Chapter 3

Methodology

3.1 Overview

Frontal systems have distinctive characteristics in two-dimensional horizontal precipitation maps. The main feature in the synoptic scale is a longish shape with a large area of several thousands square kilometres. The rain rates are mostly moderate and homogeneously distributed. In contrast, meso-scalic convective regions are systems with a smaller area and without a distinctive shape. They often exhibits more extreme rain rates with frequent inclusions of non-rain regions.

The large quantity of radar data (four slots per hour) clearly requires the development of automatic methods. Although the characteristics of the type are easily detectable by a human observer, it is not possible to distinguish frontal and convective rain by a simple threshold technique for any of those characterisations. The features a human may use to recognise the typical patterns associated with the two main types of precipitation have to be "translated" in mathematical parameters. These values will hereby classified in two subgroups: texture and shape parameters. The texture can be associated with the "inner" structure of the precipitation field, since it explains the spatial distribution of rain intensity within the area. Shape parameters are based on a binary representation of the rain field in a 2D precipitation map and can therefore referred to as the "outer" structure.

Generally speaking, the algorithm uses the structural patterns of radar reflectivity factor Z of precipitation systems in combination with an Artificial Neural Network (ANN) in order to distinguish frontal and convective precipitation. Pankiewicz (1997) suggested a similar technique for classifying shallow and deep convective air masses from satellite data as an estimate of convective precipitation. In that study the textural parameters within clip-outs of a fixed size were used as input parameters of a neural network algorithm. In contrast to that approach textural and geometrical parameters of entire contiguous precipitation systems are used here.

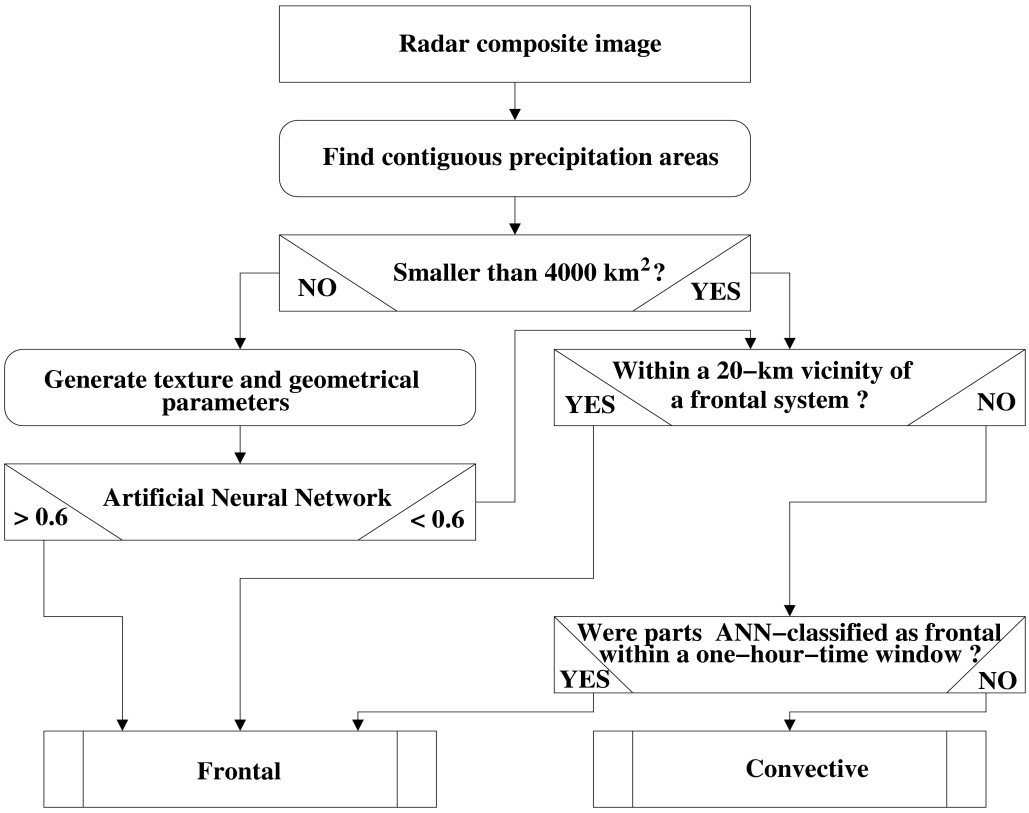


Figure 3.1: Flow diagram of the methodology do classify precipitation into frontal and convective partitions.

Firstly, the general algorithm flow are described before the single steps are discussed in detail. Fig. 3.1 shows a top level diagram of the classification approach. Thus, based on any individual radar composite image, contiguous precipitation areas are firstly identified. If these areas are smaller than a predefined size of 4000 km^2 they are initially assigned "convective". If a individual precipitation event is larger than the threshold several texture parameters are derived for each of these precipitation areas. These parameters are then used as input for an ANN that decides on whether each individual case is convective or frontal. In the next step all precipitation events within 20 kilometres distance from a frontal precipitation event are assigned as frontal. In a similar manner the temporal neighbourhood of a frontal system is assigned frontal for all images within 60 minutes around the current image. In the following, the individual steps are described in detail. The approach introduction will be accompanied by means of sample images.

Example In order to illustrate the scheme processing more descriptive it will be gone through the each single step by means of an example case at the end of each section of this

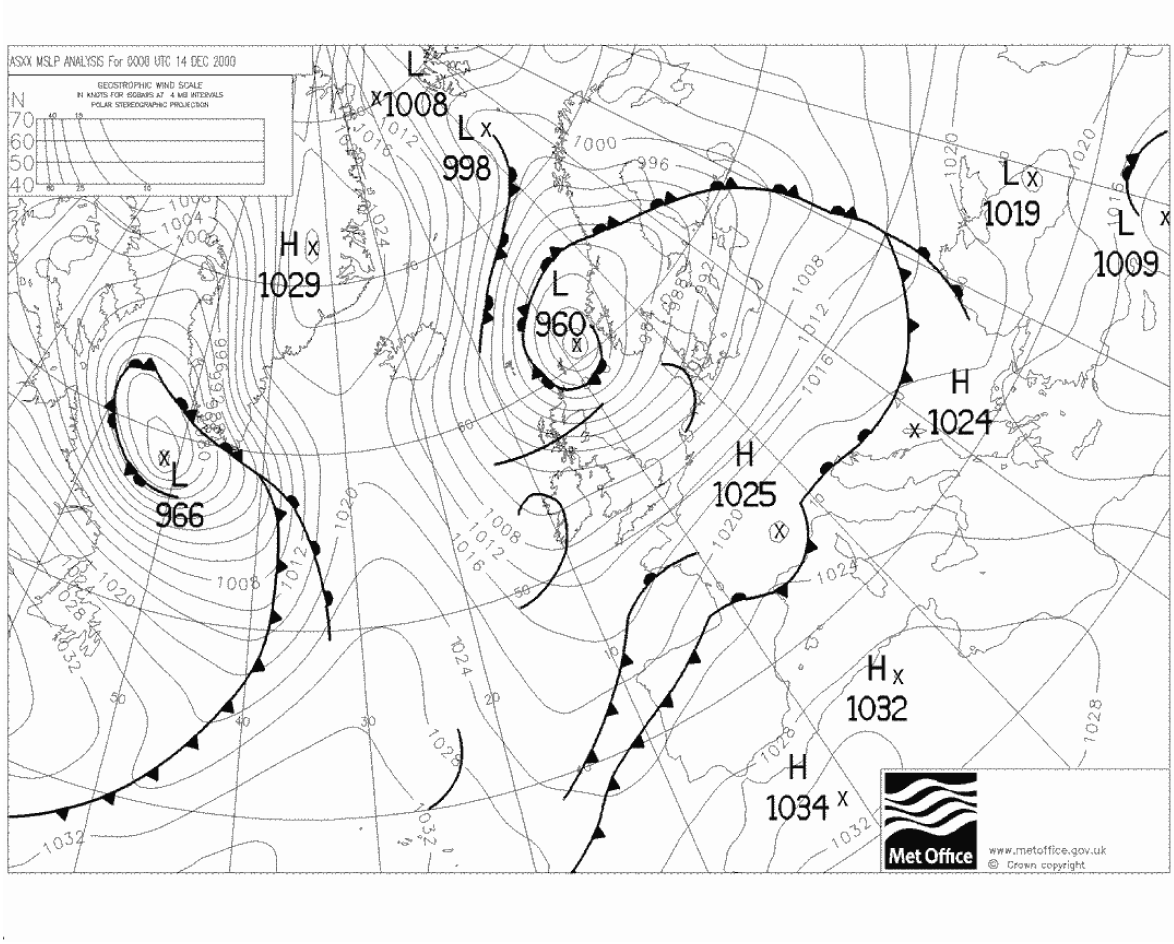
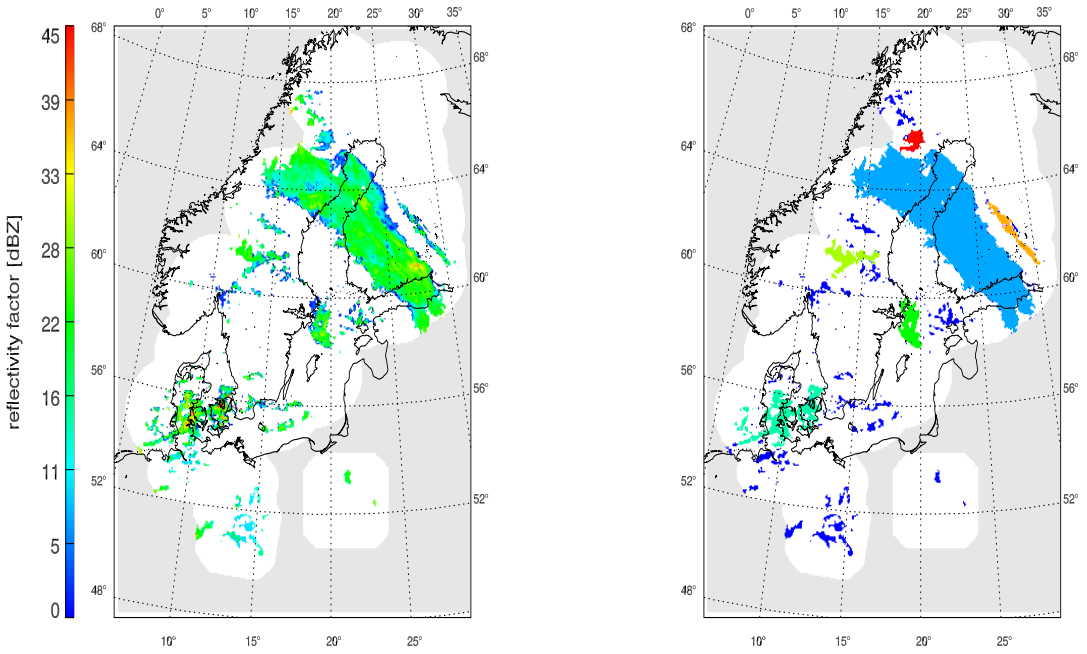


Figure 3.2: Met Office re-analyse map (14th December 2000 00:00 UTC) by courtesy of U.K. Met Office

chapter. Fig. 3.2 gives an overview over the synoptical situation in Europe at the 14th December 2000 00:00 UTC. The Baltic area is influenced by a low pressure system with its centre west of Norway. An occlusion covers the northern part of the Bay of Bothnia. Left panel of Fig. 3.3 shows, as the starting point of the algorithm, the map of corrected radar reflectivity factor Z_{cor} in decibel (dBZ) for 13th December 22:45 UTC. The radar image shows a rainband over middle Finland, northern Sweden and the Bay of Bothnia. Isolated precipitation events can be identified south of this large precipitation area. A large precipitation area is located over Denmark.

3.2 Identification of contiguous rain regions

Region identification mehods assign unique labels to image regions. In this study all pixels of a contiguous rain field get one unique number as a label. Contiguous rainfall regions are determined using an algorithm with an eight-connected neighbourhood definition, i.e. the



(a) (b)

Figure 3.3: Example case: a.) Radar reflectivity factor Z_{cor} (13th December 2000 22:45 UTC). b.) Different colours indicate different contiguous rain fields with area exceeding 40.000 square kilomtres (13th December 2000 22:45 UTC).

algorithm searches for neighbouring pixels with rain rate or radar reflectivity factor (or more generally the grey values of an image) greater than a certain threshold not only in horizontal, but also in diagonal directions (see Fig. 3.4). The examination algorithm is based on a region growing technique similar to that described in Schowengerdt (1997). Each connected rainfall area is labelled with an unique region index. Thereby, included non-precipitation pixels do not count as a part of the labelled area.

In order to delineate different areas, a threshold value for the radar refelectivity factor of Z_{thr}= 2dBZ has been set. This value corresponds to about 0.05 mm/h from May to September and about 0.07 mm/h in cold season. Application of the threshold leads to a binary image with label "zero" for non-rain areas and label "one" for rain areas.

The threshold of 2 dBZ has been found empirically by trying to obtain as stable results as possible with respect to the number of identified precipitation events per image. This is shown in Fig. 3.5, where the average number of precipitation events found in one image as a function of the threshold value is plotted for several hundred randomly chosen images. It can be seen that the number of large precipitation areas is almost constant in a range between -5 dBZ and 8 dBZ. One can conclude that the classification results are insensitive

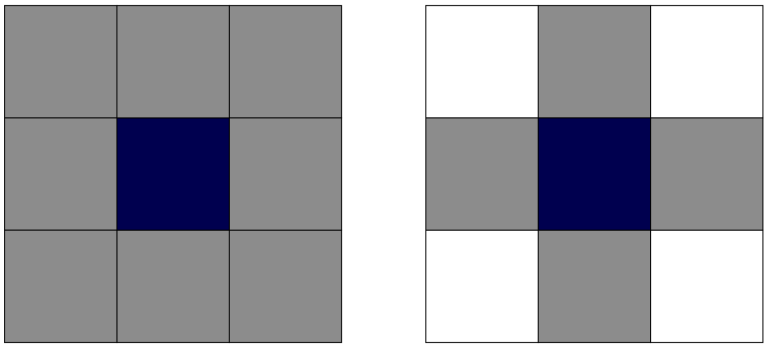


Figure 3.4: Pixel neighbourhood definitions. Left panel shows the 8-pixel neighbourhood, where all pixels are assigned "connected" (gray shaded) to the central one. For the 4-pixel neighbourhood definition only the horizontally adjacent pixels are connected to the central pixel.

to the particular choice of the threshold as long as it is not extremely high (larger than about 8 dBZ). Besides small modifications in size and shape, the choice of the threshold in the above-mentioned range leads mostly to almost identical precipitation areas.

The size of the contiguous areas is then used to preliminarily assign a class to each of the precipitation events as shown in Fig. 3.1. That means that small areas with a size in area of less than 4000 km^2 are preliminarily assigned convective, whereas large areas are further processed as described below.

Example Right panel of Fig. 3.3 shows each contiguous precipitation area labelled with a unique colour. Dark blue areas are smaller than 4000 km^2 and are initially assigned as convective. Six areas was found with a larger area size. It will be focused in the following on the large blue area over the northern Baltic sea and adjacent land regions and on the green area over Denmark. Fig. 3.6 shows cutouts of Fig. 3.3 of these two chosen areas. Note, that the aspect ratio is unchanged, but the size is different.

3.3 Parameters to describe a precipitation system

For precipitating areas larger than the aforementioned size, it is initially unclear whether they are associated with a frontal system or not. The human observer can visually classify those events, but also with a non-negligible false alarm rate as discussed below. The human observer relies on texture and shape features to identify precipitation and our classification resembles this particular capability. A large horizontal extent as well as the elongated shape and smooth variations of rain rate or radar reflectivity can be identified with frontal rain bands. In contrast, intermittent and spatially highly variable rain events are identified as convective precipitation. I prescribe eight values for each precipitation area that has been

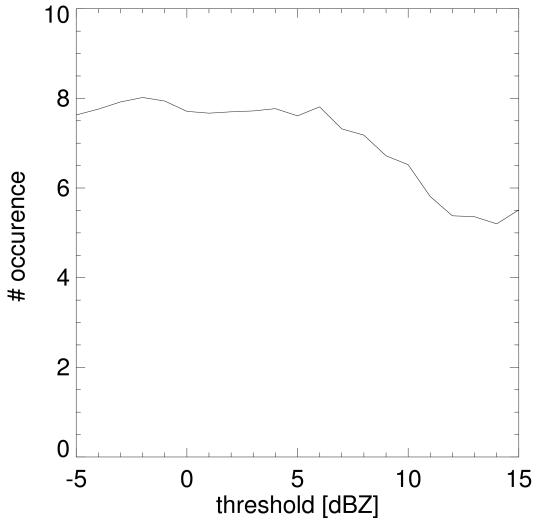


Figure 3.5: Number of contiguous precipitation areas as a function of threshold for rain pixels.

identified within the image to capture these features. They form an input vector that is subsequently used to characterise precipitation events by type.

A set of different structure measures for each precipitation area is employed to classify its type. The parameters are subdivided into two classes:

1. Shape descriptors. This group includes the total size of a feature as well as its eccentricity as characterised by the minor and major axis of the circumscribing ellipse. These characteristics resemble the human observers capability to distinguish between long and narrow features such as rainbands and more circularly shaped rain events.

2. Grey value difference (GVD) statistics. The GVD method uses histograms of the differences between the grey values of adjacent pixels to specify texture characteristics.

Example Fig. 3.6 illustrates that all connected pixels with all neighboured non-rain pixel within a 20 kilomtres vicinity will be taken in account by calculating the parameters.

3.3.1 Shape descriptors

Describibing the shape of an object in a mathematical way can prove to be very difficult. People may use terms such as rounded, elongated or jagged, however these terms are not easy to transform in quantitative values. The shape parameters aplied here are based on a binary version of the precipitation fields, i.e. after thresholding the rain rate images each

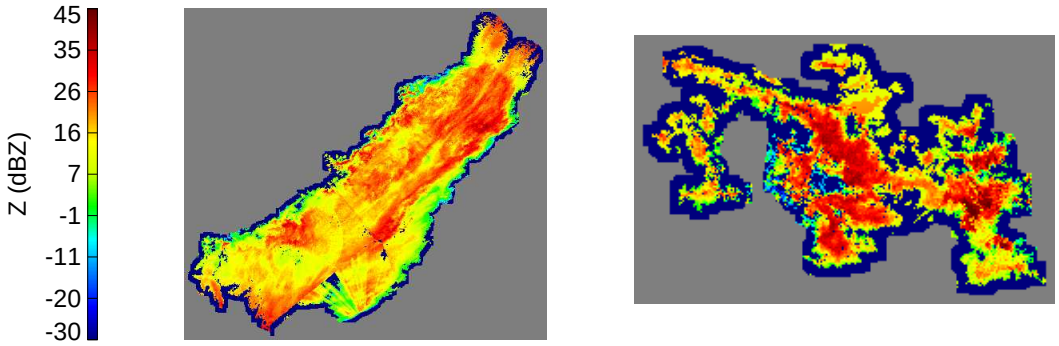


Figure 3.6: Example case: Radar reflectivity factor Z of two precipitation regions. Only pixels are shown, those are considered in textural and shape analysis.

pixel will be set to 1 ('wet') or 0 ('dry'). A shape can be described by using boundary information (i.g. chain code representation) as well as by the description of the shape itself. The latter will be used for this study. The descriptors were chosen to quantify what one can easily formulate by eye about the difference between the shape of frontal and convective precipitation systems. The most apparent features of a meteorological front are the elongated shape and an extent of at least several hundred kilometres. Thus, the chosen parameters focus on the elongatedness and the size of the area. All shape descriptors are simple scalar values for this method.

1. **Size.** The area size of a precipitation field, A, is simply the number of pixels $n_{rain}$with rain of which the contiguous region consists.

2. **Length of the major axis.** The length of the first principal component axis refers to the largest dimension of the shape.

3. **Eccentricity.** The eccentricity is here defined as the ratio of the longest chord perpendicular to the first principal axis to the first principal axis itself. The eccentricity factor may take values from almost zero (high eccentricity, long narrow features) to one (both major axes have the same length, e.g. circular shape). High eccentricity is typically associated with frontal rain areas.

4. **Compactness.** The region of interest used for this parameter includes each wet pixel identified as a part of the connected rain field and each pixel which is less than 20 kilometres to the perimeter of this particular rain field. For this study compactness is defined as the ratio of the number of rain pixels to the number of all pixels within the area circumscribing the precipitation field. Note, that there are differing definitions for compactness in the literature.

3.3.2 Texture information

In this study, grey value differences between adjacent pixels are evaluated for each precipitation field. Grey value differences (GVD) are a common method to describe the texture of spatial features (Uddstrom and Gray, 1995). However, the actual definition of GVD parameters in the literature varies. Therefore I briefly describe the parameters used.

The corrected radar reflectivity data were transformed linearly to grey values running from 0 to 255, representing a range of reflectivities from -30 dBZ to 70 dBZ.

Co-occurrence matrix A good overview over the methods introduced there is given in Haralick et al. (1973) or in Baraldi and Parmiggiani (1995). The first step is to produce a *gray value matrix* that includes the frequency of occurrence of gray level pairs. The corresponding pairs of pixel have to be located in a pre-defined fixed geometrical constellation, such as the two adjacent neighbour in the horizontal direction in an image. In case of $n = 256$ possible gray values, the grey value matrix has a dimension of 256 x 256. A gray value combination value $(gv1, gv2)$ increases the counter of the position $(gv1, gv2)$ in the gray value matrix.

For the purpose of this study a GVD function for a grey value image $\phi = f(x,y)$ has the form $P(m)$, where the m-th entry is the frequency of occurrence of the grey level difference $m = (|\phi(x',y') - \phi(x'',y'')| + |\phi(x',y') - \phi(x''',y''')|)/2$ for each pixel $\phi(x',y')$ and its two adjacent pixels in the eastern direction $\phi(x'',y'')$ and southern direction $\phi(x''',y''')$. At least one of these three pixels must be situated within the connected rain area. The normalized probability density function $\widetilde{P}(m)$ is defined as

$$\widetilde{P}(m) = \frac{P(m)}{N} \tag{3.1}$$

where N is the total number of pixels for this region.

For each large contiguous precipitation field the following parameters have been calculated:

1. **Mean grey value difference**

$$\mu = \sum_m m\widetilde{P}(m) \tag{3.2}$$

   which, if small, indicates that the GVD are concentrated near the origin. If the rain event is completely homogeneous the mean gray value difference would be zero. A large value corresponds to inhomogeneous precipitation fields.

2. **Homogeneity**

$$H = \sum_m \frac{\widetilde{P}(m)}{1+m^2} \tag{3.3}$$

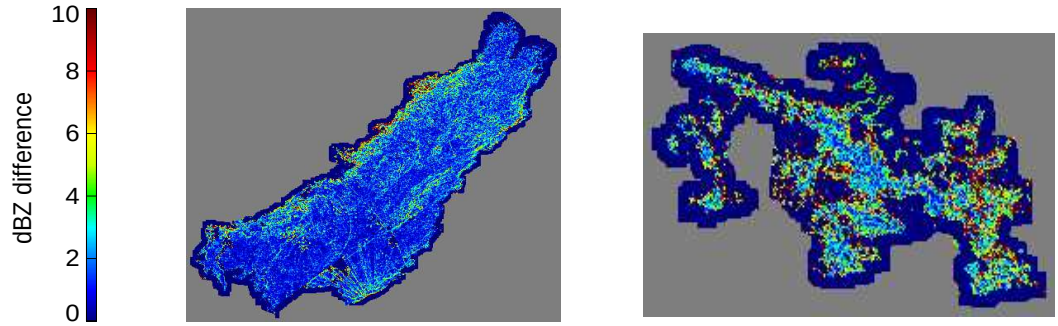


Figure 3.7: Example case: Radar reflectivity difference map: mean difference to the two direct adjacent pixels in southern and eastern direction.

will be small for large GVD due to the dominance of the squared m in the denominator. Highest results will be expected if $\widetilde{P}(m)$ is high for low values of m. Thus, if the rain field is homogeneous H, will be maximal.

3. **Entropy**

$$E = -\sum_{m} \widetilde{P}(m) log\left(\widetilde{P}(m)\right) \tag{3.4}$$

indicates whether the texture is organised. This is largest when a big number of grey value differences m occurs and the $\widetilde{P}(m)$ are uniformly distributed, but is small when they are highly variable.

4. **Contrast**

$$C = \sum_{m} m^2 \widetilde{P}(m) \tag{3.5}$$

describes the variability of the grey values in the feature. A small value of C means that small differences between adjacent pixels prevail in the image and the texture is only variable over larger distances.

These features measure various properties of the GVD probability density function, however, they are not independent from each other.

Example In Fig. 3.7 the maps of the average difference to the two adjacent pixels in southern and eastern direction is shown. The normalised histogram of the values in this difference map represents $\widetilde{P}(m)$, as shown in Fig. 3.8. The frontal region (i.e. the left image in Fig. 3.7) has only few values with large dBZ differences and most of the values concentrate near the origin. The convective case is much more variable with difference values also higher than

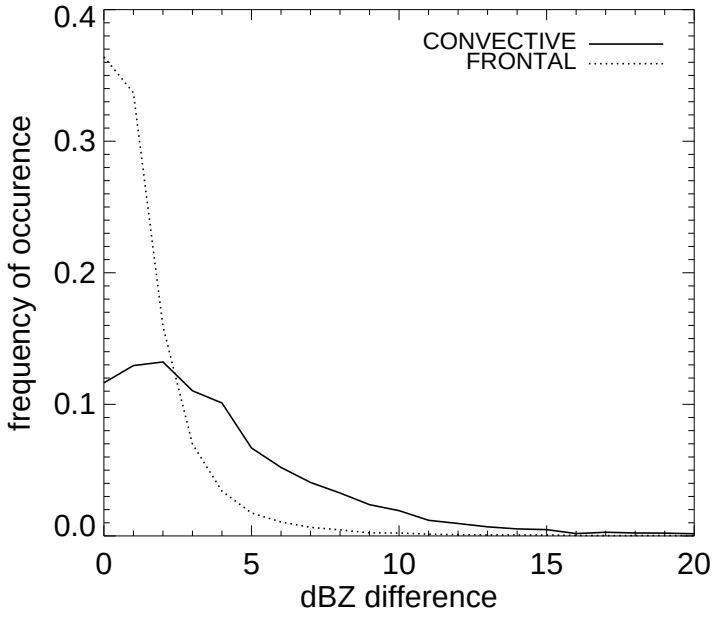


Figure 3.8: Example case: Gray value difference density function GVD $\tilde{P}(m)$ for two precipitation regions labelled as frontal and convective.

10 dBZ. It seems, that the pixel-to-pixel variability is a distinct difference feature and can be described by modi of the gray value difference function.

3.4 Development of the classification tool

A supervised multilayer perceptron ANN is used as a classification tool. Multilayer perceptron ANN are able to "learn" the functional relation between a set of i input vectors $\mathbf{x}_i$ and output vectors $\mathbf{y}_i$. A training dataset $[\mathbf{x}_i, \mathbf{y}_i]$ has to be provided to the ANN to learn the relation between $\mathbf{x}_i$ and $\mathbf{y}_i$ before it can be used to independently classify new sets of $\mathbf{x}_i$. A first test on the performance of the ANN is usually obtained by creating a second independent set of $[\mathbf{x}_i, \mathbf{y}_i]$ called a test dataset. It allows one to quantify how well the ANN is able to generalise the learned relation.

3.4.1 Selection of the training dataset

The first step in training a neural net is to find a dataset $[\mathbf{x}_i, \mathbf{y}_i]$ that can be used for training. This dataset has to be sufficiently large to span the full range of expected variability of the input vector $\mathbf{x}_i$. For this particular case, an approach is used to randomly select weather radar images from 2001 to 2002 and within each scene several precipitation events are picked again randomly that were then classified visually. The re-analyse maps of U.K. Metoffice (example in Fig. 3.2) gave additional indications of the actual synoptical situation. Each

of these events was then stored as a vector $[\mathbf{x}_i, \mathbf{y}_i]$ where $\mathbf{x}_i$ are the aforementioned eight structural parameters and $\mathbf{y}_i$ is either zero if the event was visually classified as convective or one if it was classified as frontal. Hereafter the value $\mathbf{y}_i$ is also referred to as the front index of a precipitation region. After several thousand of these visual classifications the dataset was homogenised so that frontal and convective cases occur with roughly the same likelihood. Finally, it was split into two datasets, one for training and one for testing.

3.4.2 Statistical properties of the training dataset

Before it is proceeded with the neural network training, it is worthwhile to study the statistical properties of the training dataset. This will allow to draw some first conclusion on how well the various texture and geometrical parameters are correlated with the visual classification of the training dataset. It is also possible to find high cross-correlations among some of the input features to potentially reduce the dimensionality of the input dataset to prevent the neural network from being trained with redundant input variables.

Table 3.1 shows the cross-correlation analysis for the set of training data $[\mathbf{x}_i, \mathbf{y}_i]$. The last row gives the correlation between the various input parameters and the visual classification. A positive classification here indicates that if the respective parameter has a high value, the event is more likely to be frontal.

Table 3.1: Cross-correlation matrix of input and output parameters of the neural network.

Correlation coefficient									
	1 mean	2 homogeneity	3 entropy	4 contrast	5 size	6 major axis	7 eccentricity	8 compactness	9 front index
1 mean	1								
2 homogeneity	-0.68	1							
3 entropy	0.86	-0.76	1						
4 contrast	0.81	-0.50	0.47	1					
5 size	-0.36	0.24	-0.23	-0.39	1				
6 major axis	-0.37	0.20	-0.20	-0.43	0.89	1			
7 eccentricity	0.07	0.08	0.00	0.09	-0.06	-0.25	1		
8 compactness	-0.63	0.41	-0.44	-0.67	0.72	0.76	-0.02	1	
9 front index	-0.39	0.25	0.78	-0.45	0.69	0.80	-0.16	0.78	1

The mean grey value difference is weakly negatively correlated with the visual classification which indicates that if the mean differences between the adjacent pixels in the feature

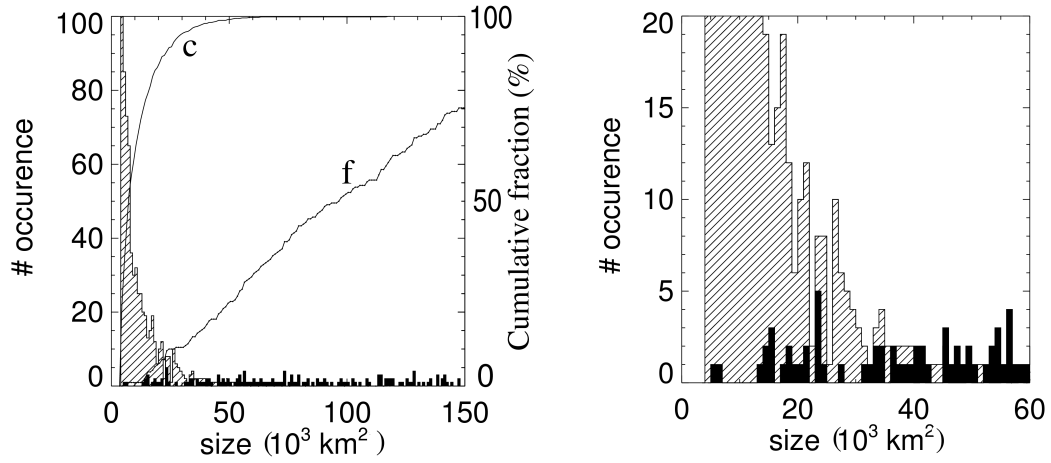


Figure 3.9: Frequency of occurrence of precipitation areas of specific type as a function of the size. The left image shows the result of the entire range of size (Filled: frontal, streaked: convective). Solid lines indicate the cumulative probability histogram for frontal (f) and convective (c) events. The right image displays the clipping of the size range smaller than 60,000 km^2.

are getting smaller, it is more likely to be frontal. The highest positive correlation is found for the size of the major axis and the relative amount of raining pixels in the box around the feature as defined by the minor and major axis. Both show a correlation of about 0.8, which indicates that larger precipitation events are more likely to be associated with fronts and that events which are more regularly filled with precipitation are more likely to be frontal.

Looking at the cross-correlations between the different input parameters one finds maximum absolute correlations slightly smaller than 0.9 between entropy and mean difference as well as between the overall size of the area and its major axis. These parameters are obviously correlated since an increase in the mean difference corresponds to an increase in the average disorder, and hence entropy, of the system. Interestingly, even though both features are correlated to some extent their correlation with the classification is very different.

I used a multi-layer perceptron neural network (Marquard, 1963; Rumelhart and McClelland, 1986) with an input layer of eight neurons for the eight texture and shape parameters, a hidden layer of 25 neurons and one output neuron that varies between zero (convective) and one (frontal). The transfer function at each node is sigmoidal.

Fig. 3.9 shows the number of classified precipitation events as a function of their horizontal extent in km^2. Small precipitation events (less than 15,000 km^2) are almost exclusively classified as convective. Between 15,000 km^2 and 50,000 km^2 the classification is not unique

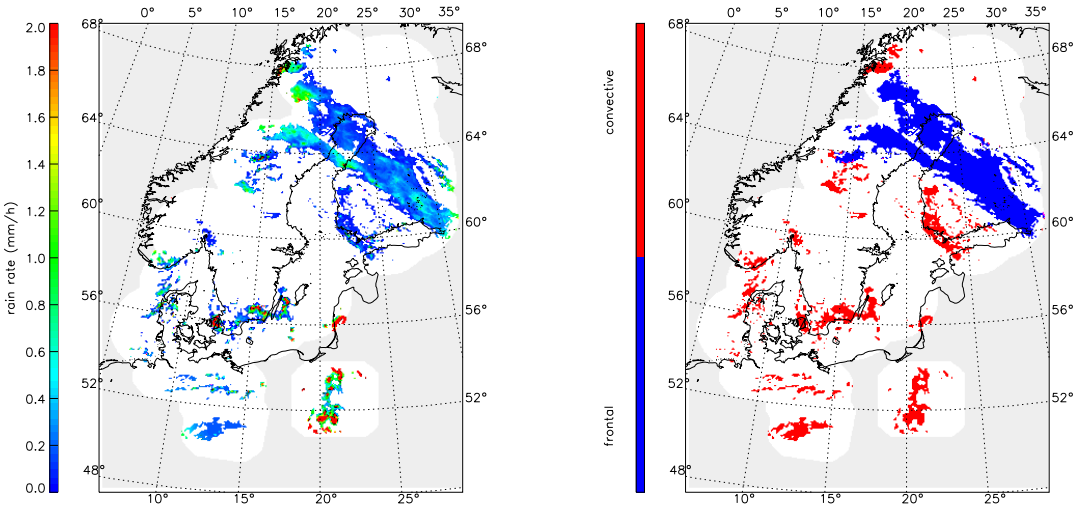


Figure 3.10: Case study of 04:00 UTC 14 December 2000. The left image displays rain rate. The right image shows the frontal/non-frontal classification.

and the precipitation event might be either convective or frontal depending on its shape and texture. Very large features (larger than 60,000 km^2) are entirely classified as frontal. For sizes around 30,000 km^2 about 50 percent of the precipitation events are classified frontal and about 50 percent convective. Note that even though most of the individual precipitation events are small, the contribution of a precipitation event to the total rainfall is to first order proportional to its horizontal extent, so that the medium-sized precipitation events contribute significantly to the total rain rate.

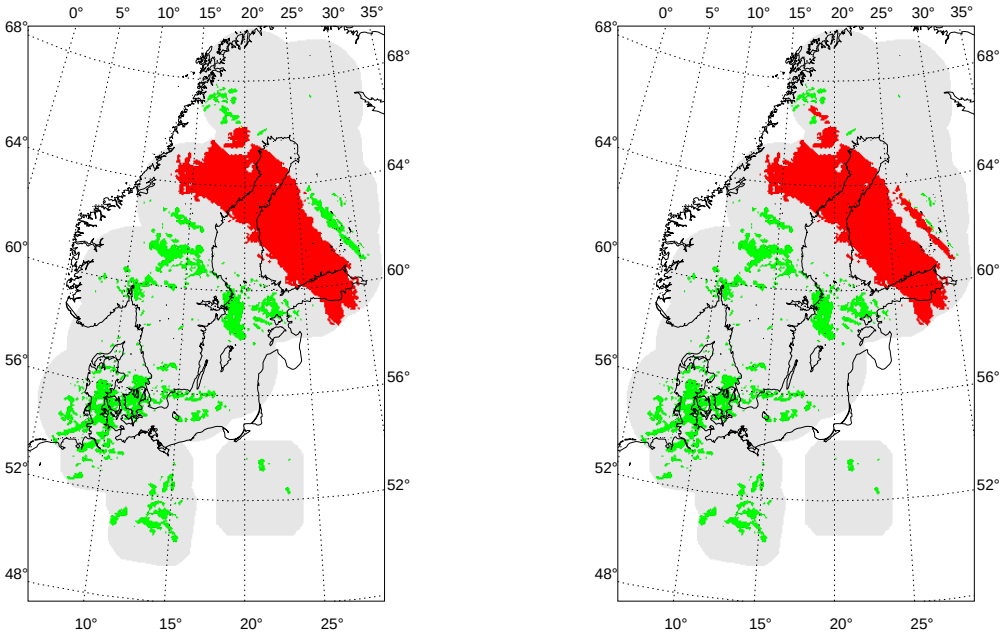
3.4.3 Temporal and spatial adjustments

After the neural network classification has been completed, two more steps are performed before the final classification is produced. In the first step, isolated precipitation events close to a frontal precipitation event are classified as a part of the latter frontal area. The search radius used for this adjustment is 20 km and hence very small given the typical extent of a frontal system. In the case study displayed in Fig. 3.4.3, some small isolated features can be identified at the north eastern part of the frontal system (64° N/ 30° E) that just by their size would have been identified as convection. It is clear, however, that they belong to the front. Similarly, at the south western part (64° N/ 15° E) a somewhat larger patch has been identified as frontal. Further south, two more patches can be identified which are not within the search radius and therefore remain convective. It is unclear whether these precipitation events belong to the front and this question could also not be answered by using auxiliary

data such as the weather map or Meteosat infrared image.

Obviously, fronts that enter the observation area might initially be misclassified as convective just because only a small fraction of the entire frontal area can be seen by the radars. It is therefore a second temporal criterion added to minimise this particular error source. For every event that has been classified as convective, it is searched earlier and later images if a frontal event has been detected in the same geographical region. If so, the event is classified as frontal, too. The search window for the temporal adjustment is one hour.

Subsequently the example case highlights some of the uncertainties associated with the classification method. From looking at this example, it is clear that a 100 percent accuracy can not be achieved and the definition of frontal and convective is fuzzy to some extent. A quantitative error analysis and evaluation of the classification is then provided in Chapter 4.



(a) Before adjustments

(b) After adjustments

Figure 3.11: Example case: Results of the classification before and after the adjustment scheme. Red areas indicate frontal zones and green areas convective zones.

Example case The classification results shown in Fig. 3.11 correctly identifies the large scale precipitation event as frontal and the small scale events as convective. Some of the small scale features close to the front are also assigned frontal due to the aforementioned spatial and temporal tests. However, several features close to the front appear to be not necessarily correctly identified. The small rainband over southern Finland is classified as

convective, but an observer would intuitively associate it with the frontal rain band and classify it as frontal. In the very northern part of the frontal area, several smaller convective events are assigned frontal. However, the northernmost precipitation events are identified as convective. Even if the observer used weather maps as auxiliary information, it would not be possible to doubtlessly assign these areas as belonging to the frontal precipitation event.