COMBINING REMOTE SENSING DATA SOURCES AND TERRESTRIAL SAMPLE-BASED INVENTORY DATA FOR THE USE IN FOREST MANAGEMENT INVENTORIES

Matthias DEES[*], Jan DUVENHORST[**], Claus Peter GROSS[**], Barbara KOCH[*]
[*] University of Freiburg, Germany
Department for Remote Sensing and Landscape Information Systems
ferninfo@uni-freiburg.de
[**] PROCUL Consulting Freiburg, Germany

Working Group VII/4

KEY WORDS: Forestry, Sample Inventory, Stratification, Data Fusion, Photogrammetrie, Landsat TM.

ABSTRACT

This paper presents two elements of a study on forest inventory and mapping in the context of forest management with a test site in the state Nordrhein-Westfalen, Germany. The first section concerns the analysis of the sample based forest inventory in a systematic grid design. The use of aerial photos represents an inexpensive, exact means of mapping the borders of the stand. This is a requirement for the analysis of the sample based forest inventory as a stratified sample. This analysis option makes it possible to reduce the sampling error for the central assessment attributes compared to an analysis using the simple random sampling approach. Along with the possibility of increasing accuracy, it is also possible to reduce the size of the sample by 25% without any loss of accuracy for the central assessment attributes. The second section concerns the k-nearest-neighbour method, in which sample data and medium resolution satellite data (Landsat TM and IRS1C LISS) are used. This method can provide a representation of the spatial distribution of central attributes. So far the mapping of main tree species has been the subject of study. This method does not provide a sufficient information basis for the standwise forest management inventory under the forest conditions that apply to the area studied. It can, however, provide a good overview of the spatial distribution of the main tree types.

1 INTRODUCTION

The study on the system of forest inventory and mapping was commissioned by Civil Forest Service of the state Nordrhein-Westfalen, Germany. The system of forest inventory and mapping is part of the system of forest management. In the civil forest enterprises, every ten years the forest management plan is revised or updated. Elements in gathering information are 1) remote sensing, 2) a terrestrial sample based inventory (whose introduction is planned), 3) surveys for the entire stands in a part of the stands, and 4) information from previous surveys. The general aim of the study is, on one hand, to examine the possibilities of introducing remote sensing into the system of forest inventory and mapping and, on the other, to examine the possibilities that arise from the introduction of the stand inventory on the basis of samples. The introduction of the stand inventory on the basis of samples makes statistically supported information possible at the level of the forest enterprise and for larger sections of the forest enterprise. The gathering of information on the level of the stand thus is no longer necessary to provide information about the entire forest enterprise that could be gained from the summation of the information of all single stands. Nevertheless, the single stands should continue to be described individually as they represent the smallest spatial element for forest management. The previous method of mapping and describing the stand, which is time-consuming and costly, should be replaced to the extent possible by less expensive survey methods. In this context, this study comprises the following elements:

1. Evaluation of the possibilities of mapping the stands using various data sources and techniques by means of visual interpretation (identification of homogenous units, registering of borders, updating of borders).
2. Evaluation of the possibilities of describing the mapped stands with various data sources and techniques by means of visual interpretation.
3. Evaluation of the possibilities of describing the stands with the k-nearest-neighbour method by linking the sample information with the information from the satellite data from the sensors Landsat TM5 and IRS1C LISS.
4. Identification of further possibilities of high quality stand descriptions with a minimum of terrestrial surveys, thus reducing costs.
5. Investigation of the option of using stratification in the analysis of the sample inventory.

The present paper presents two areas of investigation from the overall context of the study: on the one hand, the method of stratification in the analysis of sample-based inventory; on the other, the possibilities of the k-nearest-neighbour method in providing information for important attributes concerning the forest stands and within the stands.

2 TEST SITE AND DATABASE

The study is taking place in the state forest of the forest district Münster in the state of Nordrhein-Westfalen. The location of the state forest Münster in the Federal Republic of Germany is shown in figure 1. The state forest Münster consists of several relatively small, very widely scattered forest areas as shown in figure 1. The forest area takes up an area of around 2,600 ha. Oak (Quercus robur & Quercus petrea) and beech (Fagus sylvatica) stands predominates among broad-leaved stands, while spruces (Picea abies) dominate among the conifers. Within the framework of the study, the data listed in table 1 along with the times of the surveys were available.

Data type	Year of acquisition
Former stand-based inventory	1984-1995
Actual stand-based inventory	1998
Systematic inventory sample size 817, grid 125m × 250m , concentric sample plots (r = 1m to r = 12m)	1998
Forest stands with accurate surveys for 23 individual stands, based on diameter measurements of all trees	1998
Digital panchromatic Orthophotos, scale 1: 12 500	1988-1996
CIR-photos, scale 1: 12 500	1998
IRS 1C – PAN, IRS 1C – LISS	1997
Landsat TM 5, supplied already geo-referenced and topographically normalised by the forest research institution of the state Nordrhein-Westfalen LÖBF, Münster.	1997
Digital Hight Model, 50 m raster	

Table 1. Overview of the study data

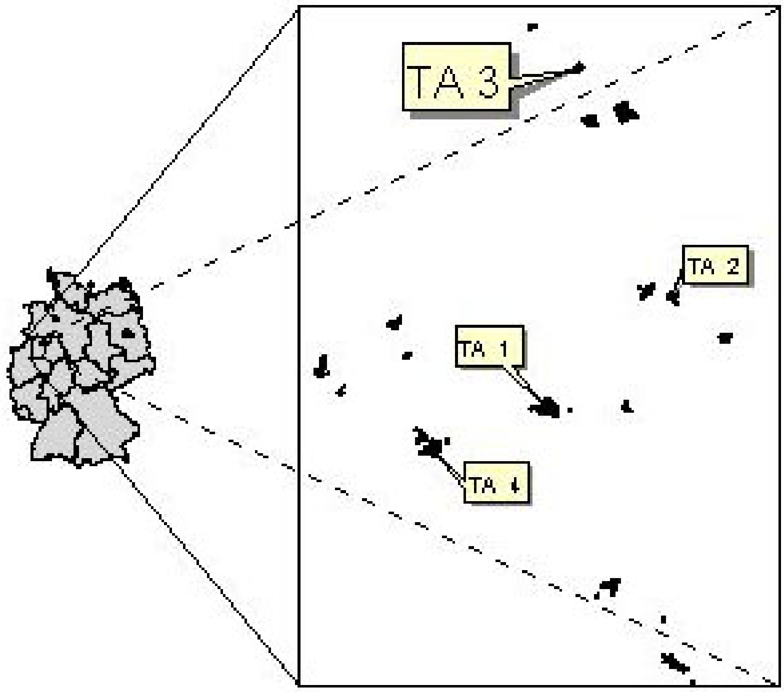

Figure 1. Location of the area under study within Germany, with test areas, selected for intensive analysis

3 STRATIFICATION

3.1 Requirements

The partition of the entire forest area in the forest district into stands of known size is a requirement for the application of stratification. The summarising of stands into strata using attributes of the stands results in strata with known borders and areas inside the forest district.

A distribution of the samples in a systematic grid makes an optimal distribution within the area under study possible and leads, along with a partition into strata, to a distribution of the samples across the strata that is approximately proportional to the area of the strata, provided that the sample size comprises the number of strata several times over. The exact localisation of the samples during field work allows all sample points to be clearly assigned to stands.

These requirements are fulfilled in the forest inventories in the public forest enterprises. A further requirement is the exact delineation of the forest stands. This can be done at little economic expense by registering the stand borders using aerial photos (Duvenhorst & Niehaus-Übel, 1996). Figure 2 shows the mapping of forest stands carried out within the framework of the study on the basis of CIR aerial photos.

3.2 Statistical methods

In the sample inventory estimates are needed for a large number of attributes. Statistically, they can be primarily derived from estimates of sums, mean values, and ratios. The forest area is known in inventories of forest enterprises. If the attributes are determined per area with reference to the single plot size and as quantities per ha, sums can be determined from the product of the characteristics mean values and the overall area in ha. The following examples should make this clear:

Attributes	volume per ha	(e.g. 250 m³/ha)
	volume spruce per ha	(e.g. 60 m³/ha)
	area spruce per ha	(eg. 0,12 ha/ha = 0,12)

Estimation of corresponding totals:

Volume = total area × mean[volume per ha]

Volume spruce = total area × mean[volume spruce per ha]

Species area = total Area × mean[area spruce per ha]

Estimation of corresponding ratios:

Volume spruce per area spruce = mean[volume spruce per ha] / mean[area spruce per ha]

Thus it is sufficient to derive estimators for means with the corresponding variances of the estimation of the mean and to derive estimators for ratios with the corresponding variances of the estimation of the ratios. An alternative is to derive estimators for totals and ratios instead of means and ratios as described by e.g. Dees (1996), but since the means and totals defined above differ only by the factor of the known area, the two approaches are equivalent.

The combination of a systematic grid with a partition of the area into strata can be viewed approximately as an independent systematic sampling within each strata. The estimation of overall estimates is based on the estimation of means, variances and covariances inside the strata. The estimation inside the strata is done assuming simple random sampling inside the strata. Thus the resulting variances will overestimate the true values resulting in conservative estimates of the errors and derived confidence statements.

To evaluate the benefits of a stratified estimate, an estimate is calculated without stratification. These estimates are calculated using estimators for simple random sampling, again resulting in conservative estimates of the errors. This is common practice in forest inventories in order to be on the safe side. Since the finite population correction can be neglected in forest inventories for forest enterprises, the following estimators do not include a finite population correction.

Figure 2. Registering of the stand borders based on aerial CIR photos

In simple random sampling, the mean is estimated by

$$\hat{\bar{y}}_{srs} = \sum_{k=1}^{n} y_k / n \tag{1}$$

and the variance of the estimator is estimated by

$$\hat{V}(\hat{\bar{y}}_{srs}) = \frac{1}{n(1-n)} \sum_{k=1}^{n} (y_k - \sum_{k=1}^{n} y_k / n)^2 \tag{2}$$

with y_k - the value of the attribute y on element *k*, that is one of the elements out of the sample *k=1*..n of size *n* (Cochran, 1976). In simple random sampling the ratio is estimated approximately unbiased by

$$\hat{R}_{srs} = \frac{\hat{t}_{y,srs}}{\hat{t}_{z,srs}} = \frac{\hat{\bar{y}}_{srs}}{\hat{\bar{z}}_{srs}} \tag{3}$$

where *t* is the symbol for sums and the variance of the estimator is estimated by

$$\hat{V}(\hat{R}_{srs}) = \frac{1}{\hat{\bar{z}}_{srs}^2} \frac{1}{n} \left[S_y^2 + \hat{R}_{srs}{}^2 S_z^2 - 2\hat{R}_{srs} S_{yz} \right] \tag{4}$$

with

$$S_y^2 = \sum_{k=1}^{n} \frac{(y_k - \sum_{k=1}^{n} y_k / n)^2}{n-1} \;;\; S_z^2 = \sum_{k=1}^{n} \frac{(z_k - \sum_{k=1}^{n} z_k / n)^2}{n-1} \;;\; S_{yz} = \sum_{k=1}^{n} \frac{(y_k - \sum_{k=1}^{n} y_k / n)(z_k - \sum_{k=1}^{n} z_k / n)}{n-1}$$

(Särndal et al, 1992, Dees, 1996) with z_k - the value of the attribute z on element *k* and where $\hat{\bar{z}}_{srs}$ is defined according to $\hat{\bar{y}}_{srs}$.

In stratified sampling with simple random sampling of n_h elements ($k=1..n_h$) within *h* strata (*h= 1.. H)* the mean is estimated as

$$\hat{\bar{y}}_{strat} = \sum_{h=1}^{H} W_h (\sum_{k=1}^{n_h} \frac{y_k}{n_h}) \tag{5}$$

with the strata specific weights W_h = [Area of Strata h] / [total area].
The variance of the estimated mean is estimated by

$$\hat{V}(\hat{\bar{y}}_{strat}) = \sum_{h=1}^{H} W_h^2 \frac{\frac{1}{n_h - 1} \sum_{k=1}^{n_h} (y_k - \bar{y}_{s_h})^2}{n_h} \tag{6}$$

(Cochran, 1976, Särndal et al., 1992).
In stratified sampling with simple random sampling within the strata, the ratio is estimated approximately unbiased as

$$\hat{R}_{strat} = \frac{\hat{t}_{y,strat}}{\hat{t}_{z,strat}} = \frac{\hat{\bar{y}}_{strat}}{\hat{\bar{z}}_{strat}} \tag{7}$$

The variance of the estimated ratio is estimated by

$$\hat{V}(\hat{R}_{strat}) = \frac{1}{\hat{\hat{z}}^2_{strat}}(S^2_{y,strat} + \hat{R}^2_{strat}S^2_{z,strat} - 2\hat{R}_{strat}S_{y,strat}) \qquad (8)$$

with

$$S^2_{y,Strat} = \sum_{h=1}^{H}[\frac{W_h^2}{n_h}\sum_{k=1}^{n_h}\frac{(y_k - \sum_{k=1}^{n_h} y_k / n_h)^2}{n_h - 1}] \; ; \; S^2_{z,Strat} = \sum_{h=1}^{H}[\frac{W_h^2}{n_h}\sum_{k=1}^{n_h}\frac{(z_k - \sum_{k=1}^{n_h} z_k / n_h)^2}{n_h - 1}]$$

$$S_{yz,Strat} = \sum_{h=1}^{H}[\frac{W_h^2}{n_h}\sum_{k=1}^{n_h}\frac{(y_k - \sum_{k=1}^{n_h} y_k / n_h)(z_k - \sum_{k=1}^{n_h} z_k / n_h)}{n_h - 1}]$$

where $\hat{\hat{z}}_{strat}$ is defined according to $\hat{\hat{y}}_{strat}$ (derived from Särndal et al. (1992), equation. 5.6.10). The sampling errors can be derived by taking the square root of the variances of the estimates.

3.3 Results and Conclusions

For the establishing of strata, a partitioning was selected that enables a reduction of the estimation errors both for the estimates according to age groups and according to tree type. Every one of the stands has a main tree type. All stands are assigned to a stratum using the main tree type and the age group of the main tree type. First, the main tree type assignments were made: class 1 'oak' ; class 2 'other broad-leaf tree species'; class 3 'coniferous species'. Within each of these classes, three age classes are defined according to table 2, resulting in 9 strata. The entire area of 2268,3 ha was divided into these 9 strata, which covered proportions of the entire area from 7.1% to 18.7%.

	age group 1	**age group 1**	**age group 1**
class 1 'oak'	<= 60 years	<= 40 years	<= 40 years
class 2 'other broad-leaf tree species'	61 – 120 years	41 - 80 years	41 - 80 years
class 3 'coniferous species'	> 120 years	> 80 years	> 80 years

Table 2. Definition of the strata

Attribute	**Estimate with absolute sampling error, stratified sampling**		**Estimate with absolute sampling error, simple random sampling**		**Potential to reduce the sample size using stratified random sampling (maintaining the sampling error) in [%] for the single attributes**
Standing cross volume [m³]	631495	± 9754	630587	± 11342	26
Standing commercial volume [m³]	484509	± 7712	483148	± 9073	26
Oak, standing cross volume [m³]	214581	± 7032	216623	± 10661	56
Beach, standing cross volume [m³]	148800	± 8393	142449	± 8846	8
Spruce, standing cross volume [m³]	69183	± 6124	75081	± 7712	38
Oak, relative area [%]	34,2	± 0,9	34,6	± 1,5	60
Beach, relative area [%]	18,2	± 1,1	17,4	± 1,1	8
Spruce, relative area [%]	9,9	± 0,8	10,7	± 1,0	40
Standing cross volume per forest area [m³/ha], main stand	253,8	± 4,3	253,9	± 4,8	23
Oak, standing cross volume per beech area [m³/ha], main stand	271,9	± 5,9	271,4	± 6,6	23
Beech, standing cross volume per beech area [m³/ha], main stand	300,9	± 10,5	299,9	± 11,0	9
Spruce, standing cross volume per spruce area [m³/ha], main stand	299,3	± 14,1	301,3	± 15,3	7

Table 3. Estimates, errors and potential to reduce the sample size for main attributes
(main stand: stand without reserving of standards and undergrowth)

Out of the 817 sample plots 735 that belonged to the state forest have been used. The results of the estimates for the main attributes is given in table 3. The estimates are given for the main attributes, the total volume, the area and volume parameters of the main broad-leaf tree species group and the main coniferous tree species group. All main attributes are estimated with smaller sampling error. The reduction of error varies from attribute to attribute. If the total volume is given the highest priority, the potential to reduce the sample size due to estimating with stratification is 26%, or roughly 25%. By reducing the grid density in one direction by 0.5 , this reduction can be easily achieved. Further analysis planned in the study comprises an alternative definition of strata (4 age classes mixed with 2 species type classes [broad-leaved / coniferous]), an analysis of different grid densities and the analysis of the option to use differing grid densities in different strata and a comparison with other stratification techniques for forest inventories for forest enterprises as developed by Böckman et al. (1998).

4 USING THE KNN TECHNIQUE

4.1 Requirements and data preparation

A large number of sample plot data that are geo-referenced and satellite data also geo-referenced and topographically normalised are a prerequisite for applying the k-nearest-neighbour method (Tompoo & Pekkarinen, 1997). The first analysis was made on base of the Landsat TM data; the analysis based on IRS 1C LISS data is under preparation. The Landsat TM 5 data have been supplied in an already geo-referenced and topographically normalised form by the forest research institution of the state Nordrhein-Westfalen LÖBF, Münster. Details on the data processing are given in Diemer & Lucaschewski (1999). The TM-channels 1 to 5 and 7 have been used. All of the 817 sample plots have been used, both the 735 that belonged to the state forest and 82 plots that have been assessed in the forest of a public foundation.

4.2 Methods

The k-nearest-neighbour method for quantitative attributes postulates that there is a context between a measurable physical attribute, such as the timber volume and the spectral signature of multi-spectral remote sensing data. If for a large number of sample points ("reference points") the spectral values of corresponding pixels and terrestrial measurements of attributes are available, an estimate can be determined for all pixels for which no information from sample data is available using a simple method that does without model assumptions: for every pixel *s*, the Euclidean distance to all reference points $v = 1..n$ (i.e., to their corresponding pixels) is determined as a measurement for the similarity of the signature (Tompoo & Pekkarinen, 1997):

$$E_{sv} = \sqrt{\sum_{i=1}^{I} (B_{is} - B_{iv})^2} \tag{9}$$

E_{sv}: Euclidean distance of pixel *s* to reference point *v*
I: number of channels
B_{iv}: value of the reference point *v* in channel *i*
B_{is}: value of the pixel *s* in channel *i*

The reference points *k* $(j=1..k)$ with the closest distance are then selected from all *n* reference points. A weight W_j is assigned to each of the selected *k* reference points so that the sum of all *k* weights is 1 and the weight is reversely proportional to the square of the Euclidean distance is (Tompoo & Pekkarinen, 1997):

$$W_j = \frac{1/E_{svj}{}^2}{\sum_{j=1}^{k} 1/E_{svj}{}^2} \tag{10}$$

Let the measurement of the attribute on the reference point *j* be *Mj*. The estimate value $\hat{M}_s$ for pixel *s* is then calculated as the balanced mean of *k* values (Tompoo & Pekkarinen, 1997):

$$\hat{M}_s = \sum_{j=1}^{k} W_j * M_j \tag{11}$$

In this process, an estimated value is also calculated for the pixels for which reference information is available. This reference area, though, is not used for these estimates so that only n-1 reference areas are available for such pixels. This enables a cross-validation as described below.

In the k-nearest-neighbour method for qualitative attributes, the sum of the weights is first calculated for each class of the qualitative attribute (Tompoo & Pekkarinen, 1997). Then the class of the qualitative attribute with the greatest weight sum is assigned to the pixel.

The verification of quantitative attributes is done by the root mean square error $RMSE_{knn}$

$$RMSE_{knn} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \qquad (12)$$

$i=1..n$ verification area
n number of verification areas
y_i: measurement at verification area i
$\hat{y}_i$: k-nearest-neighbour estimate of the attribute at verification area i

(Facakas et al. 1999).
For the further analysis of the quantitative attributes, the root mean square error addressing the overall average of the sample survey to all pixels $RMSE_{average}$ is calculated

$$RMSE_{average} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2} \qquad (13)$$

$i=1..n$ reference area
n number of verification areas
y_i: measurement at the verification area
$\tilde{y}_i$: estimated overall mean

The root mean square error based on the k-nearest-neighbour estimate $RMSE_{knn}$ is a measure of the accuracy of the estimates. The comparison of the root mean square error based on the k-nearest-neighbour estimate $RMSE_{knn}$ with the root mean square error addressing the overall average of the sample survey to all pixels $RMSE_{average}$ indicates the additional information gained by the k-nearest-neighbour method. This can be done with arbitrary sizes of verification areas, such as forest stands with data from accurate surveys. The verification and evaluation can also be done on a pixel level using cross validation (Facakas et al. 1999). The validation of qualitative attributes is made by measurements of co-occurrence.

4.3 Results and Conclusions

The qualitative attribute 'dominating tree species group of the area' (spruce, pine, oak, beech, other broad-leaved trees) and the quantitative attributes 'area proportion of a single tree species group' were studied. The visual comparison with the aerial photo shows great correspondence when single tree types dominate over large areas (see figure 2 and 3). Such dominance of single tree types does not, however, exist over large areas. In addition, in small stands mixed signatures predominate due to the influence of neighbouring stands. From the pixel-wise 'dominating tree species group of the area', the 'dominating tree species group of the stand' was calculated determining the 'dominating tree species group of the area' with the highest proportion within the stand. The comparison with the reference data from stands with accurate surveys shows that correspondence is insufficient at an overall accuracy of 47.8% (n = 23). If only stands with a size of two and more ha are included the overall accuracy is considerably higher (70%, n = 10).

	All verification stands, n = 23			**Large stands > 2 ha, n = 10**		
	Oak	**Beech**	**Spruce**	**Oak**	**Beech**	**Spruce**
$RMSE_{knn}$ [%]	28.1	26.4	35.4	25.9	15.0	21.1
$RMSE_{average}$[%]	32.7	28.1	49.2	35.8	20.5	46.3
improvement in *RMSE*[%]	4.6	1.7	13.8	9.9	5.6	25.2

Table 4. Evaluation of the quantitative attribute 'area proportion of a single tree species group'

The verification parameters for the quantitative attributes 'area proportion of a single tree species group' for the dominating species groups oak, beech and spruce are given in table 4. The accuracy provided by the KNN method is quite low for single stands but is considerably better for stands of 2 ha or more. Further attributes will be analysed within the study and a second set of reference data will be used. The preliminary conclusions are that this method thus does not provide sufficient information on the stand level for a forest management plan under the forest conditions (especially stand sizes and species mixture) that apply to the area studied. It can, however, provide a good overview of the spatial distribution of the main tree types.

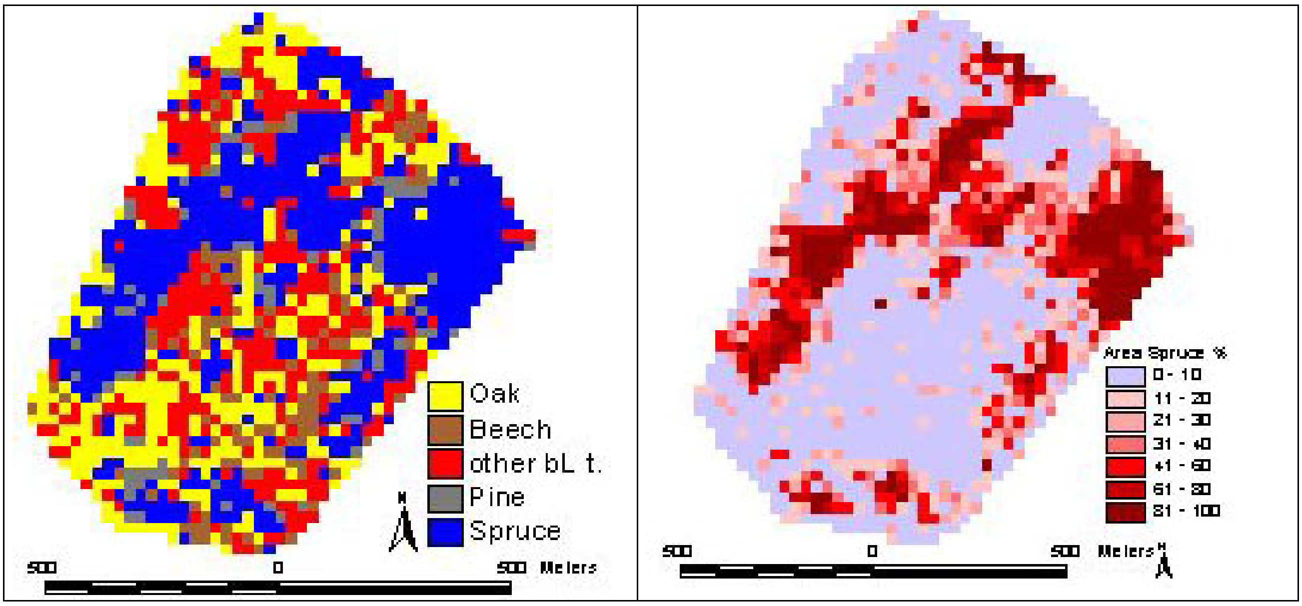

Figure 3. Dominating tree species (left), area proportion of spruce (right) in test area 3 (see Fig. 1)

REFERENCES

Böckmann, Th., Saborowski, J., Dahm, S., Nagel J. und Spellmann, H., 1998. Die Weiterentwicklung der Betriebsinventur in Niedersachsen, Forst und Holz. 3, p. 219-226.

Cochran, W. G.,1977. Sampling Techniques (3. Edition). John Wiley and Sons, New York.

Dees, M., 1996. Regressions- und Kleingebietsschätzung bei forstlichen Großrauminventuren unter Nutzung von Forsteinrichtungs- und Satellitendaten. Mitteilungen der Abteilung Forstliche Biometrie, 96-1. Forstwissenschaftliche Fakultät, Universität Freiburg i. Br.

Diemer, C., Lucaschewski, I., 1999. Verbesserung der stichprobenbasierten Landeswaldinventur Nordrhein-Westfalen mit Hilfe von flächigen Satellitenbilddaten. http://www.dglr.de/erdanwendungen/index.html. (4.4.2000)

Duvenhorst, J. und Niehaus-Übel,G., 1996. Zeitgemäße Forsteinrichtung mit Fernerkundung und GIS. Allgemeine Forst und Jagdzeitschrift - Der Wald. 2/1996, p. 68-70.

Facakas, Z., Nisson, M. & Olsson, H., 1999. Regional forest biomass and wood volume estimation using satellite data and ancillary data. Agricultural and Forest Meteorology 98-99, p. 417-425.

Särndal C. E., Swenson, B. & Wretman, J., 1992. Model Assisted Survey Sampling. Springer-Verlag, New York.

Spelsberg, G., 1997. Verdichtung des Stichprobennetzes aus dem Testlauf für eine Betriebsinventur im Staatswald des Landes NRW. In: Landesforstverwaltung NRW (Eds.): Testlauf zur Landeswaldinventur. Landesforstverwaltung NRW, Düsseldorf, p. 116-122.

Tomppo, Erki und Pekkarien, A. ,1997. Methodenerprobung der Finnischen Multiquellen-Waldinventur in Nordrhein-Westfalen. In: Landesforstverwaltung NRW (Eds.): Testlauf zur Landeswaldinventur. Landesforstverwaltung NRW. Düsseldorf. p. 52-61.