Extended Accuracy Assessment of Coherence Model

Wolfgang Wagner, Jan Vietmeier (DFD)

Heiko Balzter (NERC)

February 16, 2000

1. Introduction

This is a follow-up note to the working note "Use of the ERS Coherence for Forest Classification" from December 9, 1999. It expands the accuracy assessment through the calculation of κ which is a measure of the difference between the observed agreement and the agreement that might be solely attributed to chance. κ also allows a direct comparison with other classification results.

2. Two Forest Classes

The same methodology to determine the two forest classes (growing stock volumes smaller and greater than 70 m^3/ha) as in the previous working note is used. Table 1 shows the results for all currently available testsites. For one testsite (Irbeiskii 11-13) three different data sets produced by different people are available. Table 1 shows the percentage of correctly classified polygons in the second column from the right and κ in the last column. One can see that κ varies between values around zero (classification by chance) and values around 90 %, with an average value of 42 %.

What are the reasons for this not particularly encouraging result? One first curious observation is that for the three data sets for Irbeiskii κ varies between 6 and 50 %, i.e. the high importance of careful pre-processing (co-registration etc.) becomes obvious. Another observation is the strong relationship between κ and the dynamic range of the coherence γ given by the difference $\gamma_{0.9}$ - $\gamma_{0.1}$ (Figure 1). This shows that the classification accuracy is to a large extent determined by the spread of the coherence values.

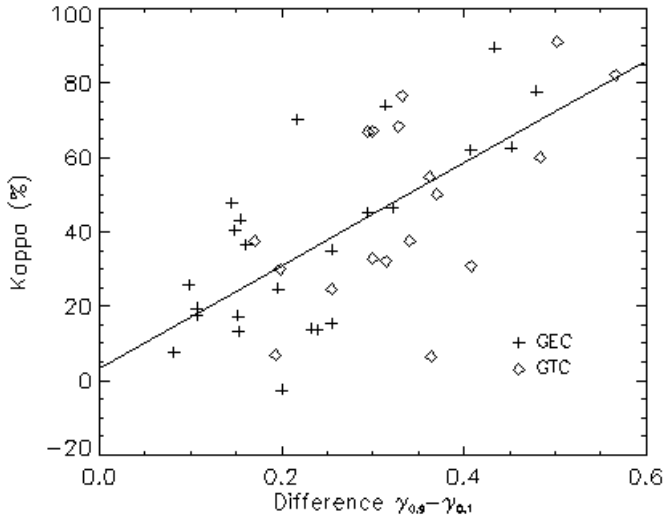

Figure 1: Accuracy coefficient κ versus the difference of the $\gamma_{0.9}$ and $\gamma_{0.1}$ percentiles of the coherence distribution for two forest classes. The solid line represent the fitted regression line.

Enterprise	No.	Year	Mask	Tr.	Fr.	Prd.	Coh1	Coh9	% Cor	Kappa
Bolshemurtinskii	1	1998	0	348	2457	GTC	0.3	0.6	84.4	66.9
Bolshemurtinskii	2	1998	0	348	2457	GTC	0.2	0.7	96.3	91.1
Bolshemurtinskii	3	1998	15	348	2457	GTC	0.2	0.4	76.5	29.9
Bolshemurtinskii	4	1998	0	348	2457	GTC	0.3	0.6	88.5	76.7
Bolshemurtinskii	1	1998	10	305	2457	GTC	0.2	0.5	84.2	68.5
Bolshemurtinskii	2	1998	10	305	2457	GTC	0.2	0.4	79.5	37.5
Bolshemurtinskii	1	1998	10	305	2457	GEC	0.2	0.5	86.8	73.7
Bolshemurtinskii	2	1998	10	305	2457	GEC	0.2	0.3	82.7	43.2
Chunsky	1	1998	50	491	2439	GEC	0.2	0.3	78.6	40.2
Chunsky	2	1998	5	491	2439	GEC	0.2	0.6	94.7	89.3
Chunsky	3	1998	0	491	2439	GEC	0.3	0.8	88.9	77.6
Ermakovsky	1	1996	100	305	2529	GEC	0.2	0.3	66.2	19.3
Ermakovsky	2	1996	50	305	2529	GEC	0.2	0.3	64.3	7.6
Ermakovsky	3	1996	100	305	2529	GEC	0.2	0.3	71.0	25.6
Ermakovsky	4	1996	20	305	2529	GEC	0.2	0.3	79.5	47.7
Ermakovsky	1	1996	20	33	2529	GEC	0.2	0.6	78.6	46.3
Ermakovsky	2	1996	100	33	2529	GEC	0.3	0.5	54.5	24.7
Ermakovsky	3	1996	50	33	2529	GEC	0.2	0.4	68.6	15.2
Hrebtovsky	1	1996	-9999	448	2421	GEC	0.2	0.4	71.1	36.4
Hrebtovsky	2	1996	-9999	448	2403	GTC	0.4	0.6	63.2	24.6
Hrebtovsky	3	1996	-9999	448	2403	GTC	0.4	0.6	63.7	6.8
Hrebtovsky	4	1996	-9999	448	2385	GTC	0.4	0.8	65.7	30.9
Irbeiskii	11	1993	50	491	2475	GTC	0.2	0.6	61.9	6.5
Irbeiskii	12	1993	-9999	491	2475	GTC	0.2	0.6	73.9	37.5
Irbeiskii	13	1993	-9999	491	2475	GTC	0.2	0.6	79.0	50.1
Irbeiskii	2	1993	-9999	491	2511	GEC	0.2	0.6	84.7	61.9
Irbeiskii	3	1993	-9999	448	2511	GEC	0.2	0.4	56.6	-2.6
Lake_Baikal_South	1	1998	-9999	462	2565	GEC	0.2	0.4	61.3	13.5
Lake_Baikal_South	2	1998	-9999	419	2565	GEC	0.2	0.6	86.9	62.5
Nishni_Udinskii	1	1997	-9999	362	2511	GTC	0.3	0.8	79.8	59.9
Nishni_Udinskii	2	1997	-9999	362	2493	GTC	0.3	0.8	91.8	82.2
Nishni_Udinskii	3	1997	-9999	405	2493	GEC	0.2	0.4	61.2	13.9
Nishni_Udinskii	4	1997	-9999	405	2511	GEC	0.2	0.4	60.3	13.2
Primorskii	1	1996	-9999	47	2475	GTC	0.4	0.7	77.7	54.8
Primorskii	2	1996	-9999	47	2475	GTC	0.3	0.6	65.4	32.8
Primorskii	3	1996	-9999	47	2475	GTC	0.3	0.6	64.9	32.2
Primorskii	4	1996	-9999	47	2475	GTC	0.3	0.6	89.2	67.0
Shestak	1	1998	-9999	47	2457	GEC	0.2	0.5	82.5	45.1
Shestak	3	1998	-9999	4	2457	GEC	0.2	0.3	68.6	17.2
Shestak	4	1998	-9999	4	2475	GEC	0.2	0.3	68.5	17.5
Ulkanskii	1	1998	-9999	147	2475	GEC	0.2	0.4	87.8	70.0
Ulkanskii	2	1998	-9999	104	2493	GEC	0.2	0.4	75.0	35.0
Average							0.2	0.5	75.3	41.7

Table 1: Accuracy analysis of coherence model for determining two forest classes. The columns are: Name of forest enterprise, testsite number, inventory year, estimated percentage of masked area (-9999 is the missing value), track, frame, product (GEC or TGC), 0.1 and 0.9 percentiles of coherence distribution, percentage of correctly classified polygons, and κ.

Other factors that might contribute to low κ values are the time gap between the forest inventory and the year of the SAR acquisitions (1998) and the topography of the testsites. To show the impact of these two factors the residuals of κ (i.e. the difference of the observed κ value and the predicted κ value using a linear regression model) are plotted in Figure 2. As expected κ tends to be lower for those testsites where the last inventory has been made some few years ago (left hand side of Figure 2). To investigate the topography the percentage of the masked test site area (Jan's masking program) was estimated for the testsites of DLR. Also in this case a relationship can be observed, but information on the percentage of masked area should also be collected for the other testsites.

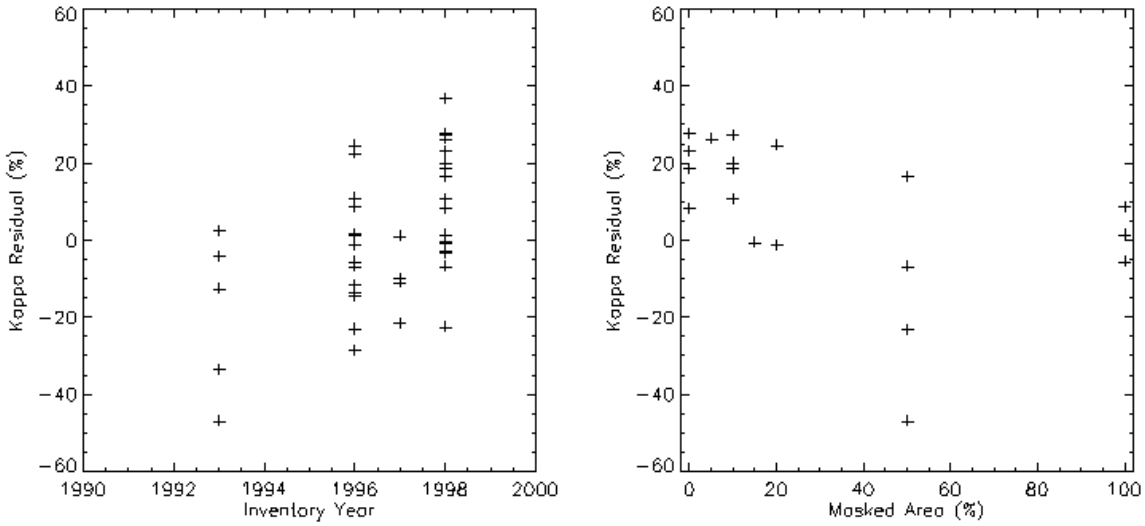

Figure 2: The residuals of κ versus inventory year and percentage of masked area.

3. Three Forest Classes

Also for the three classes case κ was calculated. As can be seen in Figure 3 κ is lower than 60 % for all sites, which means that based on the coherence image alone, one can at best determine two forest classes.

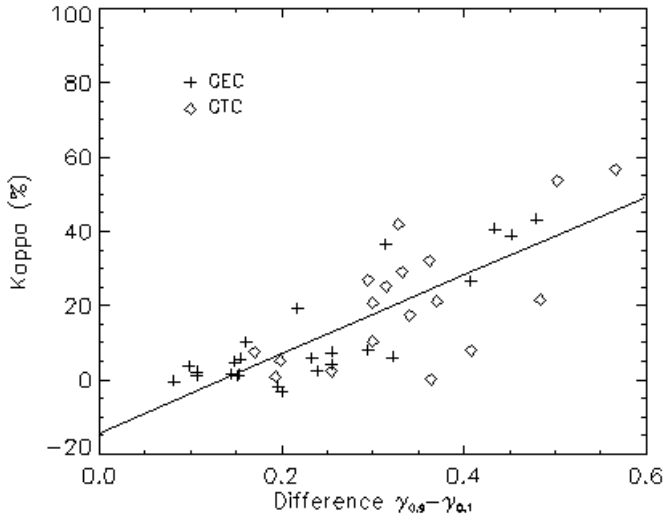

Figure 3: Accuracy coefficient κ versus the difference of the $\gamma_{0.9}$ and $\gamma_{0.1}$ percentiles of the coherence distribution for three forest classes. The solid line represent the fitted regression line.

4. Conclusions

The κ statistics suggests that using a threshold approach at best two forest classes (growing stock volumes smaller and greater than 70 m^3/ha) can be separated. Only for coherence images with a large dynamic spread of the coherence ($\gamma_{0.9}$ - $\gamma_{0.1}$ > 0.3) satisfactory results with κ values greater than about 60 % can be achieved. For coherence images with a low dynamic range much of the classification results can be attributed to chance. Also topography seems to represent a problem.