Chapter 4

Validation

While rainfall is one of the most important parameters in the atmospheric sciences, it is difficult to measure and no single rainfall estimate can be used as a calibration reference for any other one. On the one hand, the classification discussed here is easier to validate than the absolute amount of rainfall, since the only point where the absolute rainfall amount impacts the classification is the delineation of the different features within an image. As described above, this delineation is only modestly sensitive to the absolute rainfall amount and hence absolute calibration uncertainties do not affect the classification very strongly. On the other hand, the method depends on a classification performed by a human observer which makes it more difficult to validate. A third issue arises from the lack of suitable validation data that would allow the categorisation of the type of precipitation event with a higher accuracy.

The use of weather forecasts from numerical weather prediction (NWP) models might, in principle, enable identification of frontal areas, but it is unclear whether these fronts are predicted at the right location. Furthermore, a procedure to identify the location of the front in the NWP-field would have to be established. Such a procedure, though, would have a similar intrinsic uncertainty as the classification we want to validate. However, NWP analysis fields, weather maps, and geostationary satellite imagery certainly help in several cases where the location of fronts can be identified.

The second available radar-independent source of information about rainfall are surface observations (SYNOP) which might give some information about the type of precipitation. therefore a two-folded approach is used to assess the accuracy of our method by first comparing it to visually classified UKMO analysis fields. Secondly, it is compared to the SYNOP data, identifying events that are more likely to be associated with a front or convection deduced from the SYNOP.

Validation of detection algorithms that can have only two possible results need to be translated into skill scores of 2 by 2 contingency tables such as Table 4.1. The four possible combinations of the table are here

Table 4.1: The Standard Contingency table for Yes/No decisions

	Algorithm		
O		1	0
B	1	a	b
S	0	c	d

(a) correct classification of frontal precipitation (i.e. frontal in both data sets)

(b) misclassified frontal cases (frontal in “truth” data set and convective in output of algorithm)

c) misclassified convective cases (convective in “truth” data set and frontal output of algorithm)

(d) correct classification of convective precipitation (convective in both).

Woodcok (1976) discussed several common skill scores for Yes/No decision tables, such as the Heidke skill score (Heidke, 1926; Brier and Allen, 1952) or the Appleman score (Appleman, 1960). Most of them depend strongly upon the distribution of the trial conditions. However, if there is a biased distribution of mixtures of event and non-event cases, the ranking is not longer consistent. It is a fact, that there is a loss of information by compressing the four elements (a,b,c,d) of the contingency table into a single skill score. In the article of Woodcok (1976), the authors stated that almost all known discriminants (e.g. Ratio test, Skill test, etc.), except the Hanssen and Kuipers' score, are in general trial-dependent. Parts of the validation data set used in this study (here ground-based observations) biases the trial conditions in a non-randomly manner. The location of synoptical stations, for example, is very likely to influence the mixture of the trial and is not representative for the entire data set. This discrepancy can surely not be eliminated totally, but ensuring that both possible results are equally weighted in all samples leads to compatible rankings.

The hit rates for the two possible categories describe the fraction of correctly classified events, so that $hit_1 = \frac{a}{a+b}$ and $hit_2 = \frac{d}{c+d}$. The Hanssen and Kuipers discriminant V (Hanssen and Kuipers, 1965) equals the relevance of hit rates for both possible events - hit_1 and hit_2 ranging from 0 to 1 - of both possible events without respect to potential inequalities of the sample:

$$V = \frac{ad - bc}{(a+b)(c+d)} = \frac{a}{(a+b)} + \frac{d}{(c+d)} - 1 = hit_1 + hit_2 - 1 \tag{4.1}$$

where V ranges between perfect, but inverse score ($V = -1$), random score ($V = 0$) and perfect score ($V = 1$). The variance σ_V of V gives a measure for the statistical significance and was also discussed in Woodcok (1976):

$$\sigma_V = \frac{N - 4(a+b)(c+d)V}{4N(a+b)(c+d)} \tag{4.2}$$

with N being the total number of elements in the sample. Highly skewed sample distributions lead to high uncertainties in the score and have in turn a high variance value.

Table 4.2: Contingency tables for the comparison of the results with a) test data set, b) validation with UKMO analysis fields and c) validation with SYNOP-data. Second column indicates in a.) the number of contiguous precipitation areas, in b.) the number of pixels and in c.) the number of synoptic observations. The third column shows the corresponding Contingency Tables of percentage and last column the Hanssen and Kuipers' skill score (*hks*).

		Algorithm output				
		Number		Percentage		
		Frontal	Convective	Frontal	Convective	hks
a.) Test data	Frontal	86	16	19.2%	3.6%	0.82
	Convective	12	333	2.68%	74.50%	
b.) UKMO	Frontal	$3.46 \cdot 10^6$	$0.40 \cdot 10^6$	56.6%	6.6%	0.76
	Convective	$0.17 \cdot 10^6$	$2.09 \cdot 10^6$	2.7%	34.1%	
c.) SYNOP	Frontal	5244	2034	42.2%	16.4%	0.57
	Convective	661	4490	5.3%	36.1%	

4.1 Comparisons with visually classified re-analysis fields

In a first step, the results of the method are compared to classifications based on the UKMO analysis. Despite the aforementioned issues, this validation is not just a mere quality check, but it also allows to establish an error estimate associated with the visual classification of the UKMO data as a reference. For this purpose, I randomly selected 20 radar images for each month of the year 2000 (in total 240 images) and classified the type of all precipitation pixels based on UKMO analysis maps by eye. A half of the data set was examined by a second expert to avoid biased estimations. The middle part of Table 4.2 summaries the results of the comparison. In total, $6.12 \cdot 10^6$ single precipitation pixels were classified in both data sets, where 63% of the rain events were classified as frontal and 37% as convective by means of the UKMO maps. The classification method assigned 8.1% of the rain events as convective, in cases where the observer assigned them as frontal. This is a false classification rate of about 13 %. Similarly, 3.2% of the convective events are falsely classified as frontal, which corresponds to a false classification rate of about 10 %. Therefore, a gross false classification rate of 10 % can be assigned based on these results.

4.2 Comparisons with surface synoptic observations

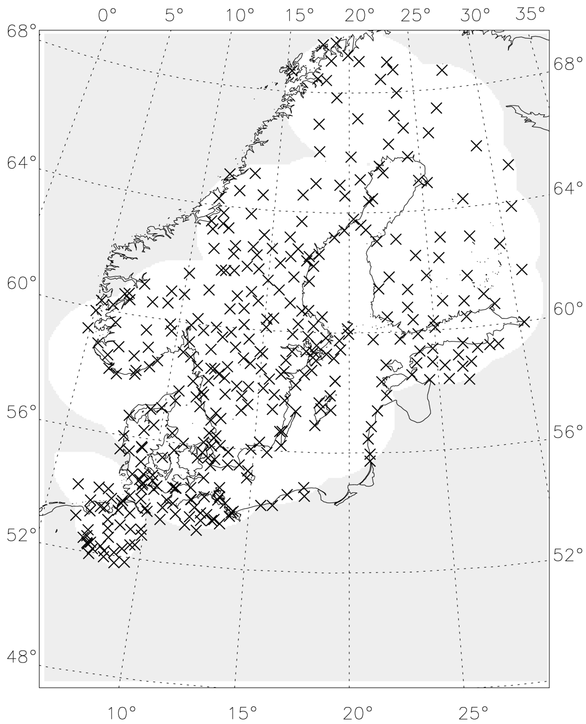

Figure 4.1: Location of synoptic stations used for comparison.

Surface synoptic observations (SYNOP) are conducted according to the standardised procedure defined by the World Meteorological Organisation at fixed UTC times of the day according to WMO key FM 12. SYNOP reports are examined and sent typically every six hours at 00, 06, 12 and 18 UTC. A SYNOP report consists of groups of values describing general weather information, such as surface temperature, barometric pressure and visibility at a weather station. The SYNOP parameters *Present Weather*, *Total Cloud Coverage* and *Cloud Type* for three atmospheric layers are chosen to evaluate the type of or evaluation of the skill of a frontal/convective discrimination algorithm. In general, all of these parameters are estimated by skilled meteorologists "by eye" and are therefore of subjective nature.

The *Total Cloud Coverage* is given in eights from zero for absolutely no cloud to eight for a totally covered sky. The *Present Weather* key includes, amongst others, information about the duration (continuous or discontinuous), intensity (weak, moderate or heavy) and type (e.g rain, snow, hail or graupel) of precipitation. The *Cloud Type* key is reported for low level, middle level and high level clouds. The key can have ten different values for each layer and includes all prevalent cloud classes, such as the different types of cumulus, altocumulus

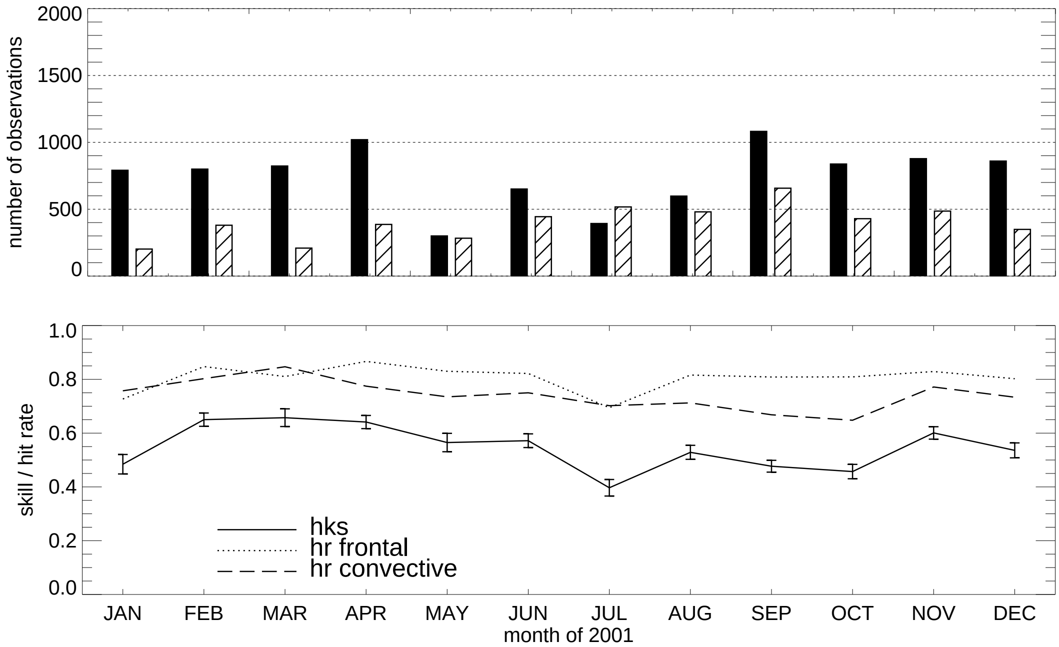

Figure 4.2: Validation results in dependency of month of 2001. (top): Number of SYNOP observations (filled: frontal events, streaked: convective events. (bottom): hit rates and Hanssen and Kuipers skill (hks) parameter V. Errorbars indicate variance σ_V of V.

and cirrus.

For the year 2001, all available data from 390 SYNOP are matched with radar estimates for the same time and location. An overview of the geographical locations of the SYNOP stations is given in Fig. 4.1.

The SYNOP data were subdivided into cases which were likely to be associated with frontal data, unclear cases, and cases which were likely to be associated with convection. The following criteria were used:

- A SYNOP observation was assumed to be frontal if the observer reported continuous rain or snow (weak, moderate, or heavy), no cumulonimbus and at least 5/8 cloud cover.

- A SYNOP observation was assumed to be convective if showers (rain, snow, graupel) were reported and cloud cover did not exceed 6/8.

- All SYNOP observations that did not fall into either category were rejected as unclassified.

Note that a classification of the SYNOP data itself is not considered. Rather it is desired to derive a subset of the SYNOP data and to have a reasonable confidence that the selected cases are convective or frontal. Obviously, the above definition rejects several observations, and the dataset was reduced to 7278 frontal and 5151 convective cases.

The results of the validation of the data set as a whole are summarised in the bottom row of Table 4.2. About 80 percent of the SYNOP observations were correctly classified, with a Hanssen and Kuipers score V of 0.54. The performance in detecting convective cases correctly was lower than that of frontal cases.

Fig. 4.2 shows seasonal variations in V. The upper panel illustrates the number of SYNOP observations for each month. The frequency of observations strongly depends upon the choice of criteria for the classification of the SYNOP data and does not represent any climatological "truth", although the smaller likelihood of frontal precipitation in summer is well recognisable in the data. V ranges between 0.4 and 0.65. A smaller number of convective observations leads to a higher V, which indicates a better performance in areas and times containing frontal precipitation.

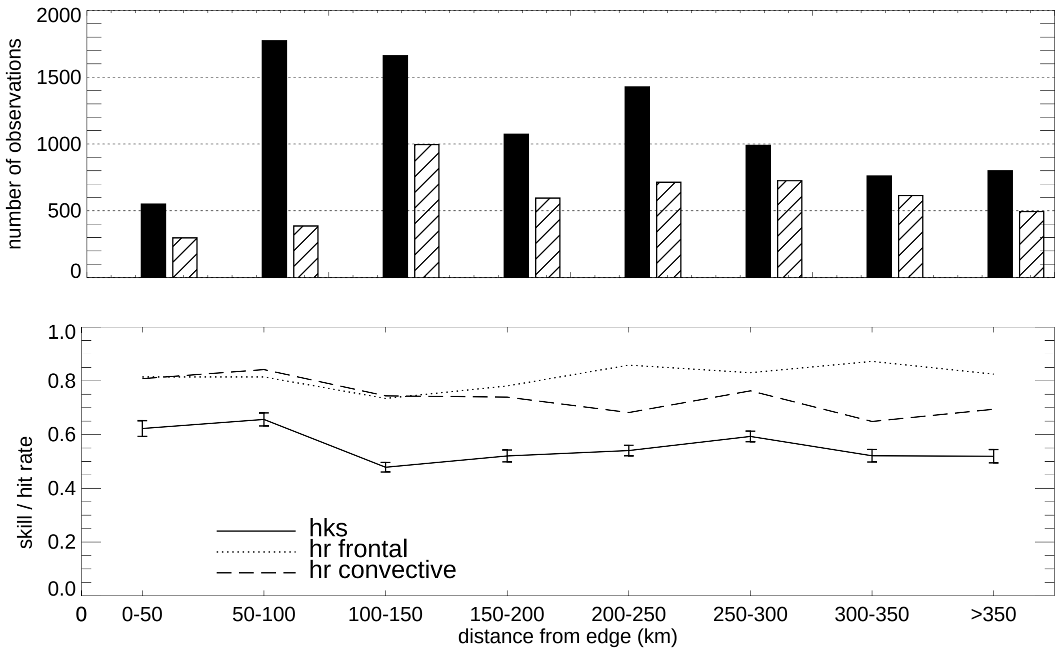

Figure 4.3: Validation in dependency of the distance to the edge of the radar range. (top): Number of SYNOP observations (filled: frontal events, streaked: convective events. (bottom): hit rates and Hanssen and Kuipers skill V (hks) parameter. Errorbars indicate variance σ_V of V.

Due to the spatial nature of the algorithm and the relatively large extent of the frontal areas, a number of incomplete frontal systems at the edge of the radar range might falsify the classification there. Fig. 4.3 shows the skill of the algorithm with respect to the distance to the edge of the radar range. It is shown that the variability of the skill score in respect to the distance to the edge of the observation domain are low.