One-class classification for monitoring a specific land cover class: SVDD classification of fenland

Sanchez-Hernandez, C., Boyd, D. S. and Foody, G. M.

IEEE Transactions on Geoscience and Remote Sensing, 45, 1061-1073 (2007)

The manuscript of the above article revised after peer review and submitted to the journal for publication, follows. Please note that small changes may have been made after submission and the definitive version is that subsequently published as:

Sanchez-Hernandez, C., Boyd, D. S. and Foody, G. M., 2007. One-class classification for monitoring a specific land cover class: SVDD classification of fenland, *IEEE Transactions on Geoscience and Remote Sensing*, 45, 1061-1073.

One-Class Classification for Mapping a Specific Land Cover Class: SVDD Classification of Fenland

Carolina Sanchez-Hernandez[1], Doreen S. Boyd[2], and Giles M. Foody[2*]

1. Research and Innovation, Ordnance Survey, Romsey Road, Southampton, SO16 4GU, UK (e-mail: *Carolina.Hernandez@ordnancesurvey.ac.uk*)

2. School of Geography, University of Nottingham, University Park, Nottingham, Nottingham, NG7 2RD, UK (e-mail: *Doreen.Boyd@nottingham.ac.uk*)

* *Corresponding author*, School of Geography, University of Nottingham, University Park, Nottingham, NG7 2RD, UK (e-mail: *giles.foody@nottingham.ac.uk*)
[*NOTE*: - contact e-mail to end January 2007 is *g.m.foody@soton.ac.uk*)

Abstract

Remote sensing is a major source of land cover information. Commonly, interest focuses on a single land cover class. Although a conventional multi-class classifier may be used to provide a map depicting the class of interest the analysis is not focused on that class and may be sub-optimal in terms of the accuracy of its classification. With a conventional classifier, considerable effort is directed on the classes that are not of interest. Here, it is suggested that a one-class classification approach could be appropriate when interest focuses on a specific class. This is illustrated with the classification of fenland, a habitat of considerable conservation value, from Landsat ETM+ imagery. A range of one-class classifiers are evaluated but attention focuses on the support vector data description (SVDD). The SVDD was used to classify fenland with an accuracy of 97.5% and 93.6% from the user's and producer's perspectives respectively. This classification was trained upon only the fenland class and was substantially more accurate in fen classification than a conventional multi-class maximum likelihood classification provided with the same amount of training data,

which classified fen with an accuracy of 90.0% and 72.0% from the user's and producer's perspectives respectively. The results highlight the ability to classify a single class using only training data for that class. With a one-class classification the analysis focuses tightly on the class of interest, with resources and effort not directed on other classes, and there are opportunities to derive highly accurate classifications from small training sets.

I. Introduction

Given that the remotely sensed response is predominantly a function of Earth surface properties, remote sensing has great potential as a source of information on land cover. Remote sensing has been used to map and monitor land cover at a range of spatial and temporal scales in order to satisfy a range of scientific and practical requirements [1,2]. Supervised classification lies at the heart of such studies as a means of converting the remotely sensed image into a thematic map depicting the land cover classes. Typically, the land cover map produced depicts many classes. However, interest is often focussed on just one specific class [3-7]. This is evident, for example, in many studies of major land cover transformations such as deforestation and urbanization. Attention has also focused on deriving information on specific crop types since pioneering major programmes such as LACIE which exploited imagery from early satellite remote sensing systems [8]. It is also the case in many ecological studies where attention is focused perhaps on an invasive species [9,10] or a rare habitat for conservation monitoring [11]. With such applications the use of conventional image classification techniques may be inappropriate and wasteful of resources, particularly when a competitive and efficient approach is required [11].

Many problems are encountered in mapping land cover from remotely sensed data by a classification analysis [12]. One concern is that the correct execution of a conventional supervised classifier for the derivation of the desired information typically requires the analyst to train the classifier on every class that occurs in the study area, even if most are of no interest to the analysis, in order to satisfy the underlying assumption of an exhaustively defined set of classes. Failure to exhaustively define the set of classes may result in substantial error which may not be captured in the evaluation of classification accuracy [13]. Moreover, with a conventional statistical classifier such as the maximum likelihood (ML) classification, the analysis may aim to optimize classification accuracy over all land cover classes rather than focus on that specific class of interest. The classification analysis may, therefore, not be focussed on the needs of a particular application. Thus, when there is a specific land cover class of interest, it may sometimes be preferable to adopt an alternative to the standard multi-class classifiers that are predominantly used in remote sensing.

A variety of approaches may be used to shift the focus from the general to specific class of interest (e.g. [14,15]). One simple approach is to adopt binary classifiers to separate the specific class of interest from all others [11,16]. A further refinement of this approach is to conduct a one-class classification. With such an approach, classification resources (e.g. training data) and processes (e.g. class allocation) can be focused on the specific class of interest only. A range of techniques have been developed for this type of analysis. For example, approaches commonly referred to as partially supervised classification [3,5] seek to separate the class of interest from others with the aid of unlabelled training samples. Although these approaches have been used with some success there are limitations to their use. For example, the use of some approaches is

based on the assumption that the probability density of the class of interest, unless known in advance, can be estimated accurately from the training set, which may require the use of a large training sample [5]. As the expense of training data acquisition is one reason for the adoption of partially supervised classifications [17] the requirement for a large sample may be a major limitation in some studies. Here attention is focused on a straightforward approach to classifying a specific class of interest that requires training in a manner similar to conventional supervised classifications but only training data for the class of interest. This paper proposes the adoption of such a one-class classification approach based on the principles of the support vector machine (SVM) for accurate classification of a single class of interest from remotely sensed data. The classifier is introduced in section III after briefly reviewing in section II a range of one-class classifiers. In section IV the study area and class of interest are introduced. The data and methods used are outlined in section V before presenting the results in section VI. The paper closes with a brief summary of the work and key conclusions in section VII.

II. One class classification

The term one-class classification is believed to have originated from [18]. Other terms have been used in the recent literature that refers to the same or similar concept, such as outlier detection [19], novelty detection [20] and concept learning [21]. One-class classification has proved valuable in a variety of research arenas, such as document classification [22], texture segmentation [23], image retrieval [24] and ecological modeling [25].

A range of approaches exist to classify a specific class of interest, including reconstruction methods [26,27], density methods [28-30] and boundary methods

[31,32]. Reconstruction methods and density methods require extensive knowledge and large amount of information about the dataset of interest. Boundary methods, however, are more flexible in that they do not require the extensive knowledge of the dataset, as they concentrate on the boundary that fits around the class of interest [33]. This makes the boundary method very attractive to use in remote sensing applications which are focussed on a specific land cover class, particularly when a competitive and efficient approach is required. Boundary methods are largely based upon the principles of SVMs. One is a model in which a hypersphere is computed to describe the specific class of interest [32]. The model is written in a form comparable to SVMs [34], and is called the support vector data description (SVDD) [33].

A variety of one-class classifiers are therefore available. Here, some key details on a set used in this research are briefly provided. In all of the one-class classification methods two elements can be identified. The first element is a measure for the distance $d(\mathrm{z})$ or resemblance (sometimes a probability) $p(\mathrm{z})$ of a case $\mathbf{z}$ to the target class. The second element is a threshold θ on this distance or resemblance. New cases may be are accepted by the description when the distance to the target class is smaller than the threshold or when the resemblance is larger than the threshold. The one-class classification methods differ, however, in their definition of $p(\mathrm{z})$ or $d(\mathrm{z})$, in their optimization of $p(\mathrm{z})$ or $d(\mathrm{z})$ and thresholds with respect to the training set [33].

The most straightforward method to obtain a one-class classifier is to estimate the density of the training data and to set a threshold on this density. It is assumed that the target data were derived from a family of known distributions (e.g. Gaussian or Parzen distributions). The probability density is then estimated from available data samples. Cases of unknown membership may then be assigned to the target class if $p > \theta$ where θ

> 0 is a chosen threshold level. Therefore, the whole classification process is defined by the choice of θ. Several data distributions can be assumed. The Gaussian distribution assumes a unimodal and convex model of the data which can sometimes be overly rigid and inappropriate. To address this issue a Mixture of Gaussians can be used which is a linear combination of normal distributions. An extension of this approach is the Parzen density estimation method where a mixture of Gaussian kernels are centred on individual training points [35].

Reconstruction methods have not been primarily constructed for one-class classification, but rather to model the data. By using prior knowledge about the data and making assumptions about the generating process, a model is chosen and fitted to the data. Most of these methods make assumptions about the clustering characteristics of the data or their distribution in subspaces. With the application of the reconstruction methods, it is assumed that outlier objects do not satisfy the assumptions about the target distribution. The reconstruction error of a test object is used as a distance to the target set. Because these methods were not developed for one-class classification, the empirical threshold has to be obtained using the training set. The simplest method is the k-means classifier. In this method it is assumed that the data are clustered and can be characterized by a few prototype cases. The distance d of a case $\mathbf{z}$ to the target set is then defined as the squared distance of that case to the nearest prototype [35]. In the self-organizing map (SOM) the placing of the prototypes is not only optimized with respect to the data, but also constrained to form a low-dimensional manifold [36]. In these methods the Euclidean distance is used in the definition of the error and the computation of the distance. Another method is the principal components analysis (PCA). The PCA mapping finds the orthonormal subspace which captures the variance in the data as best as possible. The reconstruction error of a case $\mathbf{z}$ is now defined as the

squared distance from the original object and its mapped version. Other reconstruction methods include those based on auto-encoders and diabolo networks [37-39].

In the boundary methods, only a closed boundary around the target set is optimized. In most cases the distances or weighted distances *d* to an (edited) set of cases in the training set are computed and objects are accepted or rejected according to a threshold. For example, the k-centre method covers the dataset with *k* small balls with equal radii [40]. The ball centres are placed on training cases such that the maximum distance of all minimum distances between training cases and the centres is minimized. When the centres have been trained, the distance from a test case **z** to the target set can be calculated. In the NN classifier, a test object **z** is accepted when its local density is larger or equal to the local density of its (first) nearest neighbour in the training set. This means that the distance from case **z** to its nearest neighbour in the training set is compared with the distance from this nearest neighbour to its nearest neighbour [41]. In the SVDD on the other hand, a boundary in the form of a sphere contains all the target data within the smallest radius and all the outliers will lie outside this sphere. These outliers are identified by calculating the distance of a new case **z** to the centre of the sphere [32]. As classification with the SVDD is the main focus of this article it is more fully described in the following section.

III. One-class classification by SVDD

The support vector data description (SVDD) is a one-class classifier based on the principles of the SVM. A SVM is binary classifier that may be used to separate a specific class from others. Recent studies have shown the SVM concept to have much promise in remote sensing and land cover mapping applications with the SVM approach

often shown to be more accurate than standard parametric classifiers (e.g., maximum likelihood) and popular alternatives (e.g. feedforward neural networks) for land cover mapping [42-45]. To provide an introduction to the SVDD, SVM classification will first be briefly discussed.

The basic SVM is a binary classifier that seeks to fit an optimal separating hyperplane or decision boundary between the classes. For simplicity, the discussion will assume a simple two-dimensional scenario. The classification process is based on support vectors, essentially training samples that lie on the very edge of the class distributions in feature space in the border region between the classes. Although a binary classifier, the basic SVM approach may be extended for multi-class classifications [42,45].

The binary SVM is trained on two classes of data where $\{\mathbf{x_i}\ y_i\}$, $i = 1,..,\ l$, $y_i \in \{-1,1\}$, $\mathbf{x_i} \in R^d$. A hyperplane that separates the two classes, positives from negatives, may be expressed as $\mathbf{w} \cdot \mathbf{x} + b = 0$. Where $\mathbf{w}$ is the normal to the hyperplane, $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin and $\|\mathbf{w}\|$ is the Euclidean norm of $\mathbf{w}$ (Fig. 1). Many hyperplanes may be able to separate the classes but the aim is typically to find the optimal separating hyperplane. The definition of the optimal separating hyperplane is based on the margin or distance between the classes. If d_+ (d_-) is the shortest distance from the hyperplane to the closest positive (negative) case then the margin of a separating hyperplane is $d_+ + d_-$. When the classes are linearly separable, the aim is to maximize the margin. If all the training data satisfy the following constraints:

$$\begin{aligned} \mathbf{x_i} \cdot \mathbf{w} + b \geq +1 \quad for \quad y_i &= +1 \\ \mathbf{x_i} \cdot \mathbf{w} + b \leq -1 \quad for \quad y_i &= -1 \end{aligned} \tag{1}$$

then:

$$y_i(\mathbf{x_i} \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall_i \tag{2}$$

Two hyperplanes, H_1 and H_2, may be defined (Fig. 2). Cases that lie on H_1, $\mathbf{x_i} \cdot \mathbf{w} + b = 1$, and H_2, $\mathbf{x_i} \cdot \mathbf{w} + b = -1$, are called support vectors and are the critical elements of the training set. H_1 and H_2 are parallel and no training points fall between them. Also $d_+ = d_- = 1/\|\mathbf{w}\|$ and the margin is $2/\|\mathbf{w}\|$.

A binary SVM may be used to concentrate the resources used in the collection of training data on the specific class of interest by aggregating all remaining classes in the image into a single class [11], here referred to as the class other. The analyst then has to simply train the classification on the two classes to allow a hyperplane that separates them to be defined. This type of approach results in a binary land cover map in which the class of interest may be more accurately mapped than from a standard multi-class classifier [11]. Although this type of approach may be accurate and is focused tightly on the class of interest, some effort is still required to characterize cases outside the class of interest to enable the classification hyperplane that separates the class of interest from others to be fitted to the data. The approach can, however, be further refined by adopting a classifier that requires only training data from the single class of interest. That is, it may be possible to train only upon the class of interest. It may, however, sometimes be beneficial to include training cases from other classes. Even in such

circumstances, however, training is biased toward the class of interest. This option is discussed in more detail below.

In the SVDD, a model f (**x;w**) is defined in such a way that instead of looking for a hyperplane it searches for a closed boundary around the data, a hypersphere [46]. The hypersphere is characterised by a centre **a** and radius R (Fig. 3). During the training stage of a classification only examples of the class of interest are used and, therefore, all of them are assumed to be within the radius R of the hypersphere. The aim is to find the hypersphere with minimum radius R and for that an error function F is formulated so that:

$$F(R,\mathbf{a}) = R^2 \tag{3}$$

With the constrains that all the training data, $\mathbf{x_i}$, are within that R^2

$$\left\|\mathbf{x_i} - \mathbf{a}\right\|^2 \le R^2, \qquad \forall_i \tag{4}$$

In order to allow the possibility of outliers in the training set, the distance from $\mathbf{x_i}$ to the center **a** should not be strictly smaller than R^2, but larger distances should be penalized. Therefore, slack variables, $\xi_i \ge 0$ may be introduced and the minimization problem changes to:

$$F(R,\mathbf{a},\boldsymbol{\xi}) = R^2 + C\sum_i \xi_i \tag{5}$$

with constraints that almost all objects are within the sphere:

$$\left\|\mathbf{x_i} - \mathbf{a}\right\|^2 \le R^2 + \xi_i, \qquad \boldsymbol{\xi} \ge 0, \quad \forall_i \tag{6}$$

where *C* is a parameter that gives a trade-off between the volume of the description and the misclassification errors.

Constraints (6) can be incorporated into equation (5) by using Lagrange multipliers. The resultant Lagrangian formula is,

$$L(R, \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\xi}) = R^2 + C\sum_i \xi_i - \sum_i \alpha_i \left\{R^2 + \xi_i - (\left\|\mathbf{x_i}\right\|^2 - 2\mathbf{a}\cdot\mathbf{x_i} + \left\|\mathbf{a}\right\|^2)\right\} - \sum_i \gamma_i \xi_i \tag{7}$$

with the Lagrange multipliers $\alpha_i \ge 0$ and $\gamma_i \ge 0$. *L* should be minimised with respect to $R, \boldsymbol{\xi}$, **a**, and maximised with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ with the constraints,

$$\sum \alpha_i = 1 \tag{8}$$

$$\mathbf{a} = \frac{\sum_i \alpha_i \mathbf{x_i}}{\sum_i \alpha_i} = \sum_i \alpha_i \mathbf{x_i} \tag{9}$$

$$C - \alpha_i - \gamma_i = 0 \tag{10}$$

From equation (10), $\alpha_i = C - \gamma_i$. Instead of the constraint $\gamma_i \ge 0$ and $\gamma_i = C - \alpha_i$, a new constraint on α_i can be introduced as $0 \le \alpha_i \le C$ where the value of *i* lies in the range {1,…, *N*}. With this constraint and substituting (8)-(10) into (7) gives the simplified formula,

$$L = \sum_i \alpha_i (\mathbf{x_i} \cdot \mathbf{x_i}) - \sum_{i,j} \alpha_i \alpha_j (\mathbf{x_i} \cdot \mathbf{x_j}). \tag{11}$$

The magnitude of the Lagrangian multipliers varies with the position of the case relative to the hypersphere (Fig. 4). For a case within the hypersphere $\alpha_i = 0$ and $\gamma_i = C$. For a

case on the hypersphere boundary, $\alpha_i = \{0\text{-}C\}$ and $\gamma_i = \{0 - C\}$. Alternatively, for a case outside the hypersphere $\alpha_i = C$ and $\gamma_i = 0$.

Only objects $\mathbf{x_i}$ with $\alpha_i > 0$ are needed in the description and these objects are the support vectors of the description. In the SVDD, the pixels that are support vectors lie in the border of the hypersphere. As well as with the support vectors in the SVM, these support vectors are essential for the calculation of the optimal hypersphere. The magnitude of the parameter *C* must be selected by the analyst for the application in-hand [46]. An appropriate value for this parameter may be selected from empirical trials or cross-validation [47-49] and there is a degree of robustness to the inclusion of outliers, in part through the specification of *C*.

Because it is possible to calculate the centre of the hypersphere **a**, it is easy to test if a new object, here a pixel with a particular spectral response, **z**, is accepted by the sphere description and, therefore, belongs to the class of interest. This requires the calculation of the distance from the pixel **z** to the centre of the hypersphere. A pixel **z** is accepted within the hypersphere when this distance is smaller than or equal to the radius:

$$\|\mathbf{z}-\mathbf{a}\|^2 = (\mathbf{z}\cdot\mathbf{z}) - 2\sum_i \alpha_i (\mathbf{z}\cdot\mathbf{x_i}) + \sum_{i,j} \alpha_i \alpha_j (\mathbf{x_i}\cdot\mathbf{x_j}) \le R^2 \qquad (12)$$

The squared distance from the center of the sphere **a** to (any of the support vectors on) the boundary is R^2. Support vectors which fall outside the description ($\alpha_i = C$) are excluded [46].

The one-class SVDD classification can, therefore, be expressed as:

$$f_{SVDD}(\mathbf{z};\boldsymbol{\alpha},R)=I\left(\|\mathbf{z}-\mathbf{a}\|^2\le R^2\right)=I\left((\mathbf{z}\cdot\mathbf{z})-2\sum_i\alpha_i(\mathbf{z}\cdot\mathbf{x_i})+\sum_{i,j}\alpha_i\alpha_j(\mathbf{x_i}\cdot\mathbf{x_j})\le R^2\right) \quad (13)$$

where I is defined as:

$$I(A)=\begin{cases}1\ if\ A\ is\ true\\ 0\ otherwise\end{cases}$$

As with SVMs, to extend the model to non-linear cases, to give it generality and prevent over-fitting, a mapping ϕ of the data using kernel functions may be undertaken [33]. These kernel representations were introduced by Boser *et al*. [50], based upon the work of Aizerman *et al*. [51] and they consist of projecting the data into a high dimensional Euclidean space H in which the linear learning machines can be implemented. When a kernel function maps the target data into the feature space the hypersphere model fits the data more appropriately and a more accurate classification may be obtained [33].

To avoid the trivial solution of accepting all cases, an assumption about the outlier distribution is made. If no example outliers are available, it is assumed that the outliers are uniformly distributed around the target data. Using the Bayes rule, the posterior probability for the target class can be computed by:

$$p(\omega_T|\mathbf{x})=\frac{p(\mathbf{x}|\omega_T)p(\omega_T)}{p(\mathbf{x})}=\frac{p(\mathbf{x}|\omega_T)p(\omega_T)}{p(\mathbf{x}|\omega_T)p(\omega_T)+p(\mathbf{x}|\omega_O)p(\omega_O)} \quad (14)$$

where ω_O is the outlier data and ω_T the target data. When it is assumed that $p(\mathbf{x}|\omega_O)$ is independent of **x (**i.e. it is uniformly distributed in the area of the feature space that it is

being considered) $p(\mathbf{x}|\omega_T)$ can be used instead of $p(\omega_T|\mathbf{x})$. The $p(\omega_T|\mathbf{x})$ is transformed into $p(\mathbf{x}|\omega_T)$ by a strictly increasing function. So when $p(\omega_T|\mathbf{x_1}) < p(\omega_T|\mathbf{x_2})$ also $p(\mathbf{x_1}|\omega_T) < p(\mathbf{x_2}|\omega_T)$ holds. The values of $p(\omega_T)$ and $p(\omega_O)$ have to be assumed *a priori*.

Using a uniform outlier distribution also means that when false positives are minimized, the data description with minimal volume is obtained. So instead of minimizing both false positives and false negatives, a combination of false positives and the volume of the description can be minimized to obtain a good data description. When the true outlier distribution deviates from the uniform distribution, another data description will show better generalization performance, but this cannot be checked without the use of example outliers. Therefore, the generalization can only be given on the target data. It may, however, sometimes be useful to include outlier cases in the analysis. This is an option to enhance the basic SVDD but as the number of outliers grows the analysis is turned increasingly into a binary classification. Thus,when data about outliers (i.e., cases of the other classes in the image that are not of direct interest) are available they too can be used during the training to potentially enhance the data description and to help obtain a tighter boundary around the class of interest in the areas where outlier objects are present. Within the remote sensing context, this would be the equivalent of training a SVDD on examples of two classes, the specific class of interest and others. When data on outliers are available they may, therefore, be used to train a SVDD and a more traditional binary classifier. However, the SVDD differs from a conventional classifier because it always obtains a closed boundary around the class of interest. Furthermore, it does not require a representative sample of the outliers. Also a conventional classifier typically distinguishes between two (or more) classes without special focus on any of the classes.

The research reported in this article uses the SVDD approach in both modes: (i) training on one class only and (ii) training on one class with a small amount of data on the other classes in the image. The latter was included as the outliers may sometimes help fit an appropriate hypersphere. The inclusion of outlier cases is merely an option that may sometimes help extend the basic SVDD classification. The principal aim of the research reported was to assess the potential of the SVDD, one-class classifier, for mapping a specific class of interest from remotely sensed data.

IV. Study Site

The study focussed on the area known as The Norfolk Broads, East Anglia, UK (Fig. 5). This is an area renowned internationally for its wetland ecosystem and comprises a range of different landscapes, land-uses, habitats, and a considerable richness of wildlife [52]. One habitat of major interest is fen, which supports a large diversity of plant and animal communities. The UK, in general, and the Norfolk Broads in particular, are thought to host a large proportion of the surviving fen within Europe, and as such they require management under a number of legislative directives [53]. Fenlands, however, are dynamic semi-natural systems, thus informed management is required to maintain open-fen communities and their associated species richness. Central to fenland conservation and management, therefore, are maps of fen distributions and these could be derived from remote sensing. Here, interest is focussed on the fen class, with the remaining classes that make up the land cover mosaic of the region being of no interest.

V. Data and Methods

A Landsat 7 ETM+ image of the test site acquired on 19th June 2000 was provided by the UK's NERC Earth Observation Centre. The six non-thermal spectral wavebands (wavebands 1-5, and 7) with a spatial resolution of 30 m, and the computed normalised difference vegetation index (NDVI) which is widely used to characterise land cover, were selected for use in the analyses. These data were geometrically corrected to the UK national grid with an estimated RMS error of 0.12 pixel. Aerial photography acquired at a 1:10,000 scale in June 1999, was also geo-corrected to the Landsat ETM+ image. These photographs were used to delineate eight broadly defined land cover classes that dominated the land cover of the Broads; these being fen (the specific class of interest) together with saltmarsh, grazing marsh, agriculture, forest, urban, sand and water (which represent the outliers).

Although feature reduction may not be required for SVM based classification [54] its application can increase classification accuracy and methods designed for application with SVM classifications have been proposed [55,56]. As the Landsat ETM+ data provides a relatively low dimensional data set, the degree of redundancy in the data could be evaluated through assessment of simple measures of spectral separability. Thus, here a feature selection analysis was undertaken to reduce the size of the data set and remove unnecessary variables through assessment of transformed divergence statistic values [57]. This feature selection analysis was undertaken in earlier work evaluating binary classifications [11] and used a set of 150 pixels, 75 belonging to the class of interest (fens) and 75 belonging to the rest of the land cover types amalgamated to one class. For each pixel sampled, the digital number (DN) in each of the 6 ETM+ non-thermal spectral wavebands and the NDVI, derived from the data acquired in

ETM+ wavebands 3 and 4, were derived for the investigation. The transformed divergence statistic was calculated for every possible pair of features. Evaluation of the derived transformed divergence statistics indicated that the data acquired in Landsat ETM+ waveband 2 and the NDVI (Fig. 6) offered the highest average separability and only these data were used in the research reported here.

Although attention was focused on the SVDD a small pilot study was undertaken to evaluate a series of one-class classifiers to confirm its suitability for the application. This pilot study did not aim to rigorously evaluate all the classifiers but to provide a general assessment of their ability to classify the fen and illustrate the suitability of the SVDD, in particular, for this task. In total, eight one class classifiers, outlined briefly in section II, were applied to the data. For some classifiers a range of key parameter settings were applied, using values selected on the basis of previous research. Each classification analysis was trained with a sample of 100 pixels of the class of interest, fen, only. The accuracy of each classification was evaluated with a testing set of 250 pixels, comprising 125 pixels representing fen and 125 from the other classes.

Building on the pilot study attention focused on three sets of classification analyses performed: (i) using the SVDD trained on the specific class of interest only (i.e., a one-class classification); (ii) using the SVDD trained on the specific class of interest with a limited amount of data on the outliers and (iii) a conventional multi-class maximum likelihood (ML) classification to provide a benchmark for evaluative purposes. Since training set size can influence the accuracy of classifications, including those by ML and SVM-based classifiers [42,45], and the sample used in the pilot study was very small a larger training sample was acquired. From this sample, training sets for each classification could be formed. The main focus is on training sets comprising 150

pixels, a size that is small for inexpensive acquisition but just large enough to satisfy assumptions in the benchmark analysis. The size and composition of the training set used in each analysis, however, varied between the classifications and is described below. Some analyses also used smaller training sets derived by sampling from the 150 available cases to explore the effect of variations in training set size on the analysis.

A. One-class SVDD Classification

The training dataset for the SVDD classification comprised 150 pure pixels from areas of fenland only. Using this training dataset a number of classifications were conducted with a variety of kernels. The use of kernels makes it possible to map the data implicitly into a feature space and to train a linear machine in such space. The aim was to find the kernel function that made the calculations of hypersphere in the feature space efficient. The literature provides relatively little guidance on the selection of appropriate kernel functions and associated parameters [44] and hence the kernel, kernel parameters and value of *C* were selected from a series of analyses using a range of kernel types and five-fold cross validation. These were the polynomial kernel (with degrees 1 to 10 inclusive evaluated with an increment of 1); the gaussian radial basis function kernel (with free parameters of 1 to 10 inclusive evaluated with an increment of 1) and the exponential radial basis function kernel with degrees 1 to 10 inclusive evaluated with an increment of 1). In each case, a variety of settings for the penalty for misclassification term *C*, with a focus on accuracy assessed through a five fold cross-validation approach, were performed. In a one-class classification, *C* may be expressed as a threshold that describes the fraction of target cases rejected and the values used were 0.1, 0.01 and 0.001. On the basis of these analyses, the polynomial degree function kernel of free parameter value 2 and C=0.01 was selected for all subsequent SVDD analyses undertaken for fenland classification.

B. One-class SVDD Classification with Outliers

Classifications were conducted exactly as with the one-class SVDD except that the 150 training cases used were drawn from the class of interest as well as some drawn from the other seven land cover classes. To ensure resources were focussed mainly on the class of interest, the composition of the training dataset was imbalanced. Here, 75% of the training cases were drawn from the class of interest with the remaining 25% of data composed of randomly selected cases drawn from the other 7 classes present in region (i.e., the outliers). This composition was selected arbitrarily to ensure that effort was focused on the class of interest.

C. ML Classification

A standard multi-class ML classification was undertaken to provide a benchmark analysis. This analysis represented what may be considered to be a standard approach to thematic mapping from remotely sensed imagery. Here, the training dataset of 150 pixels was composed of cases of the 8 land cover classes present in the region in approximately equal proportions. The number of training pixels ranged from 15 to 20 for each class. Although the number of training samples for each class was relatively small the sizes are close to those suggested for a two dimensional data set (represented here by the data in ETM+ band 2 and the NDVI) by a widely used heuristic for use with this type of classification [58]. Moreover, the sample size was much larger than those that may actually be required for accurate classification [59]. The sample size may, however, not be optimal but it must be noted that this ML classification was used here to indicate the degree of accuracy achievable by a standard classifier using a small training set of the same size as that used with the SVDD classifications. A further ML classification, trained with 1200 pixels (150 per-class) was, however, undertaken to

illustrate the accuracy achievable by a conventional classifier when a very large training sample was available. The first ML classification was, therefore, based on a training set the same size as that used in the main classifications by the SVDD while in the second ML classification the same number of training cases for the fen as in the SVDD classifications was available and the sample size for each class greatly exceeded the minimum requirements for a ML analysis.

D. Accuracy Assessment

All of the classifications were evaluated using an independent testing data set and the accuracy with which the class of interest, fen, was classified was computed (%).The testing set comprised 125 cases of fenland, the class of interest, and 125 cases drawn from the other 7 classes. Since the same testing set was used throughout, the statistical significance of differences in classification accuracy was assessed using a M^{c}Nemar test [60]. With this test, the assessment is based on the evaluation of the derived *Z* statistic, and two classifications differ at the 95% level of confidence if $Z > |1.96|$. Throughout the evaluation, particular attention was focused on the accuracy with which the fen class was classified. As both commission and omission error can represent a problem in one-class mapping, the exact details depending on the specific application in-hand, both were calculated along with the related user's and producer's accuracy values from each classification confusion matrix derived [12].

V. Results and Discussion

Initial attention focused on the results of the pilot study. All of the one-class classifiers evaluated were able to yield a highly accurate classification (Table 1). In particular, four

classifiers produced classifications with an overall accuracy of up to ~94%. Critically, however, the results confirmed the potential of the SVDD for one class classification of fenland. With the SVDD the largest overall accuracy observed was 94.8% and there was typically relatively little difference in accuracy when viewed from the user's and producer's perspectives (Table 1). All further one-class classification analyses were, therefore, undertaken with the SVDD and, except where stated, all discussion below relates to analyses based upon the larger training set, comprising up to 150 pixels.

The overall accuracy with which the test set was classified by the ML classification, the benchmark analysis, was 68.8%, with a commission error of 10.0% and omission error of 28.0% for fens (Table 2a). Thus, the user's and producer's accuracy for fenland classification were 90.0% and 72.0% respectively. The large degree of omission error in the ML classification is manifest in its output, with the extent of the fenland class substantially under-estimated (Figs 5 and 7a). The accuracy of this ML classification is substantially below widely used map accuracy targets and the map derived from it may, therefore, be viewed as being inadequate for use in fenland monitoring. For example, the widely used map accuracy targets discussed by Anderson *et al*. [61] suggest a minimum producer's accuracy of 85.0%, a value 13.0% above that derived. Furthermore, the accuracy of the ML classification did not change markedly when the larger training set, comprising 1200 pixels (150 per-class) was used (Table 2b). Although post-classification fieldwork could be undertaken to mitigate for the high errors of omission and commission this requires extra effort and resources, and indicates a lack of faith in the map derived by ML classification. This also negates a fundamental aim of using remote sensing for land cover mapping, which is to cover large areas with acceptable accuracy without the need for excessive post-classification fieldwork. .

The one-class SVDD approach yielded a classification (Fig. 7b) that was significantly more accurate than that from the benchmark ML analysis (comparison against Table 2a, Z=5.01, significant at 99.9% level; Table 3). This SVDD classification had an overall accuracy of 95.6%, with commission errors of 2.5% and omission errors of 6.4% for fenland (Table 3). The accuracy of the fenland classification was, therefore, 97.5% and 93.6% from the user's and producer's perspectives respectively. The accuracy of the SVDD classification was, therefore, above the widely used target accuracy of 85%. With the one-class SVDD classification, the class of interest, fen, was targeted throughout the classification procedure and a high accuracy derived. The results highlights that very accurate land cover maps of the class of interest can be produced with effort and resources directed on the class of interest.

The provision of information about outliers had no significant impact on the analysis with the output of the resulting classification (Fig. 7c) similar to that derived from the SVDD trained upon only the fenland class (Fig. 7b). Indeed, the exact same set of class allocations for the testing set were obtained when the classifier was trained on just the class of interest or with a training set containing outliers. Thus, in this research, the addition of training data on classes other than that of interest had no impact on classification accuracy.

The results of the classifications highlight a potential to focus resources, especially those required for training data collection, on the class of interest. This could lead to substantial savings in classification production, an advantageous feature for often resource-limited conservation applications. To further reduce resource effort, analyses were conducted to explore an important feature of SVM based classifications, namely their capacity to use small training data sets for accurate classification [62]. The one-

class classifications were repeated 9 times with a set of randomly selected training cases removed from the training set on each occasion. A series of classifications trained on between 5 and 150 cases of the class of interest were, therefore, undertaken. From these classifications, it was evident that variation in training dataset size had no significant effect (at the 95% confidence level) on the overall classification accuracy (Table 4) except when fewer than 10 training samples were used and the overall classification accuracy dropped to 86.0%. Thus, decreasing the training set size did not result in the loss of critical information to the description of the class of interest until the training set contained a very small sample (<10 training cases). Thus, the classifier could find the support vectors necessary to fit a suitable hypersphere from a small sample. Moreover, training sets considerably smaller than those derived from the application of conventional heuristics for training set definition, which suggest, for example, that the training set should contain in the order of 10-30 cases per-class per-waveband used [57,58], could be used for accurate classification. The use of such heuristics for a conventional multi-class classification would result in a large training set in which the cases of the class of interest could be a very small component. For example, for the mapping scenario considered in this paper involving two spectral features and 8 classes, the required training set would comprise in the order of 160-480 pixels of which only a small component, $\frac{1}{8}^{th}$, were drawn from the fenland class. From Table 4 it is apparent that very accurate classification of fen can be derived from the one-class SVDD with a small training set of fenland pixels. Indeed, if attention was on the user's accuracy of fenland classification a very small sample could be used for the production of an accurate classification (Table 4). Critically, the training set for the SVDD classification is, however, not just small but also focused entirely on the class of interest.

The findings of this research demonstrate the potential of the SVDD for the mapping of a specific class of interest from remotely sensed data. The approach may also be used for any class. For example, repeating the analyses with a focus on the other major class of high conservation value in the study area, saltmarsh, yielded similar results. Specifically, a SVDD (C=0.01, exponential kernel with a free parameter of 10 trained on 150 cases of saltmarsh) yielded a classification in which saltmarsh was classified with an accuracy of 93.6% and 98.3% from the map user's and producer's perspectives respectively [63]. Further research on the approach is, however, required to fully evaluate its potential. For example, the utility of the approach will clearly vary as a function of the separability of the class of interest from other classes. Nonetheless, the adoption of the SVDD may often be more appropriate than a conventional multi-class classifier. Moreover, the ability of the SVDD classifier to operate with small training data sets means that a competitive and efficient approach to the classification can be adopted. This is likely to be attractive for applications with limited resources. It is likely that the analysis can be further refined by training on those pixels (support vectors) that are used in the calculations to describe the hypersphere in a similar manner to intelligent training of SVM based classification [62].

VI. Summary and Conclusions

Remote sensing has considerable potential as a source of land cover information and in particular, for monitoring land cover and its dynamics at various scales. Commonly, supervised image classification is the basis of land cover and land cover change assessments. This type of analysis aims to convert the remotely sensed image into a thematic map that depicts the spatial distribution of the various land cover classes found

within the region. In standard statistical supervised image classifications, the aim is often to maximize the overall probability that a case is allocated correctly to a class. This requires that each class within the area to be mapped is included in the analysis to satisfy the implicit assumption made in standard classification analyses of an exhaustively defined set of classes. Although this is a logical basis for some classifications it is not always appropriate, especially as it treats all classes, including those of no interest, equally rather than focussing on the class that is of actual interest. To ensure efficient use of resources and time, an approach that concentrates on and maps the specific class of interest may be advantageous.

A set of one-class classifiers were evaluated and the results highlighted their potential for accurate classification of fenland. In particular, this paper demonstrated that the use of a classification approach based on SVM such as the SVDD can be used to accurately map a single class of interest from remotely sensed data. This is valuable for challenging applications such as the implementation of European Union directives focused on key habitats but for which resources are limited and awareness of new techniques low [64]. A key attraction of the SVDD approach is that it provides a very simple to use supervised classification analysis that requires only the training data for the class of interest. Moreover, the accuracy of the SVDD classification was considerably higher than that derived from a conventional multi-class parametric classification. A standard ML classification yielded a classification in which the accuracy of fenland classification was 90.0% and 72.0% from the user's and producer's perspectives respectively while the corresponding values for the SVDD classification were 97.5% and 93.6%. The SVDD was, therefore, able to provide very accurate information on the class of interest and only required training data on that class. Moreover, the high level of accuracy was obtained using a small training data set,

considerably smaller than would be suggested by conventional approaches. This research has demonstrated the viability of one-class classifiers and the SVDD in particular where only information about the class of interest is needed and its potential to reduce training set size in this particular application. The use of a true one-class classifier is a novel approach to remote sensing for land cover classification and its full potential should be explored further.

Acknowledgements

We are grateful to Dr. D Tax for the development and free use of the SVDD; the Broads Authority for providing habitats information and access to aerial photography; the NERC Earth Observation Data Centre for providing the Landsat 7 ETM+ imagery used and the School of Earth Sciences and Geography at Kingston University for the provision of a Ph.D. scholarship to CS-H. Much of the work reported was undertaken while GMF was at the University of Southampton and DSB at Bournemouth University. We are also grateful to the three referees for their comments on the article.

References

1. J. Cihlar, "Land cover mapping of large areas from satellites: status and research priorities," *International Journal of Remote Sensing*, vol. 21, pp. 1093-1114, 2000.

2. D. S. Boyd and G. M. Foody, "Changing Land Cover," in: *Global Environmental Issues*, F. Harris, editor, Chichester: Wiley, pp. 65-94, 2004.

3. B. Jeon and D. A. Landgrebe, “Partially supervised classification using weighted unsupervised clustering,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, pp. 1073-1079, 1999.

4. D. Fernandez-Prieto, “An iterative approach to partially supervised classification problems”, *International Journal of Remote Sensing*, vol. 23, pp. 3887-3892, 2002.

5. Z. Liu, W. Shi and Q. Qin, “Partially supervised classification – based on weighted unlabeled samples support vector machine”, *Lecture Notes in Computer Science*, vol. 3584, pp. 118-129, 2005.

6 X. Song, G. Fan and M. Rao, “Automated CRP mapping using non-parametric machine learning approaches,”, *IEEE Transactions on Geoscience and Remote Sensing*, vol 43, pp. 888-897, 2005.

7 X. M. Song, G. Cherian and G. L. Fang, “A *v*-insenstive SVM approach for compliance monitoring of the conservation reserve programme, IEEE Geoscience and Remote Sensing Letters, vol 2, pp. 99-103, 2005.

8. J. D. Hill, N. D., Strommen, C. M., Sakamoto and S. K. Leduc, “LACIE - application of meteorology for United-States and foreign wheat assessment,” *Journal of Applied Meteorology*, vol. 19, pp. 22-34, 1980.

9. E. W. Ramsey, G. A. Nelson, S. K. Sapkota, E. B. Seegar, and K Martella, K. D. “Mapping Chinese tallow with color-infrared photography,” *Photogrammetric Engineering and Remote Sensing*, vol. 68, pp. 251-255, 2002.

10. E. Underwood, S. Ustin, and D. DiPietro, “Mapping nonnative plants using hyperspectral imagery,” *Remote Sensing of Environment*, vol. 86, pp. 150-161, 2003.

11. D. S. Boyd, C. Sanchez-Hernandez and G. M Foody, “Mapping a specific class for priority habitats monitoring from satellite sensor data,” *International Journal of Remote Sensing*, vol. 27, pp. 2631-2644, 2006.

12. G. M. Foody, “Status of land cover classification accuracy assessment,” *Remote Sensing of Environment*, vol. 80, pp. 185-201, 2002.

13. G. M. Foody, “Hard and soft classifications by a neural network with a non-exhaustively defined set of classes,” *International Journal of Remote Sensing*, vol. 23, pp. 3853-3864, 2002.

14. R. M. Lark, “Components of accuracy of maps with special reference to discriminant analysis of remote sensor data,” *International Journal of Remote Sensing*, vol. 16, pp. 1461-1480, 1995.

15. G. M. Foody, P. M. Atkinson, P. Gething, N. A. Ravenhill, and C. K. Kelly, “Identification of specific tree species in ancient semi-natural woodland from digital aerial sensor imagery,” *Ecological Applications*, vol. 15, pp. 1233-1244, 2005.

16. G. M. Foody, A. Mathur, C. Sanchez-Hernandez and D. S. Boyd, “Training set size requirements for the classification of a specific class,” *Remote Sensing of Environment*, vol. 104, pp. 1-14, 2006.

17. P. Mantero, G. Moser, and S. B. Serpico, "Partially supervised classification of remote sensing images through SVM-based probability density estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 559-570, 2005.

18. M. Moya, M. Koch and L Hostetler, "One-class classifier networks for target recognition applications," *Proceedings of World Congress on Neural Networks*, Portland, OR: International Neural Network Society, pp. 797-801, 1993.

19. G. Ritter and M Gallegos, "Outliers in statistical pattern recognition and an application to automatic chromosome classification," *Pattern Recognition Letters*, vol. 18, pp. 525–539, 1997.

20. C. Bishop, "Novelty detection and neural network validation," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 14, pp. 217–222, 1994.

21. N. Japkowicz, C. Myers and M. Gluck, "A novelty detection approach to classification," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, Quebec Canada**,** National Research Council of Canada, NRC 40197, pp. 518-523, 1995.

22. L. M. Manevitz and M. Yousef, "One-class SVMs for document classification," *Journal of Machine Learning Research*, vol. 2, pp. 139-154, 2002.

23. D. M. J. Tax and R.P. W. Duin, "Uniform object generation for optimizing one-class classifiers," *Journal of Machine Learning Research*, vol. 2, pp. 155-173, 2002.

24. C. Lai, D. M. J. Tax, R. P. W Duin, E. Pekalska and P. Paclik, "On combining one-class classifiers for image database retrieval," *Lecture Notes in Computer Science*, vol. 2364, pp. 212-221, 2002.

25. Q. Guo, M. Kelly and C. H. Graham, "Support vector machines for predicting distribution of Sudden Oak Death in California," *Ecological Modelling*, vol. 182, pp. 75-90, 2005.

26. N. J. Pizzi, R. A. Vivanco, and R. L Somorjai, "EvIdent: a functional magnetic resonance image analysis system," *Artificial Intelligent in Medicine*, vol. 21, 263–269, 2001.

27. Y. Yang, T. Pierce, and J. Carbonell, "A study on retrospective and on-line event detection," *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 28-36. Melbourne, Australia, 1998.

28. G. Fumera, F. Roli, G. Giacinto, "Reject option with multiple thresholds," *Pattern Recognition*, vol. 33, pp. 2099–2101, 2000.

29. S. Roberts and L Tarassenko, "A probabilistic resource allocating network for novelty detection," *Neural Computation*, vol. 6, 270-284, 1994.

30. D. Y. Yeung and C. Chow, "Parzen-window network intrusion detectors," *Proceedings of the Sixteenth International Conference on Pattern Recognition (ICPR)*, Quebec City, Canada, 11-15 August 2002, vol.4, pp.385-388, 2002.

31 B. Schölkopf, J. C Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol.13, pp. 1443–1471, 2001.

32. D. Tax and R. Duin, "Data domain description using support vectors," P*roceedings of the European Symposium on Artificial Neural Networks*, Brussels: D. Facto, pp. 251–256, 1999.

33. D. M. J. Tax, "*One-class classification,"* Unpublished Ph.D. thesis, Delft University of Technology, Netherlands, 2001.

34. V. Vapnik, "*The Nature of Statistical Learning Theory,"* New York: Springer-Verlag, 1995.

35. C. Bishop, "*Neural Networks for Pattern Recognition*," Oxford, Oxford University Press, 1995.

36. T. Kohonen, "*Self-organizing Maps*," Heidelberg, Springer-Verlag, 1995.

37. N., Japkowicz, C. Myers and M. Gluck, "A novelty detection approach to classification," In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 518–523, 1995.

38. J. Hertz, A. Krogh and R. Palmer, "*Introduction to the Theory of Neural Computation*," Redwood City, Addison Wesley, 1991.

39. P. Baldi and K. Hornik, "Neural networks and principal component analysis: learning from examples without local minima," *Neural Networks*, vol 2 pp.53–58, 1989.

40. A. Ypma and R. Duin, "Support objects for domain approximation," In *ICANN'98*, Skovde (Sweden), 1998.

41. R. Duda and P. Hart. "*Pattern Classification and Scene Analysis*," New York, John Wiley, 1973.

42. C Huang, L. S. Davis and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of Remote Sensing*, vol. 23, pp. 725-749, 2002.

43. G. Zhu and D. G. Blumberg, "Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel," *Remote Sensing of Environment*, vol. 80, pp. 233-240, 2002.

44. M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *International Journal of Remote Sensing*, vol. 26, pp. 1007-1011, 2005.

45. G. M. Foody and A. Mathur, "A relative evaluation of multi-class image classification by support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 1335-1343, 2004.

46. D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 2004.

47. V. Cherkassky and F. Mulier, "*Learning from Data - Concepts, Theory, and Methods*," USA: John Wiley and Sons, 1998.

48. B. Schölkopf, R. Williamson, A. Smola and J. Shawe-Taylor, "SV estimation of a distribution's support," In *13th Neural Information Processing Systems Meeting Proceedings*, Denver, USA, 1999.

49. V. Vapnik, "An overview of statistical learning theory," *IEEE Transactions of Neural Networks*, vol. 10, pp. 988 – 999, 1999.

50. B. Boser, I.Guyon, and V. N Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of 5th Annual Workshop on Computer Learning Theory,* Pittsburgh, PA: ACM, pp. 144-152), 1992.

51. M. A. Aizerman, E. M. Braverman, and L. I Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 821-837, 1964.

52. R. K. Turner, S. Georgoiu, R. Brouwer, I. J. Bateman and I. J. Langford, "Towards an integrated environmental assessment for wertland and catchment management," *The Geographical Journal*, vol. 169, no 2, pp. 99-116, 2003.

53. Broads Authority "*Broads Plan 2004*," Witley Press, Norwich, 2004.

54. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 1778-1790, 2004.

55. L. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines", *Machine Learning*, vol. 46, pp. 389-422, 2002.

56. Y. Bazi and F. Melgani, "Towards an optimal SVM classification system for hypersepctral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, submitted.

57. P. M. Mather, "*Computer Processing of Remotely-Sensed Images*," 3rd edition, Chichester, Wiley, 2004.

58. J. Piper, "Variability and bias in experimentally measured classifier error rates," *Pattern Recognition Letters*, vol. 13, pp. 685-692, 1992.

59. T. G. van Niel, T. R., McVicar and B. Datt, "On the relationship between training sample size and data dimensionality of broadband multi-temporal classification," *Remote Sensing of Environment*, vol. 98, pp. 468-480, 2005.

60. G. M. Foody, "Thematic map comparison: evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering and Remote Sensing*, vol. 70, pp. 627-633, 2004.

61. J. R. Anderson, E. E. Hardy, J. T. Roach. and R. E. Witmer, "*A Land Use and Land Cover Classification System for Use with Remote Sensor Data*," U. S. Geological Survey Professional Paper 964, 28pp, 1976.

62. G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification," *Remote Sensing of Environment*, vol. 93, pp. 107-117, 2004.

63. C. Sanchez-Hernandez. "*Land cover mapping of one specific protected habitat under the requirements of the European Union Habitats Directive*," Unpublished Ph.D. thesis, Kingston University, UK, 2006.

64. A. Kobler, S. Dzeroski and I. Keramitoglu, "Habitat mapping using machine learning-extended kernel-based reclassification of an Ikonos satellite image," *Ecological Modelling*, vol. 191, pp. 83-95, 2006.

Figure captions

Figure 1. A hyperplane separating two classes.

Figure 2. Support vectors and the optimal separating hyperplane. The support vectors lies on the hyperplanes H_1 and H_2 and are ringed for emphasis.

Figure 3. The hypersphere containing the target data. The shaded objects on the edge of the sphere are the support vectors.

Figure 4. Magnitude of the two Lagrange multipliers for cases inside, on the boundary and outside the hypersphere (Based on [33].

Figure 5. Location of the study site with a generalised representation of fenland distribution.

Figure 6. The two features used in the classification. (a) Landsat ETM+ waveband 2 and (b) NDVI.

Figure 7. Classification outputs for the test site. (a) maximum likelihood, (b) SVDD and (c) SVDD with information on outliers.

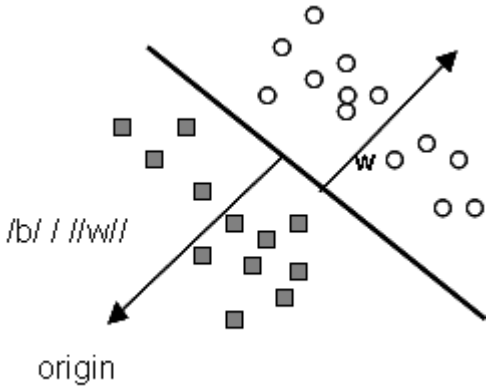


Figure 1.

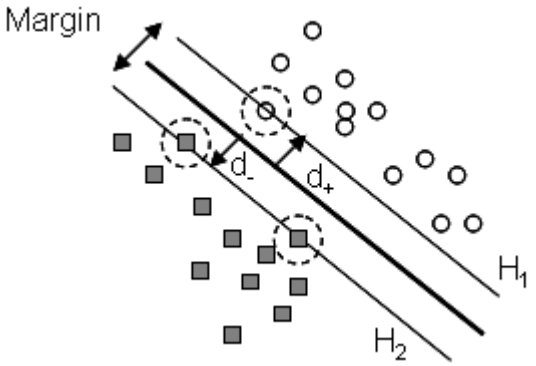


Figure 2.

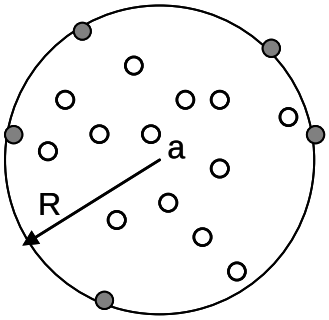


Figure 3.

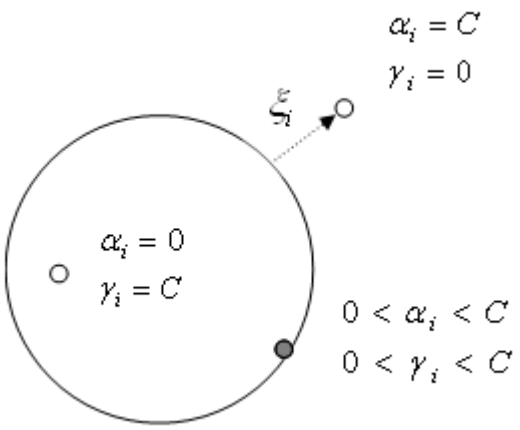


Figure 4

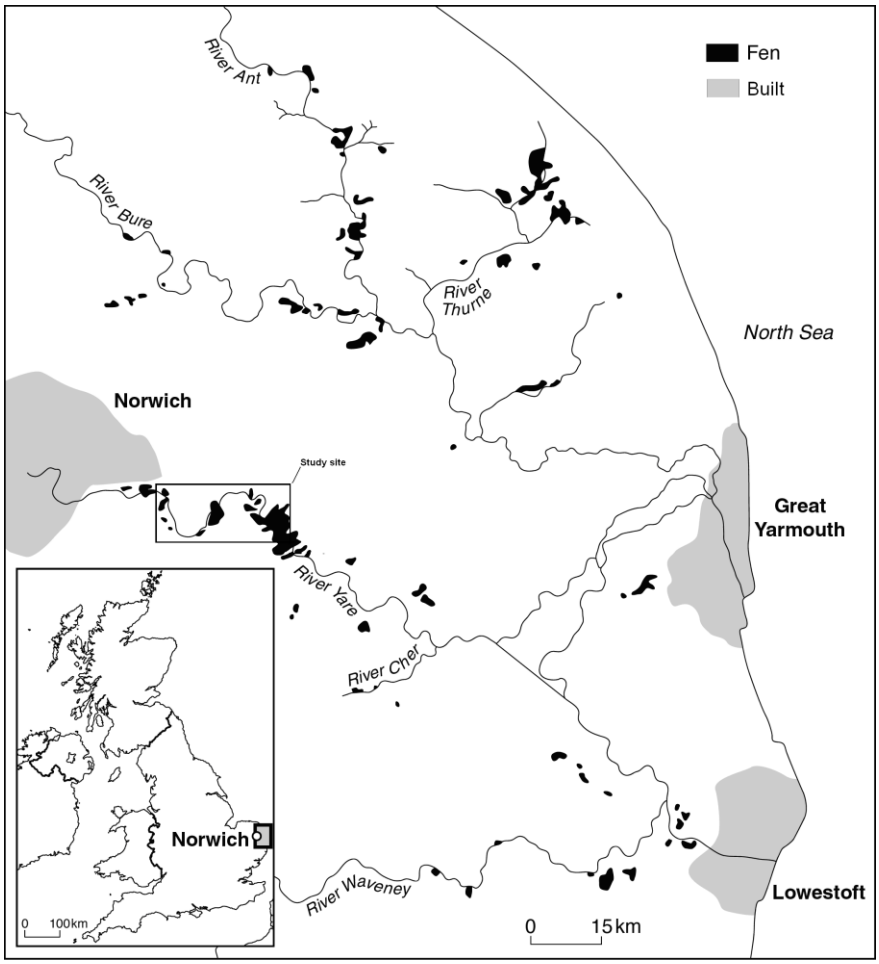
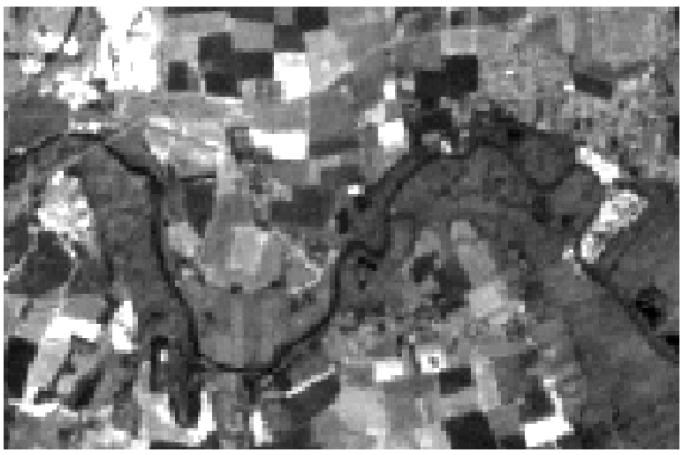

Fen
Built
River Ant
River Bure
River Thurne
North Sea
Norwich
Study site
Great Yarmouth
River Yare
River Chet
Norwich
River Waveney
Lowestoft
0 100km
0 15km

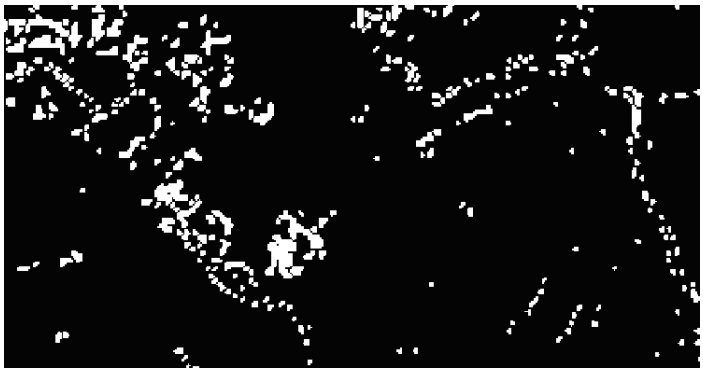
Figure 5



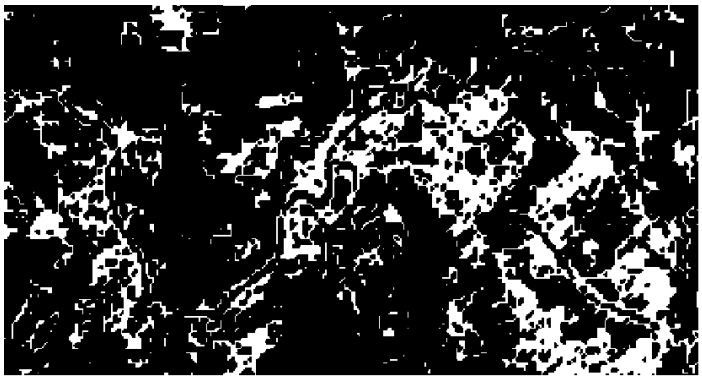
(a)

(b)

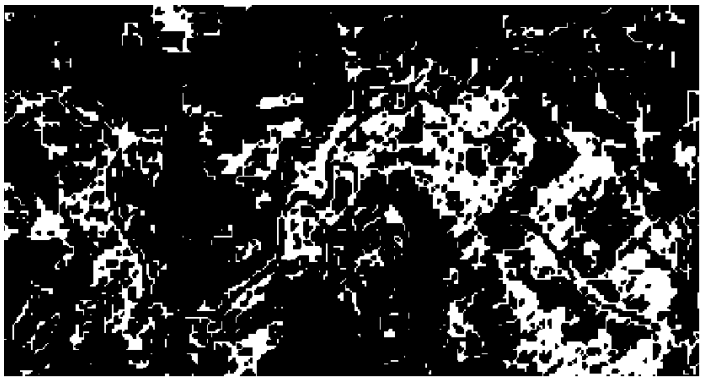
Figure 6



(a)



(b)



(c)

Figure 7

			Classification accuracy (%)			
Classifier (parameter)		Accuracy measure	C=0.1	C=0.01	C=0.001	C=0.0001
Gaussian		Overall	91.60	92.00	87.20	87.20
		Fens producer's	84.80	98.40	99.20	99.20
		Fen user's	98.15	87.23	80.00	80.00
Mixture of Gaussians	(2 clusters)	Overall	91.51	95.30	93.13	89.78
		Fen producer's	84.35	94.51	95.90	98.09
		Fen user's	99.07	96.75	91.67	84.93
	(5 clusters)	Overall	92.00	94.80	94.00	94.82
		Fen producer's	84.80	92.80	92.00	95.62
		Fen user's	99.07	96.67	95.83	94.49
	(8 clusters)	Overall	88.42	93.37	92.40	92.22
		Fen producer's	78.95	91.40	90.43	88.69
		Fen user's	99.01	96.67	95.83	97.41
	(10 clusters)	Overall	91.60	94.40	93.60	94.80
		Fen producer's	84.80	91.20	91.20	95.20
		Fen user's	98.15	97.44	95.80	94.44
Parzen		Overall	91.60	94.80	94.40	94.40
		Fen producer's	84.80	94.40	94.40	94.40
		Fen user's	98.15	95.16	94.40	94.40
k-means	(2 prototypes)	Overall	91.60	92.00	90.40	90.40
		Fen producer's	84.80	98.40	99.20	99.20
		Fen user's	98.15	87.23	84.35	84.35
	(5 prototypes)	Overall	91.60	92.80	87.20	90.40
		Fen producer's	84.80	96.80	99.20	99.20
		Fen user's	98.15	89.63	80.00	84.35
	(8 prototypes)	Overall	91.60	91.60	82.80	90.00
		Fen producer's	84.80	98.40	99.20	99.20
		Fen user's	98.15	86.62	74.70	83.78
	(10 prototypes)	Overall	88.80	93.20	93.60	94.00
		Fen producer's	78.40	96.80	92.00	89.60
		Fen user's	98.99	90.30	95.04	98.25
SOM	(2x2 neurons)	Overall	89.20	90.40	90.00	92.00
		Fen producer's	84.00	99.20	99.20	96.80
		Fen user's	93.75	84.35	83.78	88.32
	(5x5 neurons)	Overall	88.80	94.40	94.00	94.00
		Fen producer's	79.20	90.40	90.40	90.40
		Fen user's	98.02	98.26	97.41	97.41
	(8x8 neurons)	Overall	88.00	92.00	94.40	94.80
		Fen producer's	76.80	85.60	91.20	92.00
		Fen user's	98.97	98.17	97.44	97.46
	(10x10 neurons)	Overall	87.60	92.80	93.20	94.00
		Fen producer's	76.00	87.20	88.00	89.60
		Fen user's	98.96	98.20	98.21	98.25
PCA		Overall	80.00	85.60	82.80	82.80
		Fen producer's	79.20	96.00	96.80	96.80
		Fen user's	80.49	79.47	75.63	75.63
k-centre	(2 prototypes)	Overall	90.80	92.40	90.00	92.00
		Fen producer's	84.00	96.80	99.20	96.80
		Fen user's	97.22	88.97	83.78	88.32
	(5 prototypes)	Overall	87.20	92.80	93.60	90.40
		Fen producer's	79.20	93.60	92.80	89.60
		Fen user's	94.29	92.13	94.31	91.06
	(8 prototypes)	Overall	87.60	92.80	92.80	92.80
		Fen producer's	81.60	91.20	91.20	91.20
		Fen user's	92.73	94.21	94.21	94.21
	(10 prototypes)	Overall	87.60	92.80	93.60	93.20
		Fen producer's	81.60	91.20	90.40	89.60
		Fen user's	92.73	94.21	96.58	96.55
SVDD Polynomial Kernel	(degree 2)	Overall	92.80	93.20	94.80	94.80
		Fen producer's	87.20	96.80	96.80	96.80
		Fen user's	98.20	90.30	93.08	93.08
	(degree 5)	Overall	93.20	90.80	90.80	90.80
		Fen producer's	96.80	96.80	96.80	96.80
		Fen user's	90.30	86.43	86.43	86.43
	(degree 8)	Overall	93.20	88.00	88.00	88.00
		Fen producer's	96.80	97.60	97.60	97.60
		Fen user's	90.30	81.88	81.88	81.88
	(degree 10)	Overall	93.20	88.00	88.00	88.00
		Fen producer's	96.80	97.60	97.60	97.60
		Fen user's	90.30	81.88	81.88	81.88

Table 1. Summary of the one-class classification results.

Class	FE	SM	A	FO	G	S	U	W	Σ	Producer's accuracy (%)
Fen (FE)	90	18	2	15	0	0	0	0	125	72.00
Saltmarsh (SM)	3	14	0	0	0	0	3	0	20	70.00
Agriculture (A)	0	0	0	1	13	3	3	0	20	0.00
Forest (FO)	7	1	0	5	1	0	0	0	14	35.71
Grassland (G)	0	0	0	1	20	0	0	0	21	95.23
Sand (S)	0	0	0	0	0	12	3	0	15	80.00
Urban (U)	0	3	1	0	0	0	14	0	18	77.77
Water (W)	0	0	0	0	0	0	0	17	17	100.00
Σ	100	36	3	22	34	15	23	17	250	
User's accuracy (%)	90.00	38.88	0.00	22.72	58.82	80.00	60.86	100.00		

(a)

Class	FE	SM	A	FO	G	S	U	W	Σ	Producer's accuracy (%)
Fen (FE)	97	14	0	14	0	0	0	0	125	77.60
Saltmarsh (SM)	3	14	0	0	0	0	3	0	20	70.00
Agriculture (A)	0	0	1	8	5	3	3	0	20	0.00
Forest (FO)	11	0	0	3	0	0	0	0	14	21.43
Grassland (G)	0	0	2	1	18	0	0	0	21	85.71
Sand (S)	0	0	0	0	0	15	0	0	15	100.00
Urban (U)	0	3	1	0	2	1	14	0	18	77.77
Water (W)	0	0	0	0	0	0	0	17	17	100.00
Σ	111	28	4	26	25	19	20	17	250	
User's accuracy (%)	87.39	50.00	25.00	11.54	72.00	78.95	70.00	100.00		

(b)

Table 2: Confusion matrices from the maximum likelihood classification. (a) for the classification trained with a total of 150 training cases with an overall accuracy of 68.8% and (b) for the classification trained with 1200 training cases (150 of each class) with an overall accuracy of 71.6%. The rows and columns of the matrices represent the actual and predicted class of membership respectively. .

	Fen	Others	**Σ**	Producer's accuracy (%)
Fen	117	8	125	93.6
Others	3	122	125	97.6
Σ	120	130	250	
User's accuracy (%)	97.5	93.8		

Table 3: Confusion matrix for the one-class SVDD classification. The rows and columns of the matrix represent the actual and predicted class of membership respectively. The overall classification accuracy was 95.6%.

Number of training cases	Overall accuracy (%)	User's accuracy (%)	Producer's accuracy (%)
5	86.0	97.8	73.6
10	90.8	98.1	83.2
15	90.8	98.1	83.2
20	92.8	98.2	87.2
25	94.8	97.4	92.0
50	94.8	97.4	92.0
75	93.2	90.3	96.8
100	93.2	90.3	96.8
125	93.2	90.3	96.8
150	95.6	97.5	93.6

Table 4: Effect of variation in training sample size on the accuracy (%) of the SVDD classifications.

Author photographs and biographies:

Carolina Sanchez-Hernandez

Carolina Sanchez Hernandez was born in Plasencia, Spain in 1971. She graduated in geography and history in 1994 at Salamanca University, Spain. She then obtained her M.Sc. in environmental science in 2000 at Birkbeck College, University of London, UK, and has recently completed a Ph.D. on land cover mapping of specific habitats using remote sensing at Kingston University, UK. She is currently working as a Research Scientist for the Ordnance Survey, GB, investigating change detection techniques.

Doreen S. Boyd

Doreen Boyd received the B.Sc. degree in geography from University of Wales, UK, in 1992 and completed a Ph.D. degree at the University of Southampton, UK, in 1996. She is currently Associate Professor in the School of Geography, University of Nottingham, UK, having held Lectureships at Manchester, Kingston and Bournemouth Universities. Between 2004 and 2006, she held the position of Senior Research Leader in Research and Innovation at the Ordnance Survey, GB. Her main research interests are in the remote sensing of terrestrial ecosystems and the monitoring of environmental change. Doreen serves on the editorial board of *Visual Geosciences* and the advisory board of the *Journal of Maps*. She is co-founder and co-chair of the laser scanning and LiDAR special interest group of the Remote Sensing and Photogrammetry Society.

Giles M. Foody

Giles Foody (M'01) completed B.Sc. and Ph.D. degrees in geography at the University of Sheffield, UK, in 1983 and 1986 respectively. He is currently Professor of Geographical Information Science at the University of Nottingham, UK. His main research interests are in the remote sensing of land cover and biogeography. Giles currently serves as co-editor-in-chief of the *International Journal of Remote Sensing*, as a co-ordinating editor of *Landscape Ecology* and as associate editor of *Ecological Informatics* as well as a member of the editorial boards of *Remote Sensing of Environment*, *Geocarto International* and the *Journal of Applied Remote Sensing*.