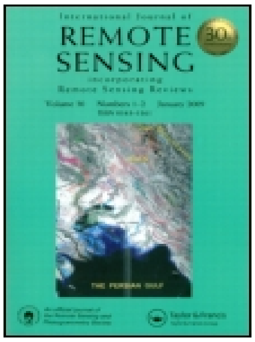

Taylor & Francis
Taylor & Francis Group

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

A novel multi-parameter support vector machine for image classification

Ce Zhang, Tiejun Wang, Peter M. Atkinson, Xin Pan & Huapeng Li

To cite this article: Ce Zhang, Tiejun Wang, Peter M. Atkinson, Xin Pan & Huapeng Li (2015) A novel multi-parameter support vector machine for image classification, International Journal of Remote Sensing, 36:7, 1890-1906, DOI: 10.1080/01431161.2015.1029096

To link to this article: https://doi.org/10.1080/01431161.2015.1029096

Published online: 07 Apr 2015.

International Journal of Remote Sensing, 2015
Vol. 36, No. 7, 1890–1906, http://dx.doi.org/10.1080/01431161.2015.1029096

A novel multi-parameter support vector machine for image classification

Ce Zhang[a,b], Tiejun Wang[b]*, Peter M. Atkinson[a], Xin Pan[c,d], and Huapeng Li[c]

[a]*Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK;* [b]*Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede 7500 AE, The Netherlands;* [c]*Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China;* [d]*School of Electrical and Information Technology, Changchun Institute of Technology, Changchun 130021, China*

(*Received 1 August 2014; accepted 24 January 2015*)

The support vector machine (SVM) classification algorithm has received increasing attention in recent years in remote sensing for land-cover classification. However, it is well known that the performance of the SVM is sensitive to the choice of parameter settings. The traditional single optimized parameter SVM (SOP-SVM) attempts to identify globally optimized parameters for multi-class land-cover classification. In this article, a novel multi-parameter SVM (MP-SVM) algorithm is proposed for image classification. It divides the training set into several subsets, which are subsequently combined. Based on these combinations, sub-classifiers are constructed using their own optimum parameters, providing votes for each pixel with which to construct the final output. The SOP-SVM and MP-SVM were tested on three pilot study sites with very high, high, and low levels of landscape complexity within the Sanjiang Plain – a typical inland wetland and freshwater ecosystem in northeast China. A high overall accuracy of 82.19% with kappa coefficient (κ) of 0.80 was achieved by the MP-SVM in the very high-complexity landscape, statistically significantly different (z-value = 3.77) from the overall accuracy of 72.50% and κ of 0.69 produced by the traditional SOP-SVM. Besides, for the moderate-complexity landscape a significant increase in accuracy was achieved (z-value = 2.44), with overall accuracy of 84.03% and κ of 0.80 compared with an overall accuracy 76.05% and κ of 0.71 for the SOP-SVM. However, for the low-complexity landscape the MP-SVM was not significantly different from the SOP-SVM (z-value = 0.80). Thus, the results suggest that the MP-SVM method is promising for application to very high and high levels of landscape complexity, differentiating complex land-cover classes that are spectrally mixed, such as marsh, bare land, and meadow.

1. Introduction

Development of robust and accurate computer-aided image classification algorithms is a key topic in geoscience. A support vector machine (SVM) is a supervised binary classifier that works on the basis of statistical learning theory (Vapnik 1995, 1998; Mountrakis, Im, and Ogole 2011). It has been applied widely in computer vision, pattern recognition, image classification, information retrieval, and data mining. Numerous studies have shown that the SVM can produce more accurate classification results than other parametric, as well as non-parametric, classifiers such as the maximum likelihood, neural network, and decision tree classifiers (Huang, Davis, and Townshend 2002; Foody and

*Corresponding author. Email: t.wang@utwente.nl

Mathur 2004; Camps-Valls and Bruzzone 2005; Waske and Benediktsson 2007; Watanachaturaporn, Arora, and Varshney 2008; Shao and Lunetta 2012). The SVM is arguably the most popular machine learning algorithm of the last decade (Su 2009). Training a SVM often requires appropriate kernel function and parameter settings to ensure model accuracy and generalization capability (Friedrichs and Igel 2005). In terms of SVM kernel choice, the Gaussian radial basis function (RBF) has been demonstrated to be an appropriate kernel in remote-sensing image classification in comparison with the polynomial kernel or linear kernel, due to the capacity of the former to solve complex non-linear classification problems and recognize complex spatial patterns without fitting a fixed parametric model (Zhu and Blumberg 2002). Kavzoglu and Colkesen (2009) compared comprehensively the Gaussian kernel SVM and polynomial kernel SVM and concluded that the former generally achieved greater accuracy and robustness with remotely sensed data. However, one of the main disadvantages of the Gaussian kernel SVM classifier is its sensitivity to the choice of Gaussian kernel width (γ) and regularization parameters (C), which are data dependent and currently lack unified standards (Ceamanos et al. 2010). Different parameter settings can give rise to variation in the classification results, thereby influencing classification accuracy.

In order to optimize parameters γ and C, several practical methods have been proposed, including exhaustive grid search using fivefold cross-validation (Min and Lee 2005), genetic-based (Wu et al. 2007), and particle swarm optimization (PSO) approaches (Bazi and Melgani 2007). These parameter optimization methods attempt to identify a single set of parameters (γ and C) for the SVM model, with successful application especially in the non-spatial domain, with crisp description and characterization.

Most natural phenomena cannot be delineated by crisp boundaries, but are bounded by fuzzy transitions or transition zones (Cheng and Molenaar 1999). Such continuous characteristics bring enormous challenge to conventional 'hard' classification methods, which are generally performed under the assumption that each pixel represents a single land-cover class with no confusion or uncertainty (Ibrahim, Arora, and Ghosh 2005). The SVM classifier also addresses the difficulties of 'hard' class allocation along the presumed 'boundaries' among different land-cover categories. The single set of optimized parameters (γ and C) adopted by current SVMs to identify such boundaries may be unrepresentative of the real world. Moreover, boundaries between neighbouring classes of land covers often overlap. Thus, an integrated decision for boundary partition should be based on fuzzy membership association or consensus theory (Pan et al. 2010).

Since dynamic regions are often difficult to assess, challenges faced within remote-sensing classification involve insufficiency and inequality for different types of sample datasets, which also affect the results of parameters γ and C. A single optimized parameter SVM (SOP-SVM), therefore, does not satisfy certain application requirements due to the potential diversity and complexity of real landscapes.

To integrate multiple optimized SVM parameters (γ and C), rather than focusing on single parameter optimization, classifier ensembles provide an alternative methodology of machine learning and provide new ideas to solve compound optimization problems. By combining diverse individual classifiers, the resulting classifiers have been demonstrated to be superior to individual classifiers (Friedl, Brodley, and Strahler 1999; Ghimire, Rogan, and Miller 2010). Classifier ensemble technology falls generally into two categories. Multiple classifier systems are based on the manipulation of training sample sets, including Boosting (Freund et al. 2003) and Bagging (Breiman 1996), Random Forests (Rodriguez-Galiano et al. 2012), etc. This ensemble generates combinations of classifiers by weighted or random resampling of their representative training datasets with replacement. However, there is hardly any improvement in classification accuracy with SVMs via this methodology, since the construction of SVMs is relatively insensitive to sample distribution (Chan, Chengquan, and DeFries 2001; Pal 2008). In contrast, an alternative classifier ensemble is derived from decision fusion through the output of multiple classifiers, achieving optimal results with SVMs. For instance, Ceamanos et al. (2010) decomposed hyperspectral remote-sensing imagery according to the similarity of spectral bands, with the corresponding components being applied on the SVM fusion ensemble. The results demonstrate that SVM ensembles have the potential to outperform a standard SVM. Ban and Jacob (2013) explored three object-based SVMs with fusion ensembles using multitemporal SAR and multispectral optical data. The results indicated that integration of multiple SVMs generally outperforms a single SVM in terms of classification accuracy. However, none of the aforementioned SVM ensembles consider parameter selection and optimization, which are crucial for the SVM performance in land-cover classification.

In this article, unlike the usual procedure of having the same parameter setting for each of the binary classifiers in SOP-SVM, we propose a novel multi-parameter SVM (MP-SVM) which optimizes the parameters of each binary classifier that form the classifier ensemble. The proposed MP-SVM solves the complex overlap issues among class boundaries and depicts intrinsic differences within complex landscapes.

2. Methods

2.1. Overview of SVMs

Assume a set of linearly separable training data $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)$, where $\boldsymbol{x}_i$ is a multidimension sample vector and $y_i \in +1,-1$ is the class label. The fundamental principle of the SVM is to construct an optimal decision hyperplane to find a maximum separating margin between disparate datasets (Vapnik 1995, 1998). Specifically, the binary classification problems associated to the SVM algorithm can be described as a quadratic programming (QP) problem:

$$\min_{w,b,\varepsilon} J(\boldsymbol{w}, b, \varepsilon) = \left(||w||^2/2\right) + C\sum_{i=1}^{n}\varepsilon_i, \tag{1}$$

$$\text{subject to } y_i[(\boldsymbol{w}(\boldsymbol{x}_i{}^{\mathrm{T}} \cdot \boldsymbol{x}_j) + b] \geq 1 - \varepsilon_i,\ i = 1, \ldots, .l. \tag{2}$$

Mathematically, $\boldsymbol{w}$ represents the normal vector of the linear hyperplane while b refers to the offset of the hyperplane in the coordinate system. Data points closest to the decision hyperplane are called support vectors that often conflict across several classifications. Therefore, support vectors are relevant directly to the construction of an optimized separating hyperplane. C and ε_i refer to the penalty value and the slack variable, respectively. Both are essential to the treatment of errors to the extent that some support vectors incorrectly fall within the margins or into other categories.

A real classification problem is normally not guaranteed to be linearly separable among different classes. Kernel functions are, therefore, used to map the input vectors into high-dimensional Hilbert space based on Mercer theory (Ban and Jacob 2013). One of the most commonly used kernel functions and its kernel matrix is the RBF:

$$k\left(x_i, x_j\right) = \exp\left(-\gamma\|x_i - x_j\|^2\right), \gamma > 0. \tag{3}$$

The dot product $\boldsymbol{x}_i^{\mathrm{T}} \cdot \boldsymbol{x}_j$ in Equation (1) is substituted by $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ in high-dimensional space, thus enabling the classification to be linearly separable. Using the Lagrange multiplier method, the QP problem can be solved by conversion into the dual problem:

$$\min W(\alpha) = -\sum_{i=1}^{n} \alpha_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j k\left(x_i, x_j\right), \tag{4}$$

$$\text{constrained to } \sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \in [0, C]. \tag{5}$$

The regularization parameter (or penalty value) C controls the trade-off between minimizing the errors between the SVM and training dataset and maximizing the margin, while γ is inversely proportional to the Gaussian kernel width that determines the computing window of the RBF kernel matrix (Ghoggali, Melgani, and Bazi 2009). These both largely determine the boundary complexity of the SVM, which reflects the final classification performance in real applications.

2.2. MP-SVM

For a RBF kernel SVM classifier, the penalty value C and kernel parameter γ need to be predetermined via the training data. A standard SVM may achieve high accuracy by setting a single set of parameters (C and γ) in certain applications (e.g. bankruptcy prediction (Min and Lee 2005)). However, in the remote-sensing domain, this is not the case. The first and foremost difficulty arises from the complexity of land-cover classes (along the boundary) that usually overlap each other due to the similarity of spectra and spatial features. Furthermore, training sample sizes obtained *in situ* often differ from class to class, which may produce biased estimation for SVM parameters. Therefore, this article proposes a multi-parameter SVM (MP-SVM) to resolve these issues.

First, the MP-SVM divides the whole training set (n land-cover classes) into $n(n-1)/2$ subsets composed of any two land-cover classes, e.g. marsh and meadow, forest and meadow, etc. The distinction among these subsets guarantees the diversity of training samples for proper construction of SVM models. The parameters of each SVM model are then optimized. Parameter optimization methods can be divided into two categories: exhaustive grid search (Min and Lee 2005) and heuristic parameter search (Wu et al. 2007), like the PSO and GA algorithms. Both methods are evaluated based on a cross-validation procedure to prevent overfitting and enhance the generalization capability. Although both PSO and GA have shown their efficiency in optimizing the parameters of the SVM (Lin et al. 2008), these methods select the best parameter combinations from the populations evolved from generation to generation with very large sample-size requirements (Wu and Wang 2009). Also, the PSO and GA algorithms require the optimization of some additional parameters such as the population number, the probability of mutation, and crossover, which increases the computational time and complexity (Lorena and De Carvalho 2008). In this research, the grid search approach is used due to its straightforwardness and robustness for parameter optimization (Hsu and Lin 2002).

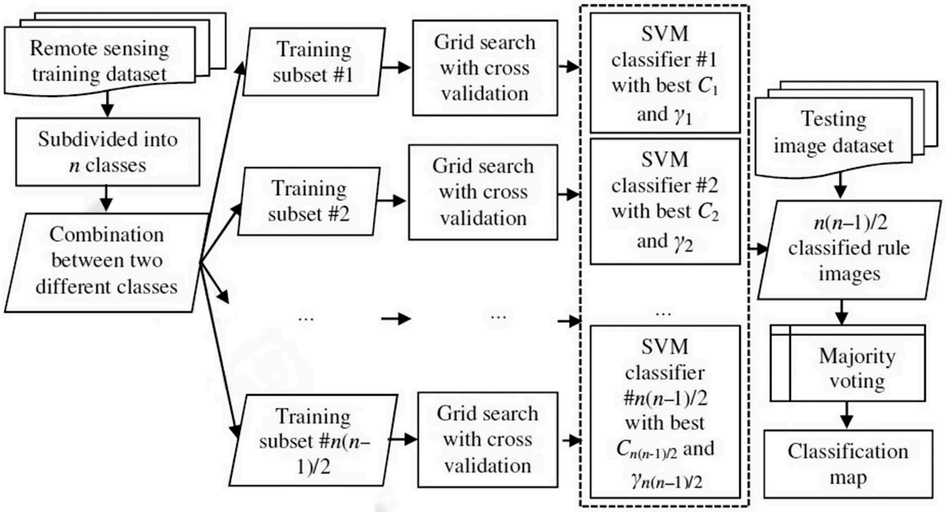

Figure 1. Flowchart of the multi-parameter SVM in remote-sensing classification.

The grid search methodology attempts to identify the pair of parameters (C and γ) with optimum cross-validation accuracy through trial and error. Since performing a complete grid search is often time-consuming, a coarse grid is used in advance to rapidly demarcate a certain region (i.e. grid); a further, finer grid search is then conducted within that area (Hsu, Chang, and Lin 2010).

Based on all the SVMs with optimized parameters searched by grid, the entire image is classified into binary images which serve as rule images for final voting. For each pixel, all the rule images participate in the decision based on simple majority voting: an intuitive statistical approach considering all the rule images as votes and selecting the majority voted class (Wang et al. 2009). Classification results are based on the 'winner-takes-all' rule (Waske and Benediktsson 2007) – to decide each pixel's land-cover class by calculating its maximum vote. Essentially, the absolute maximum distance to the hyperplane that has achieved the largest vote determines the decision result of the final classification map. Figure 1 depicts a detailed flowchart for multi-parameter SVM (MP-SVM) classification.

3. Study area and experimental data

The Sanjiang Plain, encompassing two wetlands of international importance, Honghe National Nature Reserve (HNNR) and Sanjiang National Nature Reserve (SNNR), and located in northeastern China, was chosen as a case study area (Zhang et al. 2009). Owing to excessive agricultural exploitation, the Sanjiang Plain has been developed into a hybrid ecosystem integrated by marsh, meadow, forest, and cultivated farmland. For comparison purposes, three pilot study sites (S1, S2, and S3) with distinct levels of landscape complexity and different numbers of land-cover categories were intentionally selected for testing the proposed classification algorithm (Figure 2). S1 is a mixture of pristine natural environment and human reclamation distributed throughout the area, with very high spatial heterogeneity and complexity, while S2 includes the ecologically protected HNNR and its surrounding region, with high landscape complexity. In addition, as a

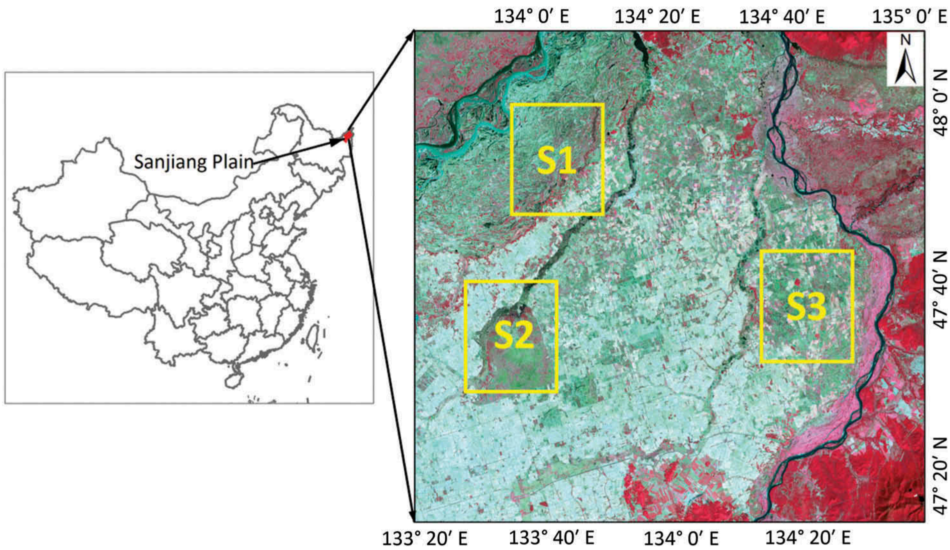

Figure 2. Three pilot study sites in Sanjiang Plain, Northeast China, S1, S2, and S3, were selected with different levels of landscape complexity.

benchmark of this experiment, S3 has low landscape complexity – primarily land reclaimed by humans.

One scene of cloud-free Landsat 5 TM imagery (WRS-2, path 113, row 27) acquired on 19 September 2010, composed of six multispectral bands (TM 1–5 and 7) with a spatial resolution of 30 m, was used in this study. The satellite sensor image was geo-rectified to the Transverse Mercator projection based on a topographic map at a scale of 1:50 000 using 70 evenly distributed ground control points. A third-order polynomial model was used for rectification implementing the nearest-neighbour algorithm with a pixel size of 30 m × 30 m for all bands. The root mean square error was less than 0.5 pixels (15 m).

A land-cover classification system was established based on three considerations: (1) the national standard land-cover classification system in China; (2) the characteristics of the vegetation distribution in the study area; and (3) the moderate spatial and spectral resolution of multispectral instruments such as the TM sensor and the spectral differences among diverse land-cover categories. S1 contains eight land-cover/use categories: paddy field (PF), dry farmland (DF), marsh (MH), bare soil (BS), meadow (MD), forest (FT), waterbody (WB), and floodplain (FP); S2 includes six land-cover/use categories that are identical to first six of S1; S3 comprises four land-cover/use categories: PF, DF, BS, and FT. Throughout the three pilot study areas, training and validation sample plots were collected during field surveys in September 2009 and September 2010 using a hand-held GPS. The number of training and validation samples in each category relies on its overall proportion to avoid over-prediction of rare classes. The mutual distance between sample plots was at least 200 m, to mitigate potential spatial autocorrelation. All sample plots in the three pilot study areas were divided randomly into 60% training and 40% testing sample sets, with 800 samples in total in S1, 595 in S2, and 320 in S3.

A wetland can be identified by three basic factors: soil, vegetation, and water regime (hydrology) (Janisch and Molstad 2004). Thus, the classification of land-cover categories of the study area (a wetland-distributed area) should consider the vegetation and water information provided by the remote-sensing imagery. The normalized difference vegetation index (NDVI), an index sensitive to leaf area index, vegetation biomass, and photosynthetic active radiation (Gao et al. 2000; Ivanova et al. 2007; Wessels et al. 2006), and tasselled cap transformation, which provides information for brightness, greenness, as well as wetness (Crist and Cicone 1984), were thus employed. Therefore, 10 bands were included (Landsat TM 1–5 and 7, NDVI, Brightness, Greenness, and Wetness) in this study.

4. Results

4.1. Model construction

For validation of the method, the LIBSVM V3.16 software based on the MATLAB R2010b programming platform was implemented (Chang and Lin 2011). Both the SOP-SVM and MP-SVM models were applied to the three study sites. In each case, 10 bands were used as input features for the SVM model and the objective variables of the model are eight, six, and four land-cover types, respectively.

The fivefold cross-validation of the SOP-SVM RBF kernel with $C \in [2^{-8}, 2^{8}]$ and $\gamma \in [2^{-10}, 2^{10}]$ was implemented (i.e. with a wide parameter search space, which can lead to high validation accuracy). Table 1 presents the optimized parameters C and γ together with the cross-validation accuracy for the traditional SOP-SVM model. The grid search approach identified the global optimized parameters with a cross-validation accuracy higher than 86% in all three pilot study sites.

Table 1. Optimized parameters chosen by grid search with fivefold cross-validation accuracy in SOP-SVM and average results for each land-cover type in MP-SVM.

	SOP-SVM	MP-SVM			
	Grid search	Grid search			
Study site	Global parameter	Land-cover types	No. of subsets with	Average (C, γ)	Average CV accuracy (%)
S1	Chosen (C, γ)	Paddy field	7	$(2^{0}, 2^{-4})$	90.12
	$(2^{0}, 2^{-4})$	Dry farmland	7	$(2^{2}, 2^{-7})$	88.47
		Marsh	7	$(2^{7}, 2^{-2})$	84.49
		Bare soil	7	$(2^{3}, 2^{-5})$	88.62
	Cross-validation	Meadow	7	$(2^{2}, 2^{-5})$	85.37
	accuracy: 90.1%	Forest	7	$(2^{4}, 2^{-4})$	91.30
		Waterbody	7	$(2^{3}, 2^{-5})$	95.60
		Floodplain	7	$(2^{5}, 2^{-3})$	87.25
S2	Chosen (C, γ)	Paddy field	5	$(2^{3}, 2^{-4})$	91.56
	$(2^{3}, 2^{-7})$	Dry farmland	5	$(2^{1}, 2^{-2})$	90.34
		Marsh	5	$(2^{5}, 2^{-7})$	84.21
	Cross-validation	Bare soil	5	$(2^{4}, 2^{-8})$	85.28
	accuracy: 86.7%	Meadow	5	$(2^{3}, 2^{-6})$	82.75
		Forest	5	$(2^{2}, 2^{-4})$	90.28
S3	Chosen (C, γ)	Paddy field	3	$(2^{2}, 2^{-5})$	90.27
	$(2^{2}, 2^{-2})$	Dry farmland	3	$(2^{3}, 2^{-5})$	88.26
	Cross-validation	Bare soil	3	$(2^{1}, 2^{-2})$	84.49
	accuracy: 90.3%	Forest	3	$(2^{2}, 2^{-4})$	89.18

In S1, 28 subsets were created based on the combination of any two categories, both of which are treated individually to form the MP-SVM model. The optimized parameters of these subsets were obtained by the grid search technique with fivefold cross-validation. The procedure was iteratively repeated in S2 with 15 subsets and in S3 with six subsets. Detailed parameters (C and γ) for each land-cover type, as well as the corresponding average cross-validation accuracies, are summarized in Table 1. Each SVM binary classifier was trained with two land-cover types, with the optimized parameters allowing the entire images to be classified into two individual classes. Thus, 28, 15, and six rule images were obtained in the three study sites and considered as contributors for the final classification outputs.

4.2. Classification results

The classification results of the MP-SVM and traditional SOP-SVM in S1, S2, and S3 were evaluated using the testing data, which comprise 40% of each original sample set. Figure 3(*a*) shows the original remote-sensing image (Landsat TM bands 4, 3, 2) with the very complex landscape in S1. The land-cover classification results of SOP-SVM and MP-SVM are illustrated in Figures 3(*b*) and (*c*), respectively. The complex landscape in S2 is shown in Figure 4(*a*), with the SOP-SVM and MP-SVM classification results shown in Figures 4(*b*) and (*c*). For comparison, a landscape with low complexity was also tested using the study site S3 (Figure 5(*a*)), with similar results for the SOP-SVM and MP-SVM classification outputs (Figures 5(*b*) and (*c*)). The corresponding confusion matrices for the three experiments are listed in Tables 2–4. Generally, the overall accuracy and kappa coefficient (κ) of the MP-SVM (82.19% and 0.80) are higher than that of SOP-SVM (72.50% and 0.69) in S1. Similar results were achieved in S2 using MP-SVM (84.03% and 0.80), with higher overall accuracy and κ than SOP-SVM (76.05% and 0.71), while for the land-cover classification with low landscape complexity the MP-SVM overall accuracy and κ were increased by only 2.35% and 0.03, respectively (Table 4).

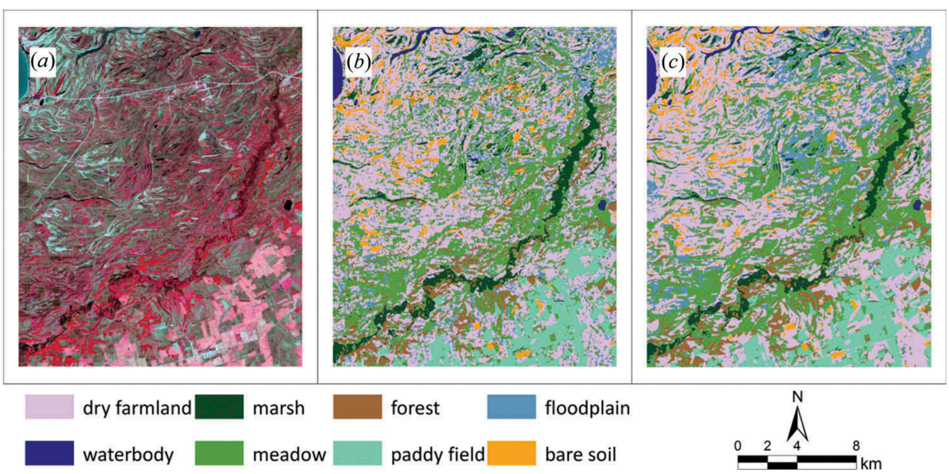

Figure 3. The land-cover classification results in S1 with very high landscape complexity: (*a*) original Landsat TM bands 4, 3, 2 based on (*b*) single optimized parameter SVM and (*c*) multi-parameter SVM.

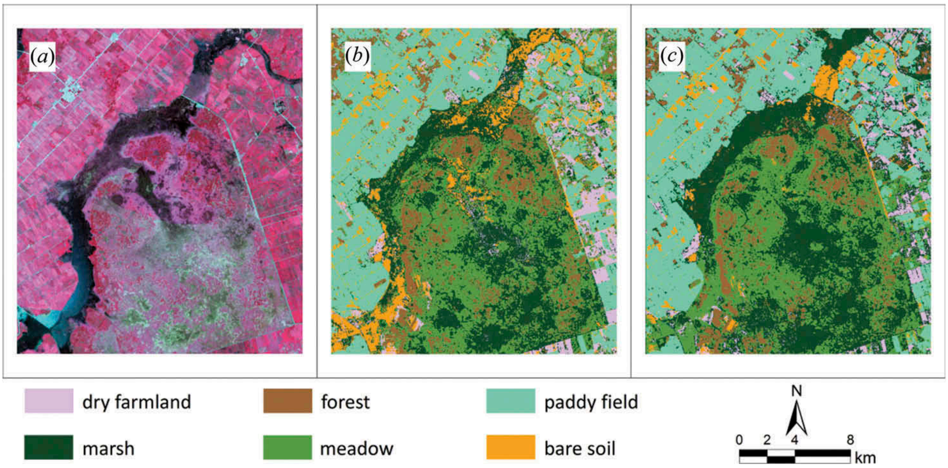

Figure 4. The land-cover classification results in S2 with high landscape complexity: (*a*) original Landsat TM bands 4, 3, 2 based on (*b*) single optimized parameter SVM and (*c*) multi-parameter SVM.

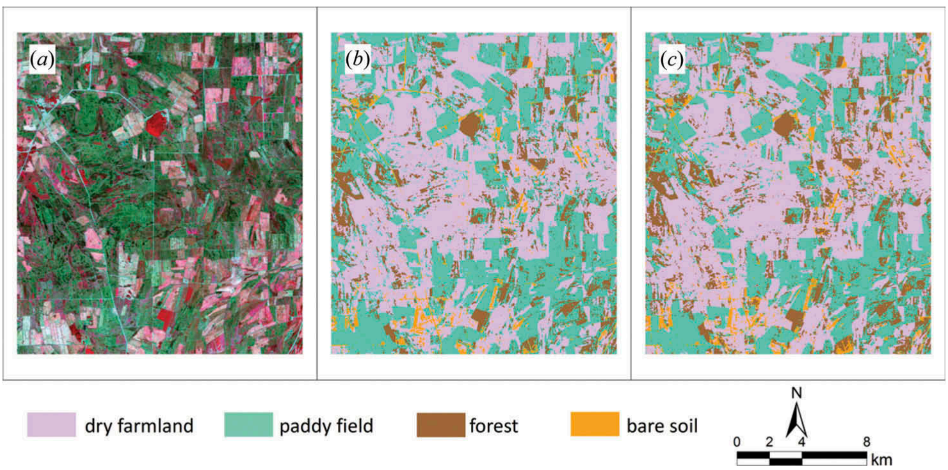

Figure 5. The land-cover classification results in S3 with low landscape complexity: (*a*) original Landsat TM bands 4, 3, 2 based on (*b*) single optimized parameter SVM and (*c*) multi-parameter SVM.

To evaluate the effectiveness and robustness of the proposed MP-SVM, the model was implemented 100 times in each study site for comparison with the traditional SOP-SVM. On each occasion the model stratified and randomly split the sample sets into 60% training samples and 40% validation samples. Kappa *z*-tests and the corresponding *p*-values for pairwise comparisons are shown in Table 5, and box plots are shown in Figure 6, demonstrating that the increases in classification accuracy in both very high- and high-complexity settings are statistically significant, with *z*-values 3.77 and 2.44 (*p*-values 0.0002 and 0.0147), respectively, both greater than 1.96 at the 95% confidence level. The increase in κ in the low-complexity landscape is not statistically significant,

Table 2. Confusion matrices for the single optimized parameter SVM (SOP-SVM) and multi-parameter SVM (MP-SVM) for eight land-cover classifiers with very high landscape complexity, including overall accuracy, kappa coefficient, producer's accuracy (PA), and user's accuracy (UA). Numbers shown in bold font are correct classifications.

Classifier	Class name	PF	DF	MH	BS	MD	FT	WB	FP	UA (%)
SOP-SVM	Paddy field (PF)	**28**	0	2	2	0	0	1	1	82.35
	Dry farmland (DF)	0	**34**	0	3	3	2	0	0	80.95
	Marsh (MH)	1	2	**22**	4	5	1	1	1	59.46
	Bare soil (BS)	2	2	5	**31**	4	2	1	2	63.27
	Meadow (MD)	1	3	4	3	**26**	2	0	2	63.41
	Forest (FT)	1	1	3	1	3	**31**	0	0	77.50
	Waterbody (WB)	0	0	2	2	2	0	**33**	1	82.50
	Floodplain (FP)	2	0	2	1	2	1	2	**27**	72.97
	PA (%)	80	80.95	55	65.96	57.78	79.49	86.84	79.41	
	Overall accuracy = 72.50%, kappa coefficient = 0.69									
MP-SVM	Paddy field (PF)	**30**	0	1	1	0	0	1	0	90.91
	Dry farmland (DF)	2	**36**	1	1	1	0	0	0	87.80
	Marsh (MH)	0	1	**31**	2	2	1	1	2	77.50
	Bare soil (BS)	1	2	4	**38**	3	2	0	2	73.08
	Meadow (MD)	0	1	2	3	**33**	2	0	1	78.57
	Forest (FT)	2	0	0	1	3	**32**	0	0	84.21
	Waterbody (WB)	0	0	0	0	1	1	**35**	1	92.11
	Floodplain (FP)	0	2	1	1	2	1	1	**28**	77.78
	PA (%)	85.71	85.71	77.50	80.85	73.33	82.05	92.11	82.35	
	Overall accuracy = 82.19%, kappa coefficient = 0.80									

Table 3. Confusion matrices for the single optimized parameter SVM (SOP-SVM) and multi-parameter SVM (MP-SVM) for six land-cover classifiers with high landscape complexity, including overall accuracy, kappa coefficient, producer's accuracy (PA), and user's accuracy (UA). Numbers shown in bold font are correct classifications.

Classifiers	Class name	PF	DF	MH	BS	MD	FT	UA (%)
SOP-SVM	Paddy field (PF)	**41**	1	3	1	0	0	89.13
	Dry farmland (DF)	2	**38**	1	2	3	0	82.61
	Marsh (MH)	3	3	**20**	4	4	2	55.56
	Bare soil (BS)	1	1	5	**29**	4	1	70.73
	Meadow (MD)	0	2	3	3	**26**	2	72.22
	Forest (FT)	0	0	1	3	2	**27**	81.82
	PA (%)	87.23	84.44	60.61	60.61	66.67	84.38	
	Overall accuracy = 76.05%, kappa coefficient = 0.71							
MP-SVM	Paddy field (PF)	**42**	0	2	0	0	0	95.45
	Dry farmland (DF)	3	**40**	0	2	2	0	85.11
	Marsh (MH)	2	0	**26**	3	3	1	74.29
	Bare soil (BS)	0	3	2	**34**	2	1	80.95
	Meadow (MD)	0	2	3	2	**30**	2	76.92
	Forest (FT)	0	0	0	1	2	**28**	90.32
	PA (%)	89.36	88.89	78.79	80.95	76.92	87.50	
	Overall accuracy = 84.03%, kappa coefficient = 0.80							

Table 4. Confusion matrices for the single optimized parameter SVM (SOP-SVM) and multi-parameter SVM (MP-SVM) for four land-cover classifiers with low landscape complexity, including overall accuracy, kappa coefficient, producer's accuracy (PA), and user's accuracy (UA). Numbers shown in bold font are correct classifications.

Classifiers	Class name	PF	DF	BS	FT	UA (%)
SOP-SVM	Paddy field (PF)	**25**	3	2	1	80.65
	Dry farmland (DF)	3	**24**	4	2	72.73
	Bare soil (BS)	3	2	**23**	4	71.88
	Forest (FT)	1	2	2	**27**	84.38
	PA (%)	78.13	77.42	74.19	79.41	
	Overall accuracy = 77.34%, kappa coefficient = 0.70					
MP-SVM	Paddy field (PF)	**26**	2	3	2	78.79
	Dry farmland (DF)	2	**25**	2	3	78.13
	Bare soil (BS)	1	3	**23**	1	82.14
	Forest (FT)	3	1	3	**28**	80
	PA (%)	81.25	80.65	74.19	82.35	
	Overall accuracy = 79.69%, kappa coefficient = 0.73					

Table 5. The z-statistic used to compare the performance of the two classifiers for three different landscape complexities, with kappa z-test and corresponding p-value in the fourth column. Significantly different accuracies with confidence of 95% (z-value > 1.96) are indicated by *.

	Kappa (standard deviation)		
Landscape complexity	SOP-SVM	MP-SVM	Kappa z-test (p-value)
Very high	0.69 (0.022)	0.80 (0.019)	3.77* (0.0002)
High	0.71 (0.028)	0.80 (0.024)	2.44* (0.0147)
Low	0.70 (0.027)	0.73 (0.026)	0.80 (0.4237)

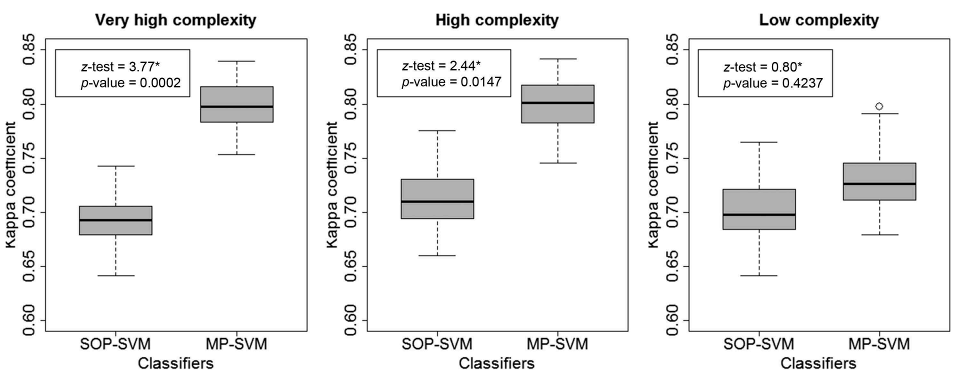

Figure 6. Statistic comparison between the SOP-SVM classifiers and the MP-SVM classifiers under three different landscape complexities with kappa coefficient, z-test, and corresponding p-value. Significantly different accuracies with confidence of 95% (z-value > 1.96) are indicated by *.

with z-value 0.80 and p-value 0.4237, less than 1.96 at the 95% confidence level (Congalton 1991). Notably, significant increases in classification accuracy were obtained for three land-cover types (i.e. marsh, bare soil, and meadow).

Regarding the result for marsh, the MP-SVM achieved a much higher accuracy than the SOP-SVM for very high and high landscape complexity. According to Tables 2 and 3, the producer's accuracy (PA) for marsh increased markedly (by 22.50% and 18.18%, respectively) and the user's accuracy (UA) increased (by 18.04% and 18.73%, respectively. As illustrated in Figures 3 and 4, marshes are successfully identified in Figures 3(*b*) and 4(*b*), whereas those in Figures 3(*a*) and 4(*a*) exhibit some conspicuous errors where large areas of marsh were falsely identified as BS. In regard to BS, a considerable increase in classification accuracy was also achieved by the MP-SVM in comparison with the SOP-SVM (Tables 2 and 3). The PA for BS increased by 14.89% and 20.34% and the UA showed a moderate increase (by 9.81% and 10.22%), respectively. Figures 3 and 4 show that BS was largely confused with marsh and meadow in the SOP-SVM classification, this being rectified in the MP-SVM classification. The classification PA of meadow was significantly increased by the MP-SVM (15.55% and 10.25%) and moderately for UA (15.14% and 4.70%), respectively. As demonstrated in Figures 3 and 4, BS was mainly misclassified as meadow by SOP-SVM.

With respect to PF and DF, the classification accuracy of both SOP-SVM and MP-SVM was relatively high, up to an average of 87.52% and 85%, respectively. This was the case mainly because the spectrum feature was distinctive due to the easily distinguishable high-level greenness and inherent regular texture. Note, from field investigations and interviews with local farmers in Sanjiang Plain, we know that the harvest seasons of major crops including rice, corn, and soybean are either at the end of September or the beginning of October, and this ruled out the possibility of any confusion between DF and BS. In regard to forest, the MP-SVM showed no obvious increase in classification accuracy in comparison with the SOP-SVM. Nevertheless, a slight increase in both PA and UA for forest was provided by MP-SVM (by 2.56% and 6.71%, respectively, in a very high-complexity landscape (Table 2) and by 3.12% and 8.5%, respectively, in a complex landscape (Table 3)).

Along with the increase in landscape complexity, two further land-cover types were present in very high-complexity landscape, namely, WB and FP (Figure 3 and Table 2). The inclusion of these two land-cover types complicates visual interpretation of the map. The results show that the MP-SVM provides an average increase in accuracy of 4.11% and 7.21%, respectively, for these two extra classes in comparison with the SOP-SVM (Table 2).

However, individual comparisons in the low-complexity landscape are not significant. Specifically, PA for forest increased by 2.94% compared with the MP-SVM, but with a 4.38% decrease in UA. Classes PF, DF, and BS all lie within a 5% increase in accuracy for the MP-SVM compared with the SOP-SVM (Figure 5 and Table 4).

5. Discussion

The results show the effect of landscape complexity on the difference between the MP-SVM and SOP-SVM in terms of classification accuracy. Low landscape complexity showed no statistically significant increase in accuracy. Greater landscape complexity led to a statistically significant increase in accuracy between the models (Figure 6). Essentially, the MP-SVM is designed to construct multiple heterogeneous SVM classifier ensembles considering different pairs of land-cover types and to estimate the best parameters of each sub-SVM classifier in a fully automatic way. Given an increasing number of landscape types and landscape patterns, it is difficult to generalize using the traditional SOP-SVM with single parameter sets. The advantage of multiple parameter sets is evident in the membership association of the class boundaries and voting results. The MP-SVM, to some extent, can represent the intrinsic differences within complex landscapes.

Originally, SVMs were defined as binary classifiers and their use for multi-class classification is more problematic, with strategies that reduce the multi-class problem to a set of binary problems typically adopted (Foody and Mathur 2004). To better solve such multi-class discriminant problems in remote-sensing classification, the SOP-SVM pair-wisely searches the global optimized parameters (C and γ), whose convenience and effectiveness have been discussed by Foody and Mathur (2004), since only a single set of parameters needs to be defined. However, it considers only the overall performance for multiple classes, attempting to identify the most balanced and global optimized parameters while paying little attention to the individual differences between diverse land-cover categories. However, the MP-SVM divides the entire training set into subsets containing either of two different classes. Each subset is trained to construct its own SVM model respectively, which considers fully the differences between and properties of the land-cover types. In addition, the MP-SVM maintains versatility and straightforward operating capability, since it is a fully automatic classifier for all kinds of land-cover classification without any expert knowledge requirement.

For some complex ecosystems (e.g. marsh) it is often not a trivial task to acquire sufficient and representative training samples, due to accessibility difficulties (Lu and Weng 2007). The SOP-SVM achieves low classification accuracies (average only 57.81%) in the classification of marsh since it is almost impossible to find globally optimized parameters for model construction. In contrast, the MP-SVM can recognize this kind of class by searching all the SVMs' optimized parameters, leading to higher classification accuracies (average 78.15%) as shown in Tables 2 and 3. It should be noted that the differentiation between natural wetlands (marsh) and BS, as well as meadow, is an ongoing challenge due to the extraordinary similarity between their spectral features. The classification accuracies of marsh are still relatively low in the MP-SVM (average 78.15%), although a 20.34% increase was achieved compared with the traditional SOP-SVM. Hence, it is necessary to incorporate textural information as well as ancillary geographical data like terrain slope and aspect to further increase the classification accuracies of marsh (wetlands) in future research (Camps-Valls et al. 2008; Na et al. 2009). Moreover, BS often overlaps with other classes because it lacks a distinguishing characteristic, and as such may be misclassified as meadow or DF. Nevertheless, the MP-SVM can analyse any subtle distinctions, and increase the classification accuracy of both BS and meadow accordingly. Thus, the MP-SVM is promising generally for land-cover classification.

One benefit of the proposed MP-SVM approach is the use of ensemble classifiers. It combines all the rule images and votes to produce a final class allocation through majority voting. Such an ensemble technique can overcome the instability of individual classifiers and take full advantage of the variety of different single classifiers to achieve robust results. Although the accuracy of each classifier in the ensemble may decrease due to individual errors, the use of many classifiers, as well as the final result of voting, eliminates the misclassifications, increasing classification accuracy. The results in this research are consistent with previous studies conducted by Waske and Benediktsson

(2007) and Pal (2008), and further demonstrate that decision fusion is crucial to effective land-cover classification (Waske and Benediktsson 2007).

6. Conclusions

Traditional SOP-SVM involves major uncertainties and inconsistencies in the distribution of data along highly dynamic natural boundaries by using single SVM models with subjective and (only) globally optimal SVM parameters. The main goal of this paper was to propose and demonstrate a novel MP-SVM by comparing it to a traditional SOP-SVM method for remote-sensing land-cover classification. The classification accuracy and robustness were tested at three different study sites using remotely sensed data with distinct levels of landscape complexity.

The MP-SVM was found to be a promising tool for land-cover classification, especially given very high levels of landscape complexity. The higher the landscape complexity and the larger the number of land-cover categories, the greater the increase in accuracy of the MP-SVM in comparison with the SOP-SVM.

For future research, a series of indices should be developed to quantify landscape complexity, such as examining the correlation between landscape complexity and the classification accuracy of the MP-SVM. In addition, the suitability and robustness of the MP-SVM for regional or even global remote-sensing classification requires further investigation.

Acknowledgements

The authors are grateful to the staff of the Honghe National Nature Reserve in Heilongjiang Province, China, for their kind assistance during the field surveys. The authors thank Andrew MacLachlan (Geography and Environment, University of Southampton, UK) for his valuable help in checking the grammar and sentence structure. The authors also thank the two anonymous referees for their constructive comments on this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China [grant number 41271196] and the European Union Erasmus Mundus Scholarship [grant number 2011-0155].

References

Ban, Y., and A. Jacob. 2013. “Object-Based Fusion of Multitemporal Multiangle ENVISAT ASAR and HJ-1B Multispectral Data for Urban Land-Cover Mapping.” *IEEE Transactions on Geoscience and Remote Sensing* 51 (4): 1998–2006. doi:10.1109/TGRS.2012.2236560.

Bazi, Y., and F. Melgani. 2007. “Semisupervised PSO-SVM Regression for Biophysical Parameter Estimation.” *IEEE Transactions on Geoscience and Remote Sensing* 45 (6): 1887–1895. doi:10.1109/TGRS.2007.895845.

Breiman, L. 1996. “Bagging Predictors.” *Machine Learning* 24 (2): 123–140. doi:10.1007/BF00058655.

Camps-Valls, G., and L. Bruzzone. 2005. “Kernel-Based Methods for Hyperspectral Image Classification.” *IEEE Transactions on Geoscience and Remote Sensing* 43 (6): 1351–1362. doi:10.1109/TGRS.2005.846154.

Camps-Valls, G., L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramon. 2008. "Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 46 (6): 1822–1835. doi:10.1109/TGRS.2008.916201.

Ceamanos, X., B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson. 2010. "A Classifier Ensemble Based on Fusion of Support Vector Machines for Classifying Hyperspectral Data." *International Journal of Image and Data Fusion* 1 (4): 293–307. doi:10.1080/19479832.2010.485935.

Chan, J. C.-W., H. Chengquan, and R. DeFries. 2001. "Enhanced Algorithm Performance for Land Cover Classification from Remotely Sensed Data Using Bagging and Boosting." *IEEE Transactions on Geoscience and Remote Sensing* 39: 693–695. doi:10.1109/36.911126.

Chang, -C.-C., and C.-J. Lin. 2011. "LIBSVM: A Library for Support Vector Machines." *ACM Transactions on Intelligent Systems and Technology* 2: 1–27. doi:10.1145/1961189.1961199.

Cheng, T., and M. Molenaar. 1999. "Diachronic Analysis of Fuzzy Objects." *GeoInformatica* 3 (4): 337–355. doi:10.1023/A:1009884730822.

Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37 (1): 35–46. doi:10.1016/0034-4257(91)90048-B.

Crist, E. P., and R. C. Cicone. 1984. "A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap." *IEEE Transactions on Geoscience and Remote Sensing* GE-22: 256–263. doi:10.1109/TGRS.1984.350619.

Foody, G. M., and A. Mathur. 2004. "A Relative Evaluation of Multiclass Image Classification by Support Vector Machines." *IEEE Transactions on Geoscience and Remote Sensing* 42: 1335–1343. doi:10.1109/TGRS.2004.827257.

Freund, Y., R. Iyer, R. E. Schapire, and Y. Singer. 2003. "An Efficient Boosting Algorithm for Combining Preferences." *The Journal of Machine Learning Research* 4 (December): 933–969.

Friedl, M. A., C. E. Brodley, and A. H. Strahler. 1999. "Maximizing Land Cover Classification Accuracies Produced by Decision Trees at Continental to Global Scales." *IEEE Transactions on Geoscience and Remote Sensing* 37: 969–977. doi:10.1109/36.752215.

Friedrichs, F., and C. Igel. 2005. "Evolutionary Tuning of Multiple SVM Parameters." *Neurocomputing* 64: 107–117. doi:10.1016/j.neucom.2004.11.022.

Gao, X., A. R. Huete, W. Ni, and T. Miura. 2000. "Optical-Biophysical Relationships of Vegetation Spectra without Background Contamination." *Remote Sensing of Environment* 74: 609–620. doi:10.1016/S0034-4257(00)00150-4.

Ghimire, B., J. Rogan, and J. Miller. 2010. "Contextual Land-Cover Classification: Incorporating Spatial Dependence in Land-Cover Classification Models Using Random Forests and the Getis Statistic." *Remote Sensing Letters* 1: 45–54. doi:10.1080/01431160903252327.

Ghoggali, N., F. Melgani, and Y. Bazi. 2009. "A Multiobjective Genetic SVM Approach for Classification Problems with Limited Training Samples." *IEEE Transactions on Geoscience and Remote Sensing* 47: 1707–1718. doi:10.1109/TGRS.2008.2007128.

Hsu, C.-W., C.-C. Chang, and C.-J. Lin. 2010. "A Practical Guide to Support Vector Classification." *Bioinformatics* 1: 1–16.

Hsu, C.-W., and C.-J. Lin. 2002. "A Comparison of Methods for Multiclass Support Vector Machines." *IEEE Transactions on Neural Networks* 13: 415–425. doi:10.1109/72.991427.

Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. "An Assessment of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 23: 725–749. doi:10.1080/01431160110040323.

Ibrahim, M. A., M. K. Arora, and S. K. Ghosh. 2005. "Estimating and Accommodating Uncertainty through the Soft Classification of Remote Sensing Data." *International Journal of Remote Sensing* 26: 2995–3007. doi:10.1080/01431160500057806.

Ivanova, Y. D., S. I. Bartsev, A. A. Pochekutov, and A. V. Kartushinsky. 2007. "The Analysis of Seasonal Activity of Photosynthesis and Efficiency of Various Vegetative Communities on a Basis NDVI for Modeling of Biosphere Processes." *Advances in Space Research* 39: 95–99. doi:10.1016/j.asr.2006.02.028.

Janisch, J. E., and N. E. Molstad. 2004. "Disturbance and the Three Parameters of Wetland Delineation." *Wetlands* 24: 820–827. doi:10.1672/0277-5212(2004)024[0820:DATTPO]2.0.CO;2.

Kavzoglu, T., and I. Colkesen. 2009. "A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification." *International Journal of Applied Earth Observation and Geoinformation* 11: 352–359. doi:10.1016/j.jag.2009.06.002.

Lin, S.-W., K.-C. Ying, S.-C. Chen, and Z.-J. Lee. 2008. "Particle Swarm Optimization for Parameter Determination and Feature Selection of Support Vector Machines." *Expert Systems with Applications* 35: 1817–1824. doi:10.1016/j.eswa.2007.08.088.

Lorena, A. C., and A. C. P. L. F. De Carvalho. 2008. "Evolutionary Tuning of SVM Parameter Values in Multiclass Problems." *Neurocomputing* 71: 3326–3334. doi:10.1016/j.neucom.2008.01.031.

Lu, D., and Q. Weng. 2007. "A Survey of Image Classification Methods and Techniques for Improving Classification Performance." *International Journal of Remote Sensing* 28: 823–870. doi:10.1080/01431160600746456.

Min, J. H., and Y.-C. Lee. 2005. "Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters." *Expert Systems with Applications* 28: 603–614. doi:10.1016/j.eswa.2004.12.008.

Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 247–259. doi:10.1016/j.isprsjprs.2010.11.001.

Na, X., S. Zhang, H. Zhang, X. Li, H. Yu, and C. Liu. 2009. "Integrating TM and Ancillary Geographical Data with Classification Trees for Land Cover Classification of Marsh Area." *Chinese Geographical Science* 19: 177–185. doi:10.1007/s11769-009-0177-y.

Pal, M. 2008. "Ensemble of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 29: 3043–3049. doi:10.1080/01431160802007624.

Pan, X., S. Zhang, H. Zhang, X. Na, and X. Li. 2010. "A Variable Precision Rough Set Approach to the Remote Sensing Land Use/cover Classification." *Computers and Geosciences* 36: 1466–1473. doi:10.1016/j.cageo.2009.11.010.

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104. doi:10.1016/j.isprsjprs.2011.11.002.

Shao, Y., and R. S. Lunetta. 2012. "Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification Using Limited Training Data Points." *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 78–87. doi:10.1016/j.isprsjprs.2012.04.001.

Su, L. 2009. "Optimizing Support Vector Machine Learning for Semi-Arid Vegetation Mapping by Using Clustering Analysis." *ISPRS Journal of Photogrammetry and Remote Sensing* 64: 407–413. doi:10.1016/j.isprsjprs.2009.02.002.

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

Vapnik, V. N. 1998. *Statistical Learning Theory*. New York: Wiley.

Wang, A., W. Yuan, J. Liu, Z. Yu, and H. Li. 2009. "A Novel Pattern Recognition Algorithm: Combining ART Network with SVM to Reconstruct a Multi-Class Classifier." *Computers and Mathematics with Applications* 57: 1908–1914. doi:10.1016/j.camwa.2008.10.052.

Waske, B., and J. A. Benediktsson. 2007. "Fusion of Support Vector Machines for Classification of Multisensor Data." *IEEE Transactions on Geoscience and Remote Sensing* 45 (12): 3858–3866. doi:10.1109/TGRS.2007.898446.

Watanachaturaporn, P., M. K. Arora, and P. K. Varshney. 2008. "Multisource Classification Using Support Vector Machines: An Empirical Comparison with Decision Tree and Neural Network Classifiers." *Photogrammetric Engineering and Remote Sensing* 74: 239–246. doi:10.14358/PERS.74.2.239.

Wessels, K. J., S. D. Prince, N. Zambatis, S. MacFadyen, P. E. Frost, and D. Van Zyl. 2006. "Relationship between Herbaceous Biomass and 1-Km^2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa." *International Journal of Remote Sensing* 27: 951–973. doi:10.1080/01431160500169098.

Wu, C.-H., G.-H. Tzeng, Y.-J. Goo, and W.-C. Fang. 2007. "A Real-Valued Genetic Algorithm to Optimize the Parameters of Support Vector Machine for Predicting Bankruptcy." *Expert Systems with Applications* 32: 397–408. doi:10.1016/j.eswa.2005.12.008.

Wu, K.-P., and S.-D. Wang. 2009. “Choosing the Kernel Parameters for Support Vector Machines by the Inter-Cluster Distance in the Feature Space.” *Pattern Recognition* 42: 710–717. doi:10.1016/j.patcog.2008.08.030.

Zhang, S., X. Na, B. Kong, Z. Wang, H. Jiang, H. Yu, Z. Zhao, X. Li, C. Liu, and P. Dale. 2009. “Identifying Wetland Change in China’s Sanjiang Plain Using Remote Sensing.” *Wetlands* 29: 302–313. doi:10.1672/08-04.1.

Zhu, G., and D. G. Blumberg. 2002. “Classification Using ASTER Data and SVM Algorithms: The Case Study of Beer Sheva, Israel.” *Remote Sensing of Environment* 80: 233–240. doi:10.1016/S0034-4257(01)00305-4.