2	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.0018}\right\}$
VI	NDTI45, NDTI46	0.1	3	91.4	90.2	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.02}\right\}$
VII	NDTI15, NDTI56	0.07	4	91.4	90.2	$y = \sum_{i=1}^{35} w_i \exp\left\{-\frac{(x_i-x)(x_i-x)^{\mathrm{T}}}{0.0098}\right\}$

Note: D15 = difference between TM bands 1 and 5; D16 = difference between TM bands 1 and 6; R35, R36, R45 and R46 = ratio of TM bands 3 and 5, 3 and 6, 4 and 5 and 4 and 6, respectively; NDTI45, NDTI46, NDTI15 and NDTI56 = normalized difference between TM bands 4 and 5, 4 and 6, 1 and 5 and 5 and 6, respectively.

where $y_i = \sigma\{y(\boldsymbol{x_i}; \mathrm{w})\}$ and $\mathbf{A} = \mathrm{diag}(\alpha)$ is a diagonal matrix with non-zero elements given by the vector of hyperparameters. Next, Laplace's method is used to obtain a Gaussian approximation to the posterior density of the weights:

$$\mathrm{p}(\mathrm{w}/\mathrm{t}, \alpha) \approx N(\mathrm{w}/\mu, \textstyle\sum), \tag{11}$$

where the posterior mean and covariance are, respectively, given by

$$\mu = \textstyle\sum \Phi^{\mathrm{T}} \mathrm{Bt} \quad \text{and} \quad \textstyle\sum = \left[\Phi^{\mathrm{T}} \mathrm{B} \Phi + \mathrm{A}\right]^{-1}, \tag{12}$$

in which $\mathrm{B} = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_N)$ is a diagonal matrix with $\beta_n = \sigma\{\mathrm{y}(\mathrm{x}_n)\}[1 - \sigma\{\mathrm{y}(\mathrm{x}_n)\}]$.

The hyperparameters are then updated to maximize their marginal likelihood, according to their efficient update formula:

$$\alpha_i^{\mathrm{new}} = \frac{1 - \alpha_i \Sigma_{ii}}{\mu_i^2}, \tag{13}$$

where μ_{i} is the ith posterior mean weight and $\sum_{ii}$ is the ith diagonal element of the posterior weight covariance and the quantity provides a measure of the degree to which the associated parameter $\boldsymbol{w_i}$ is determined by the data. This process is repeated until an appropriate convergence criterion is met. The outcome of this optimization is that many elements of α go to infinity such that w will have only a few non-zero weights that will be considered as relevance vectors. Training and testing data sets and the kernel function used in the LSSVM model were used in the implementation of the RVM model for performance comparison purposes. Both LSSVM and RVM programmes were constructed and implemented using MATLAB software (MATLAB 2010).

3.3 *Model evaluation*

For the purpose of model evaluation, error matrices (Campbell 1987) were developed for all LSSVM and RVM models to determine the percentage correct (overall classification accuracy) and kappa coefficient (k) values. Percentage correct was calculated by dividing the sum of correctly classified fields by the total number of fields examined as follows:

$$\text{percentage correct}\,(\%) = \left(\frac{\text{no. of data predicted correctly}}{\text{total no. of data}}\right) \times 100. \tag{14}$$

The percentage correct values were computed separately for the training and testing data of each of the LSSVM and RVM models. The 'k' value is a measure of the difference between two maps and the agreement that might be contributed solely by chance matching of the two maps (Congalton and Green 1999). The k value was calculated as

$$k = \left(\frac{\mathrm{O} - \mathrm{E}}{1 - \mathrm{E}}\right), \tag{15}$$

where O means observed, which is the percentage correct and E is the expected, which is an estimate of the chance agreement to the observed. A k value of $+1.0$ indicates perfect accuracy of the classification.

4. Results and discussion

Tables 1 and 2 present performances of LSSVM and RVM models, respectively, and table 3 presents comparison of their performances with logistic regression models reported by Gowda *et al.* (2008). Based on the training performance results, all the LSSVM models except III and VI performed better with a percentage correct value of 91.4%. However, application of these models to the evaluation data set indicated that only models V, VI and VII maintained their high performance level with only 5 out of 41 data points misclassified. Percentage correct and k values for these models were 87.8% and 0.75 (table 3), respectively, indicating that LSSVM models may be suitable for identifying fields with contrasting tillage practices. The developed LSSVM has also been used to develop an equation (by inputting the design values of σ and b for different models in equation (5) modified for radial basis function) for prediction of tillage information. Table 1 shows the different equations for prediction of tillage information for different models. The values of α for different models are given in figure 2.

Comparison of RVM models in table 2 indicates that models that use individual bands (I and II) performed better with a percentage correct value of 94.3% with the training/calibration data set. However, the percentage correct was reduced to 85.4–87.8% when these models were applied to the evaluation data set. Models V, VI and VII that use ratio or normalized difference tillage indices performed better than those that use individual TM bands or difference indices. The percentage correct values were consistently higher for both training (91.4%) and testing (90.2%) data sets. The k values for all three models were the same and equal to 0.8, indicating that

Table 3. Comparison between testing performance of LSSVM, RVM and logistic regression models.

		Testing performance					
		RVM		LSSVM		Logistic regression	
Model	Input parameter	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
I	TM5	85.4	0.70	75.6	0.51	73	0.35
II	TM5, TM6	87.8	0.75	85.4	0.70	83	0.52
III	D15, D16	87.8	0.75	87.8	0.75	85	0.56
IV	R35, R36	85.4	0.70	85.4	0.70	80	0.60
V	R45, R46	90.2	0.80	87.8	0.75	85	0.70
VI	NDTI45, NDTI46	90.2	0.80	87.8	0.75	85	0.70
VII	NDTI15, NDTI56	90.2	0.80	87.8	0.75	83	0.60

Note: D15 = difference between TM bands 1 and 5; D16 = difference between TM bands 1 and 6; R35, R36, R45 and R46 = ratio of TM bands 3 and 5, 3 and 6, 4 and 5 and 4 and 6, respectively; NDTI45, NDTI46, NDTI15 and NDTI56 = normalized difference between TM bands 4 and 5, 4 and 6, 1 and 5 and 5 and 6, respectively; OA is the overall accuracy.

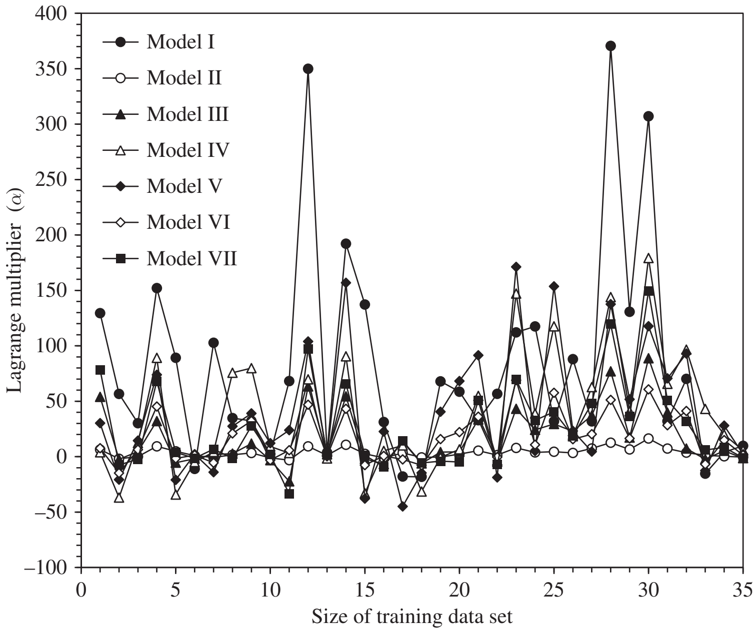

Figure 2. Values of α for different LSSVM models.

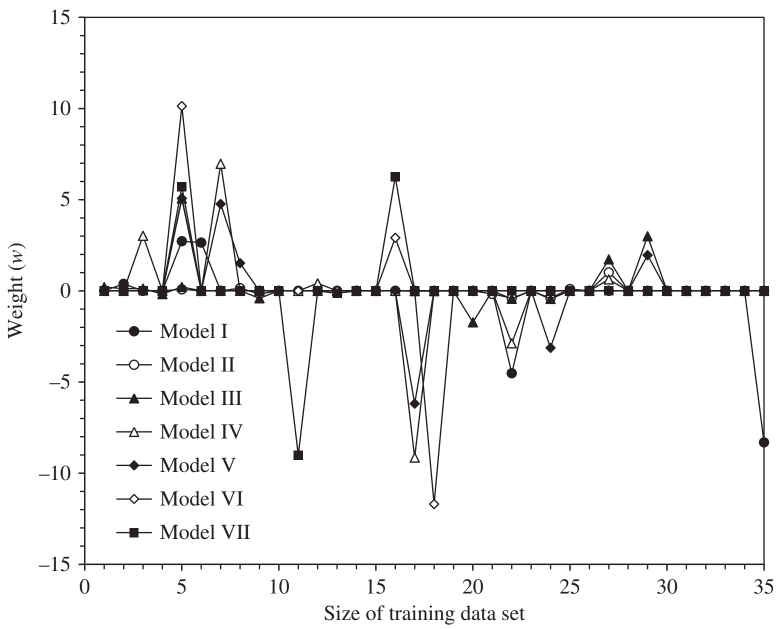

Figure 3. Values of w for different RVM models.

RVM models were superior to the best LSSVM models reported above. The equations for the RVM model were also developed for collection of tillage information. Table 2 shows the different equations for the RVM models. Figure 3 depicts the values of w for these different models.

Comparison of k values in table 3 clearly indicates that RVM models performed the best ($k = 0.8$) followed by LSSVM ($k = 0.75$) and logistic regression ($k = 0.7$) models. The RVM model uses only one parameter (σ) as a tuning parameter, whereas the LSSVM model uses two parameters (γ and σ). Furthermore, the RVM models used only 8–55% of the training data as relevance vectors. These relevance vectors were used for final prediction. So, there is real advantage gained in terms of sparsity. Sparseness means that a significant number of the weights are 0 (or effectively 0), which has the consequence of producing compact, computationally efficient models, which in addition are simple and therefore produce smooth functions. However, the LSSVM and logistic regression models do not produce sparse solution.

5. Conclusions

Tillage information on individual production fields at a regional scale is a crucial input in environmental modelling applications. In this study, two statistical learning algorithms, namely LSSVM and RVM, were evaluated to determine their ability to identify two contrasting tillage practices in the Texas High Plains and their performances were compared with logistic regression-based models. The results indicate that models that use TM band 5 or TM indices that use TM band 5 performed better with all three statistical models, indicating that near-infrared reflectance is sensitive to crop residue cover on the surface. Comparison of k values associated with tillage models indicated that RVM models performed better than LSSVM- and logistic regression-based models. The developed RVM models also produce a sparse solution, and thus users can use the developed equations for identifying tillage practices using Landsat TM data at a regional scale.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

References

Bishop, C.M., 1995, *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press).

Bricklemyer, R.S., Lawrence, R.L., Miller, P.R. and Battogtokh, N., 2006, Predicting tillage practices and agricultural soil disturbance in north-central Montana with Landsat imagery. *Agriculture, Ecosystems & Environment*, **114**, pp. 210–216.

Campbell, J.B., 1987, *Introduction to Remote Sensing*, 551 pp. (New York: The Guilford Press).

Congalton, R.G. and Green, K., 1999, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 132 pp. (Boca Raton, FL: CRC Press).

Daughtry, C.S.T., Hunt Jr., E.R., Doraiswamy, P.C. and McMurtrey, J.M., 2005, Remote sensing the spatial distribution of crop residues. *Agronomy Journal*, **97**, pp. 864–878.

DeGloria, S.D., Wall, S.L., Benson, A.S. and Whiting, M.L., 1986, Monitoring conservation tillage practices using Landsat multispectral data. *Journal of Soil and Water Conservation*, **41**, pp. 187–189.

Foody, G.M., 2008, RVM-based multi-class classification of remotely sensed data. *International Journal of Remote Sensing*, **29**, pp. 1817–1823.

Ghosh, S. and Mujumdar, P.P., 2008, Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*, **31**, pp. 132–146.

Gowda, P.H., Howell, T.A., Evett, S.R., Chavez, J.L. and New, L., 2008, Remote sensing of contrasting tillage practices in the Texas Panhandle. *International Journal of Remote Sensing*, **29**, pp. 477–3487.

Gowda, P.H., Mulla, D.J. and Dalzell, B.J., 2003, Examining and targeting conservation tillage practices to steep/flat landscapes in the Minnesota River Basin. *Journal of Soil and Water Conservation*, **58**, pp. 53–57.

Kecman, V., 2001, *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models* (Cambridge, MA: The MIT Press).

Mathur, A. and Foody, G.M., 2008, Crop classification by support vector machine with intelligently selected training data for an operational application. *International Journal of Remote Sensing*, **29**, pp. 2227–2240.

MATLAB, 2010, *Users Guide* (Natick, MA: Mathworks, Inc.).

Motsch, B., Schaal, G., Lyon, J.G. and Logan, T.J., 1990, Monitoring crop residue in Senaca County, Ohio. In *Proceedings of the ASPRS Meeting*, Cleveland, OH (Bethesda, MD: American Society of Photogrammetry & Remote Sensing), pp. 66–76.

NASS, 2005, *2004 Texas Agricultural Statistics: Texas Agriculture by the Number, Bulletin 263* (Austin, TX: Texas Field Office, U.S. Department of Agriculture, National Agricultural Statistics Service).

Park, D. and Rilett, L.R., 1999, Forecasting freeway link travel times with a multi-layer feed forward neural network. *Computer Aided Civil and Infrastructure Engineering*, **14**, pp. 358–367.

Samui, P. and Sitharam, T.G., 2008, Least square support vector machine applied to settlement of shallow foundations on cohesionless soils. *International Journal of Numerical and Analytical Methods in Geomechanics*, **32**, pp. 2033–2043.

Schwartz, R.C., Baumhardt, R.L. and Evett, S.R., 2010, Tillage effects on soil water redistribution and bare soil evaporation throughout a season. *Soil & Tillage Research*, **110**, pp. 221–229.

Sudheer, K.P., Gowda, P.H., Chaubey, I. and Howell, T.A., 2010, Artificial neural network approach for mapping contrasting tillage practices. *Remote Sensing*, **2**, pp. 579–590.

Sullivan, D.G., Shaw, J.N., Mask, P., Rickman, D., Luvall, J. and Wersinger, J.M., 2004, Evaluation of multispectral data for rapid assessment of in situ wheat straw residue cover. *Soil Science Society of American Journal*, **68**, pp. 2007–2013.

Sullivan, D.G., Truman, C.C., Schomberg, H.H., Endale, D.M. and Strickland, T.C., 2006, Evaluating techniques for determining tillage regime in the Southeastern Coastal Plain and Piedmont. *Agronomy Journal*, **98**, pp. 1236–1246.

Suykens, J.A.K., De, B.J., Lukas, L. and Vandewalle, J., 2002, Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, **48**, pp. 85–105.

Suykens, J.A.K., Lukas, L., Van, D.P., De, M.B. and Vandewalle, J., 1999, Least squares support vector machine classifiers: a large scale algorithm. In *Proceedings of the European Conference on Circuit Theory and Design (ECCTD'99)*, Stresa, Italy, pp. 839–842.

Suykens, J.A.K. and Vandewalle, J., 1998, *Nonlinear Modeling Advanced Black Box Techniques* (Boston, MA: Kluwer Academic Publishers).

Takken, I., Govers, G., Jetten, V., Nachtergaele, J., Steelgen, A. and Poesen, J., 2001, Effect of tillage on runoff and erosion patterns. *Soil & Tillage Research*, **61**, pp. 55–60.

Tang, H., Qiu, J., Van Ranst, E. and Li, C., 2006, Estimations of soil organic carbon storage in cropland of China based on DNDC model. *Geoderma*, **134**, pp. 200–206.

Thoma, D.P., Gupta, G.C. and Bauer, M.E., 2004, Evaluation of optical remote sensing models for crop residue cover assessment. *Journal of Soil and Water Conservation*, **59**, pp. 224–233.

Tipping, M.E., 2000, The relevance vector machine. *Advances in Neural Information Processing Systems*, **12**, pp. 625–658.

USDA-SCS, 1975, *Soil Survey of Moore County, Texas* (Washington, DC: Soil Conservation Service, United States Department of Agriculture).

van Deventer, A.P., Ward, A.D., Gowda, P.H. and Lyon, J.G., 1997, Using Thematic Mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering & Remote Sensing*, **63**, pp. 87–93.

Vapnik, V.N., 1995, *The Nature of Statistical Learning Theory* (New York: Springer Publications).

Vervoort, R.W., Dabney, S.M. and Romkens, M.J.M., 2001, Tillage and row position effects on water and solute infiltration characteristics. *Soil Science Society of America Journal*, **65**, pp. 1227–1234.

Vina, A., Peters, A.J. and Ji, L., 2003, Use of multispectral Ikonos imagery for discriminating between conventional and conservation agricultural tillage practices. *Photogrammetric Engineering & Remote Sensing*, **69**, pp. 537–544.

West, T.O. and Marland, G., 2002, A synthesis of carbon sequestration, carbon emissions, and net carbon flux in agriculture: comparing tillage practices in the United States. *Agriculture, Ecosystems & Environment*, **91**, pp. 217–232.