Remote Sensing of Environment 113 (2009) 957–964

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Land cover mapping of large areas using chain classification of neighboring Landsat satellite images

Jan Knorn [a,*], Andreas Rabe [a], Volker C. Radeloff [b], Tobias Kuemmerle [a,b], Jacek Kozak [c], Patrick Hostert [a]

[a] Geomatics Lab, Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[b] Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706-1598, USA
[c] Institute of Geography and Spatial Management, Jagiellonian University, Gronostajowa 7, Kraków, Poland

ARTICLE INFO

Article history:
Received 30 May 2008
Received in revised form 14 January 2009
Accepted 17 January 2009

Keywords:
Landsat
Support Vector Machines
SVM
Carpathians
Forest classification
Large area mapping
Land cover and land use change
Chain classification

ABSTRACT

Satellite imagery is the major data source for regional to global land cover maps. However, land cover mapping of large areas with medium-resolution imagery is costly and often constrained by the lack of good training and validation data. Our goal was to overcome these limitations, and to test chain classifications, i.e., the classification of Landsat images based on the information in the overlapping areas of neighboring scenes. The basic idea was to classify one Landsat scene first where good ground truth data is available, and then to classify the neighboring Landsat scene using the land cover classification of the first scene in the overlap area as training data. We tested chain classification for a forest/non-forest classification in the Carpathian Mountains on one horizontal chain of six Landsat scenes, and two vertical chains of two Landsat scenes each. We collected extensive training data from Quickbird imagery for classifying radiometrically uncorrected data with Support Vector Machines (SVMs). The SVMs classified 8 scenes with overall accuracies between 92.1% and 98.9% (average of 96.3%). Accuracy loss when automatically classifying neighboring scenes with chain classification was 1.9% on average. Even a chain of six images resulted only in an accuracy loss of 5.1% for the last image compared to a reference classification from independent training data for the last image. Chain classification thus performed well, but we note that chain classification can only be applied when land cover classes are well represented in the overlap area of neighboring Landsat scenes. As long as this constraint is met though, chain classification is a powerful approach for large area land cover classifications, especially in areas of varying training data availability.

1. Introduction

Large area land cover maps derived from satellite images play a key role in global, regional and national land cover and land use assessments, carried out for example by the United Nations (UN), the Food and Agricultural Organization (FAO), or the United States Geological Survey (USGS) (Cihlar, 2000; Franklin & Wulder, 2002; Homer et al., 2004; Vogelmann et al., 2004). Such classifications allow assessments of broad-scale forest fragmentation (Riitters et al., 2002), carbon sequestration potential (Cruickshank et al., 2000; Niu & Duiker, 2006), or the Wildland Urban Interface (Radeloff et al., 2005). Therefore, large area land cover classifications present a basic prerequisite for many scientific applications (Wulder et al., 2008).

Landsat satellite data is the most widely used data type for land cover mapping because of its 35-year data record and its relatively high spatial resolution (Cohen & Goward, 2004; Wulder et al., 2008). Landsat data will become even more valuable as the Landsat Data Continuity Mission (NASA, 2008; Wulder et al., 2008) ensures future data availability. Decreasing costs, the availability of free Landsat data in the Geocover dataset (Tucker et al., 2004), the "Mid-decadal Global Land Survey" (Khatib et al., 2007) and the USGS' decision to provide free access to all Landsat data holdings offer opportunities for large area land cover classifications using Landsat imagery.

Unfortunately, Landsat image classifications are commonly conducted on one scene at a time, which limits the rapid analysis of large areas (Cihlar et al., 1998; Cihlar, 2000) and requires that adequate ground truth data are available for each scene. For large area classifications, three approaches have been proposed and tested before: single scene classification and subsequent mosaicking, mosaicking of images and subsequent classification of the image mosaic as a whole (Cihlar, 2000), and signature extension. In signature extension, a classifier is trained on one scene and the resulting signatures are applied to different scenes in space or time (Pax-Lenney et al., 2001). Signature extension is promising, but has to account for differences in topography, phenology, illumination, landscape variability, and atmosphere that result in spectral differences among images. Tests in northwest Oregon showed that accuracy declined by 8–13% (depending on the atmospheric correction method applied) when extending the classifier from an initial training image across space to nearby scenes (Pax-Lenney et al., 2001). Across northern Canada, classification accuracy dropped approximately by 50% when

* Corresponding author. Tel.: +49 30 2093 6846; fax: +49 30 2093 6848.
E-mail address: jan.knorn@geo.hu-berlin.de (J. Knorn).

0034-4257/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2009.01.010

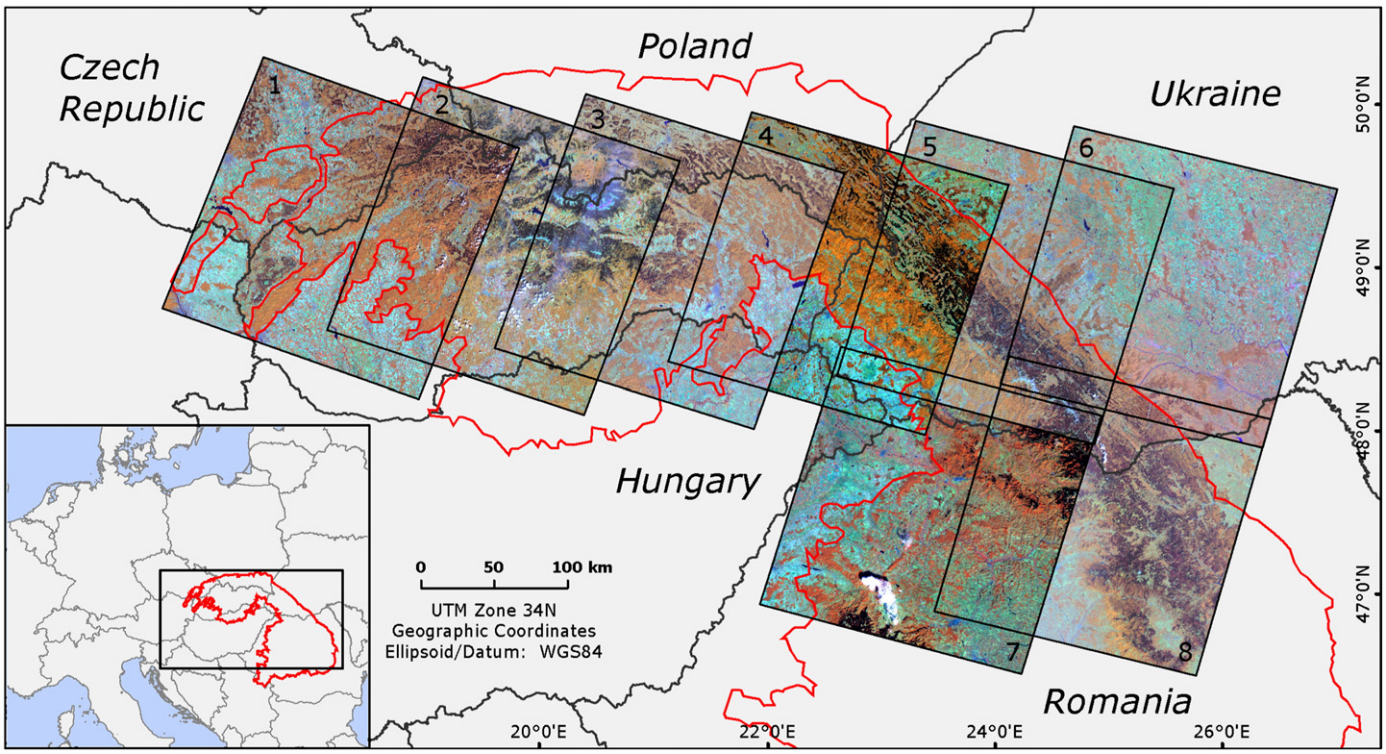

Fig. 1. The Carpathian mountain range and Landsat scenes in R–G–B bands 4–5–3 (border of the Carpathian Ecoregion overlaid in red). Numbering of Landsat scenes corresponds to numbering in the text (sources: GLCF, ESRI Data, Carpathian Ecoregion Initiative).

using signature extension for images that were about 1500 km apart (Olthof et al., 2005).

A promising approach for mosaicking images prior to classification is 'applied radiometric normalization' (Cohen et al., 2001). Here, the overlap area between neighboring Landsat images is used to extend information gained from a source image to neighboring images, thereby creating a seamless mosaic for the classification. The first step is to develop a relationship between the spectral measurements in the source image, and continuous forest variables, such as percent vegetation cover or stand age that are available from ground truth data (Cohen et al., 2001). The second step is to apply the regression equations that were developed, and predict the forest structure attributes across the entire source image. In the third step, the map with the predictions in the overlap area is used as ground truth to develop new regression equations for the neighboring image, which has most likely different phenology and atmospheric conditions. In the fourth step, these regression equations are then applied to the entire neighboring image. The resulting map of continuous forest structure attributes for the entire study area can then be classified into different forest types. When testing this approach in a 73,000 km^2 study area in western Oregon based on two Landsat TM source images, estimates for four forest cover attributes resulted in an overall accuracy of 66% (Cohen et al., 2001).

Signature extension and the mosaicking of images prior to classification have great potential for classifying large areas using Landsat imagery. However, they require considerable effort to match multiple images radiometrically. Here, we propose a new approach to large area land cover classification that fills a gap between single scene classification on one hand and signature extension or mosaicking on the other hand. We suggest the term 'chain classification' for this method.

Chain classification is similar to applied radiometric normalization in that it uses the overlap area among neighboring Landsat scenes, but we propose classifying one initial scene and then using the classification in the overlap area to train a classifier for a neighboring image. Once the second image is classified, it can be used as a new initial scene to classify a third image and so forth. One potential advantage of chain classification is that it does not require atmospheric correction or regression matching of scenes to account for radiometric differences. It can be applied both in horizontal directions (across track), and in vertical direction (along track). Furthermore, large area land cover maps often cover several countries or different land ownership regimes. The availability of spatially well distributed training and validation data is often limited in such situations. Chain classification may offer a solution to this problem by using the image with the best available ground truth data as the starting image in the image chain, and by providing training data for neighboring images from the image chain itself.

In principal, any classification algorithm could be used for chain classification. However, Support Vector Machines (SVMs), a fairly recently developed non-statistical classifier based on machine learning theory (Vapnik, 1999) offer some method-inherent advantages. Comparisons with other classification algorithms show that SVMs outperform or are at least as accurate as other parametric or non-parametric classifiers (Huang et al., 2002; Pal & Mather, 2005; Dixon & Candade, 2008).

SVMs are able to separate complex classes (Melgani & Bruzzone, 2004) such as in forest change analysis (Huang et al., 2008). In the SVM, the location of decision boundaries for optimal class separation is determined using kernel functions representing non-linear decision surfaces (Pal & Mather, 2005; Vapnik, 1995). By constructing the optimum hyperplane in feature space between two classes, an SVM is a binary classifier focusing on the classes of interest only. To determine this hyperplane, only the edges between the class distributions are described based on a relatively small amount of training data (Foody & Mathur, 2004; Foody et al., 2007; Mathur & Foody, 2008).

In summary, the overarching goal of this study was to develop a simple, robust, and reproducible method for large area land cover classification with minimal requirements for image pre-processing and training data. To do so, we tested chain classification of forest and

Table 1
Landsat images used in this study.

Id	Path/row	Acquisition date
1	189/26	08/02/2000
2	188/26	05/26/2001
3	187/26	08/20/2000
4	186/26	06/10/2000
5	185/26	06/03/2000
6	184/26	08/21/2002
7	185/27	08/22/2000
8	184/27	07/04/2002

non-forest based on the overlapping areas between Landsat scenes in the Carpathian Mountains in Eastern Europe.

2. Data and methods

2.1. Study area

We selected the Carpathians as a study area to test chain classification. The Carpathians represent a fairly homogeneous ecoregion with mostly similar environmental conditions. However, the study area includes seven countries with significant differences in forest type, non-forest land cover classes, geology, and land use patterns, and exhibits elevation-dependent vegetation gradients. This variability generates an interesting test case to investigate the feasibility of chain classification.

The Carpathians are located in central Europe and include parts of Czech Republic, Slovakia, Poland, Ukraine, Hungary, and Romania (CERI, 2001, Fig. 1). The study area covers about 185,000 km^2. The climate of the Carpathians is temperate-continental. Geology varies from Carpathian flysh, consisting of sandstone and shale layers, sedimentary rocks (mainly limestone), to a variety of crystalline rocks. Elevations range from 300 m to over 2,000 m in the alpine belt of the Tatra Mountains and the Southern Carpathians (KEO, 2007).

The forests of the study area are a patchwork of deciduous, coniferous, and mixed stands, with pronounced vegetation zones along the elevation gradient (KEO, 2007). Mixed deciduous forests, dominated by pedunculate oak (*Quercus robur*), lime (*Tilia cordata*) and hornbeam (*Carpinus betulus*), dominate the foothill zone. European beech (*Fagus sylvatica*), silver fir (*Abies alba*), Norway spruce (*Picea abies*) and sycamore (*Acer pseudoplatanus*) are typically found in the montane zone (Perzanowski & Szwagrzyk, 2001). In some places, the montane zone is almost solely covered by conifers, especially spruce plantations. At the timberline (~1500 m), stone pine (*Pinus cembra*) stands exists (KEO, 2007). Overall, about 60% of the Carpathian ecoregion is covered by forest (KEO, 2007). A history of intense land use affected most forests, transforming the landscape into a complex pattern of forests, arable land, and pastures, varying significantly between countries and regions (Turnock, 2002; Kuemmerle et al., 2006; Kozak et al., 2008). In particular, the foothill zones and plains are dominated by agricultural land use and forests are only small and scattered.

2.2. Satellite data and pre-processing

We used the optical bands of 9 Landsat Enhanced Thematic Mapper Plus (ETM+) images recorded between 2000 and 2002 to test chain classification (Table 1). Eight images were provided by the University of Maryland Global Land Cover Facility (GLCF), and one Level 1G scene (186/26) was purchased because of cloud coverage in the GLCF data. A post-processed digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) was acquired from the GeoPortal provided by the Consortium for Spatial Information within the Consultative Group on International Agricultural Research (CGIAR-CSI; Jarvis et al., 2006) and resampled to 30 m to match the resolution of the Landsat images. We orthorectified the additional 186/26 image using space resection based collinearity equations. The corresponding GLCF image of 2000 served as a base map for automatic image-matching. 359 evenly distributed ground control points (GCPs) with an overall root-mean-squared-error (RMSE) of <0.5 were selected using an Automatic Point Measurement software tool (Leica Geosystems, 2006). The image was rectified to Universal Transverse Mercator (UTM) zone 34 and the World Geodetic System (WGS) 84 datum and ellipsoid. The images 189/026, 184/026 and 184/27 were reprojected to UTM zone 34. We resampled all images to 30 m resolution using nearest neighbor resampling to ensure consistency among images. For the GLCF images, the RMSE-based geodetic accuracy is <0.5 pixels (Tucker et al., 2004). We did not screen for haze or disturbance factors other than clouds and no radiometric correction was applied. Clouds and cloud shadows were digitized and masked out for the analysis.

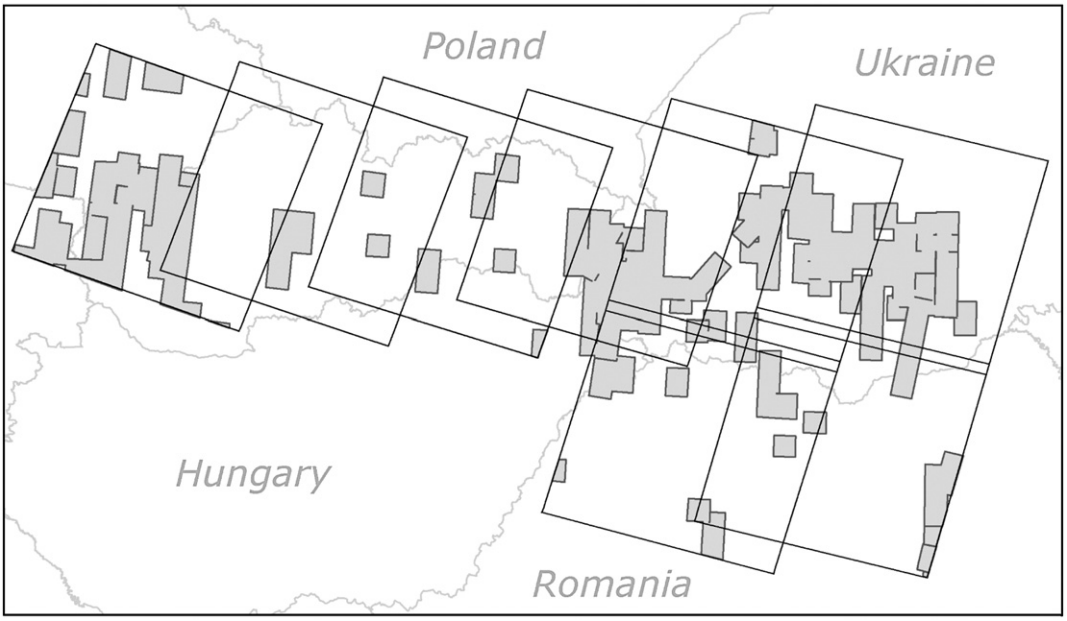

Fig. 2. Distribution of high resolution Quickbird data (gray polygons) from Google Earth™.

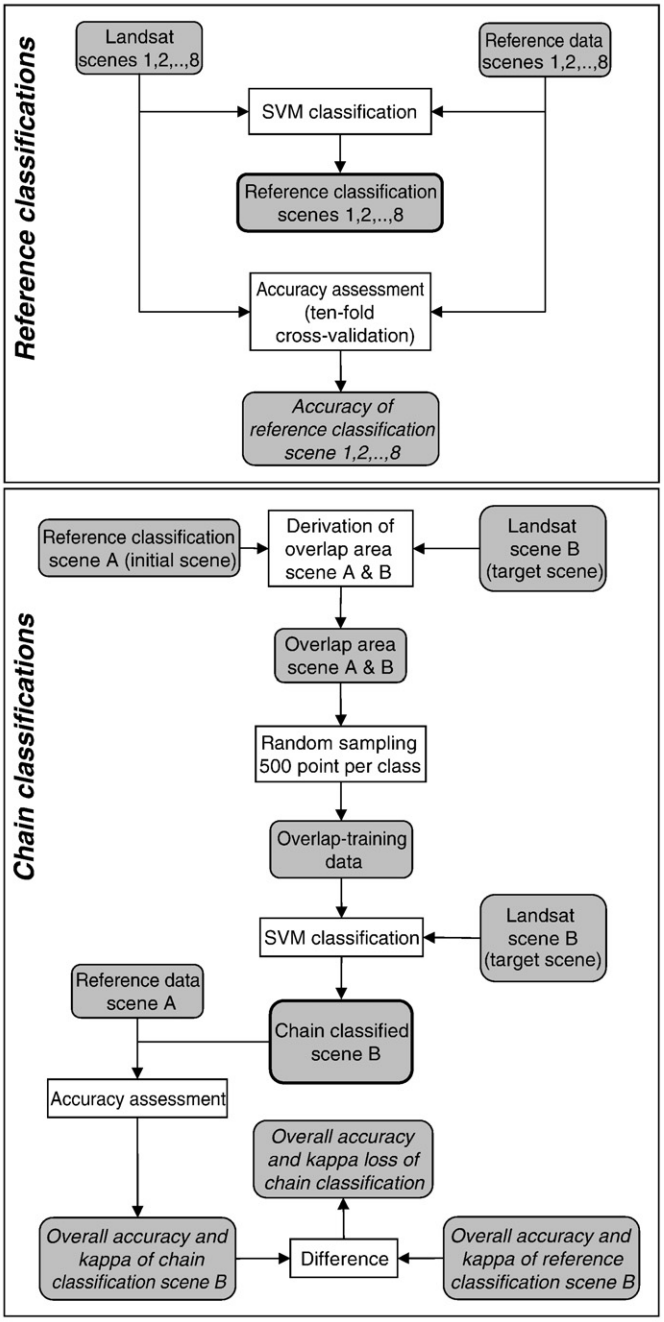

Fig. 3. Processing scheme for chain classifications. Top: Derivation of the reference classifications. Bottom: Chain classification procedure.

2.3. Image classification with Support Vector Machines

SVM classification is based on delineating two classes by fitting an optimal separating hyperplane to the training samples. The hyperplane is constructed by maximizing the margin between class boundaries and is described by a subset of the training samples, the so-called 'support vectors' (Boser et al., 1992; Cortes & Vapnik, 1995; Foody et al., 2007). SVMs need training data that optimize the separation of the classes rather than describing the classes themselves (Foody & Mathur, 2006).

Using a radial basis function, class distributions with non-linear boundaries can be mapped into a high dimensional space for linear separation (Huang et al., 2002). Training the SVM with a Gaussian radial basis function requires setting two parameters: C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the training error, while γ describes the kernel width. A small C-value tends to emphasize the margin while ignoring the outliers in the training data, while a large C-value may overfit the training data. A comprehensive description of SVMs can be found in Burges (1998) and Cristianini and Shawe-Taylor (2000). Detailed introductions in a remote sensing context are provided in Huang et al. (2002), and Foody and Mathur (2004). Training, classification, and accuracy assessment were carried out using imageSVM (Janz et al., 2007), an IDL/ENVI based tool for SVM classification of remote sensing images using the LIBSVM version 2.84 (Chang & Lin, 2001).

2.4. Training and validation data

Training and validation data, here referred to as 'reference data', was collected using Quickbird images in Google Earth™ (http://earth.google.com). About 160 Quickbird images acquired between 2002 and 2007 were available in Google Earth™ (Fig. 2) covering approximately 24% of our study area.

Reference data were collected using a random sampling design (Wang et al., 2005; Lee & Huang, 2007). A random sample of 1400 reference points per Landsat image was selected within the area covered by Quickbird imagery in each scene. We chose this number of points, after initial tests based on learning curves showed that selecting more than 500 points per class did not improve classification accuracy significantly. Points were visually classified as either forest or non-forest. The forest class in our study refers to forest as land cover and includes primary forests as well as plantations, all forest types in the study area (deciduous, mixed, coniferous forests) and all age classes. All other land cover types (e.g. settlements, cropland, pastures, and water) were defined as the non-forest class. Because SVMs can delineate multi-modal classes in feature space, we did not have to separate the non-forest training data into individual land cover classes. All reference points were also cross-checked visually on the Landsat images to account for changes that occurred between the acquisition dates of Landsat and Quickbird images. Points indistinct in Quickbird or covered by clouds were rejected from the analysis. At most, 3% of the random samples were removed.

In a first step, all scenes were classified individually based on the reference data that had been collected in each scene. We used these classifications as the benchmark against which we compared the chain classification results, and refer to the individual and independent classifications as 'reference classifications' (Fig. 3 top). We used cross-validation to obtain a robust estimate of the accuracy of these reference classifications (Steele, 2005). Using ten-fold cross-validation, we split all available ground truth points into training (90%) and validation (10%) samples and then classified each image 10 times for all 10 possible splits. Based on each classification, an error matrix, overall accuracy, user's and producer's accuracy, and kappa were calculated (Congalton, 1991; Foody, 2002). The derived accuracy measures for each classification were then averaged to calculate mean error estimates (Friedl & Brodley, 1997). The final classification was

Table 2
Accuracy assessment for reference classifications.

Id	Overall (%) accuracy	Kappa value	User's accuracy (%)		Producer's accuracy (%)	
			Forest	Non-forest	Forest	Non-forest
1	96.76	0.926	95.74	97.28	94.41	97.92
2	92.10	0.842	91.76	92.68	93.24	90.87
3	97.48	0.950	97.86	97.13	97.14	97.83
4	97.04	0.938	95.40	98.17	97.18	96.94
5	96.34	0.927	95.89	96.90	96.82	95.86
6	98.93	0.977	98.34	99.28	98.68	99.07
7	95.20	0.886	90.96	97.22	93.41	95.96
8	96.23	0.918	96.14	96.59	98.04	93.15

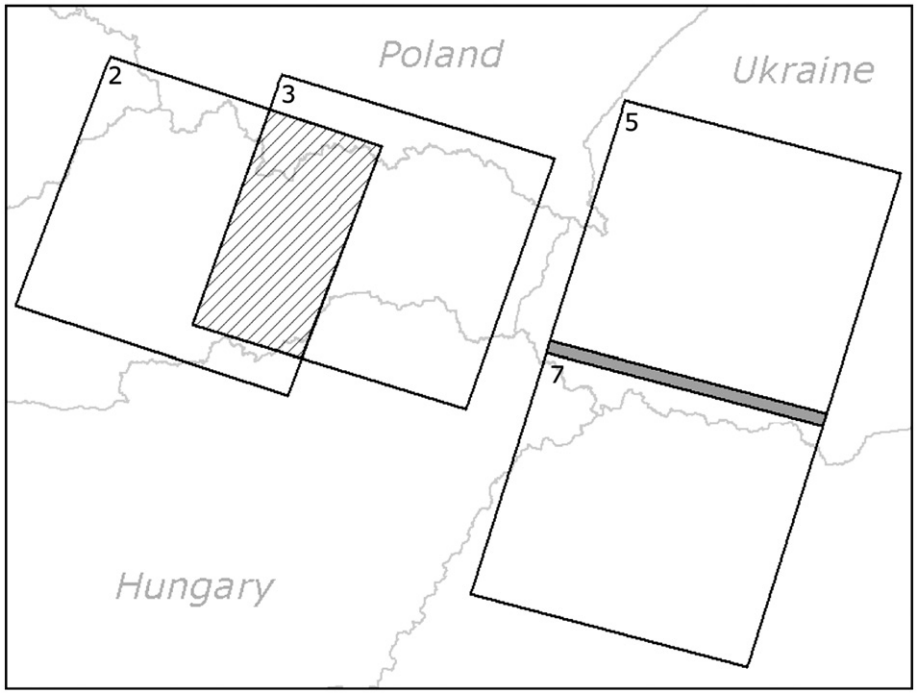

Fig. 4. Neighboring scenes with respective overlap areas. Striped: overlap area between two scenes across track; gray: overlap areas between two scenes along track.

based on an SVM trained with 100% of the ground truth data, and the mean error estimate is thus a conservative estimator of the true accuracy (Burman, 1989). Overall accuracies based on the ten-fold cross-validation ranged from 92.10% for scene 2 to 98.93% for scene 6, with an average of 96.26% (Table 2).

2.5. Chain classifications using SVMs

The first step in the chain classification was to identify the overlapping area between a reference classification serving as an initial scene and the neighboring scene to be classified (Fig. 4). Within the overlap area, 500 training points each for forest and non-forest were randomly selected. We chose 500 training points, after initial tests based on learning curves showed that selecting more training points in the overlap area did not improve classification accuracy. All training points were at least two pixels apart from forest–non-forest boundaries to account for geometric uncertainty between GLCF scenes (Tucker et al., 2004). The training data from the overlap area formed the input for the SVM classification of the neighboring target scene (Fig. 3 bottom). This procedure was repeated along a chain of overlapping, neighboring scenes. Each former neighboring scene served as a new initial scene in the next step along the chain.

2.5.1. Across track chain classification

Across track chain classification examined Landsat image neighbors in East–West direction. A total of 10 different chain classifications were possible among the six images in the northern row (Fig. 1). At the latitude of our study area, the size of the overlap area between two Landsat scenes across track was about 12,000 km^2, equaling 35% of a scene.

We also classified chains of more than one neighboring scene. The longest chain included five scenes that were classified based on one initial scene (Fig. 1). We expected decreasing chain classification accuracy with increasing chain length.

2.5.2. Along track chain classification

Four tests were possible to test chain classification along track (Fig. 1). The average along track overlap area was 3800 km^2, equaling 11% of the scene area. We expected that classification accuracy would be lower compared to across-track chains, given this rather small overlap area.

Table 3
Results of across track chain classification.

Test A	A.L.	K.L.	Test B	A.L.	K.L.	Test C	A.L.	K.L.	Test D	A.L.	K.L.	Test E	A.L.	K.L.
1–**2**	4.50	0.0882												
2–**3**	1.05	0.0214	1–2–**3**	2.41	0.0486									
3–**4**	1.43	0.0315	2–3–**4**	2.38	0.0513	1–2–3–**4**	2.08	0.0444						
4–**5**	0.71	0.0174	3–4–**5**	4.10	0.0852	2–3–4–**5**	4.45	0.0920	1–2–3–4–**5**	3.07	0.0645			
5–**6**	0.26	0.0091	4–5–**6**	0.53	0.0147	3–4–5–**6**	0.26	0.0090	2–3–4–5–**6**	0.66	0.0175	1–2–3–4–5–**6**	0.82	0.0217
6–**5**	2.95	0.0624												
5–**4**	2.33	0.0514	6–5–**4**	6.07	0.1367									
4–**3**	0.84	0.0172	5–4–**3**	1.62	0.0329	6–5–4–**3**	3.55	0.0713						
3–**2**	3.24	0.0628	4–3–**2**	3.10	0.0601	5–4–3–**2**	3.59	0.0366	6–5–4–3–**2**	13.20	0.2581			
2–**1**	1.82	0.0464	3–2–**1**	2.27	0.0566	4–3–2–**1**	0.88	0.0242	5–4–3–2–**1**	2.10	0.0534	6–5–4–3–2–**1**	9.39	0.2398
Mean	1.91	0.0408	Mean	2.81	0.0607	Mean	2.47	0.0341	Mean	4.76	0.0984	Mean	5.11	0.1307

A.L. = accuracy loss (%) of chain classified target scene compared to the respective reference classification; K.L. = kappa value loss of chain classified target scene compared to the respective reference classification. Target scenes in bold.

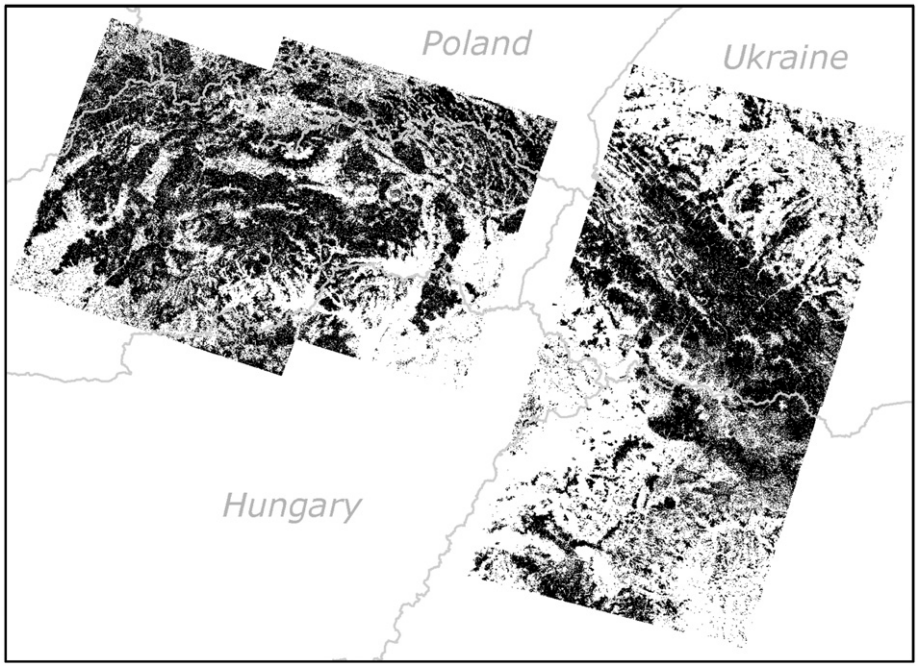

Fig. 5. Results of two chain classifications (forest in black). Scenes 2 and 5 classified initial scenes; scenes 3 and 7 chain-classified target scenes.

2.5.3. Across track chain classification based on two classified initial scenes

Last but not least, we tested if chain classification results would improve if the middle image in a chain of three images was classified based on training data from the two images at the ends of the chain. We expected that chain classification would perform better for the centered scene, because not only one but both overlap areas were used for chain classification.

The accuracy of the chain classifications was assessed independently for each scene, using the available reference data for the respective target scene. Ultimately though, the absolute accuracy was less important to us than the loss in accuracy that was caused when classifying a scene using chain classification. The performance of chain classification itself was assessed calculating the overall accuracy loss and kappa loss between a chain-classified target scene and its respective reference classification (Fig. 3 bottom). For example, in test 1–2–3–4 (Table 3 — Test C), scene 1 was the initial scene, scenes 2 and 3 were intermediate chain classifications, and scene 4 was the chain-classified target scene. The overall accuracy loss and kappa loss were calculated by comparing the accuracy of the chain-classified scene 4 and the reference classification of scene 4.

Additionally, we calculated the pixel-wise agreement between two individually derived chain classifications for the same target scene to indirectly evaluate the chain classification performance. If, for example, scene 2 was the target of a chain classification that started from scene 1 and from scene 3, respectively, pixel-wise agreement was calculated as the agreement of the two resulting scene 2 classifications.

Table 4
Overall agreement (O.Ag.) (%) between two individually derived chain classifications of the same target scene (bold) across-track.

Test A	O.Ag.	Test B	O.Ag.
1–**2**	84.95	1–2–**3**	95.94
3–**2**		5–4–**3**	
2–**3**	95.00	2–3–**4**	88.80
4–**3**		6–5–**4**	
3–**4**	92.88		
5–**4**			
4–**5**	93.30		
6–**5**			
Mean	91.53	Mean	92.37

3. Results

3.1. Across track chain classifications

The results for chain-classifying the direct neighbors had overall accuracy losses ranging from 0.26% for scene 5 to 6 (kappa loss 0.0091) to 4.50% for scene 1 to 2 (kappa loss 0.0882) with an average of 1.91% (kappa loss 0.0408) (Table 3 — Test A; Fig. 5). Both tests including scene 2 as a target scene, resulted in the highest overall accuracy and kappa losses.

Accuracy and kappa loss tend to increase as more scenes were added to a classification chain (Table 3). Average overall accuracy loss ranged from 1.91% (kappa loss 0.0408) for two scenes in a chain up to 5.11% (kappa loss 0.1307) for six scenes in a chain.

The best pixel-wise overall agreement between two different chain classifications for one scene from either neighbor was achieved for scene 3 (95%, Table 4 Test A). The chain classification of scene 2, on the other hand, exhibited only 84.95% in agreement. Average pixel-wise agreement was 91.53% (Table 4). The same tests, but with three scenes in a chain resulted in pixel-wise agreements of 95.94% with scene 3 as a target scene, and 88.80% with scene 4 as a target scene (Table 4 Test B). Resulting average agreement was 92.37%.

Table 5
Accuracy-and kappa-losses along track.

Test	A.L.	K.L.
5–**7**	0.36	0.0107
7–**5**	1.98	0.0427
6–**8**	3.76	0.0808
8–**6**	0.33	0.0105
Mean	1.60	0.0362

Target scenes in bold.

Table 6
Results of across track chain classification based on two classified initial scenes.

Test	A.L.	K.L.
1–2–**3**–4–5	2.12	0.0429
2–3–**4**–5–6	0.78	0.0189
Mean	1.45	0.0309

Target scenes in bold.

3.2. Along track chain classification

Contrary to our expectation, along track chain classification outperformed across track chain classification. Chain classification using only the small portion of overlap between scenes in North–South direction had an average overall accuracy loss of 1.60% (kappa loss 0.0362) (Table 5). Highest overall accuracy loss occurred using chain classification from scene 6 to 8 with 3.76% (kappa loss 0.0808) and lowest from scene 8 to 6 with 0.33% (kappa loss 0.0105).

3.3. Across track chain classification based on two classified initial scenes

Chain classifications of a target scene based on its two overlap areas with neighboring scenes did not result in notably higher classification accuracies (Table 6). A test based on scene 3, for example, results in an overall accuracy loss of 2.12% (kappa loss 0.0429), while for scene 4 the overall accuracy loss is only 0.78% (kappa loss 0.0189).

4. Discussion

We tested 'chain classification', a new approach to classify land cover for large areas that uses a classification in one image to train a classifier for a neighboring image. The average loss of accuracy when comparing *across track chain classifications* to reference classifications was only 1.91% with two images in a chain, 2.81% with three, 2.47% with four, 4.76% with five and 5.11% with six scenes, respectively. Average pixel-wise agreement between two individually derived chain classifications of the same scene across track was 91.53% for two scenes and 92.37% for three scenes in a chain. These outcomes highlight that chain classification is a robust way to map land cover across several scenes. Chain classification across track works well with up to four scenes in a chain in our case (Table 3 — Tests A to C). Outliers in these tests followed no decisive pattern and are likely due to irregularly distributed reference data (Fig. 2).

The small average loss of accuracy of *along track chain classification* suggests that our approach works also well in North–South direction even though the overlap area is fairly small in this case. The variance within the accuracy losses showed again no clear pattern.

Across track chain classification based on two classified initial scenes did not significantly improve with scene 3 as a target scene compared to the across track chains 1–2–3 (overall accuracy loss 2.41%) and 5–4–3 (overall accuracy loss 1.62%) (Tables 3 and 6). However, when scene 4 was the target scene, two-sided chain classification did perform better compared to the corresponding tests 2–3–4 (overall accuracy loss 2.38%) and 6–5–4 (overall accuracy loss 6.07%). These results suggest that classification accuracy can be enhanced by using two overlap areas for the target scene when one-sided chain classification does not yield optimum results. This strategy may furthermore be considered when a target scene is located in a heterogeneous landscape and is not well described by the training data of only one overlap area.

As is the case for all land cover classifications, training data should be well distributed and cover all characteristic landscape features across a scene. Due to the limited amount of Quickbird images in some regions of our study area and the lack of other ground truth data, reference data could not always be acquired with optimal distribution (Fig. 2). The higher variance in accuracy loss within the across-track tests are hence likely due to inhomogeneously distributed reference data.

Overall accuracy of the reference classifications was 96.26% on average. The stable performance across scenes with a limited amount of training data supported our assumption that SVMs are an appropriate classifier for chain classification. Nevertheless, chain classification is not restricted to SVM as a classifier and may be applied with other classifiers as well.

In chain classification, each scene is classified individually, with training data that is unique to this scene, and which is derived from the classified overlap area with its neighboring scene. This means that no signature extension from one scene to another is carried out and radiometric correction or normalization procedures are not required. This greatly simplifies image processing and eliminates a potential error source. Examining illumination differences among scenes, chain classification in combination with SVMs was robust and no illumination influence was found. It is important to note though that successful chain classification in mountainous areas requires accurate orthorectification to account for differences in viewing geometry between neighboring scenes. And screening and masking of clouds and cloud shadows is necessary to avoid gross classification errors.

The good performance of chain classification in various tests suggests that it is a valuable approach for large area land cover classifications with Landsat data. However, chain classification is by no means the only approach for this task, and it is important to understand advantages and disadvantages. In signature extension, radiometric calibration or normalization between images is an important preprocessing step (Pax-Lenney et al., 2001). This is not necessary when applying chain classification. The advantage of signature extension though, is the ability to extend classifications between single images over large distances (Olthof et al., 2005), i.e. neighboring scenes are not mandatory.

On the other hand, mapping large areas based on single scene classification presumes extensive reference data for each image as well as comprehensive resources for individual scene labeling (Cihlar, 2000). Chain classification uses only a small part of the image dataset to classify large areas. Therefore, an interpreter can focus on few well-classified initial images from where the rest is chain-classified. However, unlike large area classification based on a single scene or image mosaics, chain classification can only be applied to regions that share most spectral features.

The 'applied radiometric normalization' method developed by Cohen et al. (2001) is also based on the use of overlap areas between scenes. However, here statistical models translate the desired attributes of interest from the source image to the destination image. In comparison, chain classification based on SVMs, a faster and more straightforward process of land cover mapping, requires no radiometric adjustments. Based on the available reference data, SVMs generate an initial classification that serves for training the neighboring scene in the overlap area.

In summary, chain classification is a promising new tool for large area land cover classification. The approach is simple in that it only requires accurate georeferencing of scenes and no atmospheric correction. The accuracy loss of our classification was low (about 5% or less), even when long chains were classified. Chain classification is particularly well suited in areas where training data is only available for few scenes. This is the case, for example, in areas of different forest ownership and hence different base map availability (e.g. state forest versus private forest), in many remote areas, or in places that are inaccessible or lack high resolution information for any other reason. Chain classification is much faster and lower in work load than single scene classification, but more limited in the total area that can be classified compared to signature extension and mosaic classification. Chain classification is not restricted to any sensor as long as enough overlap area between scenes is provided and the land cover is homogenous across the images. An issue to be considered in future

chain classification approaches is the use of a "hybrid" training sample collection to overcome the limitations set by the need of homogeneity. Here, in the case of missing or insufficient representation of classes in the overlap area, additional training samples would be collected manually in the target scene and added to the SVM chain classification procedure. Chain classification could also be used to classify images from different sensors, radiometric or geometric resolution, in the same chain as long as enough overlap area exists. In this study we applied chain classification to assess forest cover, but any other cover type can be considered as well, providing the potential for further applications in the context of large area land cover mapping.

Acknowledgements

We gratefully acknowledge support by the Humboldt-University Berlin and by the NASA Land Cover and Land Use Change Program for this research. T. Kuemmerle is supported by the Alexander von Humboldt Foundation. S. Schmidt, M. Ozdogan, and two anonymous reviewers provided valuable comments that greatly improved this manuscript.

References

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In D. Haussler (Ed.), *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* (pp. 144—152). Pittsburgh, PA: ACM Press.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*, 121—167.

Burman, P. (1989). A comparative-study of ordinary cross-validation, v-fold cross-validation and the repeated learning–testing methods. *Biometrika*, *76*, 503—514.

CERI (2001). The status of the Carpathians. *A report developed as a part of The Carpathian Ecoregion Initiative* Vienna: WWF International.

Chang, C. -C., & Lin, C. -J. (2001). *LIBSVM: A library for support vector machines.* Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm [accessed 20th March 2008].

Cihlar, J. (2000). Land-cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing*, *21*, 1093—1114.

Cihlar, J., Xia, Q. H., Chen, J., Beaubien, J., Fung, K., & Latifovic, R. (1998). Classification by progressive generalization: A new automated methodology for remote sensing multichannel data. *International Journal of Remote Sensing*, *19*, 2685—2704.

Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*, 535—545.

Cohen, W. B., Maiersperger, T. K., Spies, T. A., & Oetter, D. R. (2001). Modelling forest cover attributes as continuous variables in a regional context with Thematic Mapper data. *International Journal of Remote Sensing*, *22*, 2279—2310.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*, 35—46.

Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning*, *20*, 273—297.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge: Cambridge University Press.

Cruickshank, M. M., Tomlinson, R. W., & Trew, S. (2000). Application of CORINE land-cover mapping to estimate carbon stored in the vegetation of Ireland. *Journal of Environmental Management*, *58*, 269—287.

Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote Sensing*, *29*, 1185—1206.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*, 185—201.

Foody, G. M., Boyd, D. S., & Sanchez-Hernandez, C. (2007). Mapping a specific class with an ensemble of classifiers. *International Journal of Remote Sensing*, *28*, 1733—1746.

Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 1335—1343.

Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, *103*, 179—189.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, *26*, 173—205.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*, 399—409.

Homer, C., Huang, C. Q., Yang, L. M., Wylie, B., & Coan, M. (2004). Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering and Remote Sensing*, *70*, 829—840.

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*, 725—749.

Huang, C., Song, K., Kim, S., Townshend, J. R. G., Davis, P., Masek, J. G., et al. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, *112*, 970—985.

Janz, A., van der Linden, S., Waske, B., & Hostert, P. (2007). ImageSVM— A user-oriented tool for advanced classification of hyperspectral data using support vector machines. *EARSeL SIG imaging spectroscopy* Belgium: Bruges.

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2006). *Hole-filled SRTM for the globe Version 3.* Available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org [accessed 10th July 2007].

KEO (2007). Carpathians environment outlook report. *United Nations Environment Programme Division of Early Warning and Assessment— Europe. Geneva.*

Khatib, L., Gasch, J., Morris, R., & Covington, S. (2007). Local search for optimal global map generation using mid-decadal Landsat images. *Proceedings of AAAI Workshop on Preference Handling for Artificial Intelligence, Vancouver, BC.*

Kozak, J., Estreguil, C., & Ostapowicz, K. (2008). European forest cover mapping with high resolution satellite data: The Carpathians case study. *International Journal of Applied Earth Observation and Geoinformation*, *10*, 44—55.

Kuemmerle, T., Hostert, P., Perzanowski, K., & Radeloff, V. C. (2006). Cross-border comparison of land cover and landscape pattern in Eastern Europe using a hybrid classification technique. *Remote Sensing of Environment*, *103*, 449—464.

Lee, Y. J., & Huang, S. Y. (2007). Reduced support vector machines: A statistical theory. *IEEE Transactions on Neural Networks*, *18*, 1—13.

Leica Geosystems (2006). Imagine AutoSync™ White Paper. *Leica geosystems geospatial imaging, Norcross, USA.*

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

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 1778—1790.

NASA [National Aeronautic Space Administration] (2008). *Landsat data continuing mission [online].* Available from: http://ldcm.nasa.gov/index.htm [accessed 23th February 2008].

Niu, X., & Duiker, S. W. (2006). Carbon sequestration potential by afforestation of marginal agricultural land in the Midwestern U.S. *Forest Ecology and Management*, *223*, 415—427.

Olthof, I., Butson, C., & Fraser, R. (2005). Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. *Remote Sensing of Environment*, *95*, 290—302.

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, *26*, 1007—1011.

Pax-Lenney, M., Woodcock, C. E., Macomber, S. A., Gopal, S., & Song, C. (2001). Forest mapping with a generalized classifier and Landsat TM data. *Remote Sensing of Environment*, *77*, 241—250.

Perzanowski, K., & Szwagrzyk, J. (2001). Carpathian montane conifer forest [online] Available from: http://www.worldwildlife.org/wildworld/profiles/terrestrial/pa/pa0504_full.html [accessed 23th February 2008].

Radeloff, V. C., Hammer, R. B., Stewart, S. I., Fried, J. S., Holcomb, S. S., & McKeefry, J. F. (2005). The wildland–urban interface in the United States. *Ecological Applications*, *15*, 799—805.

Riitters, K. H., Wickham, J. D., O'Neill, R. V., Jones, K. B., Smith, E. R., Coulston, J. W., et al. (2002). Fragmentation of continental United States forests. *Ecosystems*, *5*, 815—822.

Steele, B. M. (2005). Maximum posterior probability estimators of map accuracy. *Remote Sensing of Environment*, *99*, 254—270.

Tucker, C. J., Grant, D. M., & Dykstra, J. D. (2004). NASA's global orthorectified landsat data set. *Photogrammetric Engineering and Remote Sensing*, *70*, 313—322.

Turnock, D. (2002). Ecoregion-based conservation in the Carpathians and the land-use implications. *Land Use Policy*, *19*, 47—63.

Vapnik, V. N. (Ed.). (1995). *The nature of statistical learning theory* New York: Springer-Verlag.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*, 988—999.

Vogelmann, J. E., Howard, S. M., Yang, L. M., Larson, C. R., Wylie, B. K., & Van Driel, N. (2004). Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, *67*, 650—662.

Wang, J. G., Neskovic, P., & Cooper, L. N. (2005). Training data selection for support vector machines. *Advances in natural computation, Pt 1, proceedings* (pp. 554—564).

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., et al. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment*, *112*, 955—969.