SUPPORT VECTOR DOMAIN DESCRIPTION MODEL TO MAP SPECIFIC LAND COVER WITH OPTIMAL PARAMETERS DETERMINED FROM WINDOW-BASED VALIDATION SET

Shuang Zhu¹,², Jinshui Zhang², Guanyuan Shuai², Zhoumiqi Yuan²

¹Beijing Polytechnic College, Beijing 100042, China
²College of Resources Science and Technology/State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
³Earth and Environmental Sciences, Michigan State University, Okemos 48864, MI, US

ABSTRACT
This paper developed an approach to determine optimal parameters, C and s, for support vector domain description (SVDD) model to map specific land cover from integrating of training and window-based validation sets (WVS-SVDD). The validation set based on window-based approach made a tighten hypersphere because of compact constraint by the outlier pixels which were located closely to the target class in the feature space. The target land covers of wheat and bare land were considered to test the proposed method’s performance. The overall accuracy for wheat reached as high as 94.37%. However, the underestimation of wheat, only 71.12% of the user’s accuracy, attributed to the validation set covering a small portion of wheat spectra. The larger window sizes were tested to achieve more wheat pixels’ samples for validation set. The results showed that wheat accuracy were improved along with window size increasing and the overall accuracies were higher than 88%.

The bare land as more heterogeneous land cover against wheat was selected to analyze the applicability and suitability of WVS-SVDD. SVDD classification was conducted and compared to the support vector machine (SVM) method as benchmark. The producer’s and user’s accuracies for bare land were over 80% at 2.4-m resolution scale and the overall accuracies were similar to those produced by the SVM at coarser spatial resolutions, highlighting the applicability of WVS-SVDD. Therefore, the developed method showed its advantages using the optimal parameters, C and s, for mapping homogeneous wheat and heterogeneous bare land, which exhibits great potentials to achieve highly accurate specific land cover.

Key words—Support vector data description (SVDD); Optimal parameters; Window-based validation set; Simulated annealing; Land cover

1. INTRODUCTION
Remote sensing technology using different classifiers, such as maximum likelihood classifier (MLC) and support vector machine (SVM), shows excellent potential capacity to produce timely and accurate land-cover thematic maps, which could efficiently and quantitatively represent the Earth’s land surface development processes and dramatic changes, which are impelled from natural and human factors. These invaluable data are prerequisite to meet the demands of global climate change research or policy decision-makings [1-3]. For the multi-classifiers algorithm, it needs to make an exhaustively labeled training data set to represent the wall-to-wall investigated area, in order to avoid the degradation of classification performance due to negligence or omission of any class for the training sample set.

The support vector data description (SVDD) model which is one of a boundary methods projects the data into a high-dimensional feature space in order to enclose most target objects and rejects outliers [4]. This method exhibits excellent ability to describe the specific land-cover distributions and map it [5-6].

According to SVDD’s principle, the tradeoff coefficient (C) and kernel width (s) are two critical parameters need to be optimized, which have deep impacts on the performance of SVDD for specific land cover identification. The C is defined as the ratio of target objects to outlier objects in a training sample set and the s is to control the compactness of hypersphere which can separate the target objects from the outliers. When s is fixed, the hypersphere changes according to variation of the C. Therefore, more objects are rejected as outliers as C decreasing, leading to a dilated hypersphere. When C kept as constant, smaller s would contribute to an over-tight boundary around the training sample set whereas a looser one would be derived from large s. Currently, little attentions are drawn on how to determine the appropriate parameters, C and s, for SVDD model to ensure specific land-cover mapping performance.

In this paper, a novel approach is developed to optimize C and s for SVDD and ensure performance of specific land cover mapping with the smaller number of labeled training pixels against multi-class classification. The proposed method integrates two kinds of data set, training and validation set, acquired from different window scales. The
simulated annealing (SA) search algorithm as an efficient tool is employed to search the optimal parameters. Hereafter, the method is abbreviated as WVS-SVDD.

2. WVS-SVDD

The proposed method mainly includes four components (Fig. 1): 1) training set selection; 2) validation set selection; 3) optimized parameters determination using simulated annealing (SA) algorithm; and 4) SVDD-based specific land-cover classification.

2.1 Training set selection

The training set size and spectral feature are critical for supervised classification to ensure the land-cover thematic map accuracy. According to traditional principle, 120 training sample pixels in this study were chosen from the remote sensing image for further analysis. The edge training set consists of two kinds: mixed and corner pixels. The mixed pixels are situated around target land-cover parcel boundaries where pixels represent the mixed spectra of the target class and other classes, which serves for target class as the basis separated from other classes. The corner pixel set are defined as the pixels lying at the vertex of the convex data from the view of spectral feature space, which can be constructed by the first and second component generated from the minimum noise fraction (MNF) transformation. The total two part of pixels are effective training set for SVDD to map the target land cover.

2.2 Validation set selection

A window-based sampling method based on the principle of SVDD is developed in this paper to get the neighboring pixels of training set as validation set, which effectively avoids the requirement of an exhaustive training set in multi-class classification. The proposed method is based on the principle that mixed pixel straddling a boundary represents the two mixed spectral responses, which is dominated by each of the classes separated by that boundary. The mixed spectral responses of two classes are also adjacent in the feature space, which could provide potential information for constructing optimal separating hypersphere, as illustrated in Fig. 2.

2.3 Optimal C and s determination using SA algorithm

The Simulated Algorithm (SA) as a global optimization algorithm was firstly proposed by Metropolis et al., in which local optimality was avoided during each iteration. The “cooling” process, analogical to the cooling of a metal, enables SA to search gradually to accomplish global optimization. The SA algorithm has been proposed for SVM parameters and feature search, leading to higher classification performance of SA-SVM than that of grid search, another general parameters determination method for SVM. Here, SA is further introduced to determine optimal C and s, two sensitive parameters for SVDD. The main steps are demonstrated as initialization, provisional state, acceptance tests, incumbent solutions and temperature decrease.

2.4 SVDD-based specific land-cover classification

Finally, the training set and related optimal parameters determined from SA model are input into SVDD model for target land cover identification. The binary map for target land cover as the interested class is produced. In the map, the target class is set as 1, other classes are 0.

3. CASE STUDY

The study area was a sub-region of Tongzhou, a district of southeast Beijing, China, which covers approximately 40
km² (39°1'N – 39°4' N, 116°2' E – 116°8' E) (Fig. 3). The topography is flat, and the fragmented agriculture land is a representative landscape in this area. The cultivated land is dominated by wheat, corn and other crops. The other three primary land covers are buildings, bare land and water.

![Fig. 3 Study area and QB imagery (bands combination= 4, 3, and 2)](image)

A high-quality multi-spectral Quickbird (QB) imagery under cloudless condition acquired on May 2, 2006 was used for this study. The primary parameters for this sensor is 2.4-m spatial resolution, with four bands (blue: 450–520 nm; green: 520–600 nm; red: 630–690 nm; NIR: 760–900 nm).

### 3.1 Comparison between WVS-SVDD, conventional SVDD and SVM

We compared the WVS-SVDD with conventional SVDD approaches, and a binary SVM classification for mapping wheat to test the strengths and weaknesses of WVS-SVDD. The wheat mapping results from two SVDD based and SVM based classifications were exhibited in Fig. 4 and the related quantitative accuracy assessment was given in Table 1. The binary SVM yielded an ideal performance with a satisfied overall accuracy of 94.37%. For the conventional SVDD-based wheat mapping method, the overall accuracy was under 80%. It was an exciting result of omission error as 3.94%, meaning the constructed large hypersphere allowing acceptance of as many wheat pixels as possible. However, little omission error was achieved at the cost of a large commission error of 36.57%. Comparing to the conventional SVDD, the WVS-SVDD yielded a more accurate result with an overall accuracy of 89.25%, reaching the same level with SVM. In this result, less than 3% outliers were falsely accepted, which highlighted the potential of outliers established from the window-based validation set can reject other classes. However, the omission error of wheat class was relatively high because the constructed hypersphere was relatively small to enclose heterogeneous wheat spectral feature under 2.4 m resolution completely.

![Fig. 4 Comparison of wheat results from different mapping methods (a) QB image, (b) actual wheat distribution, (c) result from SVM classification, (d) result from WVS-SVDD, (e) result from traditional SVDD.](image)

### Table 1 Classification accuracy for the proposed, traditional and SVM methods

| Classification accuracy (%) | PA   | UA    | OA  |
|-----------------------------|------|-------|-----|
| WVS-SVDD                    | 71.12| 97.36 | 89.25|
| Traditional Method          | 95.33| 64.84 | 80.33|
| SVM                         | 94.57| 90.37 | 94.37|

### 3.2 The effect of untrained classes on the classification accuracy

Table 2 showed accuracy of SVDD and SVM classification to map wheat. Under the condition of the use of training set with all defined classes, both SVDD and SVM achieved good performance with overall accuracy above 95%. However, the outlier class had obvious impacts, following with reduction of classes in the non-wheat, on SVM classification accuracy. These results suggest that SVM is sensitive to untrained class and an exhaustive training set is necessary to ensure an accurate classification. The WVS-SVDD still achieved good performance with high overall accuracy, little sensitive to whether tree or bare land pixels excluded from the validation set, which could reach the same level of the SVDD classification as the original validation set. This is because the hypersphere constructed using the optimal C and s with support of window-based validation set could create the enclosed space to determine the wheat, suffering little influence from untrained non-wheat classes.

### Table 2 accuracy assessment for each classification at different spatial patterns and window sizes

| Spatial resolution & Window size | # of pixels in validation set | Optimal Parameters | Classification accuracy (%) |
|---------------------------------|------------------------------|--------------------|-----------------------------|
| 2.4 m                           | Target | Outlier | C   | s   | PA   | UA     | OA    |
| 1                               | 470    | 128     | 0.02| 1.01| 71.12| 97.36  | 89.25 |
| 2                               | 1056   | 367     | 0.02| 1.02| 79.75| 96.22  | 91.64 |
20 m

|   | 1687 | 702 | 0.05 | 3.76 | 80.49 | 97.05 | 92.34 |
|---|------|-----|------|------|-------|-------|-------|
| 3 | 2307 | 1104 | 0.02 | 1.40 | 81.18 | 95.62 | 92.13 |
| 4 | 2926 | 1580 | 0.01 | 1.03 | 80.11 | 96.11 | 91.93 |

5 m

|   | 338  | 123  | 0.11 | 1.12 | 81.07 | 86.27 | 88.89 |
|---|------|------|------|------|-------|-------|-------|
| 1 | 697  | 383  | 0.08 | 9.76 | 85.50 | 86.31 | 90.20 |
| 2 | 1090 | 748  | 0.21 | 7.37 | 81.91 | 93.66 | 91.75 |
| 3 | 1542 | 1202 | 0.21 | 1.94 | 82.32 | 93.42 | 91.80 |
| 4 | 2026 | 1744 | 0.09 | 4.37 | 85.10 | 86.73 | 90.25 |

10 m

|   | 385  | 136  | 0.02 | 1.40 | 93.53 | 92.32 | 95.01 |
|---|------|------|------|------|-------|-------|-------|
| 1 | 822  | 317  | 0.04 | 1.57 | 91.13 | 93.87 | 94.81 |
| 2 | 1302 | 554  | 0.02 | 1.51 | 92.93 | 92.53 | 94.90 |
| 3 | 1818 | 841  | 0.11 | 1.68 | 89.44 | 97.12 | 95.37 |
| 4 | 2358 | 1163 | 0.14 | 7.65 | 90.57 | 96.72 | 95.62 |

15 m

|   | 464  | 144  | 0.01 | 1.04 | 91.92 | 93.10 | 94.75 |
|---|------|------|------|------|-------|-------|-------|
| 1 | 1013 | 363  | 0.03 | 1.06 | 90.68 | 95.09 | 95.07 |
| 2 | 1657 | 905  | 0.02 | 1.08 | 92.07 | 93.48 | 94.95 |
| 3 | 2375 | 1480 | 0.02 | 1.11 | 91.63 | 94.39 | 95.14 |
| 4 | 3179 | 2176 | 0.03 | 1.09 | 90.87 | 95.13 | 95.10 |

20 m

|   | 538  | 106  | 0.03 | 1.52 | 93.30 | 93.40 | 95.29 |
|---|------|------|------|------|-------|-------|-------|
| 1 | 1198 | 302  | 0.03 | 1.47 | 93.39 | 93.41 | 95.33 |
| 2 | 1926 | 567  | 0.01 | 1.19 | 94.71 | 92.32 | 95.33 |
| 3 | 2694 | 913  | 0.03 | 1.51 | 93.33 | 93.41 | 95.30 |
| 4 | 3461 | 1329 | 0.02 | 1.21 | 93.88 | 93.87 | 95.66 |

4. CONCLUSIONS

This paper proposed a WVS-SVDD method that integrated window-based validation set and SA based optimal C and s algorithm to map specific land cover. The results indicated that the proposed method performed much better than the conventional SVDD classification for one-class classification. For wheat class, the overall accuracy of WVS-SVDD based classification was 89.25%, which was little less than that derived from the SVM classification. Related to the traditional SVDD method, the advantage of WVS-SVDD could provide effective outliers which were located tightly around the target training samples, resulting in an efficient hypersphere to reject non-target land-cover classes. Thus the proposed method was able to significantly improve the overall accuracy, which was 8.92% higher than the traditional SVDD method.

The study has shown that classification accuracy increased accompanying with window size increment to some extent. Moreover, the proposed method was test over various spatial patterns, varying from 2.4-m to 20-m resolution. Producer’s accuracy increased from 71.12% to 94.71% as the land-cover classes became more homogeneous from 2.4-m to 20-m resolution. Under such improvements, the low commission error was maintained, which was benefited from the informative outliers. Therefore, the results highlighted the suitability of the proposed method over different spatial patterns.

The efficiency of untrained classes on the one- and multi-class classification was also analyzed in this paper. The results suggested that SVDD is considerably less sensitive to the effect of untrained classes, thus SVDD is suitable for specific land-cover mapping with the limited and small validation set, compared to multi-class classification which requires exhaustive training set to ensure classification accuracy. Besides, the WVS-SVDD can achieve good performance using window-based validation set at different window scales and different pixels’ spatial resolution. Although window-based validation set can be an efficient tool to determine optimal parameters, the performance of one-class classification of high-resolution images remains challenge due to the spectral heterogeneity. And, it is still involved a hard work to determine the validation set for WVS_SVDD.

5. REFERENCES

[1] C. Giri, B. Pengra, J. Long, and T. R. Loveland, “Next generation of global land cover characterization, mapping, and monitoring,” Int. J. Appl. Earth Obs. Geoinf., vol. 25, pp. 30-37, Dec. 2013.
[2] D. Pouliot, R. Latifovic, N. Zabcic, L. Guindon, and I. Olthof, “Development and assessment of a 250m spatial resolution MODIS annual land cover time series (2000–2011) for the forest region of Canada derived from change-based updating,” Remote Sens. Environ., vol. 140, pp. 731-743, Jan. 2014.
[3] O. L. Puertas, A. Brenning, and F. J. Meza, “Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975–2010),” Remote Sens. Environ., vol. 137, pp. 112-123, Oct. 2013.
[4] D. M. J. Tax and R. P. W. Duin, “Support vector data description,” Mach. Learn., vol. 54, no. 1, pp. 45-66, Jan. 2004.
[5] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Mari, and J. L. Rojo-Alvarez, “Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection,” IEEE Trans. Geosci. Remote Sens., vol. 46, no. 6, pp. 1822-1835, Jun. 2008.
[6] M. Marconcini, D. Fernandez-Prieto, and T. Buchholz, “Targeted land-cover classification,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 7, pp. 4173-4193, Jul. 2014.