

# Satellite image based vegetation classification of a large area using limited ground reference data: a case study in the Usa Basin, north-east European Russia

Tarmo Virtanen, Kari Mikkola & Ari Nikula



Predicted global changes can be studied effectively by combining spatially explicit data sets on vegetation and other landscape properties with process models. However, detailed knowledge of the vegetation distribution of remote Arctic areas is relatively scarce. This paper shows how a mesoscale vegetation and land cover classification of a large, remote Arctic area can be conducted at a fine spatial resolution (30 m cell size) using a limited ground reference data set. The study area is the catchment of the River Usa (93 500 km<sup>2</sup>) in north-eastern European Russia. Vegetation zones in the Usa Basin range from taiga in the south to forest-tundra and tundra in the north, and to alpine in the Ural mountains in the east. Classification was done using a mosaic of spectrally adjusted Landsat TM5 images from five different dates and a semi-supervised method. Ground reference data were collected during the summers of 1998, 1999 and 2000. Accuracy of the 21-class vegetation type/land cover classification produced was tested against test points interpreted from oblique aerial photographs taken from a helicopter (logistic limitations prohibited the collection of representative ground reference data). The main vegetation types (forests, willow dominated stands and meadows, peatlands, tundra heaths, mainly unvegetated areas, and water bodies) were distinguished with relatively high accuracy: 84% of the test points were classified correctly. Spatially detailed land cover data sets like the one described here allow detailed landscape-level analysis and process modelling on many different subjects.

*T. Virtanen, Dept. of Biological and Environmental Sciences, University of Helsinki, Box 65, FI-00014 Helsinki, Finland, tarmo.virtanen@helsinki.fi; K. Mikkola & A. Nikula, Rovaniemi Research Station, Finnish Forest Research Institute, Box 16, FI-96301 Rovaniemi, Finland.*

Global changes, especially the predicted climate change, are expected to cause many changes in the nature of the Arctic regions. Changes in vegetation distribution, permafrost, hydrology and many other processes can be foreseen (Anisimov & Nelson 1996; Woodwell et al. 1998; Rupp et al. 2000; Skre et al. 2002; Van der Linden et al. 2003). These changes would, in turn, have feedback effects on the global climate system (Bonan et al. 1995; Betts 2000; Harding et al. 2002). Regional and global environmental changes can be studied

effectively by combining spatially explicit data sets on vegetation and other landscape properties with process models (Levin 1992; Running et al. 1994; Kittel et al. 2000; Plummer 2000; Rupp et al. 2000; Ranson et al. 2001; Van der Linden et al. 2003). However, present knowledge of the detailed vegetation distribution of the remote Arctic areas is relatively scarce (Walker et al. 1995; Rees et al. 2002). Remote sensing images and their classification provide possibilities to cover large areas at relatively low costs compared to traditional field

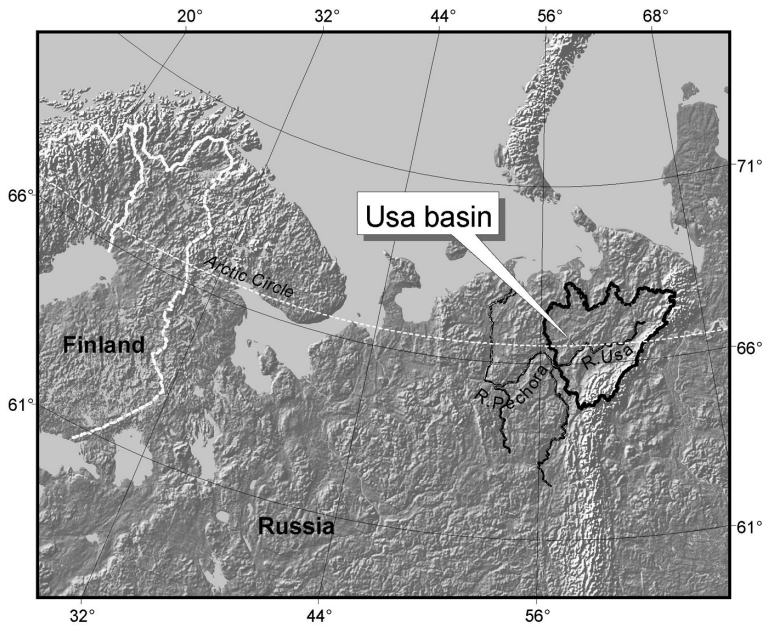


Fig. 1. Location of the Usa Basin, north-east Russia.

inventory based vegetation mapping (Walker et al. 1995; Barnsley et al. 1997; Cihlar 2000; Franklin & Wulder 2002; Rees et al. 2002).

In recent years some remote sensing based global data sets on vegetation have become available for the global change research community (Hansen et al. 2000; Loveland et al. 2000; Franklin & Wulder 2002; Bartalev et al. 2003). The minimum spatial resolution, i.e. the smallest distinguishable objects in the image, of these data sets is about 1 km. This is due to the resolution of the sensors (AVHRR, SPOT4-VEGETATION) on which these global data sets are based. To quantify, understand and model many ecological phenomena and processes, more spatially detailed data are often needed due to the fine-scaled patchiness of the vegetation (Levin 1992; Barnsley et al. 1997; Woodcock et al. 1997; Stow et al. 1998; Virtanen et al. 1998; Plummer 2000; Rees et al. 2002). Some vegetation types occur in such small patches that they cannot be recognized at  $1 \times 1$  km pixel resolution. Discrepancies between different global data sets have been detected in the transition zones between vegetation types especially, and this is at least partly due to the more fine-grained landscape mosaic characteristic of these areas (Hansen & Reed 2000). More detailed classification data like those presented in this paper could be used as a reference in evaluating present and future global land cover

data sets (Thomlinson et al. 1999; Hansen et al. 2000; Loveland et al. 2000; Franklin & Wulder 2002; Bartalev et al. 2003).

The recent improvements in computing power and increasing satellite image archives now enable production of spatially detailed classifications for large areas relatively quickly and efficiently (Homer et al. 1997; Cihlar 2000; Ma et al. 2001; Franklin & Wulder 2002; Rees et al. 2002). The most commonly used higher spatial resolution satellite images are captured by Landsat TM5 and ETM+ sensors that have a resolution of  $30 \times 30$  m, and data with a resolution of a few metres are already available, e.g. from the IKONOS satellite. In spite of the development of spatially finer resolution data, assigning relevant information about vegetation or land cover to spectral characteristics requires information on the vegetation and other land cover forms in the area. In remote areas where road networks are scarce or even completely absent, and the summer season is short and spectral properties of the plants change rapidly during that time, obtaining comprehensive ground reference data for classification is often not an easy task (Rees et al. 2002). Therefore, a hybrid of methods, such as the use of multi-source geographic and other information in classification, is often required instead of relying solely on ground reference data (Homer et al. 1997; Franklin & Wulder 2002).

The purpose of this paper is to present and discuss the methods and the results of a study in which a vegetation type and land cover classification was produced at a fine spatial resolution (30 m pixel size) for a large, remote Arctic area in northern European Russia where the possibilities of performing fieldwork are limited. The work is based on the classification of a mosaic of several Landsat TM5 images using relatively sparse ground reference data. The work was carried out within the framework of the European Commission-funded TUNDRA project, in which spatially explicit Geographical Information System (GIS)-based data sets have been built on the main vegetation types and their carbon stocks, soil properties, permafrost conditions and main climatic parameters of the catchment of the River Usa (Fig. 1). The accuracy of our results is evaluated, and the advantages of using this kind of spatially detailed data in studies on a range of subjects are discussed. The resulting classification data have in fact already been utilized in a number of studies: hydrological modelling (Van der Linden et al. 2003), soil carbon estimates (Kuhry et al. 2002), carbon fluxes (Heikkinen et al. 2004), vegetation changes during the Holocene (Kultti et al. 2003), river channel dynamics (Huisink et al. 2002) and forest line location (Virtanen et al. in press). Additional studies are in progress.

## Materials and methods

### *Study area*

The Usa Basin (93 500 km<sup>2</sup>, Fig. 1) is primarily located in the Komi Republic, except for some of the northern sections that extend up into the Nenets Autonomous Area. The region is unique in continental Europe in having such a broad, lowland tundra–taiga transition zone and extensive permafrost. The River Usa discharges into the River Pechora on the west side of the catchment. With elevations ranging from 300 to 1800 m, the Ural Mountains bound the area in the east and occupy approximately 15% of the area. The remaining part of the basin has an altitude of between 40 and 300 m, with most of the area lying below 200 m. About 75% of the Usa Basin is dominated by permafrost areas of various degrees of continuity, ranging from isolated patches in the south to continuous permafrost in the north (Oberman & Mazhitova 2003).

There is a south-west to north-east decline in the mean annual temperature in the Usa Basin lowlands; the temperature also decreases with elevation in the Urals. Mean annual temperatures for the period 1961–1990 varied from –2.5 °C in Pechora (located immediately to the south-west of the Usa Basin) to –6.1 °C in Vorkuta (located in the north-eastern part of the Usa Basin). The mean January and July temperatures were, respectively, –20.3 °C and +16.1 °C in Pechora and –21.2 °C and +13.0 °C in Vorkuta. Mean annual precipitation is around 550 mm in both Pechora and Vorkuta, but orographic precipitation results in higher values in the Urals. Meteorological data were obtained from the Komi Republican Centre for Hydrometeorology and Environmental Monitoring, Syktyvkar, Russia.

Vegetation zones in the Usa Basin range from taiga in the south to forest–tundra and tundra in the north. The northern part is covered by treeless tundra vegetation, the upland areas being occupied by dwarf shrub tundra vegetation with a well-developed lichen and/or moss layer. Peat plateau mires are common. Willow (*Salix* spp.) dominated, often paludified, vegetation occurs in depressions and river valleys. The central part of the Usa Basin consists of a mosaic of tundra and northern coniferous taiga forests, and the southern part belongs to the northern taiga forest zone (Kozubov et al. 1999). Large open mires are common in the lowlands of the taiga zone. The forest stands in lowland areas mainly consist of mixed forests dominated by spruce (*Picea obovata* Ledeb.). Pubescent birch (*Betula pubescens* Ehrh.) is the most frequent broadleaved tree. Scots pine (*Pinus sylvestris* L.) is rare and is only found tens of kilometres to the south of the spruce treeline. Pine mainly occurs around open mires, although there are some stands on sandy soils along the largest rivers in the western parts of the Usa Basin. In the alpine taiga zone, the forests consist of spruce, Siberian fir (*Abies sibirica* Ledeb.), larch (*Larix sibirica* Ledeb.) and birch (Gorchakovskii 1960). Areas above the treeline are characterized by patchy alpine meadows and dwarf shrub–lichen dominated tundra vegetation. The steep and rocky slopes and the highest altitudes are almost completely devoid of vegetation.

Apart from in the surroundings of a few villages, towns and industrial areas, there has been no forest cutting. Other human impacts on the vegetation are insignificant, with the exception of

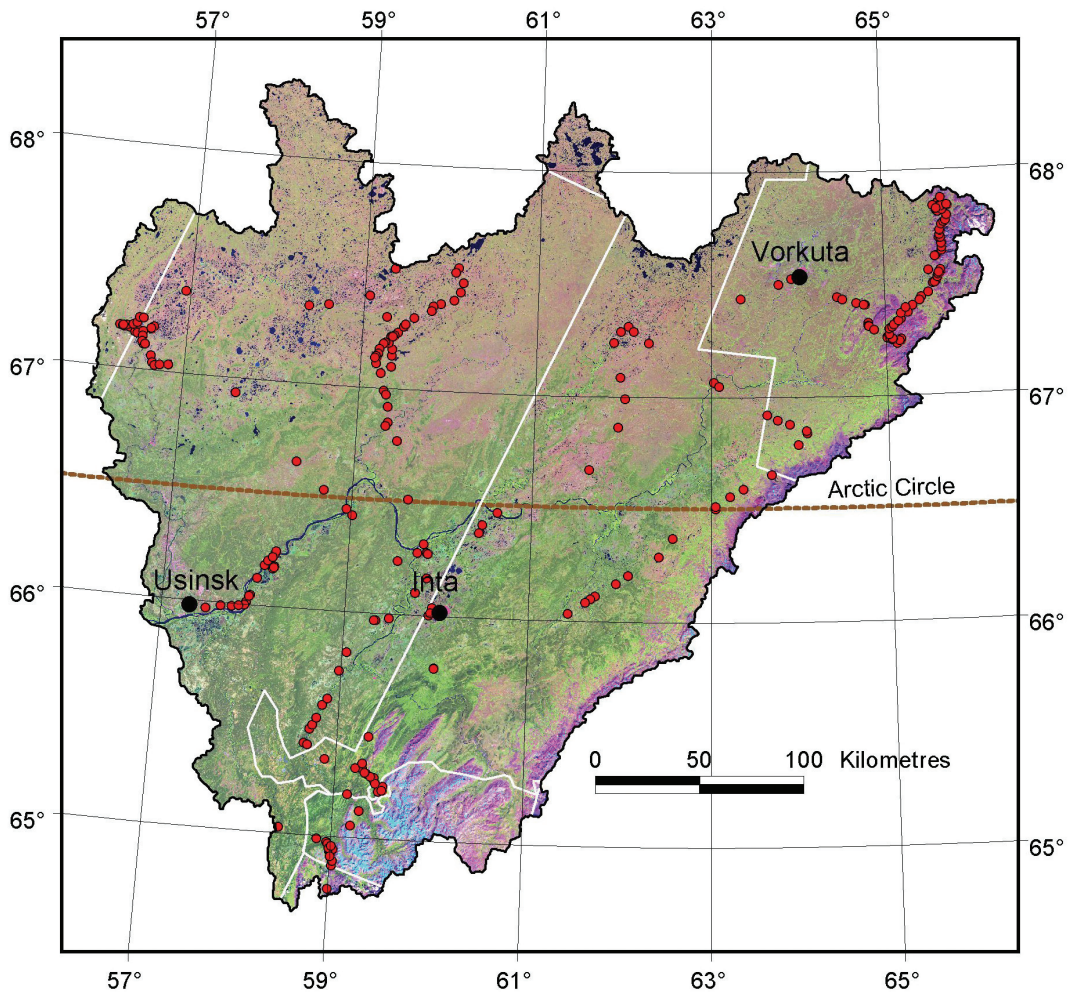


Fig. 2. Spectrally corrected Landsat TMS image mosaic for the Usa Basin. Channels 5, 4 and 3 are shown as red, green and blue. Areas covered by different images and the locations of the aerial oblique photographs (red dots) used to test the classification accuracy are indicated.

some vegetation changes around the industrial complex of Vorkuta (Virtanen et al. 2002). Apart from air traffic between the main cities (see Fig. 2), the only land connection is a railroad running through the region from the south-west to Vorkuta that continues over the Urals. There are no car drivable roads outside the surroundings of the three main cities, except for the 100 km long road from the town of Usinsk to the northern oil fields.

#### Fieldwork

During the fieldwork carried out in the summers of 1998 (4 weeks), 1999 (4 weeks) and 2000 (2

weeks), we collected a set of ground reference data from five different sub-areas covering all the main vegetation types in these areas (Fig. 3). This sub-area selection was partly determined by the research interests of the other research teams in the TUNDRA project. Most of the studied locations were reached by helicopter, except for sub-area 5, which was accessed by an all-terrain land vehicle, and sub-area 4, which was reached by boat. During the helicopter flights we took hundreds of oblique photographs (35 mm slides) with a 35 and 50 mm objective to be used in evaluating the accuracy of the classification (Fig. 2). Ground reference sites selected to present all the charac-

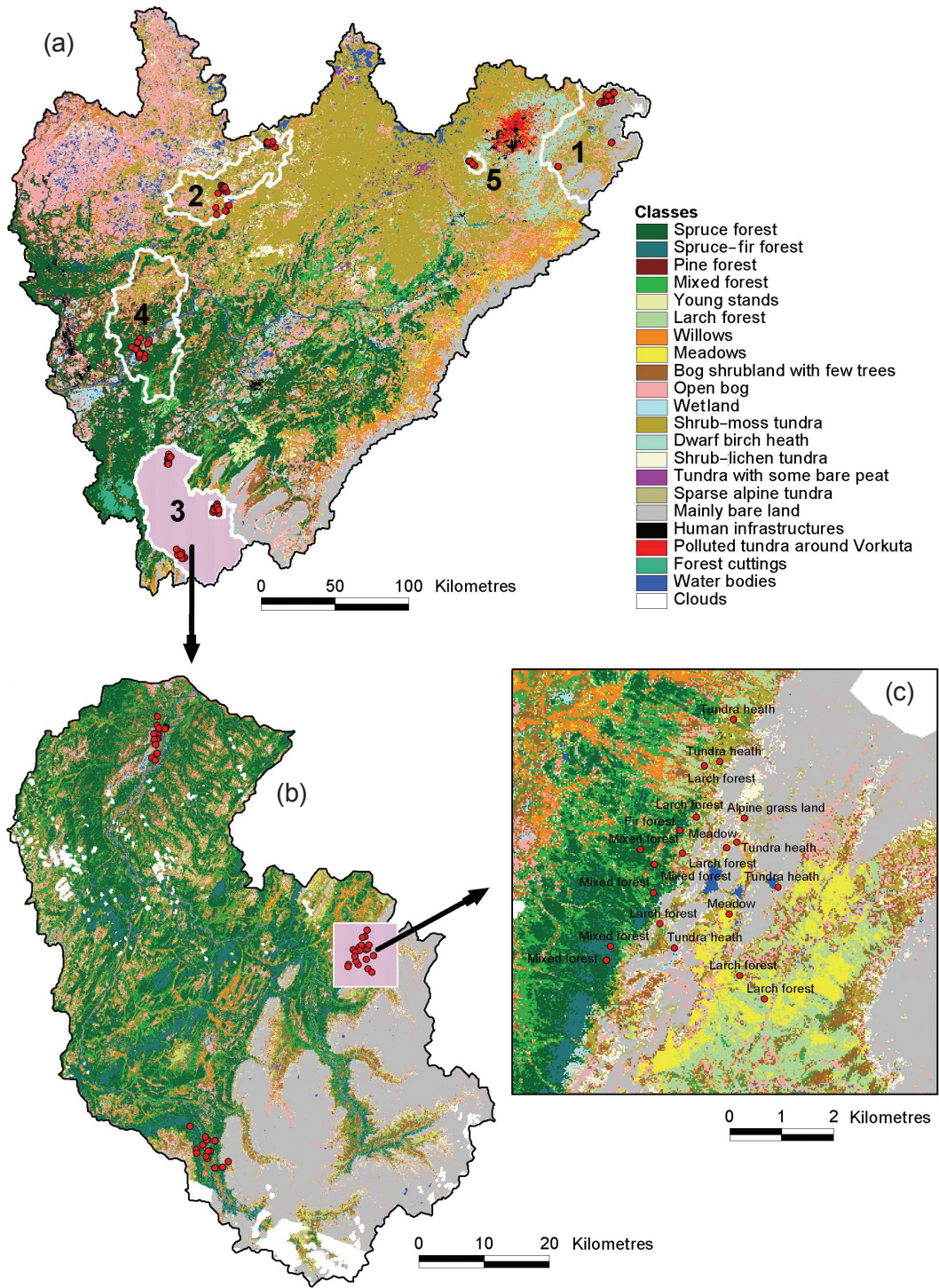


Fig. 3. (a) Classification for the whole Usa Basin resampled to a 500 m grid. Five sub-areas and ground reference locations (red dots) are indicated. (b) Classification of sub-area 3. (c) More detailed view of sub-area 3 showing individual grid cells. Ground reference locations in the area are indicated.

teristic vegetation types (avoiding mixed pixels) were located and marked on satellite image print-outs in the field. At each site (totalling 158), we estimated the percentage cover of different vegetation growth forms and the most common species and vegetation assemblages on three circular plots (10 m diameter), situated 30 m apart along a south–north line. The percentage cover was estimated visually in a vertical projection. On forested sites we also measured a number of stand parameters (tree species composition, basal area, number of stems and average height).

#### *Satellite image data*

We constructed an image mosaic based on Landsat 5 TM images (pixel size 30 m) to cover the whole Usa Basin (Fig. 2). Among the images acquired from the 1980s to the late '90s, 1988 had the largest cover of cloudless images over the study area and, therefore, the following images from five different days were acquired (ddmmyy, path, row[s]): 310788, 166, 13-14; 130788, 168, 12-15; 110788, 170, 12-14; 030888, 171, 13; 230695, 168, 14. Because of cloudiness, an additional image from 1995 was used for some parts of the southern area.

When creating the mosaic, the images were first georeferenced and rectified to UTM coordinates using Russian 1:200 000 digital map data (source: GOSGISCENTER, Moscow). The images from different dates were then spectrally standardized by multitemporal relative calibration using first or second order linear regression equations for each channel (see e.g. Olsson 1993). Regression equations were calculated for the pixel data values sampled from the overlapping areas of the images. The calibration rectangles, blocks of pixels of which values were used to calculate calibration equations, were chosen from deep water, coniferous forest (low near-infrared values) and rocky sites in order to avoid seasonal bias effects. The number of these few hundred pixel-sized rectangles was 5–7 per image pair. Channels 1 and 6 did not provide a good fit and were omitted from the classification. It is known that values of channel 1 (blue) are sensitive to varying atmospheric conditions, and calibration results between different day images are not as good as in other channels (Olsson 1993). Thermal channel 6 has coarser spatial resolution (120 m cell) than in other channels, and is typically not used in vegetation classifications (Price 1981; Homer et al. 1997).

Finally, the image mosaic was clipped according to the catchment area of the River Usa derived from a digital elevation model (DEM). This 100 m pixel DEM was calculated with ARC/INFO's TOPOGRID tool for the whole of the Usa Basin using the contours and hydrological layers of the Russian digital map data.

#### *Classification*

The classification work was carried out in a number of steps. First, the five sub-areas (Fig. 3) were each classified by a supervised method using bands 2–5 and 7. Spectral signatures for different vegetation types were derived by creating representative samples around ground reference points (N=158) with ERDAS IMAGINE's interactive seed pixel region growing technique (ERDAS 1997). We used the parallelepiped decision rule to assign the signature minimum–maximum limits in feature space for each vegetation class. Those pixels which fell outside the defined signature minimum–maximum limits (gaps in feature space between the signatures) were assigned zero values at this point. These unclassified pixel areas (10–20% of the total pixel count, depending on the area) were then visually inspected from the original image and several additional signature addition and classification rounds were done until only scattered pixels were left unclassified. The field knowledge and experience of the local landscape plus the scanned photographs taken in the field and during helicopter flights were of great help in this part of work.

In the second phase, signatures from the five sub-areas were pooled and the iterative process of classification of the whole Usa Basin was started. The first classification runs produced clearly erroneous results—for example, scattered spruce forests were found in many locations in the treeless tundra, and the human impacted tundra around the mining town of Vorkuta was classified to the category *mixed forest*. Using similar iterative approach as described above new signatures were added to the set and the procedure was repeated until there were only scattered unclassified pixels and visual inspection of the whole area showed no clear discrepancies. The final run to fill the remaining feature space gaps (unclassified pixels) was done using maximum likelihood as a parametric decision rule. The number of spectral classes increased during the process to around 100. These primary classes were then sub-

jectively combined to form 21 final classes. The insight and experience gained during the fieldwork, some general Russian thematic maps, general knowledge of the landscape structure, DEM and the several hundred photographs taken in the field were applied during this process.

During the classification work we found that mountain shadows caused obvious problems in classification. The most shadowed mountain slopes especially were systematically misclassified as water. Sophisticated algorithms for correcting satellite image radiance values in relation to slope have been developed (ERDAS 1997; Florinsky 1998) but, as the DEM available for the area had a coarser resolution (100 m) than the satellite images used, these methods were not feasible. Instead, we simply reclassified all the pixels within the mountain area that fell on slopes steeper than 10 degrees as *mainly bare land*, because most of those areas are shadowed rocky mountain slopes. The class *human infrastructures* (cities, villages, industrial areas, railroads and main roads) was separated from the spectrally similarly to the *mainly bare land* class using layers in the Russian digital map data and some manual delimitation.

The new forest cuttings found in a limited area in the south-west corner of the Usa Basin were delineated by visual examination of the image. Also distinctive clouds were delimited manually, as cloud shadows could not be easily separated from water bodies in the classification. In the final mosaic, 2.8% of the area was cloud covered. Clouded areas were masked out from the classifications and treated in the calculations as missing data.

#### *Testing the accuracy of classification*

We could not perform classification accuracy tests with real ground reference data due to the limited amount of field data. Instead, we used oblique aerial photographs taken during helicopter flights as test data for classification. With the aid of information on flight paths, we connected 271 scanned slides spatially to the satellite images (Fig. 2). (Some of these photographs [with reduced resolution] can be seen on the internet at [www.urova.fi/home/arktinen/tundra/photogallery.htm](http://www.urova.fi/home/arktinen/tundra/photogallery.htm).) With the help of located slides we identified and visually inspected 1328 test points from the available photographs of the different vegetation types on the image mosaic. The clas-

sification accuracy was tested by cross-tabulating these points with our classification data. Random sampling was assumed for calculations. Forest cuttings (no photographs) and water bodies (all non-mixed water pixels can be classified to an accuracy of almost 100%) were omitted from the test. We performed two tests, the first one for the remaining 19 classes, and the second for the clusters of five combined classes: *forests* (classes 1-7 in Table 1), *willows and meadows* (classes 8-9), *peatlands* (classes 10-13), *tundra heaths* (classes 14-18) and *mainly bare land and infrastructures* (classes 19-20).

## Results and discussion

### *Classification accuracy*

The area and proportion of each class within the sub-areas and the whole Usa Basin are shown in Table 1. (A high resolution image of the classification is posted as supplementary material at [www.npolar.no/PolarResearch](http://www.npolar.no/PolarResearch); see the table of contents for this volume. The data are also available from the authors on request.) When the aerial photograph test points are cross-tabulated for 19 classes, the overall accuracy is 53% and simple Kappa 0.50, but the fit varies considerably between the classes (Tables 1, 2a). However, when only five main groups are compared, the overall accuracy rises to 84% and simple Kappa to 0.75 (Table 2b, c). This means that the misclassifications are mainly spread over the functionally and spectrally nearest classes (e.g. different forest types were generally confused with other forest types; Table 2c). This is also indicated by the fact that weighted (by class number as in Table 1; reflects to some extent the functional similarity of the classes) Kappa values are higher than simple Kappa values (Table 2a, b). Achieved accuracies and Kappa values indicate good results when the main groups are compared, but in the 19-class comparison some individual classes are not well separated from their neighbourhood classes. Typically, the overall accuracy estimates of the corresponding cover types have varied between 65 and 87% in other studies (Fitzgerald & Lees 1994; Fuller et al. 1994; Holmgren & Thuresson 1998; Muller et al. 1998; Foody 2002; Rees et al. 2002; Tømmervik et al. 2003). Thus, when taking into account the practical restraints in the fieldwork and the large region covered, we think that our

accuracy test values are acceptable.

Generally, the use of oblique photographs as a reference for accuracy testing includes more error and uncertainty sources than is the case with real ground reference points. Even though we were able to locate the photographs on the satellite image, there may have been some location errors of the test points. Moreover, some vegetation classes were almost impossible (e.g. spruce dom-

inated vs. spruce–fir dominated forest) or at least difficult to distinguish from the photographs (e.g. dwarf birch vs. dwarf shrub moss tundra). Third, test points were not selected randomly from the located photographs, which may yield some bias on the classification accuracy estimates. However, keeping in mind that no unambiguous method for testing the accuracy existed (Foody 2002), the accuracy assessment method used gives at least

*Table 1.* Brief description and coverage percentages of the vegetation and land cover classes in each sub-area (see Fig. 3) and in the whole Usa Basin (in 30 m cell and in 1 km cell majority-class grid). Test points: % indicates the proportion of test points interpreted from photographs falling in the correct class and the second figure the number of test points in each class.

	Sub- area 1	Sub- area 2	Sub- area 3	Sub- area 4	Sub- area 5	Whole basin, 30 m	Whole basin, 1 km	Test points
Area (km <sup>2</sup> )	3477	2282	4071	3509	114	93 484	93 484	
Class description								
1. Spruce forest. Spruce crowns cover more than 60% of the coverage of the trees.	–	1.0	21.3	29.6	–	12.2	20.1	67%, 97
2. Spruce–fir forest. Spruce and fir together cover more than 60% of the tree coverage.	–	–	7.8	–	–	0.5	0.1	27%, 15
3. Pine forest. Lichen field layer dominated forests, mainly pine, near the spruce forest line.	–	–	–	0.8	–	0.6	0.2	71%, 7
4. Mixed forest. Mainly spruce forests with birch, willow, alder, aspen, larch or pine.	–	1.3	13.1	15.1	–	7.3	5.3	38%, 150
5. Larch forest. Larch crowns cover more than 60% of the tree coverage (classified only in sub-area 3).	–	–	1.1	–	–	–	–	–
6. Birch dominated stands. Birch dominated, mainly young stands, and mountain birch stands.	–	0.8	2.9	5.8	–	3.2	0.8	30%, 47
7. Forest cuttings. Some new forest cuttings found in the south-west corner of the area.	–	–	–	–	–	0.3	0.4	–
8. Willow stands. Willow dominated stands.	12.9	10.8	3.6	2.2	22.6	7.5	5.5	71%, 143
9. Meadows. Grass and herb dominated non-peatlands.	3.1	0.2	1.9	0.9	1.7	2.0	0.7	50%, 68
10. Bog partly with few trees. Paludified areas, often with a few pines, birches or spruces.	–	–	9.2	16.2	–	7.4	4.1	49%, 65
11. Open bog. Mainly shrubby, ombrotrophic bogs.	12.8	12.5	6.7	11.6	16.0	16.6	19.3	41%, 175
12. Wetland. Wet sedge fens etc. Also includes areas of water and land mosaic.	3.0	2.8	1.2	2.0	9.1	4.7	1.7	40%, 160
13. Tundra with some bare peat. Sparsely vegetated tundra–palsa–peatland complex.	–	6.4	–	–	–	1.1	0.2	60%, 15
14. Dwarf shrub-moss tundra heath. Shrubby tundra heath with mainly mosses, on mineral soils.	17.7	32.9	3.7	11.7	21.2	16.7	25.8	56%, 121
15. Dwarf birch heath. Dwarf birch dominated tundra heaths.	14.1	–	–	–	25.7	4.0	3.0	65%, 62
16. Dry dwarf shrub–lichen tundra. Tundra heath with lichen dominated field layer, on mineral soils.	–	22.9	2.6	1.6	–	4.7	2.2	52%, 70
17. Sparse alpine tundra. Some sparse dwarf shrubs, grasses, herbs, sedges, mosses and lichens.	5.1	–	0.5	–	–	1.1	0.2	45%, 38
18. Human impacted tundra. Dwarf birch tundra changed by human impacts around Vorkuta.	–	–	–	–	–	0.3	0.4	67%, 6
19. Mainly bare land. Areas dominated by sand and stones (cover >50%).	28.3	0.8	23.7	0.2	–	6.5	7.1	94%, 109
20. Human infrastructures. Cities, villages, industrial areas, railroad, and main roads.	–	–	0.0	0.1	–	0.4	0.5	82%, 34
21. Water bodies. Lakes, ponds, and rivers.	3.0	7.8	0.7	2.2	3.8	2.8	2.3	–



the approximate level of confidence of the produced classification.

Mixed pixels—i.e. pixels whose reflectance values are due to several types of vegetation/landscape types within one pixel—are a well known problem in satellite image classifications (Campbell 1987; Foody 2002). In our study area this was typical especially in the northern tundra areas and peatlands, where the vegetation patches are often so small that several patches fall within one pixel (see also Stow et al. 1998). Pixels falling on the boundaries of two vegetation types are another source of mixed pixels. Mixed pixels either yield spectral classes that mix with some uniform vegetation types or produce their own specific combinations of spectral values which, in turn, may cause significant random errors (Campbell 1987; Chen et al. 2002; Foody 2002). For example, the reflectance of pure spruce stands is often similar to that of areas with small-scale variation

of waters and broadleaved trees or grassy vegetation. In addition to introducing error into the total areas of different vegetation types, misclassifications should also be taken into account when studying spatial questions such as the location of different vegetation types in relation to environmental variables or to each other.

### Class descriptions

Next we describe and briefly discuss the problems related to each class and their potential misclassification under five main groups: *forests* (cover 24.1% of the area in the 30 m cell detailed classification), *willows and meadows* (9.5%), *peatlands* (29.8%), *tundra heaths* (26.4%) and *mainly non-vegetated areas (mainly bare land, human infrastructures and water bodies; 9.8%)*.

*Forests.* We separated seven forest classes (Table 1), using crown cover exceeding 20% as the defining feature of forest. Generally, when one tree species covered more than 60% we specified that the stand was dominated by that species. If the cover of none of the species exceeded the limit, the stand was classified as *mixed forest*. According to our field measurements, mean tree volumes in northern taiga forests vary from about 110 m<sup>3</sup> ha<sup>-1</sup> (range 65–215 m<sup>3</sup> ha<sup>-1</sup>) in the southern part of the area to about 25 m<sup>3</sup> ha<sup>-1</sup> in the forest line region, and the mean tree height of the stand varies from 9–16 m to about 3.5 m, correspondingly.

Most of the forest sites in the region are mesic and spruce dominated with a variable admixture of other trees, primarily birches (Kozubov et al. 1999). Spruce with a low proportion of birch forms the northernmost forest stands in the lowlands. The separation between mixed and “pure” forests is arbitrary, as in reality there is a total continuum from almost pure stands to very mixed ones. However, in addition to the *spruce forest* class,

Table 2. Accuracy test of the classification with test points. In weighted Kappa tests weights were calculated using Cicchetti-Allison Kappa coefficient weights (SAS Institute, Inc., 1999, Cary, NC). See the Materials and methods section for details. (a) Test for 19 classes. (b) Test for classes when grouped into five main classes. (c) Cross-tabulation of the five main classes.

(a)

Kappa coefficients	Statistic value	95% confidence limits
Simple Kappa	0.503	0.474–0.531
Weighted Kappa	0.768	0.747–0.789
Overall accuracy	53.34%	

(b)

Kappa coefficients	Statistic value	95% confidence limits
Simple Kappa	0.749	0.722–0.776
Weighted Kappa	0.820	0.798–0.842
Overall accuracy	83.58%	

(c)

Classification: Test:	Forests	Willow & meadow	Peatlands	Tundra heaths	Mainly bare land & infra- structures	Total n
Forest	<b>264</b>	21	29	2	0	316
Willow & meadow	3	<b>201</b>	5	1	1	211
Peatland	20	29	<b>294</b>	60	12	415
Tundra	4	8	59	<b>213</b>	13	297
Bare	1	0	2	2	<b>138</b>	143
Total n	292	259	389	278	164	<b>1328</b>

we decided to establish a separate *mixed forest* class, as in our field measurements the mean tree volumes had about 20% lower values in mixed than in the more homogeneous spruce forests. This separation might also have some value to biodiversity-related future studies, for example.

The *spruce–fir* forest stands that occur in the foothills of the Urals in the southern part of the study area can, to some extent, be separated from Norway spruce–birch forests in the classification. This is primarily due to the fact that they usually include only a few broadleaved trees, and firs often grow very abundantly as an understorey species. Mountain shadows probably caused some misclassification; on shadowed slopes the brightness values are lower, which indicates that dark conifer (spruce and fir) dominated forests are overestimated on such slopes.

The small number of Scots pine forests with a lichen dominated field layer are found on sandy river terraces near the Usa and Pechora rivers in the western parts of the Usa Basin. In the forest line region, outside the distribution range of pine, our *pine forest* class also contains some lichen dominated stands where the dominant tree is spruce and, in the Urals, stands with larch and/or spruce as the dominant tree species. The spectral class of this *pine forest* class could be clearly separated, mainly due to the extensive lichen coverage in the field layer rather than to spectral differences between spruce and pine crowns.

In the Ural Mountains the treeline is mainly composed of larch. Individual larch trees are found on sheltered southern slopes even in the northernmost parts of our study area. In the Polar Urals sparse larch stands are typically found at an elevation of around 200 m a.s.l. In the southern parts of the Usa Basin, the alpine treeline is located at around 550–600 m a.s.l. On some slopes the larch belt is only a few tens of metres wide, and sometimes mixed with mountain birches. Toutoubalina & Rees (1999) reported that sparse larch stands can not be distinguished from treeless sites with Landsat images because the reflected spectrum of sparse larch forests is primarily due to ground vegetation. Our findings from sub-area 3 support this result. Second, we found out that the spectral signature of denser larch forest could not be separated from that of the *Salix* dominated lowland forests or mountain birch dominated forests. Therefore, we reclassified *larch forests* below 400 m a.s.l. to the spectrally most similar class, *willow dominated stands* in sub-area

3. Depending on the area and ground vegetation, larch forests are pooled either with *willow dominated stands*, *meadows*, *mixed forests* or *birch dominated stands* in the whole area classification.

The class *birch dominated stands* mainly consists of young, birch dominated stands. Birch, or in some cases aspen (*Populus tremula* L.), is the first successional tree species after forest fire in the region (Kozubov et al. 1999; Gromtsev 2002) and therefore young, birch dominated forests are mainly generated by forest fires. Examination of the pattern and brightness showed that 30- to 80-year-old larger fire areas account for about 12% of the forests. *New cuttings* covering about 300 km<sup>2</sup> in the south-west corner of the Usa Basin could be easily distinguished from naturally regenerated young successional stages due to their rectangular shape. We did not find any signs of large fires that would have occurred during the past 10–20 years before the satellite images were recorded. It is also known that the mesic spruce dominated forests in the region often regenerate through many tree generations by means of small-scale gap dynamics without forest fires (Kuuluvainen et al. 1998; Gromtsev 2002). We also included the mountain birch dominated stands at the treeline in the Pre-Polar Urals in the class *birch dominated stands*.

*Willows and meadows.* Large willow dominated stands are typical for the region in depressions and along the rivers. Willow stands are found in almost every part of the basin and, together with *meadows*, they cover almost 10% of the Usa Basin (Table 1). Willow stands vary from stands consisting of up to 15 m high willow trees (mainly *Salix viminalis* sensu lato) along southern riverbanks, to about 1 m high bushes (*S. phyllicifolia* L., *S. glauca* L., *S. lapponum* L., etc.) in the tundra. Extensive willow stands are also often growing around mires. The class *willow stands* also includes stands containing other tree species when willow undergrowth dominates the coverage.

Meadows are regularly found adjacent to willow stands along river banks and on islands in flooded areas in the lowlands. Meadows are especially common in the Urals. A transitional belt exists at the treeline where forests and luxurious grass and herb meadows alternate. Areas located above the treeline, up to about 750 m a.s.l. in the south and about 400 m a.s.l. in the north, are characterized by a mosaic of patchy alpine meadows,

dwarf shrub–lichen dominated tundra vegetation, and rocky slopes (Gorchakovskii 1960; our observations and data).

*Peatlands.* Different types of peatland are common in the Usa Basin. Large open mires and fens are found in the lowlands of the taiga zone. Peat plateaus are more typical of the central and northern parts of the area. Peatlands become patchier in the tundra and the peat layer is generally shallower. Peatlands in flat valley bottoms in the Ural Mountains resemble those in the lowlands. On the mountain slopes they are mainly sedge dominated and typically found next to steeper areas.

Spectrally peatlands are a very diverse group due to the variable moisture conditions and patchy vegetation, which makes them relatively difficult to classify even at the resolution of 30 m. This is also reflected in the slightly lower test point fit than in the other groups (Table 1). Also, as satellite sensors record reflectance primarily from the uppermost layer, tree-covered peatlands are easily confused with wooded vegetation types. Peatlands with a dense willow cover are classified as *willow stands*, and slightly paludified, spruce dominated forests as *forests*.

We divided peatlands into four classes. In the class *bog shrubland partly with few trees* we included areas found along the peatland margins in the taiga region. These peatlands often have a few pines, birches or spruces, with a mean tree volume of around 10 m<sup>3</sup> ha<sup>-1</sup>. Most of the pines within the Usa Basin grow on peatlands.

Shrub dominated treeless peatlands were categorized as *open bogs*. Separation of the shrubby tundra heath in the tundra fraction from this class is in some locations arbitrary due to the patterned nature of the vegetation. In the class *tundra with some bare peat* we classified sparsely vegetated tundra heath and palsa or peat plateau complexes found in the northern parts of the Usa Basin. In some tundra areas lichens grow extensively on top of the peat. They could not be separated from lichen-growing tundra heaths in locations where the size of these patches exceeded the size of the image pixel.

We included wet sedge fens in the class *wetlands*. Areas of water and land mosaic are also mainly of this class. In some cases we found that the mixed land/water pixels have almost exactly the same spectral fingerprint as the densest spruce stands. These misclassifications caused by mixed pixels could not be handled by adding

training categories to represent different vegetation mixtures. However, the number of erroneous pixels attributable to this is relatively small.

*Tundra heaths.* Different kinds of tundra heath are found in better-drained soils in the central and northern parts of the Usa Basin. The most common type of tundra heath in our classification is called *dwarf shrub–moss tundra*. This class includes shrubby tundra heaths with different mosses and also some lichens on mineral soils. Typical and abundant dwarf shrub species are *Vaccinium uliginosum* L., *V. vitis-idaea* L., *V. myrtillus* L., *Empetrum nigrum* L., *Ledum decumbens* Lodd., *Arctostaphylos arctica* L. and dwarf birch. Tundra vegetation often consists of a small-scale mosaic in which drier heaths are mixed with small, wetter patches with a few willows that are also slightly paludified. In many cases the patches are not larger than the individual pixels of a Landsat image, which results in their inclusion primarily into the *dwarf shrub–moss tundra* class.

In some locations the tundra heaths are lichen dominated. Areas where the lichen cover exceeded approximately 30% (as visible from above) were classified as *dwarf shrub–lichen tundra*. In some locations the lichens were heavily grazed by reindeer (Crittenden 2000; own observations). Grazing has potentially yielded some vegetation changes during the approximately ten year interval between our image acquisition and fieldwork. Changes of this kind have been analysed in the study Rees et al. (2003) conducted in the region near our study area. This might have caused some inaccuracies to the estimates of the amount of *dwarf shrub–lichen tundra* versus *dwarf shrub–moss tundra*.

In the north-eastern part of the Usa Basin, shrub tundra dominated by dwarf birch (height 30–80 cm) is the most common type of vegetation and was classified as *dwarf birch tundra*. Further west, away from the Ural Mountains, lower dwarf shrub species (*Vaccinium* spp., *Empetrum* spp., etc.) dominate the shrub tundra, and the existing dwarf birches in the heath vegetation are mainly shorter than 30 cm (Rebristaya 1977; our field observations). The predominance of dwarf birch in the north-eastern region is probably due to the higher precipitation and related thick snow cover owing to the proximity of the Urals. More than ca. 100 km west from the Urals, dwarf birch dominated heaths are found only in some sheltered locations and they are very scattered; in these cases they are mainly included in the *dwarf*

*shrub–moss tundra* class.

We observed some specific reflectance values around the coal-mining city of Vorkuta. A separate study of this issue is presented in Virtanen et al. (2002). Only the main findings are briefly summarized here. Dwarf birch tundra has changed there due to human impacts, and two impact zones could be identified (presented as an aggregated class in Table 1 and Fig. 2). The first zone, the *pollution zone*, covers 150–200 km<sup>2</sup> around the main pollution sources. Most of the lichen species are absent, and changes in the species composition of the vegetation communities in all the main plant groups are also obvious. Willows are more dominant than in the unpolluted sites. In the second zone, *slight pollution/disturbance zone* (600–900 km<sup>2</sup>), changes in the vegetation are generally similar but are less severe than those in the first zone. The amount of herbs and grasses in particular has increased compared to the unpolluted areas. The extent of the zones is furthest to the north-east from the main emission sources, matching the prevailing winds during winter.

For the area of the Ural Mountains we separated one more vegetation type: *sparse alpine tundra*. Alpine dwarf shrub and grass vegetation are found in some locations, mainly above the meadow zone. In the northern parts this zone is about 350–500 m. a.s.l., and in the south 600–1000 m. a.s.l. This vegetation type exists on thin soil that has developed on relatively gentle mountain slopes. The vegetation consists of some low dwarf shrubs, sedges, grasses, mosses and lichens, and is sparser than that in the typical lowland tundra heath.

*Mainly non-vegetated areas.* We classified areas where the vegetation cover was less than 50% of the area into *mainly non-vegetated* classes. Mainly bare land can be found in different locations: on sandy and stony riverbanks; on sandy deflations in the tundra lowlands; and in the highest parts and on the rocky steepest slopes of the mountains. We also classified snow-covered areas as *mainly bare land*. Snow-covered areas were found almost only on the highest mountains and they covered less than 1% of the whole area. Cities, villages, and other *human infrastructures*, which were separated from the spectrally similar *mainly bare land* class using Russian GIS data, covered 0.4% of the Usa Basin.

Larger water bodies were separated very well in the classification process. However, the area of

water bodies is probably slightly underestimated, as most of the water pixels mixed with vegetation were classified in the *wetlands* class.

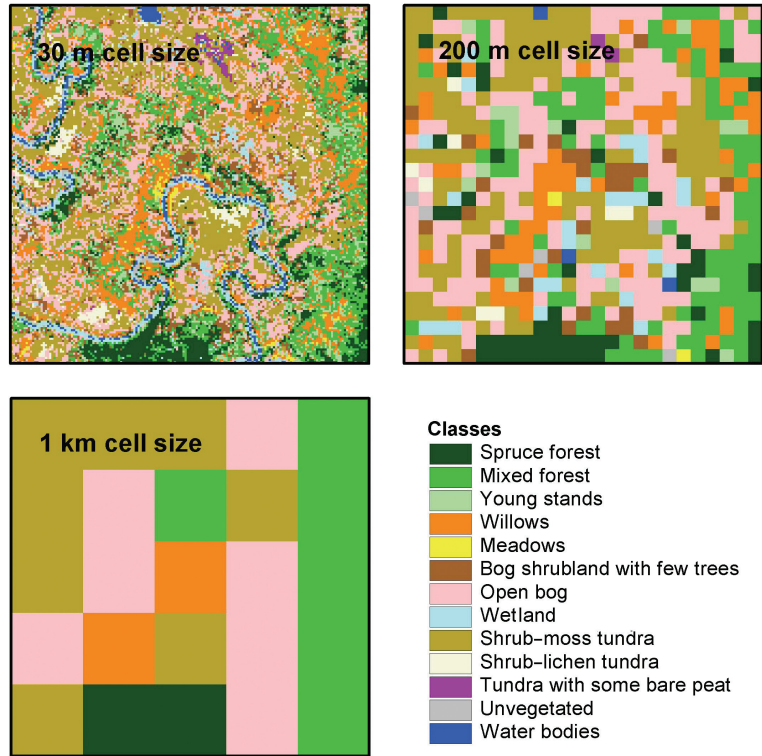
#### *Classification approach and usability of the data*

In the following section we discuss our classification method and highlight the advantages the classification data contribute to the studies on global change issues. We also briefly present the advantages that our 30 m resolution data offer for landscape-level studies when compared to low resolution data.

*Approach.* A frequently employed strategy in the use of satellite images in large-scale inventories is first to classify individual images and then to merge the classification results (Fuller et al. 1994; Cihlar 2000; Franklin & Wulder 2002). An alternative strategy is spectrally to normalize and mosaic the images before the classification (Homer et al. 1997; Cihlar 2000; Franklin & Wulder 2002). As we did not have extensive ground reference data from all the separate images, only the mosaicing method was relevant in our case. In the channel-to-channel regressions, the coefficients ( $R^2$ ) varied from around 0.85 in channel 2 to as high as 0.97 in channels 4 and 5. The coefficients of channels 3 and 7 were between these values. The values were even higher than those reported in the study of Olsson (1993), which implies that the spectral multitemporal relative calibration yielded satisfactory results. Also, according to visual inspection, the classification did not yield any obvious spatial discrepancies of the classes along the borders of the original images in the mosaic. Spectral calibration worked better in forested areas and drier tundra, but peatlands were more problematic. This is at least partly due to the more variable moisture conditions and phenological changes in green biomass of peatlands.

The main advantages of the mosaicing method are the possibility to use ground reference sites from ecologically similar areas in other satellite image scenes, and that the post-classification edge matching is not needed. The disadvantage of using a low amount of and spatially sparse ground reference data, especially with image mosaic that covers large area, is that this creates a well known signature extension problem, when the spectral signatures are also used for regions located far away from the training areas (Cihlar 2000). Some obvious misclassifications in our data could not

Fig. 4. Fragmented landscape presented using different cell sizes (30 m, 200 m and 1 km). Coordinates of the lower left corner are 66° 57' 36" N, 59° 28' 12" E.



be avoided due to this problem. To some extent, these errors could be handled by post-classification refinement using other GIS data or geographical limitations for the classes, as we did with mountain shadows. This kind of refinement could be done even more extensively if enough background information about the region's vegetation and landscapes exists, as in the work of Homer et al. (1997). Actually, many large-area classifications rely heavily on ancillary data input, using approaches like a decision or classification tree classifiers or expert classification approaches (Franklin & Wulder 2002). Within our approach the only ancillary data used was a DEM, and even it was utilized only to a small extent.

Theoretically, there are several other classification procedures in addition to the supervised one that we used. For example, classification can be produced by creating a large number of unsupervised spectral classes and then merging them to match the actual land cover categories (Homer et al. 1997; Stow et al. 1998; Cihlar 2000; Loveland et al. 2000; Ma et al. 2001; Franklin & Wulder 2002). After trying this procedure we came to the conclusion that spectral classes created by

the unsupervised method were often ecologically heterogeneous and thus not suitable for our needs. However, if the aim is only to separate broad groups, like forest, tundra and bare areas, we think that an unsupervised method could be used relatively successfully.

*Content of the classes.* Landsat TM image classification does not allow vegetation mapping to the same degree of detail as the vegetation types recognized by the traditional phytosociological school (Kalliola & Syrjänen 1991). Satellite image based vegetation classifications are useful when evaluating the frequency distribution of broadly defined vegetation classes in a specified area, but they are more limited with respect to more detailed and precise information, e.g. about forest or other vegetation structure (Holmgren & Thuresson 1998; Achard et al. 2001). However, the separation of the main functional types (e.g. forest, peatland, tundra heath) is often sufficient for the process models normally used in global change studies (Running et al. 1994; Kittel et al. 2000; Plummer 2000; Rupp et al. 2000; Van der Linden et al. 2003). The resolution used in many of these models is often of hundreds of metres

or coarser and inevitably masks the small-scale variation in the vegetation. The upscaling procedures used in various process models also largely omit the sort of small-scale misclassifications which our classification in some cases probably includes.

A common strategy for utilizing remote sensing data in modelling studies is the usage of certain indices in connection with ecological parameters. For example, the normalized difference vegetation index (NDVI) connected to the leaf area index (LAI) has often been used to study a wide range of physical and biological processes, e.g. carbon fluxes and biological production (Woodcock et al. 1997; Holmgren & Thuresson 1998; Soegaard et al. 2000; Chen et al. 2002). However, there might be some limitations in using these kinds of indices due to specific spectral properties of some vegetation types, like lichen dominated stands or dark conifer dominated forest (Rees et al. 2002). Although these indices are suitable for modelling some processes, vegetation indices lack information on functional differences between various vegetation types. The separation of different vegetation types is important if they respond in a different way to e.g. climate or other environmental changes (Kittel et al. 2000; Rupp et al. 2000; Skre et al. 2002). Examples of the processes or factors to model that require spatially explicit data of functional vegetation types are treeline dynamics (Rupp et al. 2000; Virtanen et al. in press) and various disturbance factors (Virtanen et al. 1998; Malmström & Raffa 2000; Rupp et al. 2000). Furthermore, both of these approaches can always be linked, as in the study of Liu et al. (2002), in which cover type specific formulae connecting NDVI and LAI were used for various land cover types.

*Spatial resolution.* As far as the landscape level analyses and models are concerned, it is obvious that much information is lost when using 1 km cell data instead of a 30 m cell. We have illustrated how the 30 m scale is more accurate in landscape characterization than the coarser ones, especially when the landscape pattern is fragmented (Fig. 4). The 200 m and 1 km cell sizes are calculated using majority filtering that originates from the 30 m cell size. The vegetation types occurring in small or narrow patches even completely disappear in the coarser resolutions. The same effect can be seen at the whole Usa Basin level when we compare the coverage percentages in 30 m cell resolution and in 1 km majority filtered cells

(Table 1). This finding supports the well established fact that the less abundant and more fragmented patch types in the landscape are lost at coarser resolution, and dominant types become over-represented (Turner et al. 1989).

## Concluding remarks

This paper shows how a mesoscale vegetation classification of a large and remote Arctic area can be effectively conducted with a limited ground reference data set. In our case we found out that the feasible method to produce large-area, multiple image land cover classification was supervised classification of the spectrally matched image mosaic with some post-classification refinement with additional GIS data.

Classification data like ours describing the main functional vegetation types of vegetation and land cover can be effectively used in many kinds of landscape analysis and process modelling studies (Huisink et al. 2002; Kuhry et al. 2002; Van der Linden et al. 2003; Heikkinen et al. 2004; Virtanen et al. in press). In particular, the detailed spatial resolution (30 m cell size) provides new landscape level insights compared to the currently available global vegetation data sets.

*Acknowledgements.*—T. Virtanen was affiliated with the Finnish Forest Research Institute at the time of this research. We would like to thank Peter Kuhry and three anonymous referees for their constructive comments. Tauno Luosujärvi participated in the fieldwork. John Derome improved the English of the article. We would also like to acknowledge the staff of the Institute of Biology, Komi Science Center, Syktyvkar, who gave us useful information about the nature in the region, organized the logistics and helped in the fieldwork, and all the TUNDRA partners for the excellent co-operation. This study is a part of the TUNDRA project, funded by the 4th Framework Environment and Climate Programme of the European Commission (contract ENV4-CT97-0522; Climatology and Natural Hazards).

## References

- Anisimov, O. A. & Nelson, F. E. 1996: Permafrost distribution in the Northern Hemisphere under scenarios of climatic change. *Glob. Planet. Change* 14, 59–72.
- Barnsley, M. J., Barr, S. L. & Tsang, T. 1997: Scaling and generalisation in land cover mapping from satellite sensors. In P. R. van Gardingen et al. (eds.): *Scaling up. From cell to landscape*. Pp. 173–199. Cambridge: Cambridge University Press.

- ty Press.
- Bartalev, S., Belward, A. S. & Erchov, D. V. & Isaev, A. S. 2003: A new SPOT4-VEGETATION derived land cover map of northern Eurasia. *Int. J. Remote Sens.* 24, 1977–1982.
- Betts, R. A. 2000: Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. *Nature* 408, 187–190.
- Bonan, G. B., Chapin III, F. S. & Thompson, S. L. 1995: Boreal forest and tundra ecosystems as components of the climate system. *Clim. Change* 29, 145–167.
- Campbell, J. B. 1987: *Introduction to remote sensing*. New York: Guilford Press.
- Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S. G., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymow, J. A., Swift, E., Van der Sanden, J. & Pellikka, P. K. E. 2002: Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sens. Environ.* 80, 165–184.
- Cihlar, J. 2000: Land cover mapping of large areas from satellites: status and research priorities. *Int. J. Remote Sens.* 21, 1093–1114.
- Crittenden, P. D. 2000: Aspects of the ecology of mat-forming lichens. *Rangifer* 20, 127–139.
- ERDAS 1997: *ERDAS field guide*. 4th edition. Atlanta, GA: Erdas Inc.
- Fitzgerald, R. W. & Lees, B. G. 1994: Assessing the classification accuracy of multisource remote sensing data. *Remote Sens. Environ.* 47, 362–368.
- Florinsky, I. V. 1998: Combined analysis of digital terrain models and remotely sensed data in landscape investigations. *Prog. Phys. Geogr.* 22, 33–60.
- Foody, G. M. 2002: Status of land cover classification accuracy assessment. *Remote Sens. Environ.* 80, 185–201.
- Franklin, S. E. & Wulder, M. A. 2002: Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Prog. Phys. Geogr.* 26, 173–205.
- Fuller, R. M., Groom, G. B. & Jones, R. A. 1994: The land cover map of Great Britain: an automated classification of Landsat Thematic mapper data. *Photogramm. Eng. Remote Sens.* 60, 553–562.
- Gorchakovskii, P. L. 1960: An essay on the phytogeographical subdivision of the Uralian highlands. In V. I. Sukachev (ed.): *Studies on the flora and vegetation of high-mountain areas. Problems of Botany* 5. USSR Academy of Sciences. Pp. 25–40. Translated and published in English in 1965 in Jerusalem: Israel Program for Scientific Translations Ltd.
- Gromtsev, A. 2002: Natural disturbance dynamics in the boreal forests of European Russia: a review. *Silva Fenn.* 36, 41–55.
- Hansen, M., DeFries, R., Townshend, J. R. G. & Sohlberg, R. 2000: Global land cover classification at 1 km resolution using a decision tree classifier. *Int. J. Remote Sens.* 21, 1331–1365.
- Hansen, M. & Reed, B. C. 2000: A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *Int. J. Remote Sens.* 21, 1365–1373.
- Harding, R., Kuhry, P., Christensen, T. R., Sykes, M. T., Dankers, R. & Van der Linden, S. 2002: Climate feedbacks at the tundra–taiga interface. *Ambio Spec. Rep.* 12, 47–55.
- Heikkinen, J. E. P., Virtanen, T., Huttunen, J. T., Elsakov, V. & Martikainen, P. J. 2004: Carbon balance in East European tundra. *Glob. Biogeochem. Cycles* 18, GB1023, doi: 10.1029/2003GB002054.
- Holmgren, P. & Thuresson, T. 1998: Satellite remote sensing for forestry planning—a review. *Scand. J. For. Res.* 13, 90–110.
- Homer, C. G., Ramsey, R. D., Edwards, T. G., Jr. & Falconer, A. 1997: Landscape cover-type modelling using a multi-scene Thematic mapper mosaic. *Photogramm. Eng. Remote Sens.* 63, 59–67.
- Huisink, M., de Moor, J. J. W., Kasse, C. & Virtanen, T. 2002: Factors influencing periglacial fluvial morphology in the northern European Russian tundra and taiga. *Earth Surf. Process. Landf.* 27, 1223–1235.
- Kalliola, R. & Syrjänen, K. 1991: To what extent are vegetation types visible in satellite imagery? *Ann. Bot. Fenn.* 28, 45–57.
- Kittel, T. F. G., Steffen, W. L. & Chapin III, F. S. 2000: Global and regional modeling of Arctic–boreal vegetation distribution and its sensitivity to altered forcing. *Glob. Change Biol.* 6, Suppl. 1, 1–18.
- Kozubov, G. M., Taskaev, A. I., Degteva, S. V., Martynenko, V. A., Zaboeva, I. V., Bobkova, K. S. & Galenko, E. P. 1999: *Lesa Respubliki Komi. (Forests of the Komi Republic)*. Moscow: Design, Information, Cartography Printing House.
- Kuhry, P., Mazhitova, G. G., Forest, P.-A., Deneva, S. V., Virtanen, T. & Kultti, S. 2002: Upscaling soil carbon estimates for the Usa Basin (northeast European Russia) using GIS-based vegetation and soil classification schemes. *Dan. J. Geogr.* 102, 11–25.
- Kultti, S., Väiliranta, M., Sarmaja-Korjonen, K., Solovieva, N., Virtanen, T., Kauppila, T. & Eronen, M. 2003: Palaeo-ecological evidence of changes in vegetation and climate during the Holocene in the Pre-Polar Urals, northeast European Russia. *J. Quat. Sci.* 18, 503–520.
- Kuuluvainen, T., Syrjänen, K. & Kalliola, R. 1998: Structure of a pristine *Picea abies* forest in northeastern Europe. *J. Veg. Sci.* 9, 563–574.
- Levin, S. A. 1992: The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967.
- Liu, J., Chen, J. M., Cihlar, J. & Chen W. 2002: Net primary productivity mapped for Canada at 1 km resolution. *Glob. Ecol. Biogeogr.* 11, 115–129.
- Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, J., Yang, L. & Merchant, J. W. 2000: Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR Data. *Int. J. Remote Sens.* 21, 1303–1330.
- Ma, Z., Hart, M. M. & Redmond, R. L. 2001: Mapping vegetation across large geographic areas: integration of remote sensing and GIS to classify multisource data. *Photogramm. Eng. Remote Sens.* 67, 295–307.
- Malmström, C. M. & Raffa, K. F. 2000: Biotic disturbance agents in the boreal forest: considerations for vegetation change models. *Glob. Change Biol.* 6, Suppl. 1, 35–48.
- Muller, S. V., Walker, D. A., Nelson, F. E., Auerbach, N. A., Bockheim, J. G., Guyer, S. & Sherba, D. 1998: Accuracy assessment of a land-cover map of the Kuparuk River Basin, Alaska: considerations for remote regions. *Photogramm. Eng. Remote Sens.* 64, 619–628.
- Oberman, N. G. & Mazhitova, G. G. 2003: Permafrost mapping of northeast European Russia based on the period of climatic warming, 1970–1995. *Nor. Geogr. Tidsskr.* 57, 111–120.
- Olsson, H. 1993: Regression functions for multitemporal relative calibration of thematic mapper data over boreal forest.

- Remote Sens. Environ.* 46, 89–102.
- Plummer, S. E. 2000: Perspectives on combining ecological process models and remotely sensed data. *Ecol. Model.* 129, 169–186.
- Price, J. C. 1981: The contribution of thermal data in Landsat Multispectral classification. *Photogramm. Eng. Remote Sens.* 47, 229–236.
- Ranson, K. J., Sun, G., Knox, R. G., Levine, E. R., Weishampel, J. F. & Fifer, S. T. 2001: Northern forest ecosystem dynamics using coupled models and remote sensing. *Remote Sens. Environ.* 75, 291–301.
- Rebristaya, O. V. 1977: *Flora vostoka Bol'shezemel'skaya tundra.* (Flora of the eastern part of the Bol'shezemel'skaya tundra.) Leningrad: Nauka. In Russian with English abstract.
- Rees, G., Brown, I., Mikkola, K., Virtanen, T. & Werkman, B. 2002: How can the dynamics of the tundra–taiga boundary be remotely monitored? *Ambio Spec. Rep.* 12, 56–62.
- Rees, G., Williams, M. & Vitebsky, P. 2003: Mapping land cover change in a reindeer herding area of the Russian Arctic using Landsat TM and ETM+ imagery and indigenous knowledge. *Remote Sens. Environ.* 85, 441–452.
- Running, S. W., Loveland, T. R. & Pierce, L. L. 1994: A vegetation classification logic based on remote sensing for use in global biogeochemical models. *Ambio* 23, 77–81.
- Rupp, T. S., Chapin, F. S. & Starfield, A. M. 2000: Response of subarctic vegetation to transient climatic change on the Seward Peninsula in north-west Alaska. *Glob. Change Biol.* 6, 541–555.
- Skre, O., Baxter, R., Crawford, R. M. M., Callaghan, T. V. & Fedorkov, A. 2002: How will tundra–taiga interface respond to climate change? *Ambio Spec. Rep.* 12, 37–46.
- Soegaard, H., Nordstroem, C., Friborg, T., Hansen, B. U., Christensen, T. R. & Bay, C. 2000: Trace gas exchange in a High-Arctic valley 3. Integrating and scaling CO<sub>2</sub> fluxes from canopy to landscape using flux data, footprint modeling, and remote sensing. *Glob. Biogeochem. Cycles* 14, 725–744.
- Stow, D., Hope, A., Boynton, W., Phinn, S., Walker, D. & Auerbach, N. 1998: Satellite derived vegetation index and cover type maps for estimating carbon dioxide flux for Arctic tundra regions. *Geomorphology* 21, 313–327.
- Thomlinson, J. R., Bolstad, P. V. & Cohen, W. B. 1999: Coordinating methodologies for scaling landcover classifications from site-specific to global: steps toward validating global map products. *Remote Sens. Environ.* 70, 16–28.
- Tømmervik, H., Hogda, K. A. & Solheim, I. 2003: Monitoring vegetation changes in Pasvik (Norway) and Pechenga in Kola Peninsula (Russia) using multitemporal Landsat MSS/TM data. *Remote Sens. Environ.* 85, 370–388.
- Toutoubalina, O. V. & Rees, G. W. 1999: Remote sensing of industrial impact on Arctic vegetation around Noril'sk, northern Siberia: preliminary results. *Int. J. Remote Sens.* 20, 2979–2990.
- Turner, M. G., O'Neill, R. V., Gardner, R. H. & Milne, B. T. 1989: Effect of changing spatial scale on the analysis of landscape pattern. *Landscape Ecol.* 3, 153–162.
- Van der Linden, S., Virtanen, T., Oberman, N. & Kuhry, P. 2003: Sensitivity analysis of the discharge in the Arctic Usa Basin, east-European Russia. *Clim. Change* 57, 139–161.
- Virtanen, T., Neuvonen, S. & Nikula, A. 1998: Modelling topoclimatic patterns of egg mortality of *Epirrita autumnata* (Lepidoptera: Geometridae) with a Geographical Information System: predictions for current climate and warmer climate scenarios. *J. Appl. Ecol.* 35, 311–322.
- Virtanen, T., Mikkola, K., Nikula, A., Christensen, J. H., Mazhitova, G. G., Oberman, N. G. & Kuhry, P. in press: Modeling the location of forest line in NE European Russia with remote sensed vegetation and GIS-based climate and terrain data. *Arct. Antarct. Alp. Res.*
- Virtanen, T., Mikkola, K., Patova, E. & Nikula, A. 2002: Satellite image analysis of human caused changes in the tundra vegetation around the city of Yorkuta, north-European Russia. *Environ. Pollut.* 120, 647–658.
- Walker, D. A., Bay, C., Daniels, F. J. A., Einarsson, E., Elvebakk, A., Kapitsa, B. E., Kholod, S. S., Murray, D. F., Talbot, S. S., Yurtsev, B. A. & Zoltai, S. C. 1995: Toward a new Arctic vegetation map: a review of existing maps. *J. Veg. Sci.* 6, 427–436.
- Woodcock, C. E., Collins, J. B. & Jupp, D. L. B. 1997: Scaling remote sensing models. In P. R. van Gardingen et al. (eds.): *Scaling up. From cell to landscape.* Pp. 61–77. Cambridge: Cambridge University Press.
- Woodwell, G. M., MacKenzie, F. T., Houghton, R. A., Apps, M. A., Gorham, E. & Davidson, E. 1998: Biotic feedbacks in the warming of the Earth. *Clim. Change* 40, 495–518.