

The accuracy and relative efficiency of Landsat data and orthophotos for determining area and volume of spruce

D. A. Stellingwerf¹ and Sann Lwin²

¹ Forest Department, ITC, P.O.Box 6, 7500 AA Enschede, Netherlands

² Forest Department, Burma

Key words: forest inventory, remote sensing

Summary

The accuracy of measurement of forest class areas and non-forest areas from Landsat CCTs, using a principal component transformation analysis, was tested in a 12 841 ha area (72 blocks of ca. 178 ha each). The following classes were considered: S (pure spruce) > 40 years, M (mixed spruce-beech forest) > 40 years, B (pure beech forest) > 40 years, Y (young forest \leq 40 years) and N (non-forest). A comparison with areas from 1:10 000 scale black-and-white infrared orthophotos of the same date revealed that differentiation of species and age classes with the Landsat data was very inaccurate. Also non-forest areas, especially the smaller patches, showed great deviations. All deviations displayed a random character. The combined, unclassified forest areas in the 72 blocks were reasonably accurate, however.

The volumes of spruce, applying a two-stage sampling method, using Landsat and field plot data against orthophoto and field plot data, were also compared. Selection of primary units (PU) was based on probability proportional to size with replacements, using areas in the PUs of pure and mixed spruce in a cumulative list. In each selected PU, five secondary units (circular 0.05-ha plots) were randomly selected and enumerated in the field.

The Landsat combination required 52.6 % more PUs to obtain a sampling error of 4 %, which required 23.3 % more man-days. This increase plus additional costs make Landsat field sampling more costly for the given area.

Introduction

During the past five to ten years a great amount of information has appeared in the literature about the use of Landsat for forest vegetation mapping and timber volume determination, and about its easier availability and advantage over line and orthophoto maps with respect to computerized area determination. It was, therefore, considered worthwhile to test whether Landsat computer printouts could be used efficiently to determine forest areas. The accuracy of the measurements was determined by comparing them with areas from orthophotos, and comparing both with

field data. The standing timber volume of a tree species was also determined with Landsat data. For this purpose, a two-stage sampling design for timber volume of spruce was used, combining Landsat printouts and field data. Using the same design for the orthophotos in combination with field data opened the possibility of comparing the results of both methods.

A test area in Austria was selected for which Landsat cover and aerial photographs taken at the same data were available. The selected area covers 12 841 ha and is located in Upper Austria. The forest on this area consists mainly of Norway spruce (*Picea abies*) in pure and in mixed stands with beech (*Fagus sylvatica*). These stands cover large areas. Pure beech stands also occur in smaller areas, just as mixtures of spruce-silver fir (*Abies alba*). Non-indigenous species such white pine (*Pinus strobus*) and European larch (*Larix decidua*) are also found in mixture with spruce, but they occur less frequently. The mean elevation of the terrain is approximately 650 m above sea level.

Computer classification of Landsat MSS data

The method used for the supervised computer classification of the Landsat MSS images was one of the maximum likelihood classifiers (sigma programme). Each pixel reflectance vector in this system is measured against the sample parameters of various class spectral signatures and assigned to one or none of them. Bands 5 and 7 were selected for training because band 5 is suitable for the separation between forest and non-forest and band 7 is superior for the differentiation of classes within the forest.

Eight well-identifiable pixels on the printouts of bands 5 and 7 (scale 1:50 000) were selected outside the study area. The printouts were adjusted for skew distortion and they had an automatically built-in histogram equalization of reflectance levels for nine grey scales, ranging from black to white. The corresponding positions of the pixels were marked on the orthophotos. A transformation of the Landsat printout area equal to that of test area (covering 12 841 ha) on the orthophoto was obtained by using a multiple linear regression equation between Landsat and orthophoto coordinates of these eight points. The study area was divided into 72 equally sized and shaped blocks, each covering 20×20 pixels (= 178.3 ha).

A number of training pixels were selected in the pre-defined classes with the orthophotos as reference. These classes were S (pure spruce ≥ 40 years), M (mixed spruce-beech ≥ 40 years, with ≥ 10 % beech and ≥ 10 % spruce), Y (young forest ≤ 40 years), B (pure beech forest ≥ 40 years) and N (non-forest). These training pixels were the centres of 3×3 pixel formations. The feature space of the training set with band 5 (x-axis) and band 7 (y-axis) was less satisfactory with respect to the class differentiation than the one with the principal component (PC) transformation analysis.

The clusters of the forest classes became more 'solid' and the number of unclassified pixels were reduced accordingly when a sigma value of 0.4 was used instead of 0.2 or 0.3. The areas which resulted from the computer classification with the feature spaces of the principal components PC 1/PC 2 and of band 5/7 as inputs of the

LANDSAT DATA AND ORTHOPHOTOS FOR DETERMINING SPRUCE

Table 1. Areas of classes in % of the computer classification with a sigma value 0.4.

Class	PC1/PC2	Band 5/7
S	32.81	17.30
M	23.37	43.50
S + M	56.18	60.80
B	24.93	18.00
Y	10.75	13.30
B + Y	35.68	31.30
N	8.14	7.90

programme, at a value for sigma of 0.4, for the 72 blocks, are summarized in Table 1.

Areas of classes from orthophotos and Landsat printouts

Delineation of the five classes S, M, B, Y and N in each of the 72 blocks on the orthophotos was done under a stereoscope with additional information from the management map and from field checks where necessary. The areas resulting from the interpretation of the orthophotos and the Landsat printouts are given for the combined 72 blocks in Table 2.

Pure spruce and non-forest classes are the only ones that show a relatively small difference. With Landsat data, most of the young stands (Y) are classified as B and M, indicating that Y cannot be separated from B and M. The combinations S + M and B + Y show remarkable agreement in their respective areas for both methods. A comparison of the area of the classes from both methods for the individual 72 blocks, however, plainly shows that great deviations occur. This is demonstrated in Table 3, where the areas of the five classes and of one combination (S + M), for only ten blocks, are given. It can be seen that the deviations have a random character.

Table 2. Areas from orthophotos and from Landsat data (PC1/PC2).

Class	Orthophoto		Landsat		Difference %
	ha	%	ha	%	
S	5483.39	42.70	4240.1	32.81	- 22.67
M	1479.74	11.53	3020.0	23.37	+ 104.09
S + M	6963.13	54.23	7260.1	56.18	+ 4.26
B	398.05	3.10	3221.8	24.93	+ 709.40
Y	4253.10	33.12	1389.7	10.75	- 67.33
B + Y	4651.15	36.22	4611.5	35.68	- 0.85
N	1226.25	9.55	1052.4	8.14	- 14.18
Total	12840.53		12924.0		+ 0.65

Table 3. Areas in ha of classes and of one class combination resulting from orthophotos and Landsat data

Block No	Orthophotos						Landsat					
	N	Y	S	M	B	S+M	N	Y	S	M	B	S+M
1	63.90	52.81	54.98	5.00	0.00	59.98	53.4	26.9	55.3	23.3	20.6	78.6
2	2.12	72.98	83.40	20.92	0.00	104.32	3.4	22.1	84.9	33.2	35.9	118.1
3	0.00	88.3	50.55	34.21	3.09	84.76	0.4	11.1	70.5	48.1	49.4	118.6
4	0.00	48.60	24.69	97.01	7.87	121.70	1.6	11.7	75.9	45.8	44.5	121.7
5	2.40	39.71	104.54	24.39	5.66	128.93	1.3	14.0	76.3	47.0	40.9	123.3
6	27.91	50.96	97.14	1.66	0.00	98.80	20.1	23.0	59.8	33.2	43.4	93.0
7	4.54	81.46	73.56	13.05	4.56	86.61	1.6	22.1	65.5	40.0	50.3	105.5
8	1.64	51.72	99.64	25.41	0.29	125.05	0.0	11.7	76.8	43.1	47.9	119.9
9	4.75	94.92	56.77	21.54	0.00	78.31	0.0	20.3	64.3	44.5	50.4	108.8
10	1.95	73.33	49.13	52.62	0.00	101.75	0.0	26.1	36.4	56.0	61.0	92.4

The differentiation between the species present and between the age classes, with the described technique for handling the Landsat data, is very inaccurate. The agreement between the totals of the S + M areas and the totals of the B + Y areas from both techniques is purely accidental.

It can be concluded that forest class areas determined from Landsat data with the described transformation method will be very unreliable. Deviations in non-forest areas also occur, which can be very great for smaller patches. Maps derived from Landsat data can therefore not be used for management purposes. The deviations for total forest, without considering species and age classes, are smaller. The total forest cover from Landsat data in the given example is 11 871.6 ha, which is 2.2 % higher than the 11 614.3 ha from the orthophotos. Individual block differences for this total area are smaller than 12 %.

Determination of timber volume of spruce

Though the transformation technique used for Landsat MSS data provides unreliable information for the construction of forest management maps, it was worthwhile to test whether the area data could be used for the determination of the volume of standing timber. The volume of spruce occurring in the class combinations S + M in the 72 blocks was considered for this purpose and the results were compared with the volume of the spruce when using the orthophoto class combination of S + M.

A two-stage sampling method was used for both, in which the selection of the primary units (PU) were based on probability proportional to size (pps) with replacement and the selection of the secondary units (SU) in the selected PUs were random. With this method, the total volume is independent of the occurring errors in the S + M areas of the PUs.

Determination of timber volume with orthophoto and field data

The areas of the class combination S + M, as occurring in the individual blocks, served as PUs in the two-stage sampling design. A cumulative list of these 72 PU areas was made for the pps selection, with replacements of the PUs in the sample. This means that a good correlation is assumed to exist between spruce timber volume on the S + M areas per PU and the respective S + M areas per PU. Random figures were used for this selection. Circular 0.05-ha plots were used as SUs; in total, 5 SUs per PU were randomly selected in the PUs in the sample. Their positions were marked on the 1:10 000 scale orthophotos and they were enumerated in the field.

With an equal subsampling rate of five SUs per PU, an equal overall sampling fraction was obtained, which made the sample self-weighting.

Calculation of the number of PUs needed to arrive at a pre-defined sampling error requires determination of the standard deviation with sufficient precision. In total, 20 PUs were involved in determining the standard deviation to fulfil this requirement. An estimate of the total spruce volume (yt) was calculated with the formula:

$$y_t = (1/m) \sum^m y_j/p_j,$$

in which:

p_j = the probability of the selection of PU_j,

m = number of selected PUs in the sample

y_j = total volume of spruce on PU_j.

For $p_j = X_j/\sum^M x_j$, the total volume

$$y_t = (1/m) \sum^m (y_j/x_j) \sum^M x_j \\ = (\sum^M x_j/m) \sum^m (y_j/x_j).$$

In this expression, M = total number of PUs in the entire 12 840 ha, x_j = total S + M area in PU_j and X_j = total S + M area in M PUs. The sample mean $\bar{y} = (\bar{y}/\bar{x})$ is an unbiased estimate of the corresponding population mean.

An estimate of the standard deviation is derived from the square root of the variance, which is expressed as:

$$V y_t = [(\sum^M x_j)^2/(m-1)] [\sum^m (y_j/x_j)^2 - (1/m) (\sum^m y_j/x_j)^2]$$

Twenty PUs were selected from the cumulative S + M area list. After random selection of the five SUs per sample PU and the enumeration of the plots in the field, the total volume (y_t) and the standard deviation (s), both expressed in m³, were calculated. The latter expressed in % of the volume is the coefficient of variation (cv). The correlation coefficient (r) between the spruce volumes and the S + M areas for the PUs were also calculated.

$$y_t = 3\ 101\ 311\ m^3$$

$$s = 555\ 527\ m^3$$

$$cv = 17.4\ \%$$

$$r = 0.81$$

Total area of S + M of all 72 PUs from the cumulative area list is 6 963.13 ha. The number of PUs in the sample, for a fixed sampling error, can then be calculated with the help of this standard deviation (or cv). For a sampling error of SE% = 4, the number of selected PUs in the sample was $m = 17.4^2/4^2 = 19$.

It can be seen that this is one PU less than used to determine the standard deviation. In practice, usually less than 20 PUs are used to determine the standard deviation but, for reasons explained above, 20 PUs were taken in this test. The number of PUs required for a fixed sampling error can thus be determined with sufficiently high precision and used as a standard for comparison.

For comparison of the methods, it is essential to know the number of man-days involved in all office and field activities. The following analysis was therefore made:

<i>Office work</i>	<i>Time in man-days</i>
– Delineation of the outer rectangular boundary of the 12 841 ha	0.14
– Delineation of the exterior forest boundary within the rectangle + the delineation of the young stands	3.00
– Delineation of 72 equally sized rectangles of 178.3 ha each	0.36
– Delineation of the S + M areas in all 72 rectangles; at 20 PUs per man-day	3.60
– Determination of the S + M areas in all 72 PUs; at 20 PUs per man-day	3.60
– Preparation of the cumulative area list of the S + M areas from 72 PUs	0.50
– Selection of $m = 19$ PUs (because of replacements of selected PUs 2 repetitions occur of 3 PUs); at 400 PUs per man-day	0.05
– Random selection of 5 SUs per selected PU. In total there are $19 \times 5 = 95$ SUs; at 400 SUs per man-day	0.24
Total	11.49

Fieldwork

Some assumptions were made for calculating the number of man-days involved in the completion of the various steps. The assumptions are:

- The field crew consists of two men
- One man-day is equivalent to 8 man-hours
- The enumeration of 1 SU (circular 0.05-ha plot) takes 0.33 man-hour
- Daily travel from base camp to first SU and return is done by car at an average speed of 50 km per hour. The average daily distance to be covered by car is 13 km. This involved $(13/50)(2/8) = 0.065$ man-day
- Travel from SU to SU per PU and from PU to PU is done on foot at an average speed of 1.8 km per hour. Systematic distributions of the PUs and SUs/PU are considered for the calculation of the average distances between PUs and between SUs per PU in the sample. The average distances in metres were calculated as follows:

$$d(\text{PU})^2 = (\text{total area in m}^2)/(\text{number of PUs in the sample})$$

and

$$d(\text{SU/PU})^2 = (\text{total area in m}^2 \text{ per PU})/(\text{number of SUs per PU}).$$

The average distance between the selected number of 19 PUs (with repetitions) is 2833 m, or $(12\,841 \times 10^4/16)^{0.5}$, the average distance between the five SUs is 597 m, or $(178.3 \times 10^4/5)^{0.5}$ and the average distance between 10 SUs per PU is 422 m, or $(178.3 \times 10^4/10)^{0.5}$. The number of PUs that can be completed in the field per man-day is $m = 0.7904$. The following equation was used for this calculation:

$$0.065 + 0.68 m(5)(597/1800)(2/8) + 0.16 m(10)(422/1800)(2/8) + m(2833/1800)(2/8) + m(5)(0.33)(2/8) = 1.$$

In the equation, the first term indicates the time involved in travelling by car from base camp to first SU and return; the second term the time for walking between five SUs per PU; the third term the time for walking between the 10 SUs per PU (repetitions); the fourth term the time for walking between the PUs, and the last term the time for the enumeration of the SUs.

The factors 0.68 in the second term and 0.16 in the third term indicate the ratio of the PUs that are selected once and twice, respectively. For the completion of the $m = 19$ PUs 24.04 man-days (19/0.7904) are involved.

<i>Final calculations</i>	<i>Time in man-days</i>
Calculation of the volume of the spruce trees in the SUs (in total $5 \times 19 = 95$) with the local volume table of Krenn (1948); at 100 SUs per man-day	0.95
Calculation of the total volume, the standard deviation and the correlation coefficient	<u>0.50</u>
Total	1.45

For the completion of all three activities 36.98 man-days (= 11.49 + 24.04 + 1.45) are required.

Remark

Additional expenses include:

- flying the area for the 1:30 000 scale aerial photographs
- construction of the orthophotos, scale 1:10 000 from these aerial photographs.

Determination of the timber volume with Landsat data and fieldwork

The S + M areas of the 72 rectangles (178.3 ha each) were determined by using the computer transformation of the Landsat digital data as described above. Again a cumulative S + M area list was composed, followed by the selection of the PUs in the sample. In fact, the selection of the PUs was based on incorrect areas because of the errors (see Table 3).

The centres of the pixels which formed the corner points of the selected 20 PUs were then transferred to the orthophotos. The areas of the S + M class combination on these PUs were measured on the orthophotos after delineation under a stereoscope. Five SUs per sample PU were selected for the S + M areas in these 20 PUs on

the orthophotos. Enumeration of the SUs in the field was followed by the calculation of volume, standard deviation and correlation coefficient.

The same formulas given above were used again. In the formula

$$y_t = (\sum^M x_j) / m \sum^m (y_j / x_j),$$

$\sum^M x_j$ is the total S + M area of $M = 72$ PUs from the Landsat data,

$x_j = S + M$ area in ha of PU_j (taken from Landsat),

$y_j =$ total S + M spruce volume of PU_j (resulting from the product of the mean volume per hectare from five SUs per PU_j and the total S + M area of PU_j (from the orthophotos)).

The orthophoto thus facilitates tracing the location in the field of the SUs in the selected PUs and determining the correct S + M areas of the selected PUs.

The following results were calculated for the total volume (y_t), the standard deviation (s) and the correlation coefficient (r).

$$y_t = 3\,112\,029 \text{ m}^3$$

$$s = 684\,695 \text{ m}^3$$

$$cv = 21.4 \%$$

$$r = 0.48$$

Total S + M area from the cumulative Landsat area list = 7 260.10 ha. For a sampling error of $SE\% = 4$, the required number of PUs in the sample was $m = 21.4^2/4^2 = 29$. In this case, there were two repetitions of six PUs because of replacements.

The following analysis was made for the number of man-days needed for the completion of the activities.

<i>Office work</i>	<i>Time in man-days</i>
– Determination of the corner pixels of the 12 841 ha study area and the 72 rectangles	0.40
– Compilation of the five classes with Landsat data and information from management map and orthophotos	4.00
– Determination of the S + M areas per PU from the Landsat data	0.00
– Preparation of the cumulative S + M area list from 72 PUs	0.50
– Selection of 29 PUs (including repetitions) for $SE\% = 4$; at 400 PUs per man-day	0.07
– Transfer of the centres of the corner pixels of the 23 selected PUs from the Landsat printout to the orthophotos	2.00
– Delineation of S + M areas on these 23 PUs (20 PUs/man-day)	1.15
– Determination of the S + M areas on the 23 PUs on the orthophotos; at 20 PUs per man-day	1.15
– Random selection of five SUs per selected PU. In total $29 \times 5 = 145$ SUs; at 400 SUs per man-day	<u>0.36</u>
Total	9.63

Fieldwork

The same assumptions made for the calculations using orthophotos were valid here. The average distance between the 23 PUs is 2364 m and the average distances between the five SUs and the ten SUs per selected PU are still 597 m and 422 m, respectively.

The number of PUs that can be completed per man-day is $m = 0.8462$, which follows from the equation:

$$0.065 + 0.58 m(5)(597/1800)(2/8) + 0.21 m(10)(422/1800)(2/8) + m(2364/1800)(2/8) + m(5)(0.33)(2/8) = 1.$$

The completion of 29 PUs thus takes $29/0.8462 = 34.37$ man-days.

Final calculations

	<i>Time in man-days</i>
Calculation of the volumes of spruce trees in $5 \times 29 = 145$ SUs using the local volume table of Krenn (1948); at 100 SUs per man-day	1.45
Calculation of total volume, standard deviation, cv and r	<u>0.50</u>
Total	1.95

For the completion of all three activities *45.58 man-days* ($9.36 + 34.27 + 1.95$) are involved.

Remark

Additional orthophotos were required to supplement the Landsat printouts for the 29 PUs in the sample. For terrain conditions with little or no elevation differences, only additional aerial photographs with a forward overlap for stereoscopic study are required.

The extra expenses are:

- flying part of the area for the aerial photographs (scale 1 : 30 000) covering the selected PUs
- the purchase of the Landsat CCT
- the use of the computer for handling the Landsat tapes.

Comparison of the number of man-days for the two methods

The calculated results for the two methods can be compared. It is assumed that the volumes, as from the 20 PUs for each method, remain the same as from the actual number of PUs for $SE\% = 4$. The standard deviation is sufficiently precise, because 20 PUs were used. The total volume of spruce on the S+M area from Landsat data is only 0.35 % larger than that of the standard method, which is consistent with expectations. The standard deviation for the Landsat method is much larger and required 29 PUs in the sample (against 19 PUs for the orthophoto method) for a sampling error of $SE = 4\%$. The higher standard deviation results from the lower correlation coefficient (r), and the lower correlation coefficient is caused by the random errors in the S+M areas of the PUs in the sample. This results in the higher

number of PUs involved in the Landsat method and, consequently, a higher number of man-days for fieldwork. In total, 23.3 % more man-days are required. Considering also the additional expenses which are not analysed here, it may be concluded that it is not advisable to use Landsat data for the determination of the total spruce volume applying the method as described in a test on areas of the given size. The use of Landsat data for determining spruce timber volume, with a two-stage sampling method, is expected to become less expensive than the use of orthophotos for areas of approximately 1 500 km² and larger for similar conditions (Stellingwerf, 1979).

Bibliography

- Cochran, W. G., 1977. Sampling techniques, 3rd ed., p. 428. John Wiley and Sons, New York.
- Krenn, K., 1948. Tarife zur Massenberechnung von Beständen. Badische Forstliche Versuchsanstalt, Freiburg im Breisgau.
- Sann Lwin, 1984. The application of two-stage sampling with PPS selection of PUs and of two-phase stratified cluster sampling, using aerial photography and Landsat imagery. ITC thesis. ITC, Enschede, Netherlands, 238 pp.
- Smit, G. Sicco, 1980. Use of Landsat imagery for forest management. *ITC Journal* 3: 563-576.
- Stellingwerf, D. A., 1979. Orthophoto maps and/or Landsat printouts for forestry; aspects of a quantification problem. *ITC Journal* 4:499-518.