Impervious surface mapping in Southern Norway

Dagrun Vikhamar Schuler [a,*], Leif Kastdalen [b]

[a] Gjøvik University College, P.O. Box 191, 2802 Gjøvik, Norway
Now at: Department of Geosciences, University of Oslo, Norway – dagrunv@geo.uio.no
[b] Hedmark University College, Faculty of Forestry and Wilderness Management, 2480 Koppang, Norway – leif.kastdalen@hihm.no

Abstract – Impervious surfaces are mapped by applying regression tree classification on Landsat ETM+ data of Østfold County located in Southern Norway. Reference data consist of high precision polygon vector data of roads, parking lots and buildings for a selected region (Fredrikstad Municipality) within the study area. The output from the classification using the regression tree technique is a map showing the imperviousness on a continuous scale from 0 to 100%.

Keywords: Impervious surfaces, Landsat, regression tree, high resolution map data.

1. INTRODUCTION

Over the past, urban impervious areas have been continuously increasing at the sacrifice of green areas. Information about the actual area coverage of the two surface types, impervious surfaces and green areas, is of great importance for management of landscape and recreation areas. As impervious surfaces we understand any man-made construction that does not allow natural infiltration of water, such as buildings (residential areas, industrial areas), roads and parking lots made from e.g. asphalt, concrete or bricks. The green areas comprise parks and woodlands within and surrounding the urban areas. Imperviousness may be interpreted as an environmental indicator of an area. Increasing impervious areas lead to reduced infiltration and thereby increased surface runoff within catchments. The local climate may also be affected.

This study has been carried out within a more comprehensive land-cover mapping project of Østfold County called SatNat (Vikhamar et al., 2004). The Norwegian Directorate of Nature Management seeks new methods to estimate the coverage of the man-made areas and the green areas at a given time, and also changes in these areas over time. The sizes of these areas are reported every year as national key numbers for specific geographical regions.

The main goal of this study is to map impervious surfaces based on remote sensing data, and doing it by only using existing digital vector data as training. This also includes the challenge of separating impervious surfaces from natural grey areas such as bare rocks. Similar spectral signatures often result in misclassification. This problem has been discussed in e.g. Yang et al. (2003). A full description of our analysis is found in Vikhamar and Kastdalen (2004).

In the next sections we first describe the study area and the data set used in the analysis. Then we present the regression tree analysis, and the results we obtained in this study.

* Corresponding author.

2. STUDY AREA AND DATA SET

The location of the study area is shown in Fig. 1. Østfold County (4183 km^2) is located in Southern Norway, with the Swedish border to the east, and the Oslo fjord to the west. The area is characterized by farming landscape, forests and cities with up to 83 000 habitants. The topography is relatively flat ranging from sea level to 336 m.a.s.l. About 2/3 of the County is located below the marine limit. Typically, the farming landscape is located below the marin limit, providing clay, silt and sandy soils, while the areas above the marine limit are dominated by hilly forests due to the previous subglacial till coverage.

Figure 1. Study area is the Østfold County in Southern Norway.

Three Landsat ETM+ images acquired over the study area were used in the analysis. To take different stages of the vegetation growing season into account, we selected three cloud-free images to represent a typical spring, summer and autumn situations: 9 May 2001, 5 July 2001 and 11 October 1999. Each image includes 7 spectral bands (visible, near-, mid- and thermal infrared), with 30 m (60 m) spatial resolution. In addition, a digital elevation model of 25 m spatial resolution served to derive terrain slope.

To train the classification algorithm and to validate the results, highly detailed polygon vector data containing buildings (residences, industrial buildings and shopping centres) and roads (car roads, walkways, bicycle paths and parking lots) of Fredrikstad municipality were used. The data is photogrammetrically constructed from aerial photos of scale 1: 3000 (Norwegian FKB-A standard). Additionally, aerial photos of Fredrikstad municipality were available. Topographic maps (1: 50 000) with land cover information were also at our disposal.

Preprocessing of the Landsat ETM+ images consisted of geometric correction and radiometric calibration to at-satellite reflectance (Irish, 2000; Markham and Barker, 1986). Furthermore, we performed a tasseled cap transformation according to Huang et al. (2002). The high resolution polygon data of buildings and roads were converted from polygon data to raster data of 30 m resolution, where each pixel had a value between 0 and 100, representing the percentage of impervious areas within the pixel (Fig. 2). This was performed first by generating high resolution (1 m) raster dataset (impervious/non-impervious), and secondly by calculating the percentage of impervious areas inside the corresponding 30 m Landsat pixels. This dataset was used to train and test the classification.

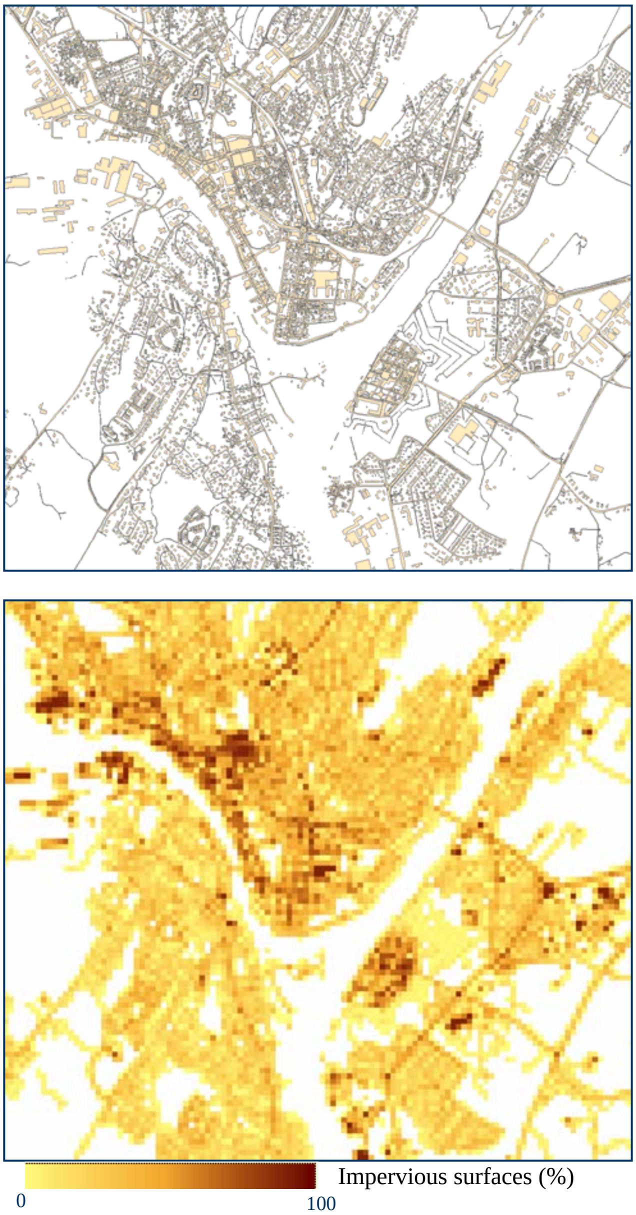

Figure 2. Reference data generated by rasterizing high resolution polygon map data of buildings and roads to 30 m pixels. The image extract shows the center of the city Fredrikstad.

3. REGRESSION TREE ANALYSIS

The classification scheme is based on decision tree methods, since these techniques can handle large and heterogeneous datasets (Breiman et al., 1984). Input data can be of various formats (numbers and/or categories). Decision trees predict categorical values (classes), while regression trees predict continuous values. We chose the regression tree technique since in this study we aimed to map the impervious surface concentration. We followed a similar procedure as described in Yang et al. (2003).

Decision tree methods strongly depend on large amounts of reference data representing the classes and their spectral variability. Software used for the regression tree analysis was Cubist (Quinlan, 1993).

4. EXPERIMENTS

The classification scheme used in this study is shown in Fig. 3. Five experiments were conducted, of which each successively included more input data (Table 1). The first experiment (A) contained a single summer Landsat image, the second (B) included all three Landsat images, the third (C) included two thermal channels, and the last two experiments (D, E) contained different terrain parameters and land-cover maps. The aim was to identify important and less important datasets by evaluating the effects on the classification accuracy. The output of the classification using the regression tree technique is a map showing the imperviousness on a continuous scale from 0 to 100%. The estimate represents the area coverage within a pixel.

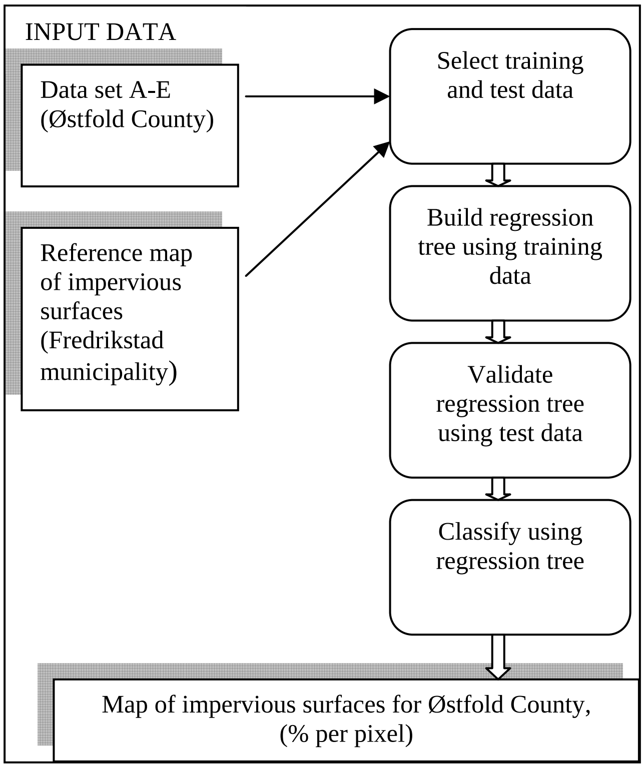

Figure 3. Scheme for the regression tree classification applied in the study.

Table 1. Data Sets A-E used in the Five Experiments. TC = Tasseled Cap Channels, T = Thermal Infrared Channel, DEM = Digital Elevation Model, LC= Land Cover Map.

Data	ETM+	TC	T	DEM	LC
A	Summer	6	-	-	-
B	Spring Summer Autumn	6 6 6	- - -	-	-
C	Spring Summer Autumn	6 6 6	yes - yes	-	-
D	Spring Summer Autumn	6 6 6	yes - yes	elevation slope	-
E	Spring Summer Autumn	6 6 6	yes - yes	elevation slope	yes

5. RESULTS AND DISCUSSION

Results of the regression tree classification are shown in Table 2 and Fig. 4. High average accuracies and correlation coefficients represent a good result. From the table it is clear that data set E gives the best result, while data set A gives the poorest result. By increasing the number of input data layers to the analysis, both the average accuracy and the correlation coefficient increase. Through the inclusion of thermal channels and topographic information and land cover maps, main groups of land cover are easier identified. The number of decision rules varied between 22 and 30. Most of the rules concern areas with 0% imperviousness. This includes sea, lakes, forests and agricultural areas. For these areas the average accuracy is high: 99.4% (Fig. 4). The average accuracy reduces with increasing imperviousness (from 93% to 78% accuracy). Areas having high concentration of imperviousness may be underestimated. This is due to the skewed distribution of the training data. Areas having low imperviousness were well represented, while areas having high imperviousness were poorer represented. To improve the classification of high imperviousness, more training data from densely populated areas should be included in future analysis.

Table 2. Average Accuracy (%) for the Regression Trees applied on Training and Test Data from Data Sets A-E. Corr. = Correlation Coefficient.

Data	TRAINING DATA		TEST DATA	
	Accuracy	Corr.	Accuracy	Corr.
A	97.9	0.59	97.9	0.59
B	98.0	0.65	98.1	0.64
C	98.2	0.70	98.2	0.70
D	98.3	0.71	98.3	0.70
E	98.4	0.74	98.4	0.75

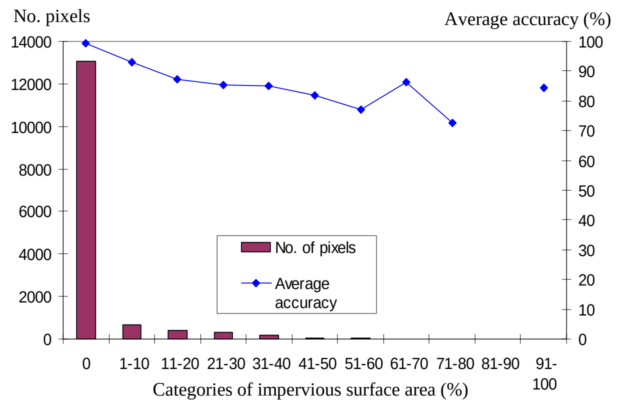

Figure 4. Average accuracy (%) presented for categories of imperviousness, obtained using data set E. It is also shown that the majority of the training and test areas (Fredrikstad municipality) contain 0% imperviousness.

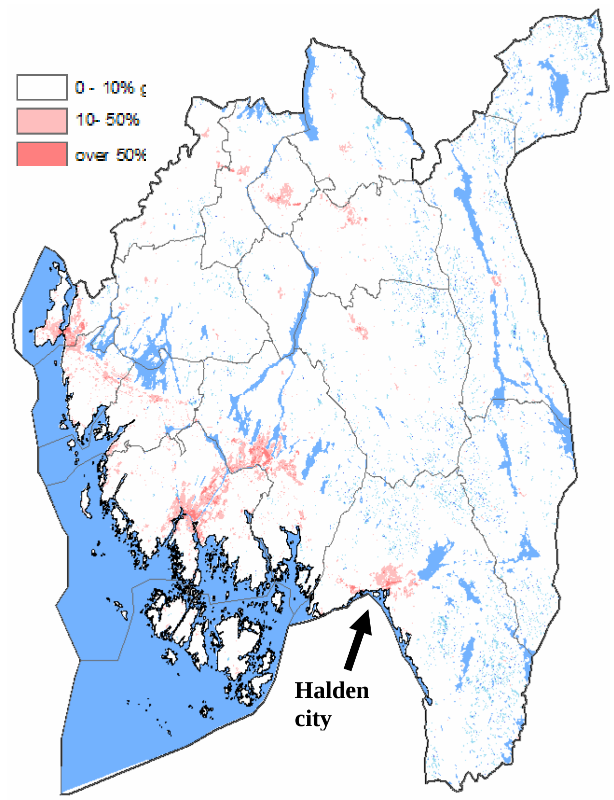

Figure 5. Map showing imperviousness of Østfold County. The imperviousness values are compacted to 3 categories.

A map of impervious surfaces in Østfold County is presented in Fig. 5. Originally, the map contains continuous values from 0-100% imperviousness, but the values are here categorized into 3 classes. In Fig. 6 the classification results for Halden city is shown in detail.

It seems that the sub-pixel classification method solves some of the problems of separating impervious surfaces from natural grey areas such as bare rocks. We conducted two tests, one by including only artificial, impervious areas in the training data set and another one that included natural, pervious surfaces of the entire Fredrikstad municipality in the training data set as well. The outcome of the first test strongly overestimates the extent of impervious surfaces while the result of the second test shows that the method performs much better. We relate the better performance to the fact that natural, impervious areas such as bare rock are often located in a natural environment and hence, the spectral mixture of such an area is significantly different from such of an artificial, impervious area that is typically located in man-made environments.

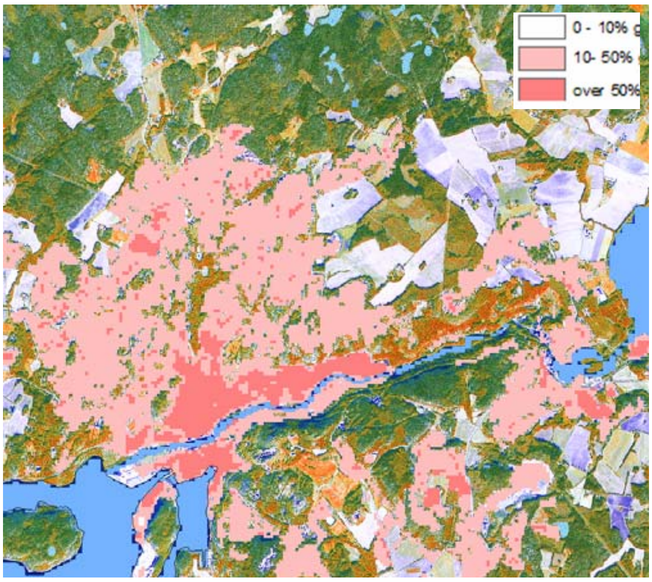

Figure 6. Map showing imperviousness of Halden city in Østfold County. For visualization purposes only, a SPOT image is used to present the area.

6. SUMMARY AND CONCLUSIONS

The main goal of this project has been to map impervious surfaces in Østfold County in Southern Norway without including natural areas such as bare rocks. Regression tree classification was applied on three Landsat ETM+ scenes, each of a spring, summer and autumn situation, in addition to a digital elevation model. Training data for the classification was derived from existing high precision polygon vector data of roads, parking lots and buildings for a selected region within the study area. Five experiments were conducted, of which each successively included more input data. The output of the classification using the regression tree technique is a map showing the imperviousness on a continuous scale from 0 to 100%. The results show that highest classification accuracy (approximately 98%) is obtained for areas with 0% imperviousness. With increasing area coverage of imperviousness the accuracy slightly reduces. For areas with 70-80% imperviousness the accuracy is lowest with approximately 75%. Few areas reach up to 100% imperviousness. We conclude that: 1. Using existing detailed vector data as the only source of training data worked well, and, 2. by including areas without artificial impervious surfaces in the training data we were able to exclude most of the natural impervious areas from the analysis.

7. ACKNOWLEDGEMENTS

The work has been carried out within the national program SatNat, funded by the Norwegian Space Centre and the Directorate for Nature Management. Aerial photo and map data were made generously made available by the Fredrikstad Municipality.

8. REFERENCES

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. Classification and regression trees. The Wadsworth Statistics/Probability Series. Wadsworth International group, California, USA, 358 pp.

Huang, C., Wylie, B., Yang, L., Homer, C.G. and Zylstra, C., 2002. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. International Journal of Remote Sensing, 23(8): 1741-1748.

Irish, R.R., 2000. Landsat 7 science data user's handbook. National Aeronautics and Space Administration.

Markham, B.L. and Barker, J.L., 1986. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. EOSAT Landsat Technical Notes, 1: 3-4.

Quinlan, R., 1993. C4.5: Programs for Machine Learning, Morgan Kaufmann. USGS, San Meteo, California.

Vikhamar, D., Fjone, G., Kastdalen, L. and Bolstad, J.P., 2004. Satellittdata til kartlegging av arealdekke. Utprøving av beslutningstremetodikk i Østfold fylke, Direktorat for Naturforvaltning (DN), Trondheim, Norway.

Vikhamar, D. and Kastdalen, L., 2004. Nøkkeltall i miljøvernpolitikken: Metodikk for bruk av satellittbilder til kvalitetssikring av kartlagte rekreasjonsarealer og nærturterreng: -Nøkkeltallene 2RM4-1 og 2RM4-2. 5, Høgskolen i Gjøvik, Gjøvik.

Yang, L., Huang, C., Homer, C.G., Wylie, B.K. and Coan, M.J., 2003. An approach for mapping large-area impervious surfaces: Synergistic use of Landsat 7 ETM+ and high spatial resolution imagery. Canadian Journal of Remote Sensing, 29(2): 230-240.