2.10 Control Survey of the Aerial Photo Interpretation

Ingrid Paschedag, Markus Keller

The goal of this study was to obtain detailed information about the reproducibility of the aerial photo interpretation. The results of this study were used to assess the quality of the interpretation and to improve the aerial photo interpretation of future inventories.

2.10.1 Selection of the Control Samples

After the end of the regular aerial photo interpretation, a second interpretation of a test sample was conducted. This test sample was put together partly from randomly drawn sample plots and from partially random drawn sample plots only. All four interpreters analyzed these test samples. The four interpretations that were conducted for a second time were compared with the first interpretation as well as to each other.

In order to check as many uncertain or contradicting forest/non-forest decisions as possible, a pure random selection of the control sample was not used.

The control sample was compiled considering the following selection criteria:

For the selection of the sample plots, care was taken to ensure that as many suitable areas as possible were in the same stereo model. As a result, frequent change of the aerial photographs in the analytic stereoplotter was avoided (Chapter 2.2.) and the expenditure for the interpretation of the control sample was minimized.

Plots were considered suitable if the terrestrial forest/non-forest decision did not agree with the ones from the aerial photo interpretation. Furthermore, all stereo models (Chapter 2.2) were considered if different forest/non-forest decisions were made in the first NFI for at least three of the aerial photo plots. The test sample was drawn at random from these plots.

About 20 forest and 10 non-forest plots were taken at random from each of the five production regions. With this, the size of the test sample was increased.

Since in this second interpretation "critical" interpretations were examined in particular, the calculated results are only valid for this test sample but not for the entire population. Problems observed in the test sample indicated, nonetheless, possible misinterpretations, especially for difficult interpretations (e.g., for the stage of development).

A possible source of error (position error) for aerial photo interpretation consisted in the orientation of the images (Chapter 2.2). In order to prevent such varying interpretations caused by position error, all interpreters had to use the same stereo model.

2.10.2 Studied Attributes

The attributes studied were divided into three categories: continuous, ordinal, and nominal. A detailed description of the individual attributes can be found in Chapter 2.2 and Chapter 6.

The following attributes were studied:

Forest/non-forest decision	Nominal
Object decision from the grid measurements	Nominal
Relief	Nominal
Stage of development	Ordinal
Canopy cover density	Ordinal
Crown coverage	Continuous
Crown height	Continuous

2.10.3 3 Comparison of the Frequency Distribution between the First Interpretation and the Control Interpretation

The goal of this examination was to test if there were significant differences between the frequency distribution of the first interpretation and the control interpretation. Since the test described in the following compared the expected with the observed frequencies, all examined attributes had to be available as classified data.

To study the frequency distributions the χ^2 test of homogeneity was used (FAHRMEIR *et al.* 1997). For this, all assessed values were compiled in a contingency table (Figure 1).

Interpreter \ Attribute value	1	...	m	
1	h_{11}	...	h_{1m}	n_1
2	h_{21}	...	h_{2m}	n_2
k	h_{kl}	...	h_{km}	n_k
	h_1	...	h_m	

Figure 1. Example of a contingency table.
m: Number of categories for an attribute
k: Number of interpreters
h: Marginal values

The null hypothesis (H0-hypothesis) of this test means that the five determined frequency distributions were equal or similar. Thus, the number of times a certain quantity was detected by each of the interpreters was the same. $\chi 2$ is a measure of deviation between the true frequencies and the expected ones.

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(h_{ij} - \frac{n_i h_j}{n})^2}{\frac{n_i h_j}{n}} \tag{1}$$

For h_{ij}, h_j, n_i, n, k, and m: see Figure 1.

After calculating χ^2, the proposed null hypothesis was tested. The error probability was calculated for this. The error probability is the probability of being wrong when the null hypothesis is accepted (BORTZ 1993). In the following, 5% was considered the maximum acceptable error probability.

2.10.4 Results

Forest/Non-forest Decision

The forest/non-forest decision was the most important attribute of the aerial photo interpretation. On the one hand, this decision helped with the stratification for the statistical analysis of the NFI; on the other hand, the decision as to which sample plots were measured in the field was made from an aerial photograph. Therefore, it was important that the forest/non-forest decision was highly reproducible.

Table 1: Forest/Non-forest decision. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Forest/Non-forest decision
38	36	34	37	35	Non-forest
72	74	75	73	74	Forest
1	1	1	1	2	Shrub forest
0	0	1	0	0	Not interpretable

As seen in Figure 2, the frequency distributions of the individual interpreter visually differs only slightly. In addition to this, the χ^2 value of 0.957 calculated with (1) indicated that the interpretations were comparable. In the present case, the null hypothesis was accepted.

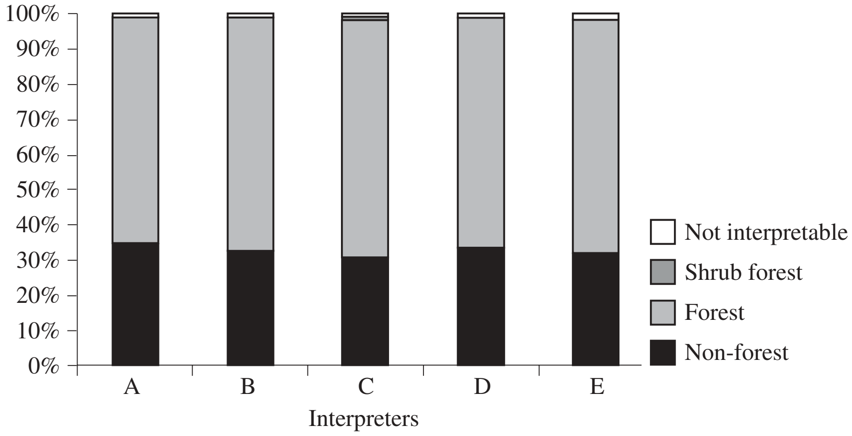

Figure 2. Forest/non-forest decision. Relative frequencies of the attribute values for the five interpreters.

Object Decision from the Grid Measurements

In the first decision, each of the aerial photo plots was classified as forest/brushwood or non-forest. In the next step, the entire interpretation area of 50 m x 50 m was covered by a 25 dot grid and an object decision was made for each of the dots (Chapter 2.2 Aerial Photography). As Table 2 clearly shows, large differences existed between individual objects. The determined frequency distributions suggested that it was not always clear whether or not the floating mark "missed" the tree and, therefore, the forest ground was interpreted. However, it was clear that it could not always be decided with certainty that the interpreted tree was a broadleaf tree or a conifer.

The null hypothesis of the χ^2 test was rejected.

Table 2: Object decision from the grid measurements. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Object decision from the grid measurements
72	70	67	112	49	Non-forest
811	748	709	921	787	Broadleaf
665	616	640	532	602	Conifer
0	4	15	0	7	Larch
145	270	240	142	228	Forest ground stockable
9	8	6	0	2	Forest ground not stockable
19	6	45	18	44	Shrub forest
4	3	3	0	6	Forest road

Crown Coverage

The crown coverage was measured with the help of the dot-grid measurements (Chapter 2.2). The values were determined by the inner nine dots. These nine dots represented approximately the larger circular sample plot of the terrestrial survey (see Chapter 2.3). The reproducibility was also investigated for the crown coverage of the entire interpretation area (25 dots). But since the results showed similar values as the study for the inner nine dots, the following results of one study only are presented.

Even though Figure 3 suggests that the comparability was poor, the calculated value for the χ^2 test was, nevertheless, clearly above the error probability of 5%. Because of this value of 0.74, the null hypothesis could not be rejected.

Since the crown coverage was derived directly from the dot-grid measurements, the question was why the hypothesis should not be rejected, since it was not significantly different than that of the determined crown coverage of the individual interpreter, while the null hypothesis of the object decision for the dot-grid measurements was rejected.

One explanation for this was most likely because the trees were divided into separate classes (broadleaf trees, conifers and larch) for the object decision, while for the calculation of the crown coverage the division did not matter, since the only thing distinguished was “tree” and “forest ground.” The unequal number of interpreted classes for the dot-grid measurements (Table 2) was also traced back to the difficulty of identifying the tree species.

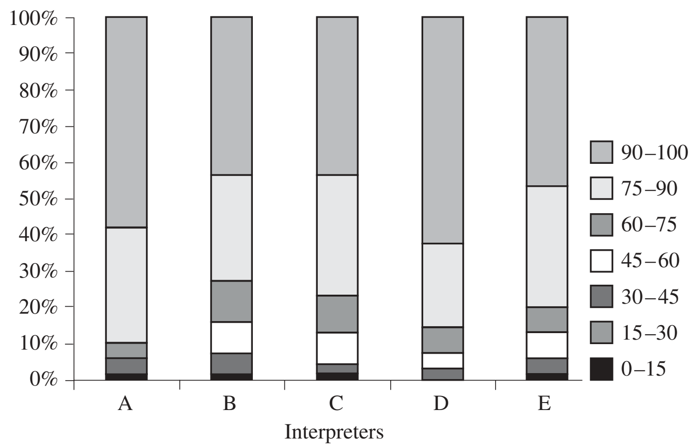

Figure 3. Crown cover. Relative frequencies of the attribute values for the five interpreters.

Table 3: Crown closure. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Crown cover
1	1	1	0	1	0–15
0	0	0	0	0	15–30
0	4	2	2	3	30–45
3	6	6	3	5	45–60
3	8	7	5	5	60–75
22	20	23	16	23	75–90
40	30	30	43	32	90–100

Stage of Development

As Figure 4 shows, the interpretations by the individual interpreters were hardly comparable. Noticeable in particular are the interpreters B and D. For B, a shifting towards a higher stage of development can be seen, while the interpretations of D were limited to two classes.

The value of 0.001 for the χ^2 test suggested rejecting the null hypothesis.

The (aerial photo) development stage was primarily defined by the dominant stand height in the NFI (see Chapter 2.2). As additional help in making the decision, the interpreter looked at other features such as the tree species, the exposition, or the elevation. The interpreter took these features into account and came to the decision about the development stage. Even though these features were taken into account, the decision about the choice of development stage was at the discretion of the interpreter. Because of this, it was difficult to reliably reproduce this decision.

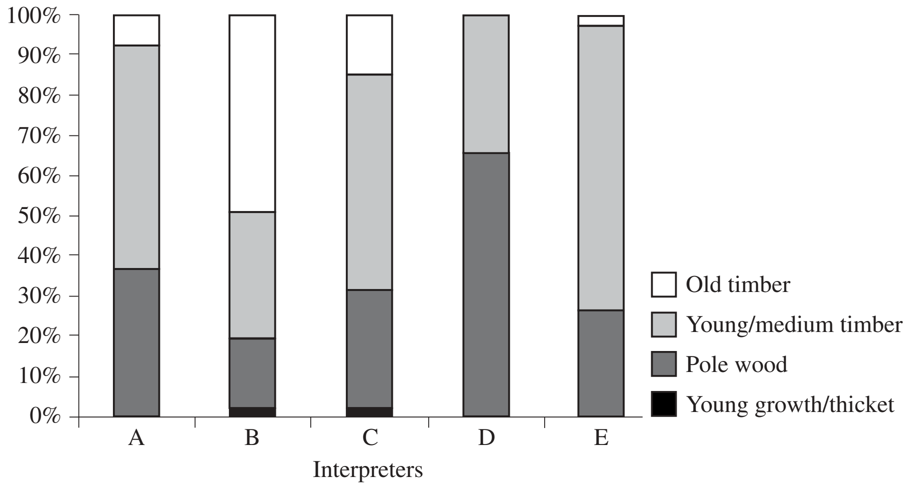

Figure 4. Development stage. Relative frequencies of the attribute values for the five interpreters.

Table 4: Development stage. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Development stage
0	1	1	0	0	Young growth
15	7	12	27	11	Pole wood
23	13	22	14	29	Young/Medium timber
3	20	6	0	1	Old timber

Canopy Cover Density

Based on Figure 5, it can be seen that the distinction between the classes “crowded” and “normal” was particularly difficult. Nonetheless, it must be noted here that for ordinal attributes it was not possible to define an exact dividing line. The value of the χ^2 test was 0.036, so that the null hypothesis was rejected with an error probability of 5%.

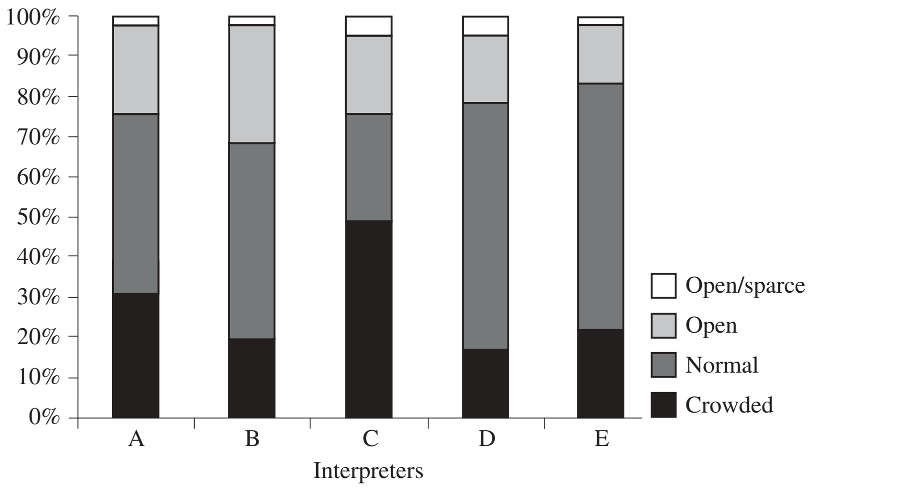

Figure 5. Closure. Relative frequencies of the attribute values for the five interpreters.

Table 5. Closure. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Crown Closure
16	8	20	7	9	Crowded
15	20	11	25	25	Normal
9	12	8	7	6	Open
1	1	2	2	1	Open/sparse

Crown Height

Since crown height is a continuous variable, all values had to be classified in the first place. The class width was 10 m. The class limits corresponded to the classification that was used during the analysis of the NFI for the stratification (Chapter 2.1).

As can be seen in Figure 6, interpreter D measured a smaller tree height more often than the other interpreters did. This was most likely the reason why interpreter D decided upon lower development stages more frequently (Table 4). The value calculated with equation (1) led to the rejection of the null hypothesis.

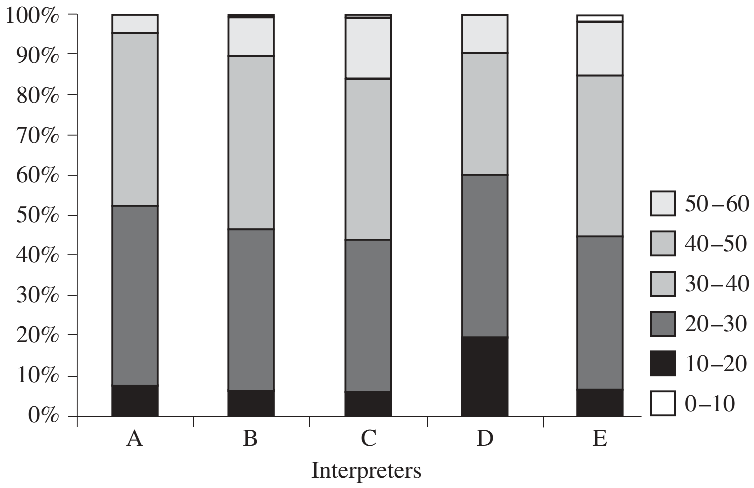

Figure 6. Crown height. Relative frequencies of the attribute values for the five interpreters.

Table 6: Crown height. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Crown height (m)
1	0	0	2	1	0-10
58	53	48	155	53	10–20
358	318	304	324	304	20–30
345	342	319	239	321	30–40
35	83	121	78	107	40–50
1	2	6	0	12	50–60

Relief

The determination of the relief was rendered from the absolute oriented aerial photograph by measuring the four corner points of the interpretation area (Chapter 2.2). Based on these measurements, the interpretation program came up with a suggested value. This suggestion was either accepted or rejected.

Figure 7 shows that the frequency of the categories “plain,” “middle slope,” and “steep slope” was approximately the same for all of the interpreters.

It was also striking that interpreter D used only three classes.
The null hypothesis was also rejected here.

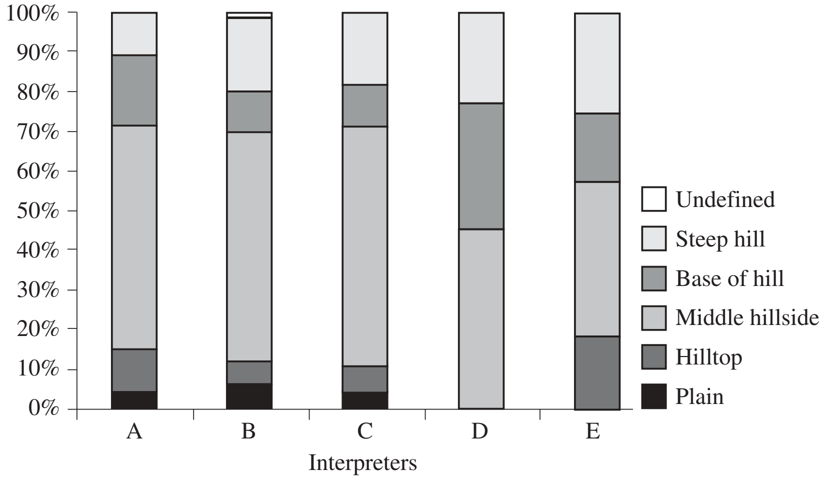

Figure 7. Relief. Relative frequencies of the attribute values for the five interpreters.

Table 7: Relief. Frequencies of the attribute values for the five interpreters.

Interpreter A	Interpreter B	Interpreter C	Interpreter D	Interpreter E	Relief
3	4	3	0	0	Plain
7	4	4	0	12	Hilltop
37	38	40	30	26	Middle hillside
12	7	7	21	11	Base of hill
7	12	12	15	17	Steep hill
0	1	0	0	0	Undefined

2.10.5 Conclusions / Outlook

In this chapter the reproducibility of the results for the aerial photo interpretation was studied, which was based on a selected aerial photo sample. The majority of the control sample plots were aerial photo plots that were difficult to interpret. It was, therefore, not surprising that in many cases the χ^2 test suggested rejecting the null hypothesis.

As expected, attributes that were measured with clear defined measurement instructions were better reproduced than ocular interpretations.

The forest/non-forest decision or the crown coverage was used as an example. Both attributes were, in essence, not interpreted but measured. A pure interpretation was completed for the attribute “development stage.” This example shows clearly that an interpretation, which was only based on describing definitions, was accordingly poorly reproduced.

For future surveys, it is advisable to define attributes, so that they do not have to be based on an expert’s opinion. Furthermore, the control interpretation should be conducted similarly to the terrestrial control (Chapter 2.9) during the survey period.

2.10.6 Literature

BORTZ, J. 1993. Statistik für Sozialwissenschaftler, Springer Lehrbuch. Berlin u.a.: Springer. 753 p.

FAHRMEIR, L.; KÜNSTLER, R.; PIGEOT, I.; TUTZ, G. 1997. Statistik, der Weg zur Datenanalyse, Springer Lehrbuch. Berlin: Springer Verlag. 594 p.