Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 110 (2007) 420–430
Remote Sensing of Environment
www.elsevier.com/locate/rse

Accuracy of forest mapping based on Landsat TM data and a kNN-based method

Arnt Kristian Gjertsen

Norwegian Forest and Landscape Institute, P.O. Box 115, N-1431 Ås, Norway

Received 9 August 2005; received in revised form 31 July 2006; accepted 4 August 2006

Abstract

A multi-source forest inventory (MSFI) method has been developed for use in the Norwegian National Forest Inventory (NFI). The method is based on a k-nearest neighbour rule and uses field plots from the NFI, land cover maps, and satellite image data from Landsat Thematic Mapper. The inventory method is used to produce maps of selected forest variables and to estimate the selected forest variables for large areas such as municipalities. In this study, focus has been on the qualitative variables 'dominating species group' and 'development class' because these variables are of central interest to forest managers. A mid-summer Landsat 5 TM scene was used as image data, and all NFI plots inside the scene were used as a reference dataset. The relationship between the spectral bands and the forest variables was analysed, and it was found that the levels of association were low. A leave-one-out method based on the reference dataset was used to estimate the pixel-level accuracies. They were found to be relatively low with 63% agreement for species groups. An independent control survey was available for a municipality and estimates from the MSFI were compared to it. The levels of error were quite high. It was concluded that the large area estimates were biased by the reference dataset.

Keywords: Multi-source forest inventory; kNN; Accuracy; Landsat TM

1. Introduction

1.1. The national forest inventory in Norway

The national forest inventory (NFI) in Norway is based on a regular 3 km by 3 km network of permanent sample plots and supplemental temporary plots. Each year 1/5 of the permanent plots is visited, and in 5 years all permanent plots are measured. Every 15 years, all temporary plots inside a county are measured in order to produce more accurate estimates at the county level. Each NFI plot is defined as a circle with a 250 m^2 area (radius = 8.92 m), and several forest variables are measured and calculated (e.g. tree species, age, volume) for each plot. In addition, forest stand variables, such as site index, development class, stand species class, are assessed based on a 1000 m^2 area of a homogenous stand that the plot falls inside. Development class is a silvicultural concept based on site index and stand age. There are five different development classes: I under regeneration, II regenerated area and young forest, III young thinning stand, IV advanced thinning stand, and V mature forest. The NFI delivers statistical reports for counties every 15 years and for the whole country each year. The system has limitations and is neither able to produce reliable statistics for small areas such as municipalities or wall-to-wall maps showing where inside the inventory area forest resources are located.

1.2. Combining NFI datasets with ancillary datasets

This limitation is common to all sample based NFIs, and in Finland a multi-source forest inventory (MSFI) method has been developed as a solution (Tomppo & Siitonen, 1991). The method uses NFI plot data together with remotely sensed images and other ancillary datasets such as digital elevation models (DEMs) and map data of general land cover classes. The datasets are combined, and the NFI plots with forest variables and ancillary data are used as reference data for a nonparametric k-nearest neighbour (kNN) rule that for each pixel in the inventory area finds the *k* nearest NFI plots in feature space

E-mail address: arnt.kristian.gjertsen@skogoglandskap.no.

0034-4257/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2006.08.018

(Tomppo & Halme, 2004). The NFI plots are typically located outside the inventory area; however, using ancillary data the method calculates the representativeness of the external NFI plots in the form of new area weights. The advantage of the kNN method is its simplicity and that it estimates all variables simultaneously. Another advantage is that the method better preserves the relationship or covariance structure between variables than methods where predictions are made separately for each variable (Tomppo & Halme, 2004).

The development in Finland has inspired a similar development of the Norwegian NFI (Gjertsen et al., 1999). The kNN-based MSFI is a very attractive method because it works in a manner familiar to the NFI by making estimates based on sample plots with associated area weights. In addition, MSFI also produces wall-to-wall maps showing the location of different forest resources. The basic difference is that in NFI all plots have the same area weights, while in MSFI, the plots receive different area weights according to how similar they are to the pixels of the inventory area. Similarity is not based on forest variables, but rather on the vector of spectral values from the image pixel covering the plot.

1.3. Experiences with the kNN-based forest inventory

In Finland the kNN-based MSFI has been used for more than 10 years as a part of the NFI. Wall-to-wall maps are being used by the forest industry for timber procurement and ecologists use the maps for habitat analyses (Tomppo, 2005). In Sweden, the method has been used to produce a complete map database for the whole country, named kNN Sweden. It has been used to improve forest statistics from the NFI by using post stratification based on stem volume strata derived from the database. The standard errors for estimates of total stem volume, stem volume for spruce, stem volume for pine, and woody biomass have been reduced by 10% to 30% at the county level (Nilsson et al., 2005). In New Zealand the kNN-based MSFI has been tested for preharvest inventory of a forest plantation with pine trees. Estimates were made at pixel level and stand level. A cross validation test showed that the estimates were unbiased but with high root mean square errors (RMSEs) (Tomppo et al., 1999). The same conclusion was made in a study in south-east Norway (Gjertsen & Eriksen, 2004). The pixel level accuracy of the kNN-based MSFI method has been tested on several areas and satellite datasets in Finland (Tokola et al., 1996). They found that the errors at pixel level for volume estimates were relatively high, with RMSE for total volume around 75 m^3/ha or between 62 and 68% of the mean estimated value. For volumes by species groups the relative errors were even higher and above 100%. The bias was however found to be low, and they concluded that for large area estimates the accuracy would improve. Also in Fazakas et al. (1999), where a study from Sweden was reported, the pixel-level accuracy was reported to be poor. However they found that the mean RMSE was reduced as a function of area of aggregation and obtained RMSE of 8.7% for biomass and 4.6% for wood volume for an inventory area of 510 ha. In Tomppo and Katila (1991) kNN-based volume estimates for three municipalities were compared with independent surveys of the municipalities made for forest planning purposes. The estimates from the latter survey varied more from municipality to municipality than the kNN-estimates. In Dees et al. (2000) a kNN-based method based on Landsat TM data was tested on a site in Germany. Area proportions of single tree species groups were estimated for forest stands. It was found that the kNN method improved the estimated values from 1.7% to 25.2% relative to estimates based on the mean values of the sample of reference plots. Reduction of RMSE was used as indicator of improvement. It was concluded that the method does not provide sufficient information for a forest management plan but that it provides a good overview of the spatial distribution of the main tree types.

1.4. MSFI and estimation errors

One problem with MSFI is that the mapping from forest variable space to spectral feature space is a many to one process (Tomppo et al., 2004). That is, forest plots of different nature can have similar spectral values. In an ideal situation, the mapping would be one to one. However, we know from experience (Tokola et al., 1996) that this is not the case and we expect inaccuracies caused by this mapping problem. The accuracy of the MSFI method is determined by the error of the pixel level estimates. A common model for measurements or estimates is that every measurement or estimate is an additive composite of two components: true value and random error. Random error affects the variability around the average but not the average. However, the real situation is often different and not all errors are random. We often have to deal with systematic error components caused by factors that systematically affect the measurements in some direction across the sample, and therefore also influence averages for large areas.

What are the likely causes of systematic error (bias) in the kNN-based MSFI method? One cause is that all forest conditions in the inventory area may not be represented in the reference dataset. Another cause is that the k nearest neighbours in feature space will not always be the k nearest neighbours in variable space. With a many-to-one mapping into feature space a certain form of 'randomness' is introduced in picking the nearest neighbours. This may cause the estimates to be biased if the distributions of the forest variables in the reference dataset are different from their true distributions in the inventory area.

A kNN-based MSFI has been tested on an area in Norway. The objectives were to test the accuracy of the estimates both at pixel and municipality levels and to study the error components and in particular the extent to which the reference dataset introduces a bias for large area estimates.

2. Data and study area

2.1. Study area

The study area is located in the southeast part of Norway and lies between 59 and 61° north centred on Oslo (Fig. 1). This is a nemo-boreal area dominated by conifers, Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*), with only small patches