University of Southampton

Faculty of Engineering, Science and Mathematics
School of Engineering Sciences

The Use of the Texture and Motion of Clouds from Geostationary Satellite Images in Rain Rate Estimation and Prediction

by

Aimamorn Suvichakorn

Thesis for the Degree of Doctor of Philosophy

April 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ENGINEERING SCIENCES

Doctor of Philosophy

THE USE OF THE TEXTURE AND MOTION OF CLOUDS FROM GEOSTATIONARY SATELLITE IMAGES IN RAIN RATE ESTIMATION AND PREDICTION

by Aimamorn Suvichakorn

This thesis addresses the problem of estimating rainfall rates from satellite imagery. The potential for using cloud motion and texture to estimate rain rates has been examined. The main types of textural information, i.e. statistical, structural, frequency and spatio-temporal, have been used to derive features from the satellite measurements and then used to determine a relationship to the radar-observed rain rates. These features were ranked by two scoring functions that were devised to assess their relationship to rain rates. The first scoring function selected a feature set for classifying samples into three rain rate classes. The selected features successfully classify rain rates of a mid-latitude cyclone seen on Meteosat7 with 84.8-99.3 % accuracy with a significant Hanssen-Kuipers discriminant score when a probabilistic neural network was used. A similar accuracy was found when a support vector machine (SVM) was used. Another scoring function was used for the selection of the features for estimating rain rates of each class. A Gaussian-kernel SVM that has been trained by these features produced visually agreeable rain estimates, which were much better than those produced by other methods that used only spectral information. Using the same types features at different time also achieved the similar accuracy.

The method was robust and continuous rain estimates were obtained. Unlike other techniques in which additional information has always been required, the results showed that textural information alone can be used for rain estimation. This is preferable when only satellite measurements are available. Frequent updating of the observed rain rates can be done to improve the accuracy of the estimation.

The potential for using cloud motion to predict rain rates was also examined. It was found that a combination of the maximum cross correlation and optical flow techniques provided the best estimate of the velocity of clouds. A cloud's displacement derived by the maximum cross-correlation technique was used for the approximation of the future location of its corresponding rain and the final velocity derived by the optical flow technique predicts how the rain rates would change. The rain rates predicted by this novel method provided good correlation to the observed rain rates at an hour later.

Dum vita est spes est— while there is life, there is hope.

Anonymous

Contents

List of Figures

List of Tables

List of Abbreviations

AMSR	Advanced Microwave Scanning Radiometer
AMV	Atmospheric motion vector
ANN	Artificial neural network
ASM	Angular second moment
AVHRR	Advanced Very High Resolution Radiometer
BADC	British Atmospheric Data Centre
BMVA	British Machine Vision Association
CCD	Cold cloud duration
coif	Coiflet
CMORPH	CPC MORPHing technique
CPC	Climate Prediction Centre
CSI	Critical Success Index
db	Daubechies wavelet
DC	Digital counts
DFT	Discrete Fourier Transform
FFBP	Feed-forward backpropagation
FFT	Fast Fourier Transform
GLM	Generalised Linear Model
GOES	Geostationary Operational Environmental Satellite
GPI	GOES Precipitation Index
GRNN	Generalised Regression Neural Network
GS	Gilbert skill score
HK	Hanssen-Kuiper discriminant
HSB	Hue-Saturation-Brightness
HSS	Heidke skill score
HSV	Hue-Saturation-Value
IGARSS	International Geoscience and Remote Sensing
IPWG	International Precipitation Working Group
IR	Infrared
ISDR	International Strategy for Disaster Reduction
JMA	Japan Meteorological Agency
MCC	Maximum cross correlation
MLP	Multi-layer perceptron
MSG	Meteosat Second Generation
MSS	Multispectral Scanner
MW	Microwave
MVIRI	Meteosat Visible and Infrared Imager

NCEP	National Centers for Environmental Prediction
NESDIS	National Environmental Satellite, Data, and Information Service
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
PCA	Principal Components Analysis
PGS	Primary Ground Station
PrNN	Probabilistic neural network
RBF	Radial basis function
RGB	Red-Green-Blue
RMS	Root-mean-square
RR	Rain rates
SAFNWC	Satellite Application Facility in support to Nowcasting and Very Short-Range Forecasting
SFS	Shape from Shading
SIST	Solar-Infrared Infrared Split-Window Technique
SSCC	SEVIRI Solar Channel Calibration
SSM/I	Special Sensor Microwave Imager
SST	Sea surface temperature
STD	Standard deviation
STFT	Short-time Fourier Transform
SVM	Support Vector Machine
TAMSAT	Tropical Applications of Meteorology using SATellite and other data
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measuing Mission
USAID	U.S. Agency for International Development
UTC	Coordinated Universal Time
VIS	Visible
VISST	Visible Infrared Solar-Infrared Split-Window Technique
WV	Water vapour

Declaration of Authorship

I, Aimamorn Suvichakorn, declare that the thesis entitled

The Use of the Texture and Motion of Clouds from Geostationary Satellite Images in Rain Rate Estimation and Prediction

and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

  1. A. Suvichakorn and A. Tatnall. Flow analysis of cloud images from geostationary satellites. Spatiotemporal Image Processing, One Day BMVA symposium at the Royal Statistical Society, London, 2004.

  2. A. Suvichakorn and A. Tatnall. The application of cloud texture and motion derived from geostationary satellite images in rain estimation—a study on mid-latitude depressions. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, volume 3, pages 1682–1685, Seoul, 2005.

Signed:

Date:

Acknowledgments

This thesis is the result of four years of work whereby I have been accompanied and supported by many people. First and foremost, I would like to thank my supervisor Dr. Adrian Tatnall. He was very generous with his time and knowledge and assisted me in each step to complete the thesis. Without his dedicated support, advice and a lot of his patience, this work seemed impossible to finish.

I am grateful to the British Meteorological Office for Radarnet data, Eumetsat for Meteosat7 images and the British Atmospheric Data Centre for distributing these data.

I would like to thank Dr. Hugh Lewis for the materials contributing to this work. I would also like to thank other members of the Astro research group : François, Nick, John, Riaz, Dr. Sabrina, Dr. Steffen, Dr. Balaji, Dr. Christoph, Dr. Scott, Dr. Frank, Dr. Gareth, Dr. Ismat, Dr. George, Dr. Paulo, Dave, Nick Bevan, Rodrigo, Sam and Tarek for welcoming me here. It was a great fun working (and playing badminton) with them.

I would like to express a deep gratitude to my family for their untiring supports and prayers and a deep apology to my grandfather and my grandmother that I could not present this thesis to them *in time*.

My PhD was a big decision in my life–being far away from home, leaving friends, going to a place that I have known no one and most importantly people there do not speak my language. Along this journey, I have made a lot of new friends. Some of them make me realise that such good people still exist. I would like to thank all of them. They make me feel that this journey is not so difficult (but still challenging anyway) and that I am not very far from home. I would like to put all their names in here but this would make the acknowledgement thicker than the rest of the thesis. Nevertheless, all their names are written in my heart.

Thank you all.

Chapter 1

Introduction

The estimation of precipitation by the use of geostationary satellite images has been studied in this thesis. Precipitation is a key parameter in the hydrological cycle. It plays a major role in agriculture and can be the main cause of flooding events. According to the International Strategy for Disaster Reduction (ISDR), during the period of 1991-2005 floods have the greatest damage potential of all natural disasters worldwide especially in developing countries after the tsunami and affect the greatest number of people. Rain rates or the volume of rain drops per time also correlate very closely with rain fade, which is a weakening of the satellite communication signal as it passes through raindrops. Therefore, information about rain rates is crucial for communications. However, in remote areas retrieving such information from sparse numbers of weather stations is impractical. In such cases, satellite imagery becomes a potential solution to the monitoring of large-scale systems e.g. supercell thunderstorms, cyclones, or mesoscale convective complexes. The movement and patterns of clouds, observed from satellite images, can be used to indicate the instability of the weather (Bader et al., 1995). Moreover, useful information about rain clouds can be derived or inferred e.g. cloud depth and cloud droplet size from visible (VIS) channels, cloud top temperature and height from infrared (IR) channels, and cloud phase from water vapour (WV) channels, for instance (Brown and Wickham, 1995).

Satellite-based rainfall retrieval has become one of the most intense research topics in the discipline of satellite meteorology. Numerous studies have been dedicated to understanding rainfall from spaceborne observations. Among them are many attempts to empirically relate the brightness temperature from satellites, particularly infrared channels, to rainfall, e.g. cloud-indexing, multi-spectral, life history and cloud model-based (Barrett and Martin, 1981). These techniques are summarised in Figure 1.1.

The cloud indexing approach has a long history and is still useful. It classifies

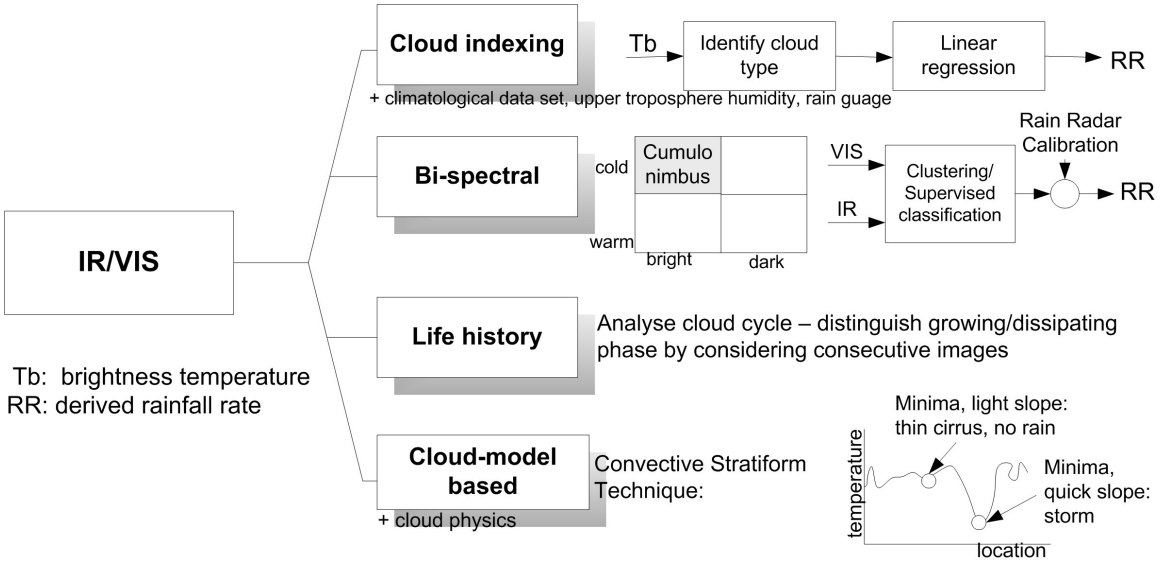


Figure 1.1: The four categories of rain estimation techniques using VIS and IR images.

images into regions according to the probabilities of the cloud producing rainfall and the nature of the rain that is produced. Subsequently, a predetermined rain rate is assigned to each cloud type. For example, Barrett (1970) derived rainfall coefficients, which afterwards were related to actual rain volumes by a series of polynomial equations, based on the monthly mean cloud cover, the probability and the intensities of rainfall. The contingency table of the rain estimates showed a good correlation with the actual rainfall, on a monthly basis. Follansbee (1973) estimated the average daily precipitation from the fraction of areas covered by cumulonimbus, nimbostratus and cumulus congestus. In the paper, cloud classification was done by expert operators. There has been extensive research on cloud classification in the past, however until now there has been no efficient methodology that can be done by computers alone. This includes the operational Meteosat Second Generation satellite's Spinning Enhanced Visible and Infrared Imager (MSG/SEVIRI) cloud mask and type products distributed by EUMETSAT[1]. The products of 21 cloud categories are derived by using brightness temperature differences, visible reflectance and ancillary data sets such as mean air temperatures and minimum sea surface temperatures by means of threshold methods (Derrien and Gléau, 2005). Although 8 MSG channels are used, there is still a misclassification, e.g. fog, stratus and stratocumulus are sometimes misclassified as

[1]The software to derive these products is developed and maintained by Satellite Application Facility in support to Nowcasting and Very Short-Range Forecasting (SAFNWC), which is a collaboration between EUMETSAT and El Instituto Nacional de Meteorologa de España (INM).

medium clouds (Gléau and Derrien, 2006). Scofield and Spayd Jr. (1984) proposed a technique for deriving rain rates for extra-tropical cyclones using Geostationary Operational Environmental Satellite (GOES) images, radar and conventional data such as surface and upper air observation. The speed of rain clouds was also included. The result, though mostly overestimated, achieved 51% and 45% accuracy for rain rate estimation and 6-hour prediction when compared with the observation using a 5×5 contingency table. Arkin (1979) identified an area colder than a given temperature as a rain-bearing cloud. The technique, known as the GOES Precipitation Index (GPI), estimated global precipitation for periods of five days (pentads) to one month. The algorithm worked well for large-scale rainfalls over the tropical oceans, but diminished over land—it underestimated rainfall over coastal and mountainous regions but overestimated rainfall over other land surfaces. This reflects the influence of surface characteristics. The Climate Prediction Centre (CPC) of the U.S. Agency for International Development (USAID) has improved this technique by integrating satellite rainfall estimates with Global Telecommunications System (GTS) automated rain-gauge data to create a bias-reduced daily precipitation estimate. The comparison of the monthly accumulated rainfall of $0.1^\circ \times 0.1^\circ$ latitude-longitude grids over Africa estimated by this method increased the correlation to the station-gauge observations from 0.34 of the conventional GPI method to 0.47 and decreased the bias from 2.25 to -0.04 mm/day (Love et al., 2004). It should be noted that the large time interval (30 days) can average the estimation errors to quite low numbers. The cloud-indexing technique is simple, yet more stable than other complex methods. However, using only temperature as a basis of rain estimation is valid only for convective rain and promising only at a coarse spatial and temporal scale. An altostratus cloud, for example, produces light to moderate precipitation although it sometimes appears as cold as a convective cloud in infrared images but with smoother texture.

Instead of using only IR images, multi-spectral techniques are based on a relationship between cold-bright clouds, which is a characteristic of cumulonimbus, and the high probability of precipitation. The technique works well in convective rain cases, better than considering visible or infrared images separately. By using a VIS/IR two-dimensional histogram, rain areas during *daylight* hours are determined by classifying only pixels (Lovejoy and Austin, 1979a; Griffith et al., 1978; Bellon et al., 1980) or supervised clusters of pixels (Tsonis and Isaac, 1985). For instance, Bellon et al. (1980) created a contingency table to obtain the bivariate rain distribution probability of the rain and no-rain events as a function of VIS and IR intensities. The probability, when assigned to each pixel according to its IR and VIS intensities, effectively filtered out

cirrus and stratus clouds. The result was also consistent such that its trajectory was similar to that of the rain radar image. However, the major difficulty of the multi-spectral technique arises when raining clouds, e.g. warm-top nimbostratus, are only slightly different from non-raining areas but they can produce considerable amounts of rain, and vice versa. There are also considerable overlaps in the distribution; an appropriate threshold must be selected to produce the least classification error, as described in Seze and Desbois (1987). O'Sullivan et al. (1990) enhanced the multi-spectral techniques by adding textural characteristics of daytime VIS data and IR temperature patterns to label rainfall into no rain, light rain, moderate and heavy rain. The probability of detection achieved 81%, compared to 66% when using only spectral features.

Life history techniques take into account a cloud's life cycle, e.g. the total cloud area and its rate of change. These methods are based on a finding that the total rain amount falling from a convective cloud is insensitive to the cloud's structure; if raining areas of convective clouds can be defined by the satellite data, the rain amounts should possibly be estimated by inference from past assessments of similar cloud events and life cycles. For instance, the area of a growing cloud system has a stronger relation to rain volumes than that of the decaying system (Lovejoy and Austin, 1979b; Griffith et al., 1978). In addition, precipitation peaks while the cloud area is rapidly growing and reduces at the time of maximum cloud area (Stout et al., 1979). The daily total duration of cold cloud has also been used in the Tropical Applications of Meteorology using SATellite and other data (TAMSAT) cold cloud statistics method to derive rainfall intensities by a linear function (Bonifacio et al., 1993). The method has been employed by the Food and Agriculture Organisation of the United Nations and a number of African Meteorological Services to provide information to the agricultural sector. The idea of the life history technique is applicable to convective rain and storms. The weakest part of the techniques seems to be the process of cloud tracking — cirrus shields might cover active convective clouds and make them invisible from the images. This leads to an overestimate in rain-bearing cloud areas, though the associative rainfall has decreased (Richards and Arkin, 1981). On the other hand, growing clouds in close proximities could have anvil interactions that can cause the rain volumes to be underestimated (Negri et al., 1984). In addition, the technique requires a period of time to observe the life cycle of clouds which sometimes are too short for processing.

Cloud model-based techniques estimate rain rates as a function of a cloud's characteristics. The relation can be found empirically using radar calibration or numerical

weather models. Adler and Negri (1988) developed the *convective-stratiform* technique (CST) using a one-dimensional cloud model to predetermine rain rates. An interesting feature of CST is the attempt to eliminate thin, non-raining cirrus by screening an IR temperature histogram for slight slopes. The remaining minima with higher temperature gradients are assumed to be convective elements. In addition, to improve precipitation estimation, the cloud physics could be included. Lensky and Rosenfeld (1997) proposed a method for categorising rain clouds into convective, intermediate and stratiform by their areas, texture and vertical velocities. The corresponding rain rate is then assigned to each cloud type, e.g. for an intermediate cloud, defined by the mixture of convective clouds and extensive stratiform clouds, the rain rate was a function of a cloud droplet's effective radius (r_e), an updraft velocity, the water vapour mixing ratio, and the air pressure. In the paper, r_e was derived from channel 3 ($3.7\mu m$) of the Advanced Very High Resolution Radiometer (AVHRR) data to indicate large water drops and ice crystals which initiated the coalescence precipitation. For a stratiform cloud, the paper suggested assigning a uniform rain rate of 3 mm/h. On the other hand, the rain rate of a convective rain was the second order parabolic function of its area. This might not be applicable to areas larger than 2 000 km^2, when cloud clusters probably protect each other from the detrimental effects of mixing with dry, mid-level ambient air. Reudenbach and Bendix (2002) derived r_e from tri-spectral bands at 0.7, 3.7 and 10.8 (μm) received from NOAA-AVHRR channels and estimated rain rates using 3-dimensional cloud model (MM5). The reliability of the technique was tested by assimilating the estimates into a hydrological drainage model. The results revealed a good visual agreement with the measured drainage graph. They proposed using this method with the Meteosat Second Generation satellite, which was not available at the time. Recently there has been a number of techniques that can derive r_e from the satellite. For example, Lensky and Rosenfeld (2006) generated a cloud mask from 8.7, 10.8, and 12.0 μm channels to ensure that only optically thick clouds were selected. Then, r_e of these clouds was derived from the 3.9-μm channel via a look-up table. Alternatively, the Visible Infrared Solar-Infrared Split-Window Technique (VISST) and Solar-Infrared Infrared Split-Window Technique (SIST), which retrieves cloud optical properties by matching calculations to observations, can be used for daytime and night-time, respectively. The two techniques have been used operationally for GOES[2] and showed a compatibility when adapted to the MSG (Minnis et al., 2004).

Passive microwave-based rain estimation techniques have the advantage of being more physically direct than visible and infrared observations as they involve absorption

[2] `http://www-angler.larc.nasa.gov/satimage/products.html`

and scattering interaction between rain droplets and microwave radiation in the atmosphere. The precipitation-sensitive properties can provide information about the total precipitable water and the cloud liquid water content. However, currently available information from polar-orbiting satellites is not frequent enough to provide reliable weather monitoring. Recently, several methods have been proposed to hybridise the microwave information with the geostationary satellite data, i.e. to calibrate the IR brightness temperature with derived precipitation rate from microwave (MW) imagers (Adler et al., 1994; Xu et al., 1999; Turk et al., 2000; Todd et al., 2001). An example of the hybrid rain estimation is illustrated in Figure 1.2. How to combine such disparate information at different spatial and temporal scales is still an open issue. Moreover, the sensitive reaction of MW to the variabilities of surface conditions, e.g. soil moisture and vegetation, can result in seriously erroneous rain rate estimation (Ferraro, 1997). Recent investigation of these hybrid techniques can be found in Kidd (2001) and New et al. (2001).

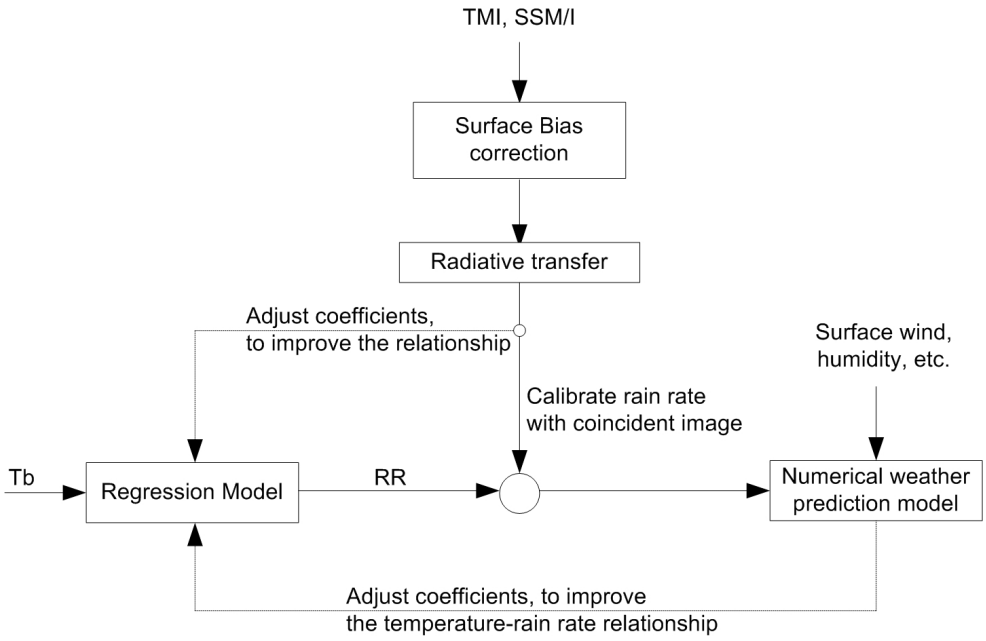


Figure 1.2: The hybrid rain estimation combining IR and MW images in Turk et al. (2000).

1.1 Motivation for Research

In addition to extensive research that shows a potential for deriving rain amounts from satellite measurements, a number of observations also reported a good correlation between rain amounts and cloud appearance. For example, stratiform clouds which appear as broad flat clouds with a very smooth texture in VIS images produce low but widespread rain rates (100 km in horizontal dimension) whereas convective clouds which appear bright in VIS images due to the thickness of these clouds, often exhibiting a lumpy texture caused by the shadows from the irregularly-shaped cloud tops, give high, localised rain rates (> 5 mm/h) (Houze Jr., 1993).

The relation between cloud appearance and rain rates also depends on particular types of meteorological events. For example, Krennert et al. (1999) observed that a cold-front dominated cyclone (aka Norwegian cyclone model, emerging cloud head cyclogenesis or warm conveyor belt type) which is meridionally elongated as seen from satellites embeds a combination of precipitation types. For example, the dry intrusion at the rear edge (the inner boundary of the spiral) makes it a preferred area of instability for cumulonimbus development. This can be seen as a sharp gradient from white to black in WV images. There may be showers, thunderstorms or hail at the rear part of the spiral. On the contrary, ahead of the surface front the precipitation is broad (5-50 km) with its maximum close to the occlusion point, which appears brightest in IR images.

Figure 1.3 shows a sample of cloud pattern sequences that are commonly associated with an occluding depression, from Kurz and Young (1995). The image interval is 3 hours. The comma shape in the first image is caused primarily by an incursion of the dry air. It suggests that the depression may already be partly occluded. Dry air which descended from upper atmosphere overruns the surface front at D, resulting in a gap between the surface front and high clouds. Later on during the occluding process, the basic pattern of the higher cloud does not change very much. Rather, only changes in texture and location of the cloud band are the main observation during this time. These patterns may cause homogeneous rain estimates if only low temperatures are used as a threshold to estimate rain rates. The IR images and their associated weather diagrams in Figure 1.4 show that these textures associate with different rain types. For example, areas of continuous precipitation represented as shaded areas in the surface pressure diagram[3] indicate that most of the precipitation areas collocate with the thick cloud on and ahead of the frontal cloud band. Showers and thunderstorms

[3]Symbols used in schematic diagrams and charts can be found in Appendix A.

are found during a development of thick, bright convective cloud band marked as X-X in Figure 1.4 (b) and (c) whereas hail, light rain and showers are more concurrent with smoother, darker clouds.

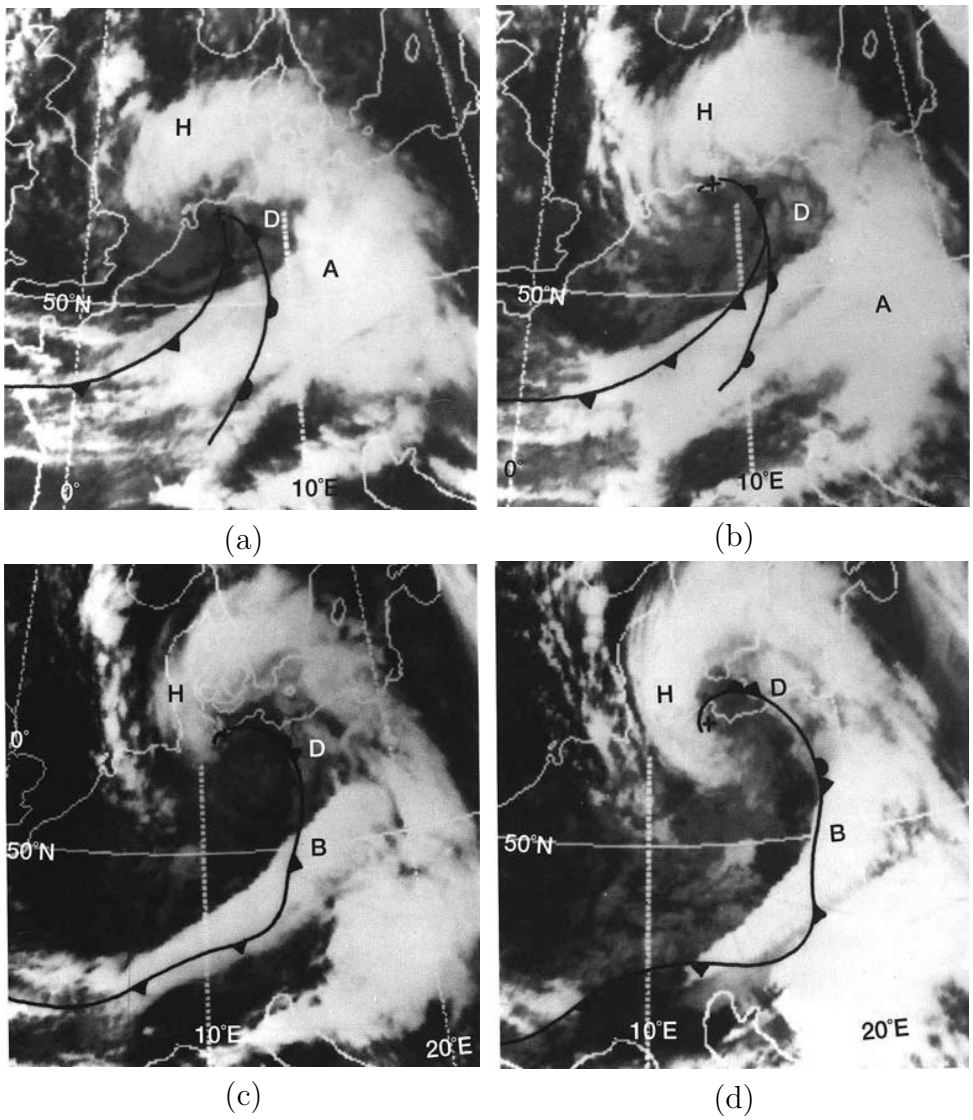


Figure 1.3: Meteosat IR image series and frontal analysis over part of western Europe on 20 October 1986 at (a) 1500, (b) 1800, (c) 2100 and (d) 2400 UTC. The symbol + is the surface low centre. The weakening warm front was omitted in the last two pictures. (Source: Kurz and Young (1995)).

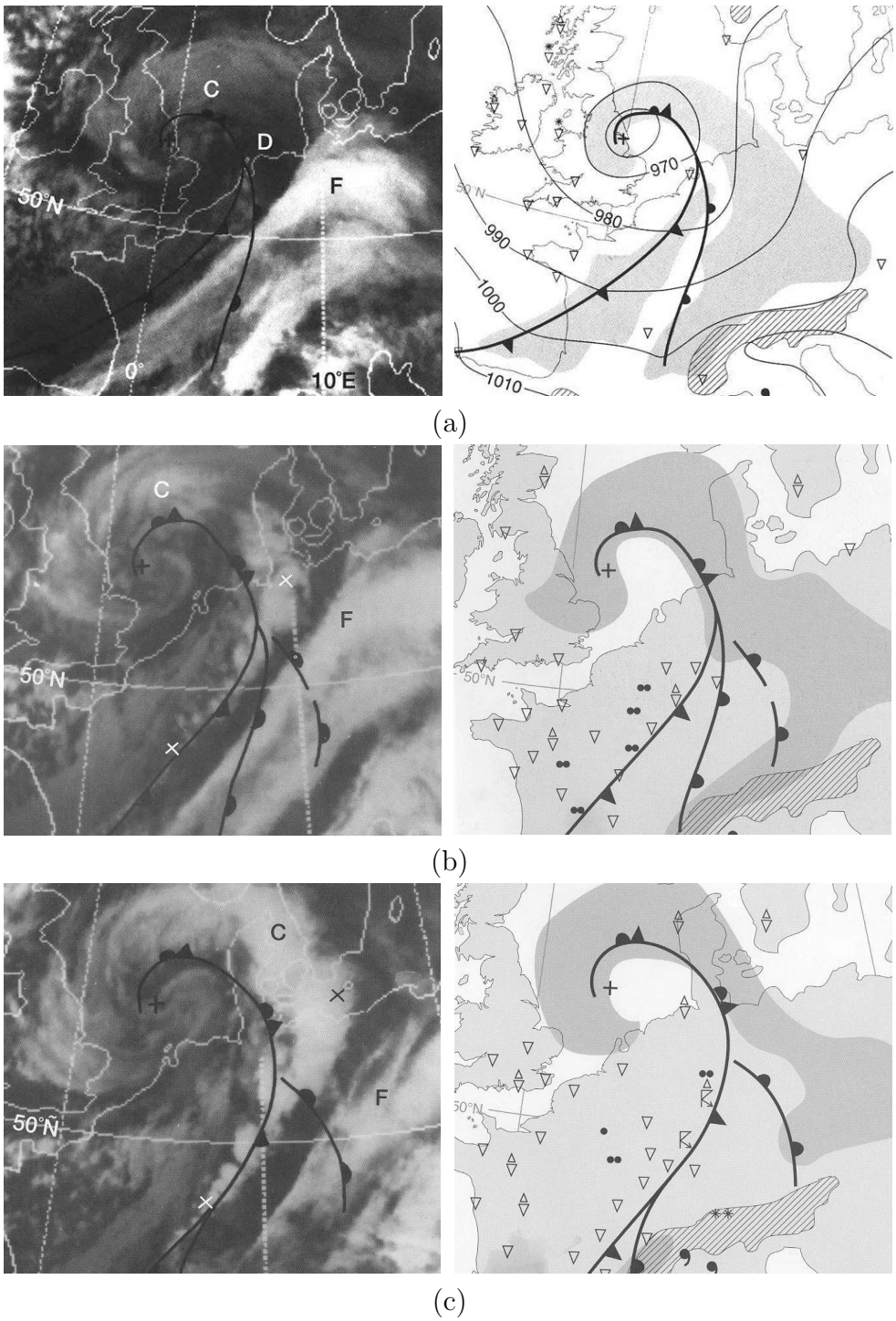
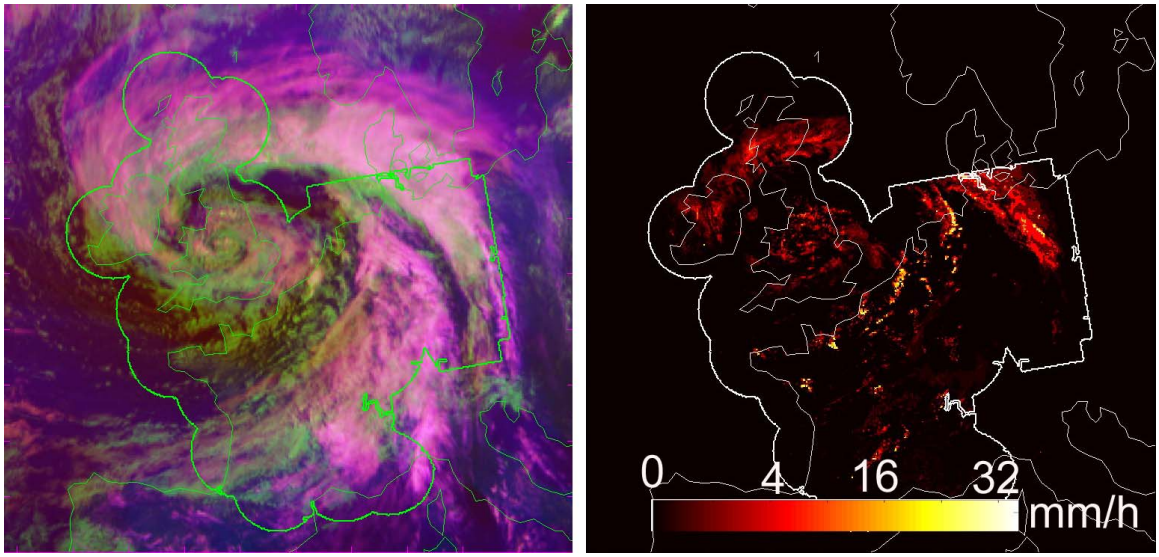


Figure 1.4: IR images and their associated weather diagram of an occluding depression over western Europe on 24 March 1986 at (a) 1200, (b) 1500 and (c) 1800 UTC. Hatched areas are mountainous and shading indicates areas of continuous precipitation. (Original source: Kurz and Young (1995)).

A better resolution satellite image and the corresponding rainfall information of a mature mid-latitude cyclone are presented in Figure 1.5. The combination of IR, VIS and WV images shows several patterns that resemble those of the rain radar image.



(a) (b)

Figure 1.5: (a) RGB false image of a summer mid-latitude depression over the British Isles and western Europe on 23 June 2004 at 1030 UTC from Meteosat7 (IR ⇒ red, VIS ⇒ green and WV ⇒ blue). The corresponding rain rates are presented in (b). The images are polar-stereographic projection and the thick white line indicates the area where the rain information is available.

These observations show an obvious relation between precipitation and cloud appearance from geostationary satellite images. Therefore, the cloud appearance is expected to provide some useful clues to help the rainfall estimation and prediction. This assumption motivates an idea to study the relationship of cloud texture to rain rates. A challenge is the diversity of textural appearance which makes it difficult to provide a precise definition. Cloud textures could either be homogeneous or have spatial arrangements. They may contain both statistical or structural properties or even random placement of elements. Fortunately, there are several classes of image processing algorithms available. These techniques are unique and will be analysed in this thesis. In addition, the correspondence between cloud and rain movement implies that the motion of clouds could be used to predict rain movement, which is required for rainfall prediction. Therefore, an analysis of cloud movement is also included in the investigation of rain estimation and prediction.

1.2 Objective

By the observation that there is a strong relation between the pattern of clouds and their corresponding rainfall, the appearance of the clouds and how they move are expected to be helpful for rainfall estimation and prediction. This study aims to extract the appropriate motion and textural information of clouds from a geostationary satellite that can be related to the corresponding rain rates. Geostationary satellite images are selected for the study because they are spatially continuous, highly repetitive and relatively inexpensive to receive and process. Though indirectly related to rainfall, they are still the most promising source for rainfall information and are the richest resource that is available. However, a particular meteorological situation can affect rainfall rate differently. Therefore, only mid-latitude cyclones are chosen as a subject of the study in order to confine various meteorological effects. Based on the geostationary satellite images that are available, the objectives of this work are to

- analyse textural information of clouds seen from the geostationary satellite,
- investigate techniques to derive motions of clouds for mid-latitude cyclones,
- study the texture of clouds from the geostationary satellite and its relationship to the intensity of rainfall of the mid-latitude cyclones on the surface of the Earth,
- use the derived cloud motions and textures to predict rain rates.

For clarity, the work does not intend to study all physical processes within clouds or their effects at the surface, but it tries to exploit the available satellite data and aims to be a step towards a better estimation and prediction of rainfall rates.

1.3 Scope of Work

1.3.1 Data

One of the inherent problems in rain rate estimation from satellite measurements is a scarcity of ground data to verify instantaneous satellite retrievals. Therefore, the mid-latitude region over the British Isles and western Europe is selected as an area of study. The main reason is the density of rainfall information that is available from the Radarnet reception system. The rain rate information is retrieved from the British Atmospheric Data Centre (BADC). It is derived by the British Meteorological Office every 15 minutes from a network of radars across the United Kingdom, France, Germany, Belgium and the Netherlands and then combined into a single pan-European product on an (equal area) polar-stereographic projection[4]. The data is collected at a nominal 5 km resolution, which provides a better resolution than using the interpolation of data from weather stations. Using this information for estimation and verification has a number of advantages, especially the avoidance of errors encountered when comparing satellite estimates with rain-gauge measurements. However, the limitation of the radar information is that there may be cases where the cloud particles are too small or the cloud is too thin and too far, e.g. more than 10 km away, to be detected (Derrien et al., 2005). There can also be spurious echoes resulting from anomalous propagation of the radar beam, errors resulting from variations in the vertical profile of reflectivity and radar sensitivity errors(Harrison et al., 2000). The rain information used in this thesis is derived from .GIF images due to the restricted access to digital data at the time the research was initiated. However, later inspection found that the two data are highly correlated, i.e. the correlation coefficient between them is approximately 0.79-0.89. The .GIF images, therefore, should be adequate for use in rain estimation and prediction.

For geostationary satellite images, infrared, visible and water vapour images of Meteosat 7 have been used in the study. The images were received from BADC and converted into radiances by calibration coefficients provided by Eumetsat. The visible images are also normalised by solar zenith angles to account for different illumination of the sun at different latitude and longitude coordinates and different times of day. The resolution of the satellite image in the mid-latitude regions increases from 5 km/pixel at the nadir to 8 km/pixel. Afterwards, all the images are reprojected from orthographic to polar stereographic projection to align with the rain radar images.

[4]At present no additional quality control is performed on the data.

Details of the image pre-processing are in Appendix B. Average resolution of the reprojected image is 5 km/pixel. The resolution per pixel of a rain radar image and the coverage of the rainfall information are illustrated in Figure 1.6.

Although the radar and satellite measurements are finely collocated with a displacement error not exceeding 1 pixel when considering coastlines and latitude-longitude grids, there can be displacement errors owing to viewing angles of the satellite relative to cloud height[5] which should be within 12 km and the error contribution arising from the non-temporal coincidence of the satellite and radar observations which can differ by at most 7.5 minutes (Uddstrom and Gray, 1996).

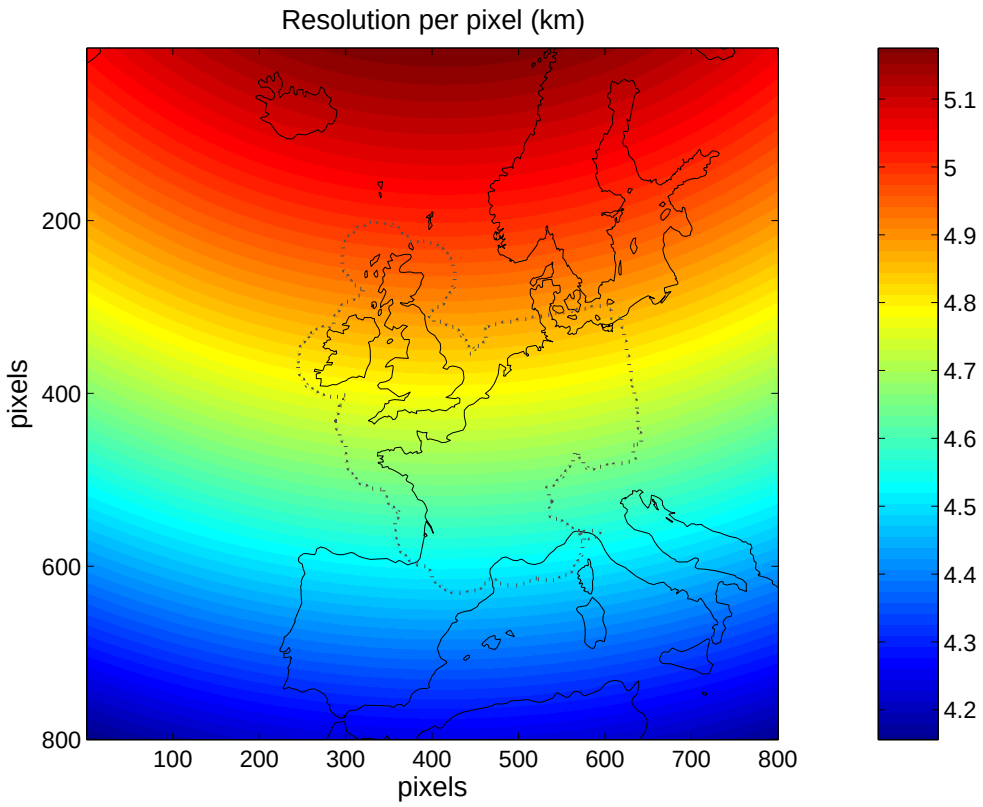


Figure 1.6: Area of the study on the polar stereographic projection. The dotted line shows the area where the rain information is available. Resolution (km per pixel) is represented in colours.

Current microwave instruments such as the Special Sensor Microwave Imager (SSM/I) over oceans could also be used as a ground truth data. Its precipitation

[5]details in Chapter 3

is calculated using the relationship to temperature derived from the measured microwave radiation and is well-calibrated with actual rain rates. However, the images have a pixel resolution of 12.5×12.5 km and are available only 3 times a day. A finer passive microwave imager at 4×4 km of the Advanced Microwave Scanning Radiometer (AMSR) at 85 GHz is also available but only once a day. On the other hand, a more recent sensor that can provide better rain information such as the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) has coverage extending only from 40°S to 40°N, which is not practical for the mid-latitudes.

1.3.2 Verification and Comparison

There has always been a problem trying to quantify the accuracy of rain estimates and compare them with different rainfall estimation methods due to different scoring schemes, different space-time resolution and different meteorological events (Smith and Austin, 2000). The methods may also relate differently to variations in topography, atmospheric instabilities, synoptic conditions or other random variations that may occur. Moreover, each statistic gives only a piece of information about the error, and so it is necessary to examine a number of statistics in combination in order to get a more complete picture. Therefore, the performance of the proposed algorithm should be carefully quantified. The statistics recommended by the International Precipitation Working Group (IPWG) validation/intercomparison project are used. These statistics include root mean square (RMS) error, Hanssen and Kuipers score and Heidke skill score. Details of these statistics and their definition are in Chapter 4.

1.4 Scheme of Work

The rain estimation procedure used in the study is described as follows. First, the information that relates to rainfall is extracted from the satellite images. Next, the most relevant information is selected as an input of an estimator, which is trained to find the relationship between the information and the rain rates acquired from the rain radar images. The output from the estimator is an estimate of the rain rate. Subsequently, the trend of the movement from a previous moment is incorporated to predict rain rates in a future period. The rain radar data is also used to validate the performance of the prediction. The flowchart of the work is presented in Figure 1.7.

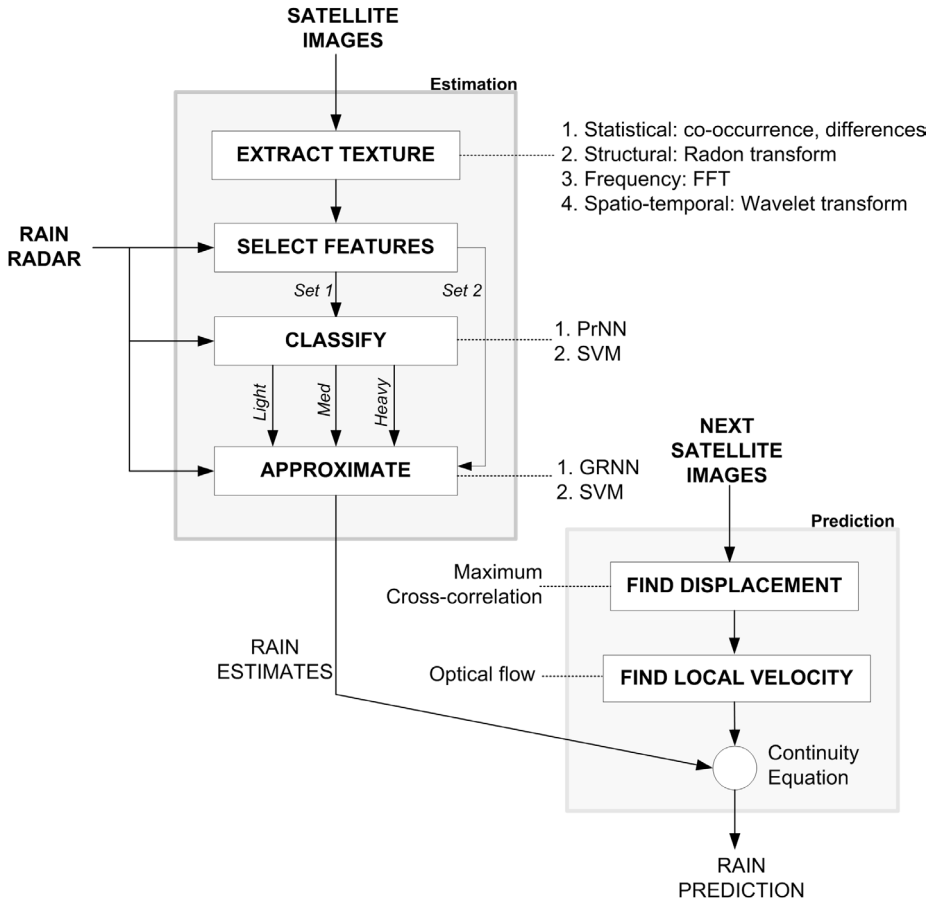


Figure 1.7: Schematic diagram of the proposed algorithm to estimate and predict rainfall. Text linked with dotted lines indicates the algorithms used in each process.

1.5 Organisation of Thesis

In this chapter the objectives of the thesis to relate texture and motion of mid-latitude cyclones to rain rates have been outlined. Relevant research has been reviewed to show the feasibility of these objectives. In Chapter 2, literature surveys of the past research have been done. The surveys focus on how texture and motion can be extracted from the satellite images and what methods can be used for the estimation. In Chapter 3, several techniques to extract characteristics of the mid-latitude cyclones from the satellite images into quantitative parameters have been studied. The implementation of rain estimation using these parameters together with the verification of the estimation results are in Chapter 4. The method of combining cloud motions to the derived rain estimates in order to predict rainfall is described in Chapter 5. Finally, the conclusion and future research topics of the work are in Chapter 6 to complete this thesis.

Chapter 2

Literature Survey

In the past, there have been many attempts to derive rain rates directly from radiances of the satellites. However, the similarities of the spectral response of different types of clouds such as ice clouds and snow make such methods likely to produce erroneous results. In this chapter, the extraction of more informative features that have been used for cloud images is investigated. The information that can be extracted from satellite images are cloud texture, cloud movement, spatial features such as shape and size of clouds and contextual features such as sea surface temperatures. Details of these features are in Section 2.1 Methods that have been used for the rain rate estimation are also discussed in Section 2.2.

2.1 Information Extraction

2.1.1 Textural features

Instead of only a pixel being considered in isolation, the correlation to its neighbours can be used to better characterise clouds. The relationship in terms of the repetitiveness of spatial arrangements is called a textural feature and is used to signify the structure of an image. While the spectral characteristics of types of clouds may change, their textural properties are often characteristic and tend to be less sensitive to atmospheric attenuation or instrumental noise (Lamei et al., 1994).

The textural information is often computed as statistical measures of the spatial distribution of image intensities. This is applicable to fine textures which have no regular spatial arrangement but obey some statistical rules. The most common method to derive the relation of any two pixels, referred as the second statistics, is to derive a set of statistical measures from a co-occurrence matrix. The matrix represents a

conditional probability of image intensities between a pixel and its neighbours at a given displacement $(\Delta i, \Delta j)$, namely

$$C(p,q) = \Pr\{I(i,j) = p \quad \text{and} \quad I(i+\Delta i, j+\Delta j) = q\}, \tag{2.1}$$

where I is an intensity at the position (i,j) and p and q are the values of the intensity[1]. This technique provides good results in cloud classification of high-resolution, low-orbiting satellite images (Welch et al., 1988; Gu et al., 1991; Pankiewicz, 1995). By an observation that the information close to the diagonal of a co-occurrence matrix from Meteosat VIS images could distinguish cloud types, Lewis (1998) created a histogram of a combination of diagonal and near-diagonal co-occurrence values in order to define modes of the histogram that corresponded to different cloud components, e.g. cloud edges, low clouds, middle clouds, high clouds and cloud tops. The clouds were then classified into high, middle and low altitude.

Instead of the probabilities of co-occurrences, the difference method describes the spatial arrangement, e.g. variability/homogeneity, of a scene by using the probability that the intensities of any two pixels in the scene differ from each other by a specific value. Using a difference array, defined by

$$D_{d,\theta}(i,j) = |I(i,j) - I(i+\Delta i, j+\Delta j)|, \tag{2.2}$$

where d is a distance between the pixel (i,j) and $(i+\Delta i, j+\Delta j)$ and $\theta = \tan^{-1}(\Delta i/\Delta j)$ is an angle between a line connecting the two pixels and the x axis, a probability density function can be estimated, namely,

$$f_{d,\theta}(k) = \Pr(D_{d,\theta}(i,j) = k). \tag{2.3}$$

Statistics of this function are ranked as one of the most useful textural features for nephanalysis (Wu et al., 1985; Gu et al., 1989; Welch et al., 1990; Khazenie and Richardson, 1991; Uddstrom and Gray, 1996). The statistics of the difference matrices, e.g. means and maxima, when added to determine rain rates from GOES infrared and visible images in Wu et al. (1985), improved the Panofsky-Brier skill score[2] by 27% more than using the radiances alone.

[1] There are 2 types of position indexes of an image used throughout the thesis. First, (i,j) indicates row and column indexes of the image. On the contrary, $\vec{x}$ refers to the position in x and y coordinates.

[2] Panofsky-Brier skill score is defined by $\frac{R-E}{T-E}$, where R is the number of correct estimates, T is the total number of estimates, and E is the expected value of R.

The co-occurrence and the difference matrices could be calculated at every point or within non-overlapped windows to reduce the allocated memory and time for the computation. However, defining neighbours by using a fixed-size window has a disadvantage when applied to objects that can vary in size. This is because the extent to which textural properties may be recognised is related to the resolution at which the image is observed, therefore the resolution plays an important role in the performance of cloud classification (Welch et al., 1989). The fixed-window method also suffers from the localising problem if a large window size is applied. Grazzini et al. (2003) gave a scale-invariant weighting function to pixels in the window so that pairs of pixels further away from the base pixel contribute less. The method was expected to enhance localisation and make the results independent of window size. Nevertheless, the result of cloud segmentation when this improved method was applied to a Meteosat IR image was rather poor. The authors gave a reason that both traditional and improved methods assumed the regularity condition of the image, e.g. each type of clouds exhibits the same textural characteristics. The assumption is violated due to thermodynamic properties of a turbulent, chaotic flow of clouds, which can make the same type of cloud appear differently in satellite images.

Another way to describe texture is via transform techniques such as the Fourier transform. It assumes that a cloud field can be modelled by a wide-sense stationary stochastic process[3] — the intensity distribution of a cloud object is assumed to be invariant with respect to spatial coordinate translations. The transformed data can then be characterised by a sum of energies, entropy measures which quantify the spread of spectral-frequency entries, or numbers of dominant frequencies. For example, the texture pattern of cumulus has a large number of dominant frequencies, which signals that it is highly textured, and its small value of entropy indicates that the texture is structured or regularly arranged (Gu et al., 1989). Nevertheless, the global processing nature of the Fourier transform means that the method fails to identify local structures in an image.

Many psychophysical studies on human texture perception, performed extensively from 1960-80, revealed that a human can clearly differentiate two types of texture if they differ in their second order statistics (Thomson and Foster, 1997). However,

[3]A stationary stochastic process is a random process whose probability distribution at a fixed time or position is the same for all times or positions. Here, *wide-sense* implies that it only requires that the first and second moments do not vary with respect to time.

in some cases, we are also able to discriminate two textures even if they have the same second statistics (Lamei et al., 1994). This observation suggests that there are not only spatial but also frequency analyses being performed in the visual system. Thus, a number of techniques involving discrimination of texture by modelling the human visual system have been proposed. A more recent spatial-frequency analysis technique is the wavelet transform. The idea behind these time-frequency joint representations is to cut the signal of interest into several parts and then analyse the parts separately in frequency domains. In the wavelet analysis, a fully scalable modulated window function is shifted along the signal to calculate the corresponding spectrum for every position. The process is repeated several times with a slightly shorter (or longer) window as the frequency changes. The output is a collection of time-frequency joint representations of the signal, all with different window sizes. This can also be considered as a multi-resolution approach. Amato et al. (2001) found that applying the orthogonal Daubechies wavelet transformation to Meteosat images as a textural feature can facilitate the discriminant analysis approach. The discriminant analysis is one of the powerful classification tools that prefers simple probability densities of the input variables. The transformed data in high-frequency, high-resolution (H-H) regions is uni-modal and, therefore, suits the discriminant analysis.

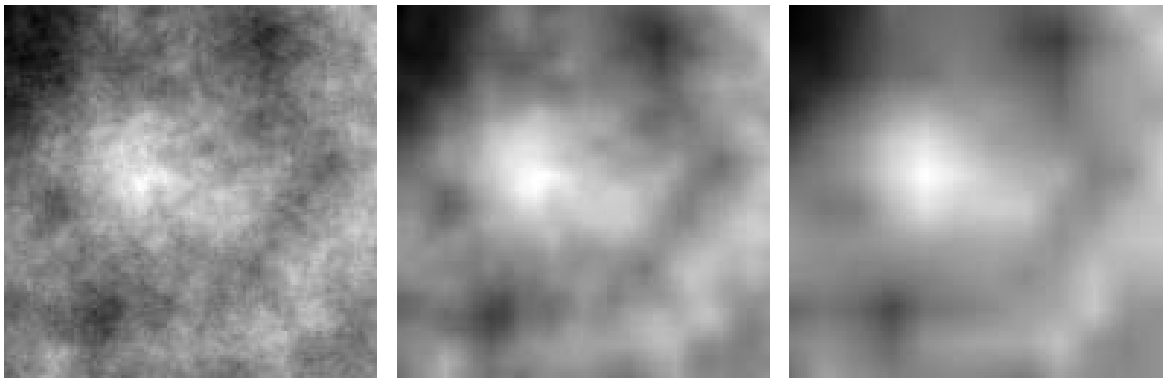
Alternatively, there have been attempts to construct a parametrically generative model that could have created the observed intensity distribution to describe texture. For example, a stochastic spatial interaction model treats an image intensity as an output of a stochastic process, whose input is a sequence of independent random variables, i.e. the observed intensity. The output intensity is modelled as a linear combination of neighbouring intensities plus an additive noise. On the contrary, a random field model utilises spatial variations. It is a probabilistic graphical model that represents the joint distribution of a set of random variables as a factorisation over a graph. The model consists of nodes that represent random variables, and arcs that represent conditional independence assumptions. The widely used random field model is the Markov random field model which assumes that an image pixel possesses a Markov property, i.e. the present state only depends on the current state. In other words, the conditional distribution depends on values of the neighbouring pixels only. However, Kostinski (2000) found that the statistical characterisation of the spatial distributions of cloud particles is a doubly stochastic Poisson process, which can result in complications in estimating parameters of the texture model.

Another method to describe the texture of an image is the fractal concept (Mandelbrot, 1982, 2001). It arises from an attempt to explain the dimension of a complex object. To measure a coastline, for an example, a number of measuring tools that are different in shape, scale and size would give different lengths of the same coastline. Therefore, the *fractal dimension* is used as the measurement of an object in order to cope with the measuring inconsistency. Using the idea that the object contains N distinct copies of itself, or fractals, each of which has been scaled down by a ratio r, the fractal dimension is defined as

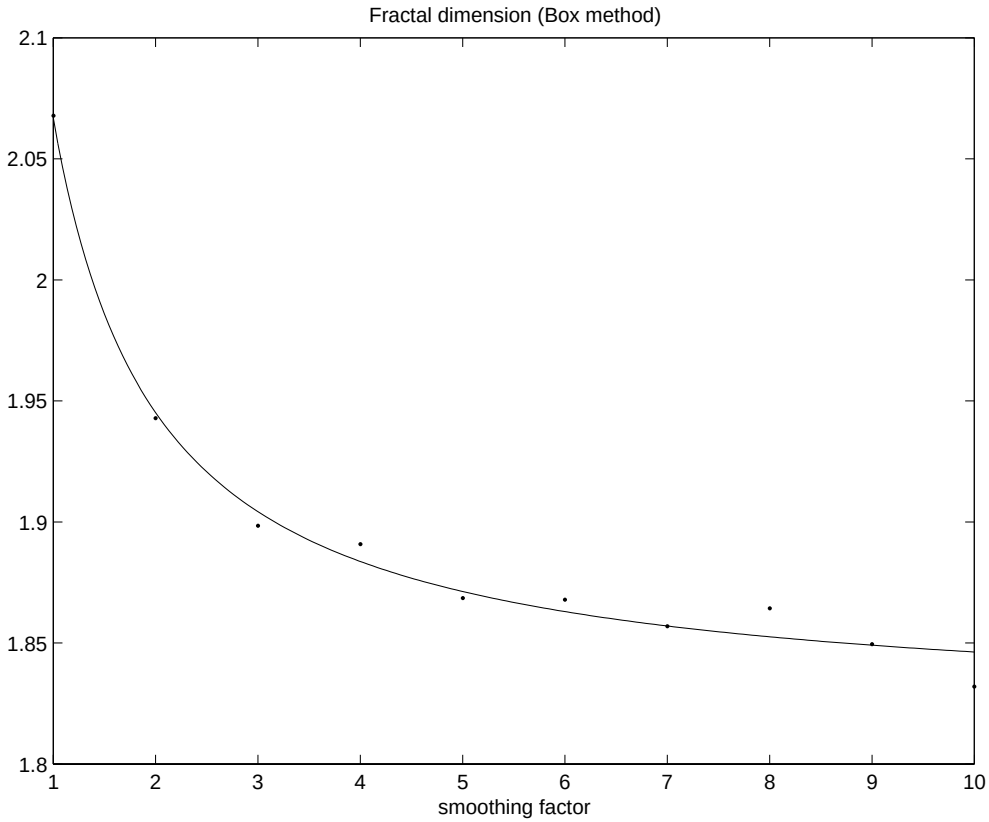
$$D_f = \frac{\log N}{\log(1/r)}. \tag{2.4}$$

Theoretically, D_f has a certain value according to the roughness of texture, despite changing scales. This property makes it appealing to be a texture descriptor. As a consequence, the idea of the fractal is now widely used in image processing and image compression (Tso and Mather, 2001). Figure 2.1 (d) shows the fractal dimension as a function of the smoothness of the texture. The smoothness is represented by sizes of a mean filter applied to a simulated image in Figure 2.1 (a). The size of the filter is labelled as a smoothing factor, e.g. a smoothing factor of 8 refers to a mean filter of size 8×8 pixels. As seen from the figure, the fractal dimension of the *whole* image inversely correlates to the smoothness of the image.

However, in nature an image can contain various types of texture. Therefore, it is impossible to find an exact replica that can precisely reproduce the whole object. Instead of using a single fractal dimension for the whole image, it can be calculated for each overlapped sub-image and then assigned to the central pixel of the sub-image. A composition of many fractals is known as a multi-fractal. Lovejoy and Schertzer (1986) produced very realistic simulations of many different cloud types using fractal generation algorithms. Grazzini et al. (2005) applied the multi-fractal concept to the Meteosat IR images in order to extract their main singularities, which were referred to as the main current of the turbulent flows. Then, the gradient of the singularities was used to identify where the vertical movement took place. In this manner, the convective characteristics of clouds were localised. The result gave a very good agreement with the TRMM ground truth data. However, selecting an appropriate size of each sub-image becomes a matter of concern when using the multi-fractal method as it should be large enough to allow sufficient numbers of window size to calculate the fractal dimension but not too large that regions with different textures are combined. The solution can be found by trial and error as different applications involve different image textures.



(a) (b) (c)



(d)

Figure 2.1: (a) A simulated image and its smoothed versions using a smoothing factor of (b) 5 and (c) 10. Fractal dimension as a function of smoothing factors applied to the original image is in (d).

Christodoulou et al. (2003) classified clouds by nine different sets of textural features. The sets included the first-order statistics, e.g. mean, standard deviation, skewness, and kurtosis; the second-order statistics, which are co-occurrence matrix, difference statistics and pair-wise statistical feature matrix, when any two pixels were considered; neighbourhood (within a specific-size window) difference matrix; Laws's texture energy measures; fractal dimensions using the Brownian motion model; and Fourier power spectrum. Each set had its own classifier. Then, the class that won a dominant voting among the nine classifiers was selected as a final output. The classes consisted of altocumulus-altostratus, cumulonimbus, cirrus-cirrostratus, cumulus-stratocumulus, stratus and clear conditions. Using 366 samples from 98 Meteosat IR images, the technique achieved 64% accuracy. The class selection was also done by the weighted averaging approach but yielded a slightly better accuracy. However, the paper did not clarify how the textural features contribute to the classification—why these features were chosen, or what was the correlation between each feature. Eliminating dependent features can boost the performance of the classifiers and therefore attain higher accuracy.

It should be noted that some assumptions are made in cloud texture analysis. Commonly, the algorithms assume that clouds exist in a single layer; they are at the same height and do not overlap with each other. The assumption is severely violated in the Arctic and near the core of cyclones where cloud layers are well-separated in a vertical column (Ebert, 1989). The window-based algorithms always regard the considered window as one object; it can be either one type of clouds, land or sea—but one only. Therefore, the chosen window size should be small enough to contain appropriate homogeneity but large enough to embrace adequate information. In addition, it implicitly assumes that the atmospheric water vapour absorption is negligible; the radiance received comes directly from clouds or the earth's surface when the sky is clear. However, based on the observation of similar texture between top of the cloud and rain radar images, these effects should be trivial when texture information, i.e. the relation between neighbouring pixels rather than the actual value of the pixels, is used.

2.1.2 Temporal features

Temporal information[4] derived from geostationary satellites has been proved to be an effective resource for retrieving wind information. Description of current operational wind retrieval procedures can be found in Schmetz et al. (1993) for Meteosat, Nieman et al. (1997) for the NOAA/ NESDIS GOES, and more recently Holmlund (2000) for the Meteosat Second Generation. In summary, the detection of the atmospheric motion vector (AMV) is done by a three-step procedure: tracer selection, height assignment and cloud motion retrieval.

For instance, at the Japan Meteorological Agency (JMA)[5], the data is derived four times a day from three consecutive infrared, visible and/or water vapour images at half-hourly intervals. First, cirrus (for high-level wind) and cumulus (for low-level wind) are detected by using histogram analysis of the infrared images within an area of 32×32 pixels for use as a target/tracer cloud. Next, a displacement between the target cloud of an image and the following image is calculated by a pattern-matching algorithm. Two of the most familiar matching techniques are the sum of squared differences and the cross correlation methods, both of which seek to find the maximum correlation between an initial target and the subsequent match scene. In the second step, the heights for high-level wind and water vapour wind are assigned to cloud top heights calculated from the brightness temperature, whereas 850 hPa is uniformly assigned to low-level cloud[6]. Afterwards, the positions of the tracked clouds are transformed into latitude and longitude coordinates, and the speeds and directions of winds are calculated by taking into account the displacement and the time difference between the images. A motion vector derived from the first and second images is compared with a vector derived from the second and third images to check the validities of AMV derivation. The variation of cloud-top temperature is also checked in order to reject vectors derived from growing or decaying cloud. However, most of the high-level clouds especially semi-transparent clouds cannot be tracked appropriately in the automatic processing, therefore a subjective quality control is also needed. Finally, AMV data are disseminated to JMA Headquarters in operational use for numerical weather prediction.

[4]For clarification, a temporal feature means a piece of information extracted from a sequence of images. On the contrary, a textural feature describes the relationship between a pixel and its neighbours whereas a spectral feature means a particular value of that pixel.

[5]`http://mscweb.kishou.go.jp/general/activities/products/amv.htm`

[6]The average pressure altitude for 850 hPa is about 1500 metres above sea level.

The AMVs, however, are not a direct measurement of the wind field— clouds are not always passive tracers, and their occurrence may not always be in areas of the strongest wind speed. In addition, the cloud motion represents a layer-mean flow rather than a wind vector at a specific level; it represents cloud-top movements rather than the bulk movement of the atmosphere (Schmetz et al., 1993).

Endlich et al. (1971) found that the efficiency of the maximum cross-correlation technique when applied to cloud images depends on the resolution of the data, the translational speed of the clouds, and the relative importance of random, non- translational changes in cloud locations and brightness. There are also cases where cloud motion changes with altitude. Therefore, it seems appropriate to include cloud-height information in order to improve cloud motion estimation by the cross-correlation method. In addition, the method suffers from expensive computation and limitation of searching directions, i.e. horizontal or vertical, such that *only* uniform translational displacements can be properly measured. Moreover, the result which is a speed, or a displacement, may not be the true velocity of the object in the real scene.

To derive velocity of an object in three-dimensional space from a sequence of two-dimensional images, Horn and Schunck (1981) proposed the idea of optical flow method exploiting a fluid dynamics constraint. This technique pretends to see a connection of the object in two sequential images and describes this invisible connection by (optical) fluid flows. The idea is also applicable to cloud motion, which is a special case of fluid motion and consists of very complex motion dynamics (Zhou et al., 2000). The main assumption used in the optical flow method is that the brightness of an object remains constant with respect to time, i.e.

$$\nabla I \cdot \vec{v} + I_t = 0, \tag{2.5}$$

where $I = I(\vec{x}, t)$ denotes the image intensity as a function of space ($\vec{x}$) and time (t), I_t the first order derivative of I with respect to time and $\vec{v}$ the image velocity vector. The optical flow method appears to work well with infrared images on several applications such as the recovery of vertical motions of a hurricane and a temporal interpolation to generate missing frames in the Meteosat sequences (Zhou et al., 2000; Larsen et al., 1998).

Due to a limitation that infrared images from geostationary satellites are the only source to classify low clouds at night, Papin et al. (2002) introduced temporal informa-

tion to perform contextual labelling to each pixel. A motion-related feature is derived from the optical flow constraint equation, including its measure of confidence. A clear-sky pixel was then identified by its non-movement. For large motions, the multi-scale analysis was performed by minimising an energy function which is a combination of two potential functions that accounted for low clouds and clear-sky characteristics, e.g. motion vectors and temperature of pixels. These labelled regions were then used as initial seeds for the minimisation process of the uncertain areas. The results, compared to the classification of AVHRR based on multi-spectral method, gave a visual agreement.

2.1.3 Spatial feature

Shape and size of clouds have always been an important source of information to meteorologists. The simple shape descriptions such as centroids, elongation ratios, local polynomial approximation or Fourier shape descriptors can be used. A precise description of shape definitions and changes during a meteorological event could give clues to its future movement and structure. Despite a rich diversity of spatial features, there is very little research in this area. Papin et al. (2002) applied distinct shape features of clouds to enhance the cloud classification, especially in the cases that fog and stratus would have been mislabelled as low clouds. Seddon and Hunt (1985) proposed shape and shadow features for the use in Meteosat images. These features were the thinness ratio and circular deficiency which were derived from cloud area and perimeter length. The remaining features were size, fullness and complexity of the skeleton. However, the method is efficient for rigid bodies but is useless to describe complex structures; a suitable shape description must be able to discard local perturbations to the original shape of an object and allow the evolution of the same object in different time frames.

There is a positive correlation between sizes and the lifetime of convective systems (Machado et al., 1998). For example, most of the low temperature (less than $245°K$) systems with 100 km equivalent radius have lifetimes shorter than 3 hours. As the systems evolve, splitting of the clouds may occur. This results in changes of the precipitation pattern and three-dimensional wind structure. The process occurs very often and can be considered as an important part of the cloud's life cycle. It also indicates the intensification of precipitation and the decay of the upward movement of the storm (Cotton and Anthes, 1989). This information is applicable to convective systems which have well-defined shapes. For other types of meteorological conditions such as mid-latitude cyclones, these characteristics are unlikely to happen. A cyclone

usually consists of several types of clouds. Therefore, although having a distinct spiral shape, the true shape of individual events occurring within the cyclone is difficult to derive.

2.1.4 Contextual features

Supplementary environmental parameters can be added to characterise clouds. One example is sea surface temperature (SST) which can be derived from microwave or infrared channels. Uddstrom et al. (2001) presented the possibility of using the SST difference derived from brightness temperatures of AVHRR's channel 4 images to recognise four cloud classes, i.e. stratocumulus, cumulus, thick and transmissive cirrus. However, sensing SST through the atmosphere by means of the thermal infrared is subject to several environmental factors that can degrade its accuracy, such as sun glint, water vapour absorption through the atmosphere, trace gas absorption, and episodic variation in aerosol absorption (Brown and Minnett, 1999). Therefore, thermal infrared measurements of SST require atmospheric correction and can only be made for cloud-free pixels. On the contrary, passive microwave measurements are largely unaffected by clouds and generally easier to correct for atmospheric effects. However, its SST has less accuracy and poorer resolution due to lower signal strength of the Earth's Planck radiation in the microwave region (Castro et al., 2004). In addition, the satellite observation relates to the temperature of the *skin* of the ocean, i.e. at top few microns, for the infrared SST and the *sub-skin*, i.e. at a depth of approximately 1 mm below the sea surface, for the microwave SST. It can be significantly different from the bulk temperature of the traditional in situ measurements which is typically sampled at around 1 m depth.

On the other hand, an atmospheric temperature structure can be exploited in addition to brightness temperature to acquire cloud-top heights to categorise clouds into high, medium and low. For example, Garand (1988) derived cloud heights (η) using the relation between cloud-top temperature (T_c) and sea surface temperature (T_s), i.e.

$$\eta = \frac{T_s - T_c}{6.5}, \tag{2.6}$$

where $6.5°Ckm^{-1}$ is a vertical lapse rate.

More parameters that probably relate to the amount of the rainfall involve geographical location of the events such as latitudes of the centre and the size of the cyclones, pressure gradient, surface winds and atmospheric winds at different levels, heat from the earth surface, seasonal and diurnal changes, cloud base level, humidity,

height of the tropopause and orographic profiles.

2.2 Rain Estimation

There are two common approaches of pattern recognition algorithms to estimate rain rates, as shown in Figure 2.2. First, the problem can be seen as function approximation, i.e. to find a relationship between the features and the rain rates directly. As a result, the estimated rain rate is continuous. However, the relationship between the two data is likely to change with time. This requires the function to be adjusted constantly. The overlapping of the feature space and outliers also demand an appropriate training approach for the approximation. Second is the classification. The feature space is partitioned according to statistical details of clusters which are then labelled to appropriate rain classes. The method is superior to the function approximation in the sense that it can tolerate the overlapping of the features and some irrelevant data that may be regarded as noise and suffer less from a severe nonlinear mapping between the features and the rain rate. However, the estimated rain rate is not continuous, rather it is defined separately by its class membership. For example, a class membership can be defined according to cloud type and then each cloud type is assigned to a specific rain rate. However, there is no assurance a priori that such cloud classes will be separable given the satellite measurements and it is possible that some analyst biases may be introduced into the definition of classes (Uddstrom and Gray, 1996). In addition, each cloud type may not correspond to the same rain rates. Otherwise, it can be categorised according to the amount of rain rates. Due to the limitation of discrete rain estimates, the classes should be well-defined to support all possible rain rates. This implies that the training data should cover all possibilities of the rain rates, which may be impossible in practice.

Next, details of function approximation and classification methods that have been widely used in remote sensing applications are described.

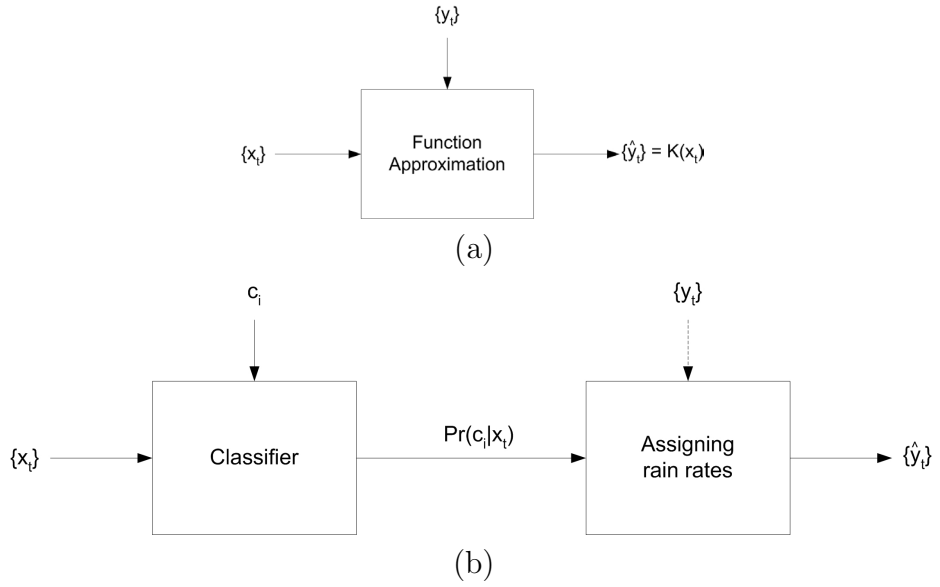


Figure 2.2: Two common ways to estimate rainfall: (a) function approximation and (b) classification. The symbols $\{x_t\}$, $\{y_t\}$ and $\{\hat{y}_t\}$ denote features, actual rain rates and rain estimates, respectively. In addition, c_i is a class in the feature space and $K(x_t)$ is a kernel function of x_t.

For cloud classification, it is always assumed that the feature space satisfies a Gaussian class-conditional probability density function[7] (Wu et al., 1985; Ebert, 1987; Garand, 1988). Then, the boundary function is defined by a linear combination of features which best separate classes. The separation is measured by a distance between the centre of each class with respect to the compactness of the within-class distribution of the features. This classification method is known as the Fisher's linear discriminant (Fisher, 1936).

In many applications, the linear discriminant cannot effectively handle complicated problems such as nonlinear boundaries between classes. In order to learn and find an appropriate boundary function from the nature of a feature space, the classifier should be able to *think* by itself. An artificial neural network (ANN) is a concept of such a thinking machine that has been extensively studied. The original inspiration for this concept stems from bioelectrical networks in the brain formed by neurons and their synapses. For example, a feed-forward neural network, which is the simplest ANN, is composed of these neurons arranged in layers. Each of the neurons in a layer is connected to every neuron in the next layer. The output of the neuron is obtained by

[7]refers to a normally distributed probability that a feature falls into a given class.

a weighted linear combination of the input values, i.e.

$$\hat{y}_j = f(\sum_i w_{ji} x_i + w_0), \tag{2.7}$$

where x_i is the input value, $\hat{y}_j$ is the output of the j^{th} unit, $f(\cdot)$ is a nonlinear function of the weighted summation of the input values to determine how it should activate other neurons in the next layer, w_{ji} and w_0 denote a weight coming from the input i and a bias of the unit, respectively. Typically, there are at least three layers in the network: input, hidden and output. After an arrangement of perceptrons, i.e. a set of neurons and their connection, the network is trained until the weights of the neurons converge to some values that give a solution nearest to the desired output. The output of the network can be a class label for classification or a real value for function approximation.

The most common learning technique for MLPs is the feed-forward backpropagation (FFBP) network. First, it determines the weights of the output layer that minimise the error from the desired output. The error is then propagated back to the previous layer to find out how much error each neuron has produced. The neuron then adjusts its weight to achieve the least possible error. Lee et al. (1990) applied the FFBP neural network to classify Landsat Multispectral Scanner (MSS) visible data into stratocumulus, cumulus and cirrus. The network achieved an overall accuracy of 93%, compared to 85% of the k-nearest neighbour approach and 84% of the linear discriminant analysis. MLPs have been widely used for rain estimation. For example, Grimes et al. (2003) applied a feed-forward MLP to estimate rain rates over Zambia. The features used for training the networks were IR brightness temperature and its first statistics, cold cloud duration (CCD), which is defined as the number of hours for which the pixel temperature is colder than some specified temperature thresholds (213, 223, 233 and 243 K in the paper), principal components of surface horizontal wind velocities, surface relative humidity, and higher-level vertical wind velocities and geometric locations. The paper claimed a better result than standard CCD method which exploits a linear relationship between CCD and rain rates. Alternatively, Tapiador et al. (2004) trained the network by IR brightness temperatures, 2-hour-and-a-half historical measurements and the first-order statistics of cloud texture. An additional feature was the number of pixels with a brightness temperature less than $225K$ to provide information on the heaviest rain areas. The estimation over mid-latitude regions achieved a visual agreement of rain rate histograms when compared to the SSM/I and rain gauge measurements. However, there were no reports of

the significance of the training features towards the networks' performance in either work.

Unlike MLP that tries to find a relation between a desired output and a real value input, a radial basis function (RBF) network gives the output as an approximation of the probability density function of the input. It uses the suitable proportion of the Gaussian mixture of the input vectors, rather than using all feature vectors. That is, the output at the j^{th} unit is a linear combination of the Gaussian mixture,

$$\hat{y}_j = \sum_i w_i \cdot f(x|\theta_i) + w_{i0}, \tag{2.8}$$

where $f(\cdot)$ is a basis function, e.g. a Gaussian function with parameter θ, w_i and w_{i0} are the weight and bias of each Gaussian. Therefore, the input vectors can be represented by a set of mean (location), variance (width) and weighting parameters (significance) of the Gaussian functions. Theoretically, the Gaussian response which decreases monotonically to the centre point is efficient for approximating target functions without being trapped in local minima of the error surface. It has been applied to rainfall tracking (Dell'Acqua and Gamba, 1998) and rainfall prediction (Dell'Acqua and Gamba, 2003) from weather radar images. By considering the weights of the network as the parameters describing the structure of the rain field, the results showed that the network better modelled the nonlinearities of the rain rate series than using linear functions. The network has an advantage of being much simpler than the MLP approach while keeping the major property of the input vectors. This attractive characteristic makes it a widely used method for both function approximation and classification.

For classification, a basis function can be regarded as the class-conditional distribution $\Pr(\mathbf{x}|c_i)$ that links to the output unit with the value $\Pr(c_i)$. The output of the network, as a result, represents an approximation to the posterior probability $\Pr(c_i|\mathbf{x})$. That is,

$$\Pr(c_i|\mathbf{x}) = \sum_i \Pr(c_i) \cdot \Pr(\mathbf{x}|c_i). \tag{2.9}$$

Clearly, a network is required for each class of the data. Moreover, more than one basis function is always needed to represent $\Pr(\mathbf{x}|c_i)$.

Alternatively, a combination of several radial basis function networks can be used to approximate the posterior probability of every class. Such a network is known

as a probabilistic neural network (PrNN)[8], first proposed by Specht (1990). While the RBF network chooses the input pattern that corresponds to a specific class for training, the PrNN weights all the inputs to find appropriate parameters for the basis functions, then select only a set of basis functions to represent a specific posterior probability. The benefit of non-iterative training procedure is that it gives rise to very fast computation. Welch et al. (1992) used 100 bootstrap sample sets[9] to compare an FFBP network, a PrNN and a traditional stepwise discriminant analysis to classify 5 spectral and 15 textural features derived from AVHRR images (channel 1, 3 and 4) into 10 classes, e.g. land, open water, sea ice, stratus over ice, stratus over water, cumulus, and multi-layered cloudiness. The experimental results revealed that the FFBP gave the best overall accuracy of 87.59±1.27%, whereas the PrNN and the discriminant analysis yielded 87±1.00% and 85.57±1.08%, respectively. However, in the paper, the FFBP required 13 hours for training, compared to 3 minutes for the PrNN and 90 seconds for the discriminant analysis. Bankert (1994) applied PrNN to classify AVHRR images (channel 1 and 4) in maritime regions into 5 classes: low clouds, altostratus, high clouds, precipitation clouds and clear. By using 5 spectral and 11 textural features, together with IR cloud fraction and latitudes, the system achieved an accuracy of 90.3±1.00% when applied to 100 bootstrap sample sets. However, the PrNN can possibly become a very large network since every training pattern needs to be stored. This leads to extensive storage and computation.

Radial basis functions can be modified to approximate a nonlinear function by applying a kernel function to each radial basis. As a consequence, the modified method which is known as the generalised regression neural network (GRNN) can complete the approximation in only one pass (Specht, 1991). While PrNN aims to estimate the posterior probability using Gaussian functions, GRNN uses the functions to approximate the response surface as the desired output. The ability to give output instantly makes it attractive and widely used in many signal-processing applications and financial prediction (Huang et al., 2004).

The support vector machine (SVM) has been recently proposed as an efficient method for pattern classification and nonlinear regression. The approach is gaining popularity due to many attractive features. For example, SVMs are computationally efficient by using only a few numbers of selected input features, known as support vectors, to

[8]The probabilistic neural network is commonly abbreviated as PNN. However PrNN has been used in this thesis to avoid any confusion with a recently-proposed pulse neural network.

[9]Bootstrapping is a general purpose validation approach to statistical inference, i.e. each sub-sample set is a random sample set with replacement from the whole sample set. Then, the validation of the algorithm is done by repeatedly analysing each sub-sample of the data.

generate a decision boundary between classes or a surface response for the regression problem. Therefore, they are able to deal with a large number of sample sizes (Cristianini and Shawe-Taylor, 2000). A kernel function can also be applied to the inner product between these support vectors and other input features to create a nonlinear boundary.

A commonly encountered problem in classification is related to the case that some classes are represented by a large number of samples while others are represented by only a few. As a consequence, the minority classes might be ignored, resulting in inaccurate classification especially if the ignored classes are significant. Such an imbalance class problem noticeably occurs in rain estimation where some events such as high rain rates of convective clouds rarely happen. If we assume that, among all types of rain of a mid-latitude cyclone, only 1% of the clouds are convective, the rain estimation can still yield 99% accuracy if the convective clouds are ignored despite their high rain rates which can be more destructive than a large area of light rain. Japkowicz and Stephen (2002) conducted a systematic comparison between three classifiers: a decision tree learning system (C5.0), an MLP and a support vector machine (SVM). It is found that the MLP is less prone to the class imbalance problems than the decision tree system because of their flexibility. However, the SVM appear to be 'absolutely not sensitive' to the class imbalance problem because boundaries between classes are calculated from the support vectors which are the data points located close to the other classes. Therefore, based on the requirement of rain estimation that the algorithm should be able to deal with large amounts of data and tolerate the class imbalance problem, the SVM is a potential candidate for the task either via approximation or classification.

Moreover, the nature of its quadratic cost functions results in a globally optimized solution, in contrast to the neural networks, in which their gradient based training algorithms only guarantee finding a local minima. When applied an SVM with a radial basis function kernel to classify GOES's IR spectral features into 10 cloud types, Azimi-Sadjadi and Zekavat (2000) achieved overall classification rate of 78.5 %.

2.3 Conclusion

Information from satellite images can be extracted in several ways. For example, cloud texture can be quantified by co-occurrence and difference matrices to capture the relation between two intensities in an image. Other texture descriptors of a cloud are its frequency and spatio-temporal components via the Fourier transform and wavelet

transform. Image texture can also be modelled by a stochastic spatial interaction model or a random field model. However, the spatial distributions of cloud particles as a doubly stochastic Poisson process can result in a complication in estimating parameters of the models. Alternatively, a fractal dimension can be used to describe the dimension of a complex object. The fractal concept assumes that the object contains distinct copies of itself. However, a cloud can contain various types of textures and therefore it is impossible to find an exact replica for the calculation.

On the other hand, shape and size can also be used to describe a cloud. However, an ill-defined edge seen in satellite images makes it difficult to derive the true shape and size of individual processes embedded in mid-latitude cyclones. A method that can extract obscure edges of clouds is needed to describe the spatial characteristics.

The information extracted from the satellite images can be used in rain estimation. The estimation can be done in two ways, i.e. function approximation and classification. A number of methods that have been used for cloud applications are analysed. The methods that can be used for rain estimation should be fast in computation and tolerant of uneven distribution of samples.

The review shows that the support vector machine tends to be the most potential method for rain estimation either via function approximation or classification. Using a small subset made up of informative points, namely support vectors, makes it computationally efficient, i.e. capable of dealing with large sample sizes, and tolerant to the class imbalance problem. Therefore, the support vector machine will be used for rain estimation in this thesis. Another method that is selected for the task is the radial basis function neural network, which has notable advantages of computational speed and class imbalance tolerance. In addition, the network can be used for both classification (PrNN) and regression (GRNN) tasks.

Chapter 3

Feature Analysis

A key question in the use of cloud texture measures for rain estimation is whether the textural information provides enough information to estimate rain rate by itself; or whether it provides any useful information at all. This can be done by analysing the correlation between the two data. In Section 3.1, a direct relation between the spectral information of satellite images to rain rates is analysed to find out how the two data mutually relate in the case of a mid-latitude cyclone. The relation is also used as a means for ranking the textural features. An analysis has been done in Section 3.2 on how well the textural features previously reviewed can describe the characteristics of meteorological processes and how these features relate to rain rates. The feasibility of deriving cloud structures using cloud shadow, which is another piece of information from geostationary VIS satellite images, is included in Section 3.3. The techniques of extracting information from satellite images used in this chapter are summarised in Figure 3.1.

3.1 Relationship of spectral features to rain rates

Research on rain estimation from geostationary satellite images generally hypothesises that the images have a correlation to rain rates. It attributes this to the fact that the *spectral features*[1] are dependent on many factors that have direct effects on rainfall. These factors include the spectral distribution of physical properties of clouds, atmospheric characteristics, and the like (Gu et al., 1989). Therefore, the dependency between spectral features and rain rates are analysed in this section to examine this

[1] In this thesis, a spectral feature refers to a particular value received from the satellite at different frequency bands. No information of neighbouring pixels is included, in contrast to the textural information.

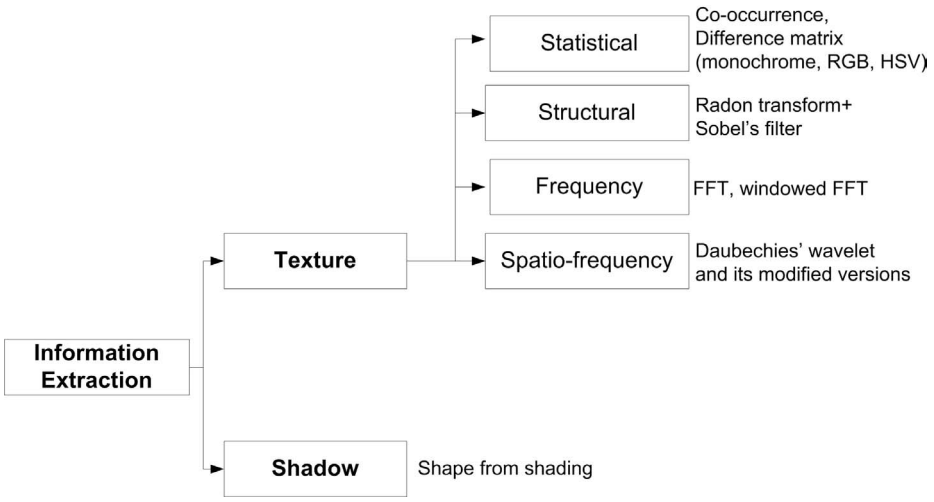


Figure 3.1: Diagram of information extraction used in this chapter.

hypothesis and to determine the extent that rain rates can be deduced directly from the image intensities[2].

Dependency can be examined via the correlation concept. The correlation can provide a clue as to how well the value of one variable can be predicted from the value of the other. A measure of such a tendency is called a correlation coefficient. It denotes the degree of a straight line or a linear relationship between two sets of random variables. Its best known form is the Pearson product-moment correlation coefficient, which is obtained by dividing the covariance of the two variables by the product of their standard deviations, namely

$$r_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}, \tag{3.1}$$

where X and Y are any two random variables, $E(X)$ is the expected value of X and likewise for Y. The correlation coefficient is 1 for a perfect increasing linear relationship and -1 for a perfect decreasing linear relationship. The values in between indicate the degree of linear dependence between the two variables. Haldar and Mahadevan (2000) suggested that the two random variables should be considered to be statistically independent if the correlation coefficient is less than ± 0.3 and perfectly correlated if the value is greater than ± 0.9. However, the interpretation of a correlation coefficient also depends on the context and purposes and should not be observed too strictly. For example, there are situations in which a correlation coefficient of two variables is zero in spite of a strong nonlinear relationship that exists among them. The significance

[2]The image intensities refer to calibrated radiance values/temperatures of Meteosat's channels.

of a correlation coefficient of a particular magnitude also depends on the size of the sample from which it was computed. For instance, a small correlation can be proven significant in large samples.

The correlation between an infrared and a radar image of the mid-latitude cyclone on 23 June 2004 at 0030 UTC is illustrated in Figure 3.2 (a). The correlation coefficient at each pixel is calculated from a window of 21×21 pixels in size and is presented in colour from blue to red. The corresponding value of each colour is given in Figure 3.2 (d). The white patches in the figure indicate either where no rain radar information is available or where all pixels in the window have zero rain rates, which result in a zero denominator in Equation (3.1). From the illustration, a band curling inward the centre of the cyclone with moderate rain rates appears to have positive correlation (0.4 to 0.6) with the IR temperature. The band corresponds to the frontal rain which usually develop during the mature stage of cyclones. On the contrary, at the front and rear edge of the rain band and in some locations with heavy rain, the correlation is strongly negative, i.e. -0.8 to -0.6. Figure 3.3 shows the positive and negative relations between the two information at the position **A** and **B** in Figure 3.2 (a) using scatter plots. The rain rates are linearly related to the IR temperatures at the position **A** where the rain rates are likely to be influenced by orographic factors of the Lake District. In contrast, the relation to rain rates is exponentially decreasing with the temperatures at the position **B**. When considering the whole image, these strong relations are evened out. The correlation coefficient of the whole image drops to -0.001. The coefficients of the time series of the cyclone between 22-23 June 2004 presenting in Figure 3.4 also demonstrate this effect. That is, the correlation between the rain rates and the satellite information only varies within ± 0.15. With this small correlation coefficient, the two information may be regarded as independent variables according to Haldar and Mahadevan (2000). This implies the difficulty to relate spectral information directly to rain rates.

However, it is noticed in Figure 3.2 (c) that the similar patterns to those of the rain rates can be observed in the IR image e.g. the shape of the precipitation area and the cold clouds at the centre of the cyclone and the fibrous texture at the outer edge of the cyclone in both radar and IR images. This effect is quite clear from visual inspection even though the overall correlation is low. Therefore, these resembling texture patterns in the satellite images should be able to help in estimating the rain rates. The textural information of the satellite images is investigated in the next section.

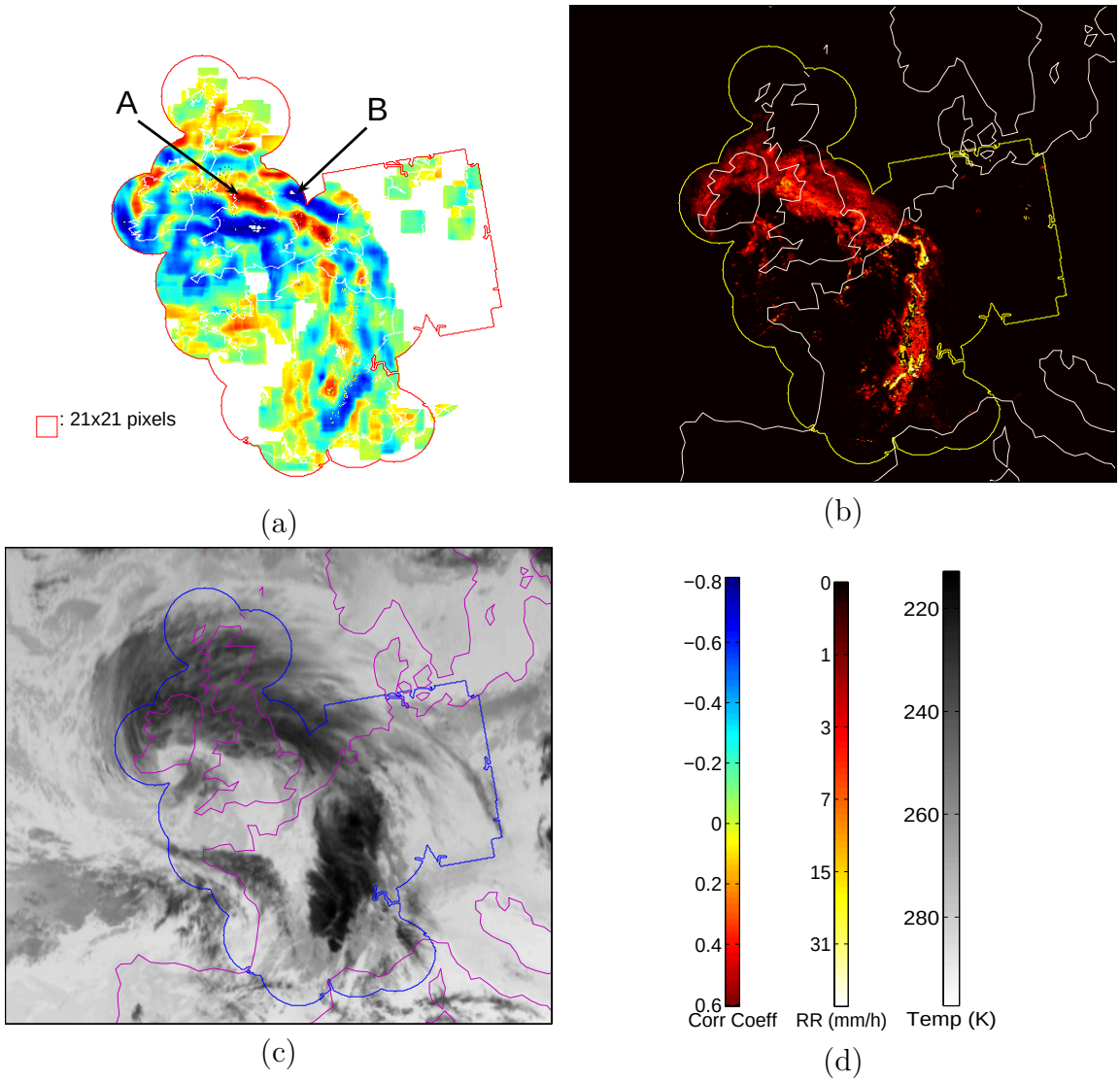


Figure 3.2: (a) Correlation coefficients between a rain radar image (b) and an infrared image (c) of a mid-latitude cyclone on 23 June 2004 at 0030 UTC, calculated using a window of 21 × 21 pixels. The size of the window is shown as a square at the bottom left of the image (a). The red, yellow and blue solid lines indicate areas where the rain radar data is available. The colour bar of the three images are presented in (d).

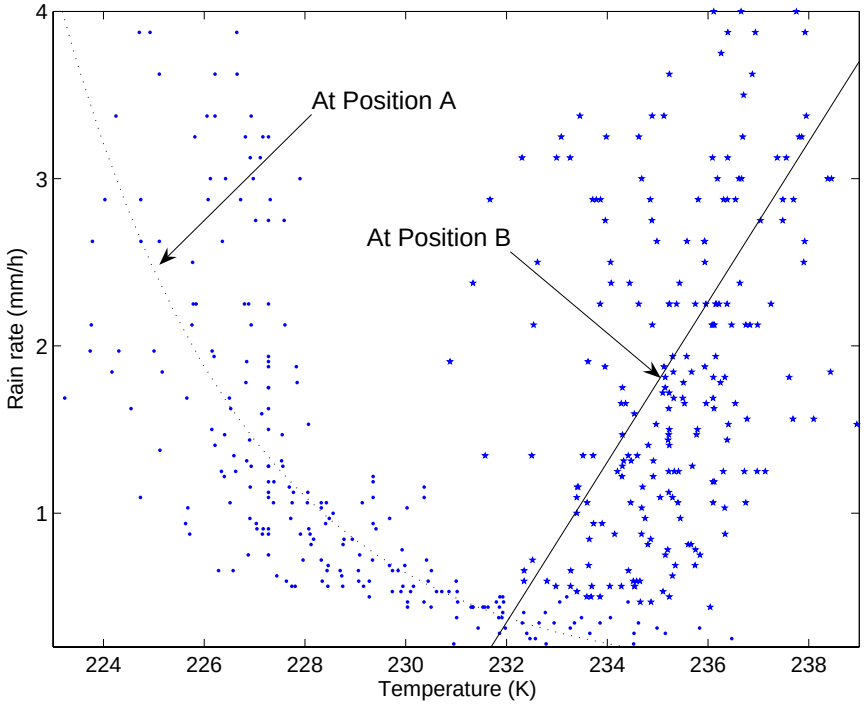


Figure 3.3: Scatter plots between IR temperatures and rain rates of area **A** and **B** in Figure 3.2.

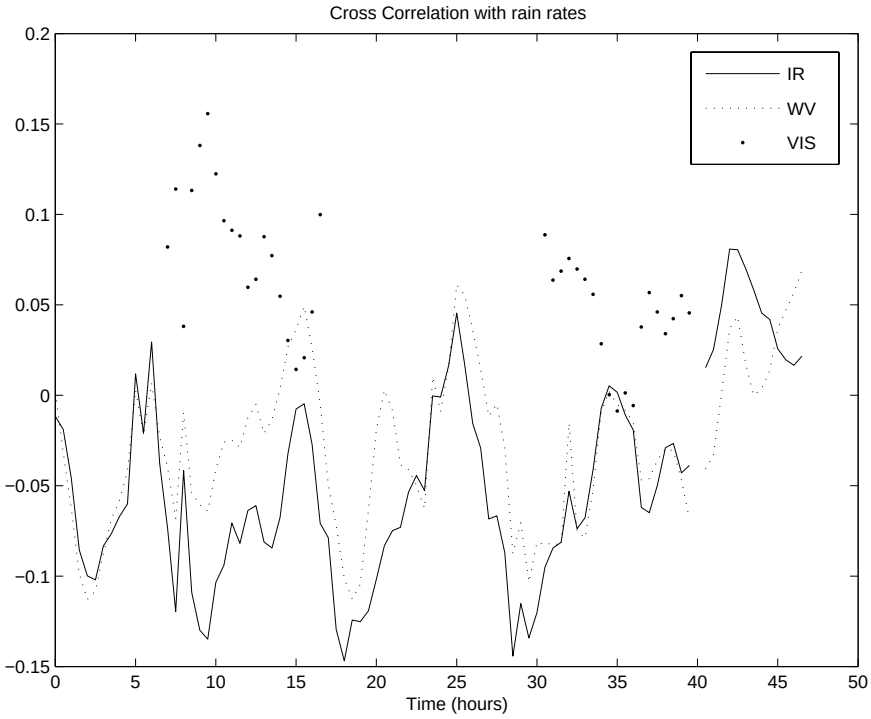


Figure 3.4: Correlation between the three channels and the rain rates for a mid-latitude cyclone on 22 June 2004 with the time line starting at 0030 UTC.

3.2 Textural features

Quantifying textural information can be done in several ways. Besides, the technique that is the best for one application might not necessarily be the best for another application. Difference in locations and environments of the events can also affect the performance of each textural feature. For the case of rain estimation, the rotation or the size of cyclones, including noise that can occur during the operation should not affect the texture parameters. Moreover, the parameters should be distinct for the interpretation and classification purpose. The required computational time should also be taken into consideration as there are many images to be processed. Nearly real-time processing is also demanded on account of frequent generation of the images every fifteen minutes by the Meteosat second generation satellite. In this section, several texture extraction methods are investigated in order to find the appropriate quantitative forms of the cloud texture that relate to the rain rates. These are generally categorised into four types: statistical, structural, frequency analysis, and spatio-temporal analysis.

The images corresponding to common meteorological processes in mid-latitude cyclones are selected for the study of the texture extraction. Here, the four events which include convective, stratiform, vortex-central and front will be used to refer to these images. The test images of the three channels and the corresponding rain rates are presented in Figure 3.5. The size of these four events are approximately $150 \times 150\ km^2$ for the convective, the stratiform and the front, and $500 \times 500\ km^2$ for the vortex-central. The different events are selected to study the classification performance of each textural descriptors at the same scale. However, the size of the vortex's central needs to be enlarge to cover the spiral shape at the centre. All images are resampled to 50×50 pixels on the polar-stereographic projection. The grey levels in the IR and WV images correspond to cold clouds for dark shades and warm clouds for brighter shades. The rain rates are shown in colours, ranging from white for heavy rain, yellow and red for moderate and light rain, to black for zero rain rates.

From the illustration, it can be seen that the four events possess distinguishing textures. The convective cloud appears as an individual cloud object. It is usually colder than its surroundings and brighter when seen in the VIS image. The corresponding rain rate is heavy and similar in shape to the cloud in the satellite image, though not as smooth. This is due to the fact that the satellite images have coarser resolution and need interpolation in order to match that of the rain radar images. Generally, higher resolution data should be averaged to the resolution of the lower. This would

be applicable to the case of rain radar images if the rain information at each pixel has related to its neighbours. However, there always are cases where rain in one area is not dependent to its neighbours due to surface conditions. Under-sampling the rain information will certainly introduce errors to the data. On the other hand, clouds are considered to be more correlated to its neighbours, making interpolation become more appropriate for this task.

By observing the images, a mesoscale convective storm that is well shaped in an IR image may become distorted and noisy in the radar image as the large region is affected by many factors at the ground surface. Size of the events also affects the pattern of the brightness temperature and the corresponding rain rate. For example, a convective cell that can be depicted by a rain radar image is usually $5 - 15\ km^2$ in size and so is too small to be delineated in the IR image. For the stratiform cloud, its gradient in the IR and WV images is hardly noticeable. However, it is observed that there is a diagonal line that appears in the VIS image that coincides with the rain band in the radar images. Contrary to other events, there is only slight rain at the centre of the cyclone, though the VIS image reveals that the cloud in that area is thick—enough to provide heavy rain in other circumstances. The rain is well spread only over the cold area that appears dark grey in the IR and WV images. In contrast, rain in the frontal cloud band agrees with the thick cloud in the VIS image in general. The IR and WV images, although possess the same diagonal band, have a shift in position when compared with that of the rain radar images; half of the rain band lies on the cold cloud and the other half on the warm cloud.

Figure 3.6 shows the relation between the IR intensities and the corresponding rain rates for each event via the probability distribution. It can be seen that a wide range of IR intensities give the same rain rate. On the contrary, the rain rates of the stratiform cloud varies within a narrow range of IR temperatures. Both vertical and horizontal stretchings of the distribution are found in the case of the frontal cloud. This is the effect of the rain band that lies across both cold and warm bands in the IR image that is previously described.

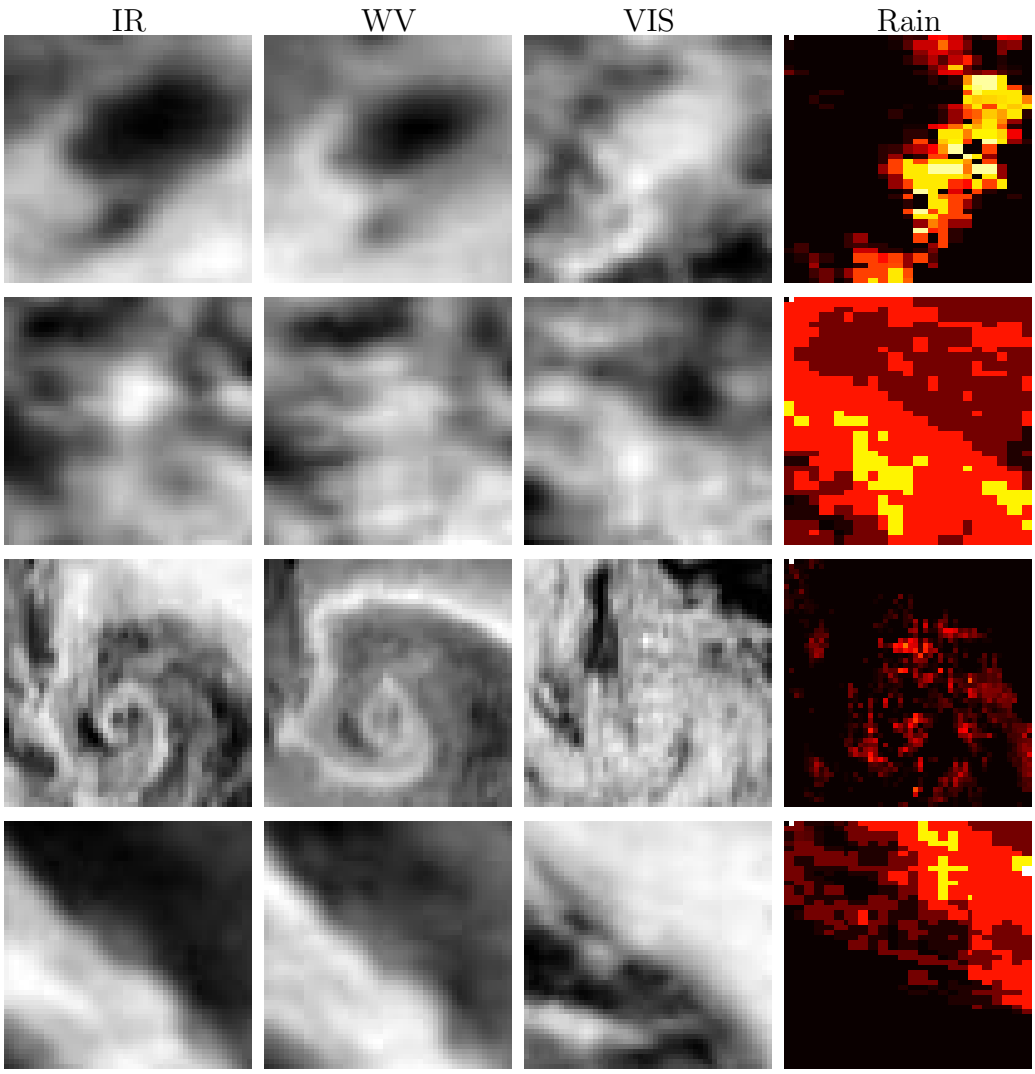


Figure 3.5: Samples of images for the study on texture. (Top→bottom): convective (150×150 km), stratiform(150×150 km) , vortex-central (500×500 km) and frontal clouds (150×150 km). Each row corresponds to the same event. Original source: a mid-latitude depression on 23 June 2005 at 1030 UTC from Meteosat7.

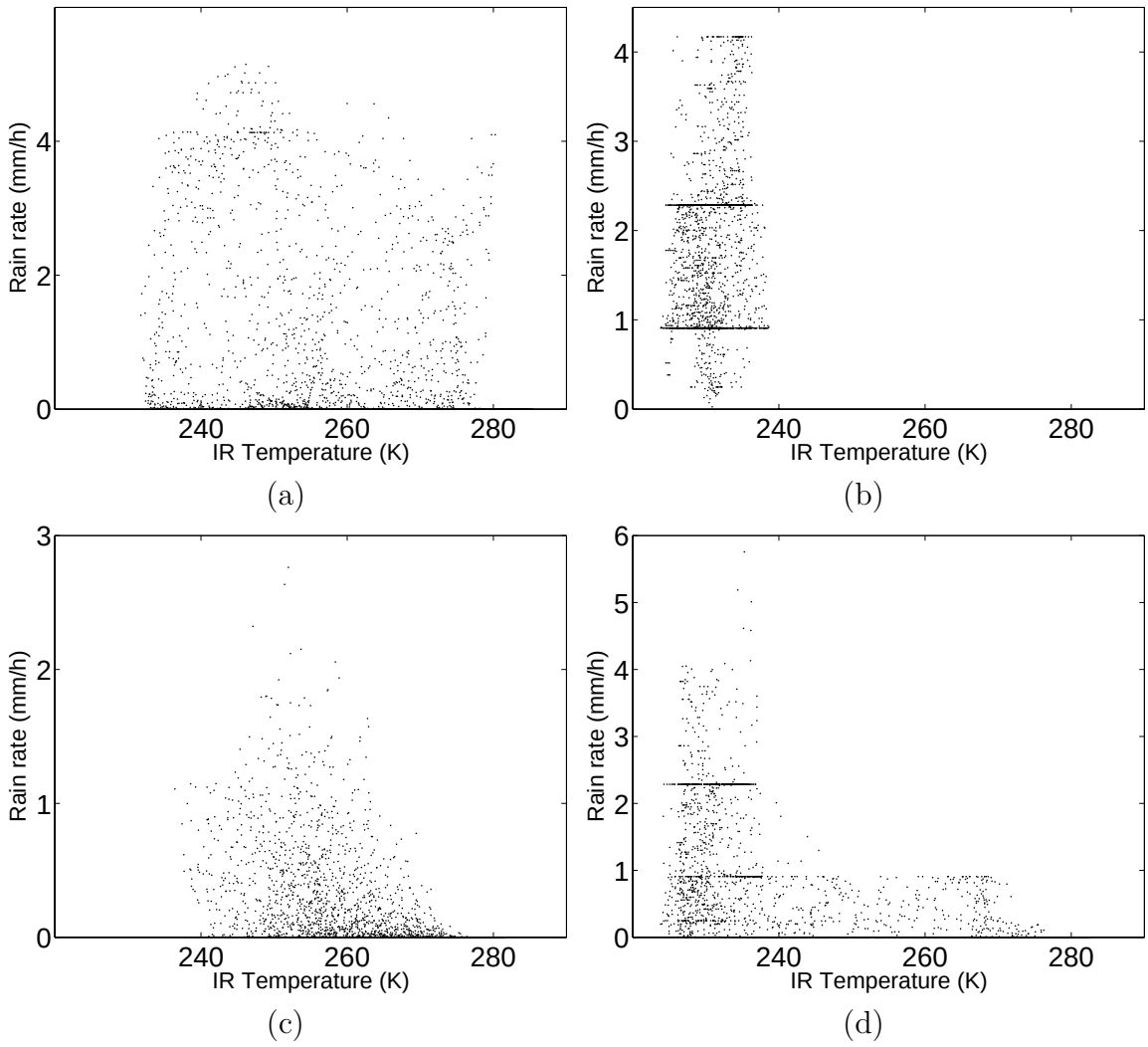


Figure 3.6: Distribution of rain rates versus temperatures from IR channels for (a) convective (b) stratiform (c) centre of the cyclone and (d) front.

Image registration error?

When comparing the satellite image and the radar image for the convective cloud, it is possible that there might be errors from registration of the two images. However, as shown in Appendix B, the two images are finely matched with a displacement error not exceeding 1 pixel when considering coastlines and latitude-longitude grids. For clouds, there can be displacement errors owing to viewing angles of the satellite. These errors increase with the distance between the considered location and the 0° latitude and 0° longitude coordinate. Figure 3.7 shows that the errors are approximately 12 km for a 15-km-height cloud at as far as 70° latitude and 30° longitude. This means that the displacement errors should be within 12 km or 4 pixels for the test images.

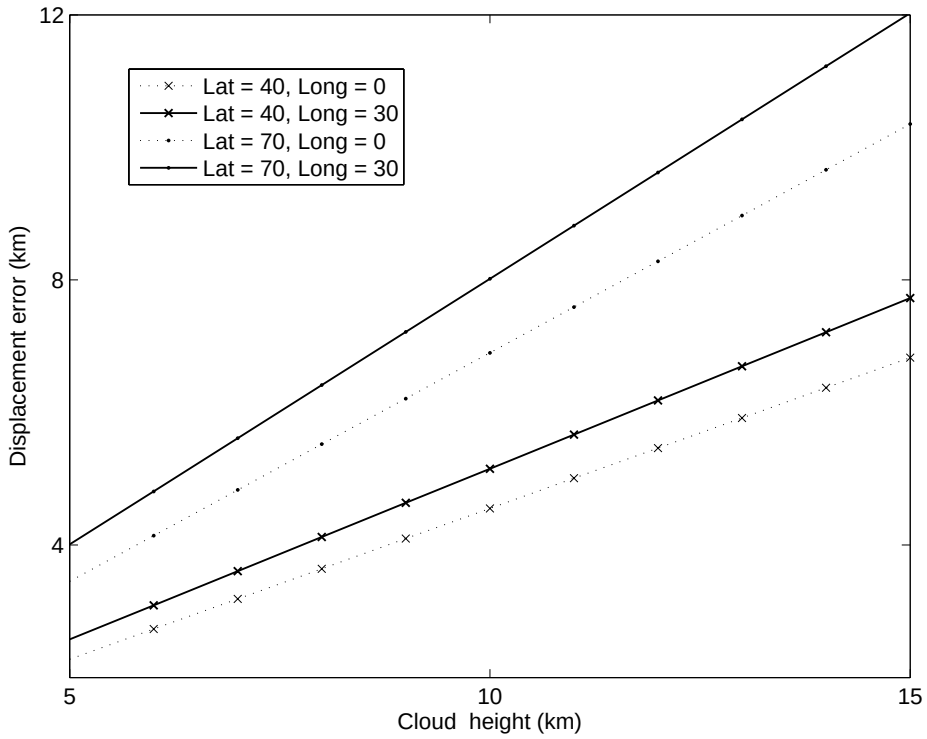


Figure 3.7: Displacement errors due to viewing angles of the satellite as a function of cloud heights.

In addition, considering a sequence of the problematic convective cloud with a rain contour (of rain rates more than 1.5 mm/h) superimposed on it in Figure 3.8, there seems to be a good correspondence between the rain rates and the clouds—the right edges of the cloud generally correspond with the rain contours. However, the rain rates behind these edges disappear. There are three possibilities to explain this.

First, the cloud might pass its mature stage and begins to die down, producing less or zero rain rates. However, the comparison between the averaged temperature of

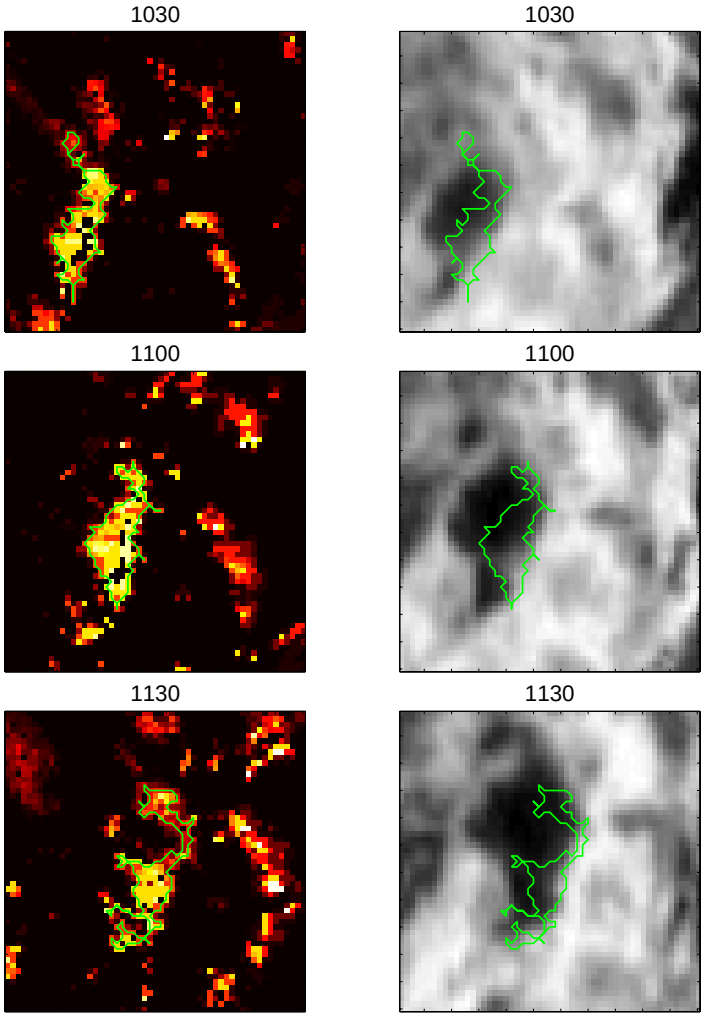


Figure 3.8: A sequence of (a) rain radar images and (b) infrared images of the convective case on 23 June 2004. The green line is a contour of rain rates higher than 1.5 mm/h. The time interval is half an hour between each image.

the cloud and the averaged rain rates within the rain contour in Figure 3.9 does not support this assumption. Despite the colder temperature which implies that the cloud is growing, the rain rate decreases. The further observation found that the cloud gets warmer after 1400 UTC which is three hours and a half later from the first image of the sequence. This implies that the cloud during 1030-1230 UTC in Figure 3.8 is still at its early stage and, theoretically, the colder cloud at 1230 UTC should produce heavier rain than at 1030 UTC.

The second possibility is that heavier rain exist behind the right edge of the contour but these could not be depicted by the radar. The location of the cloud was at the coastline of the Netherlands. Therefore, it is presumed that the radar station would be

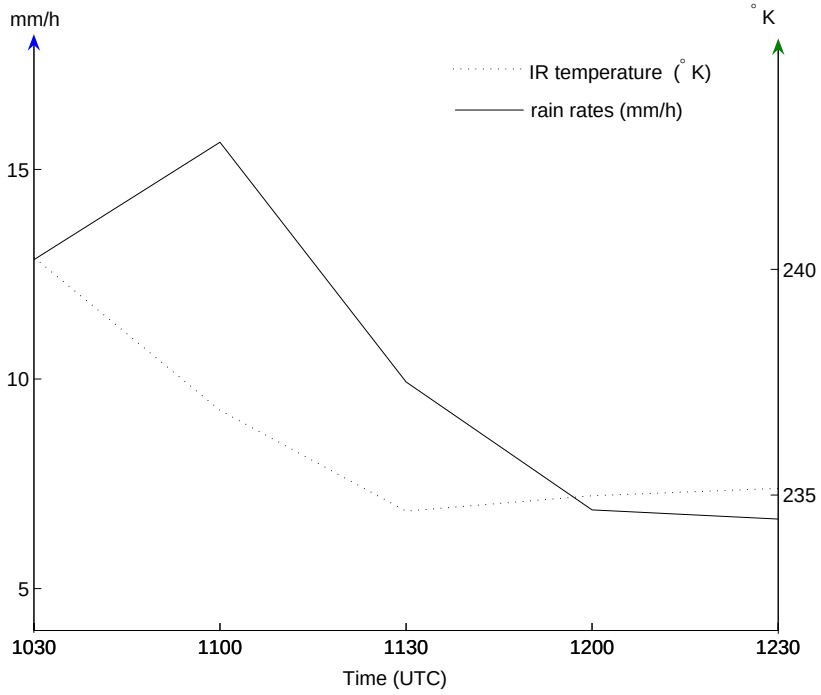


Figure 3.9: A comparison of averaged temperatures of the cloud object and averaged rain rates in Figure 3.8.

on the east or south-east side of the cloud. It is possible that the radar signal from the station might not be able to penetrate through heavy rain, resulting in underestimated rain rates behind the right edge of the cloud. Another possibility is that the cold cloud is being blown off at the cloud top to the north and west, resulting in the mismatch between the observed rain rates at the surface and the temperature at the top of the cloud.

Although, rain information from the radar stations is not a perfect ground truth for rain estimation, it is the best rain information, covering full spatial and temporal resolutions, that is available. It is good also to use the selected convective cloud image to study the texture information relative to the deficiency of radar information as this can commonly happen for the case of heavy, localised rainfalls.

Next, four types of textural features which are statistical, structural, frequency and spatio-frequency will be studied. These four types are selected to cover all characteristics of image texture.

3.2.1 Statistical texture

The first type of textural features used in the study is based on the principle of using statistics to describe the inter-pixel relationship in an image. An example of this approach is the use of a co-occurrence matrix. It is determined by a matrix of frequencies $C_d(p,q)$, which indicates how frequently two pixels with grey levels p and q and separate by a given distance d (in pixels) appear in an image. As shown in Equation (2.1), the co-occurrence can be defined as

$$C_d(p,q) = \Pr\{I(i,j) = p \quad \text{and} \quad I(i+\Delta i, j+\Delta j) = q\},$$

for a relation between any pixels with a displacement $d = ||(\Delta i, \Delta j)||$ between the pixel (i,j) and $(i+\Delta i, j+\Delta j)$. The displacement is defined as $\sqrt{(\Delta i)^2 + (\Delta j)^2}$ for the Euclidean distance. The argument $I(i,j) = p$ denotes the occurrence that the intensity at position (i,j) is equal to p in value.

The co-occurrence matrix reveals certain properties about the spatial distribution of the grey levels in an image. Examples of textural features at each displacement d that can be computed from the co-occurrence matrices are Angular Second Moment (ASM), entropy, contrast, homogeneity and correlation. These five statistics are commonly used because of their sufficiency to give good results in a classification task (Conners and Harlow, 1980). The statistics are defined as

$$\begin{aligned}
\text{ASM} &= \sum_p \sum_q C_d^2(p,q), \\
\text{Entropy} &= -\sum_p \sum_q C_d(p,q) \log C_d(p,q), \\
\text{Contrast} &= \sum_p \sum_q C_d(p,q) \cdot (p-q)^2, \\
\text{Homogeneity} &= \sum_p \sum_q \frac{C_d(p,q)}{1+|p-q|}, \\
\text{Correlation} &= \sum_p \sum_q C_d(p,q) \frac{(p-\mu_{C_p})(q-\mu_{C_q})}{\sigma_{C_p}\sigma_{C_q}},
\end{aligned}$$

where μ_{C_p} and μ_{C_q} are the row- and column-average of the matrix and σ_{C_p} and σ_{C_q} are the corresponding standard deviations. Whereas ASM describes the regularity of texture patterns, entropy gives a measure of the degree of disorganisation of the image. It yields the highest value of 1 when all elements in the matrix are equal, which means that all the intensities appear evenly within the image. Contrast uses the difference

of intensities as a weighting parameter for the probabilistic distribution. It increases when large differences of intensities occur with high probability. On the contrary, homogeneity uses an inverse of the differences to signify the similarities of intensities. Finally, correlation measures the uniformity of the distribution of the matrix. All statistics range from 0 to 1, except correlation which is 1 or -1 for a perfect positive or negative correlation between two intensities. Next, the robustness of these statistics on 4 aspects: scalability, gradient, direction and noise will be tested on synthetic images.

First, a circular object of different sizes is generated as shown in Figure 3.10 (a)-(e). The statistics corresponding to each image as a function of a displacement between two pixels (d) are presented underneath. The size of images (a) to (e) is 50×50 pixels, equal to that of the test images. It can be observed that all features except entropy vary noticeably with object sizes. However, contrast, homogeneity and correlation are less sensitive to the smoothness when the images vary from sharp to soft gradients as shown in Figure 3.11. Considering the object that changes both in size and intensity gradient, Figure 3.12 gives a clear evidence that the values of the statistics are more dominated by the sizes of the object. When the method is applied to an object that is rotated by 30°, 60°, 90°, 120° and 150° as shown in Figure 3.13, the statistics of each rotated angle are almost the same. Therefore, the feature derived by this method is a very powerful rotation invariant measure of local image texture. It is also noticed that contrast and homogeneity of asymmetric images in Figure 3.11 are different to those of other symmetric ones where both statistics have peak values at d equals to the size of half-length of the image. Accordingly, contrast and homogeneity can also be used to reflect the symmetry of the texture.

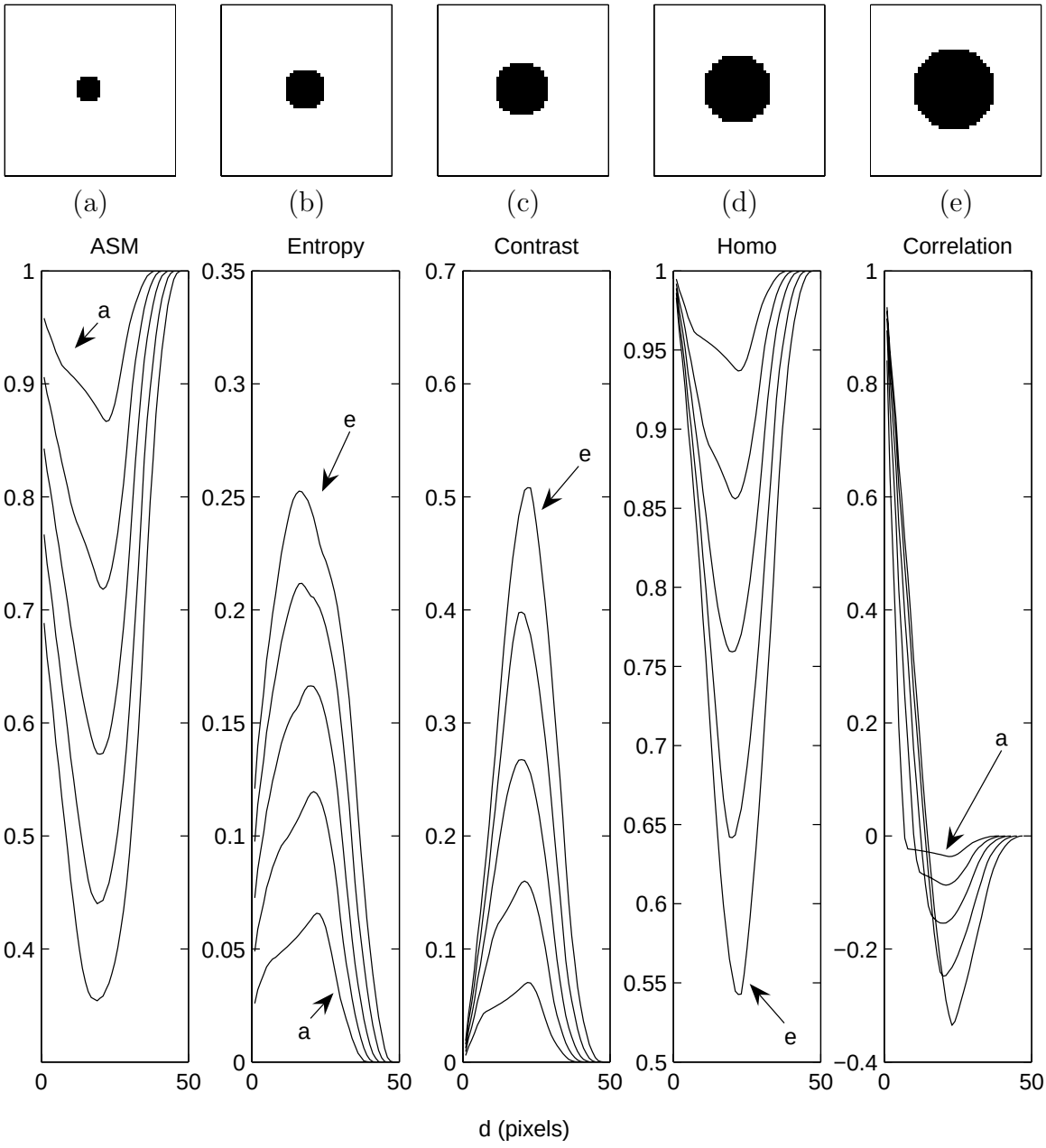


Figure 3.10: Co-occurrence statistics of objects with different sizes.

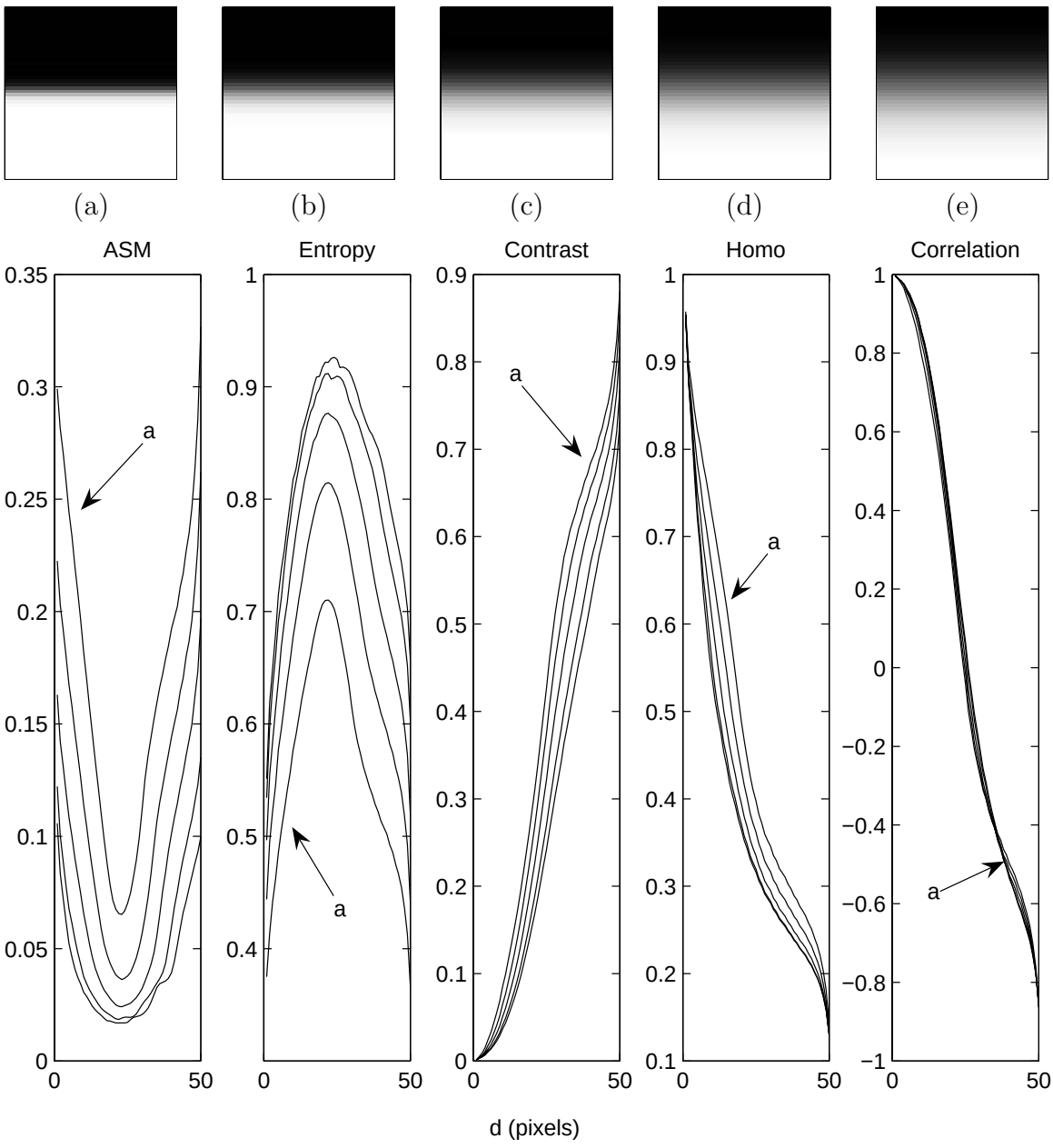


Figure 3.11: Co-occurrence statistics of objects with different gradients.

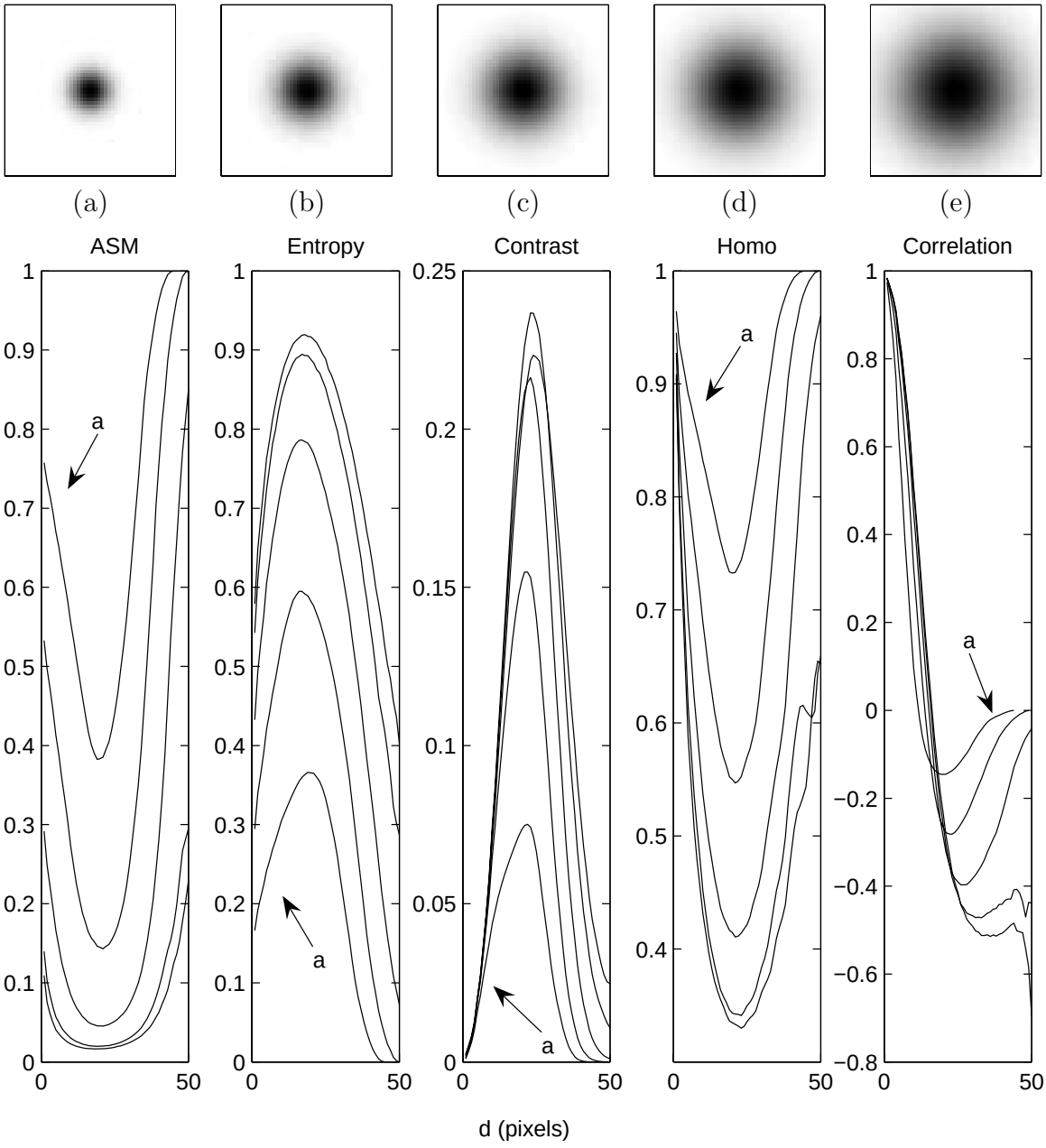


Figure 3.12: Co-occurrence statistics of objects with different sizes and gradients.

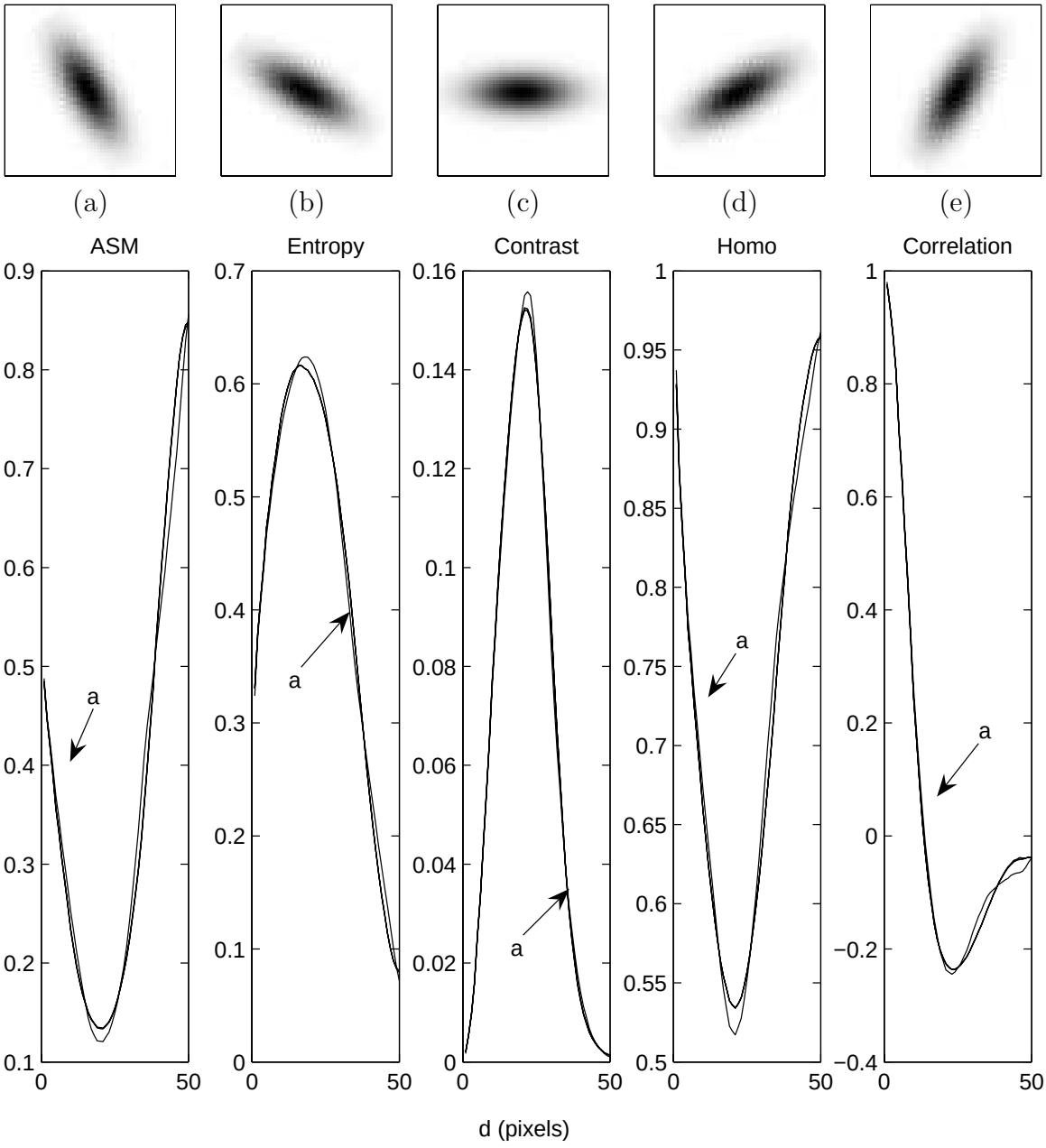


Figure 3.13: Co-occurrence statistics of objects in different directions.

Figure 3.14 illustrates the sensitivity of the statistics to additive zero-mean Gaussian noise, when monitoring at a pair of pixels with 25 pixels apart, of 20 iterations[3]. The noise ratio (n_i) is calculated by $\frac{\sigma_n}{\sigma_i}$, where σ_n and σ_i denote the standard deviations of noise and the image, respectively.

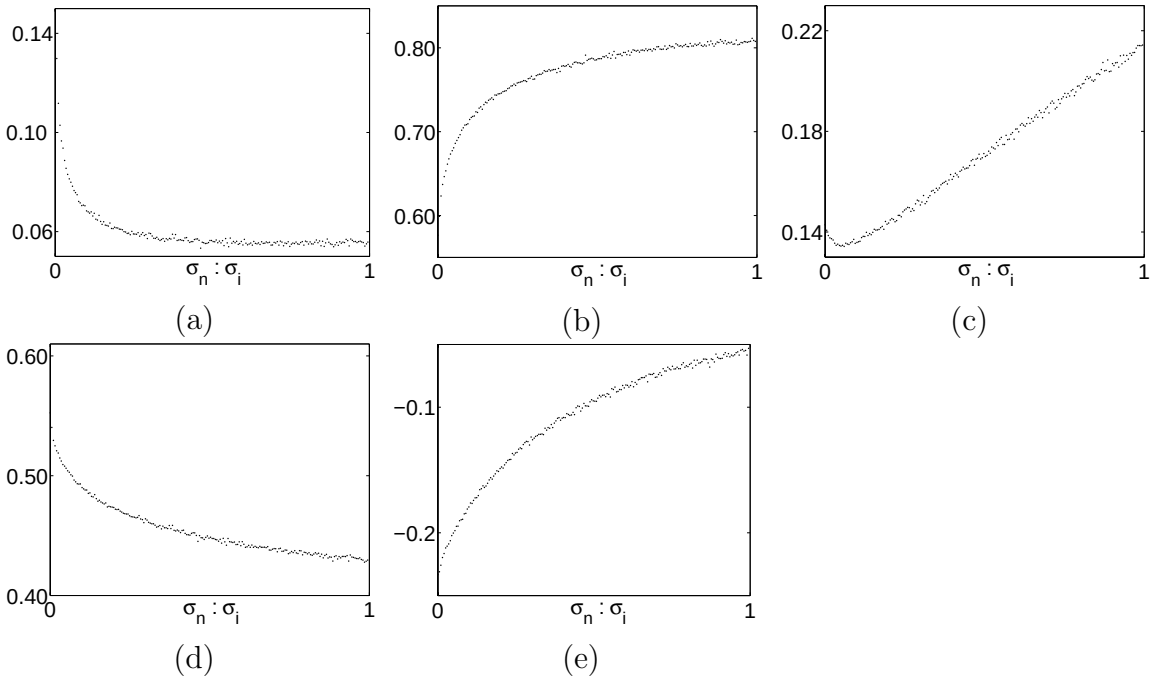


Figure 3.14: Co-occurrence statistics: (a) ASM, (b) entropy, (c) contrast, (d) homogeneity and (e) correlation of the object in Figure 3.13 with respect to noise and signal ratio.

From the illustration, adding an amount of noise introduces an independence between each intensity and therefore results in a decrease in ASM, homogeneity and correlation. On the contrary, entropy increases. Increasing further noise ($n_i > 0.15$) does not affect the values of ASM, entropy or homogeneity. These statistics stay almost constant. Contrast and correlation are the most sensitive to noise—contrast continuously increases whereas the absolute value of correlation decreases when more noise is introduced. The effect can be described as follows. When introducing noise, the distribution of intensities becomes more spread. An increase in noise pushes the distribution away from the diagonal to the edge of the matrix. The intensities become more independent of each other and therefore their correlation between each others decreases. This also results in an increase in contrast which intensifies the components

[3]For each iteration, a new zero-mean Gaussian noise is generated and added to the image then the statistics are calculated. The result in Figure 3.14 is an expectation of the statistics of all iterations.

off the diagonal line. On the other hand, a small drop in ASM and homogeneity is a result of the decreasing near-diagonal values and the increasing of off-diagonal values.

Typically, p and q in Equation (2.1) are defined as ranges of the intensity or bins that the intensity falls into e.g. $p = \{0 \rightarrow 1, 1 \rightarrow 2, \dots\}$. However, numbers of the bins, which determines the size of the co-occurrence matrix, also affects the value of the statistics. The result when using different numbers of bins, i.e. 10, 20, 30, 40 and 50 is illustrated in Figure 3.15. ASM and homogeneity which so far are the most robust features to the changes in size, gradient, direction of the objects and least sensitive to noise change noticeably in values. Both statistics increase when the number of the bins decreases. Using a smaller number of bins (large bin width) increases the probability for the intensities of any two pixels falling in the same category, giving a similar effect as if the texture is smoother and therefore resulting in a larger value of homogeneity and ASM. Therefore, it is recommended that the number of bins is fixed when analysing the texture.

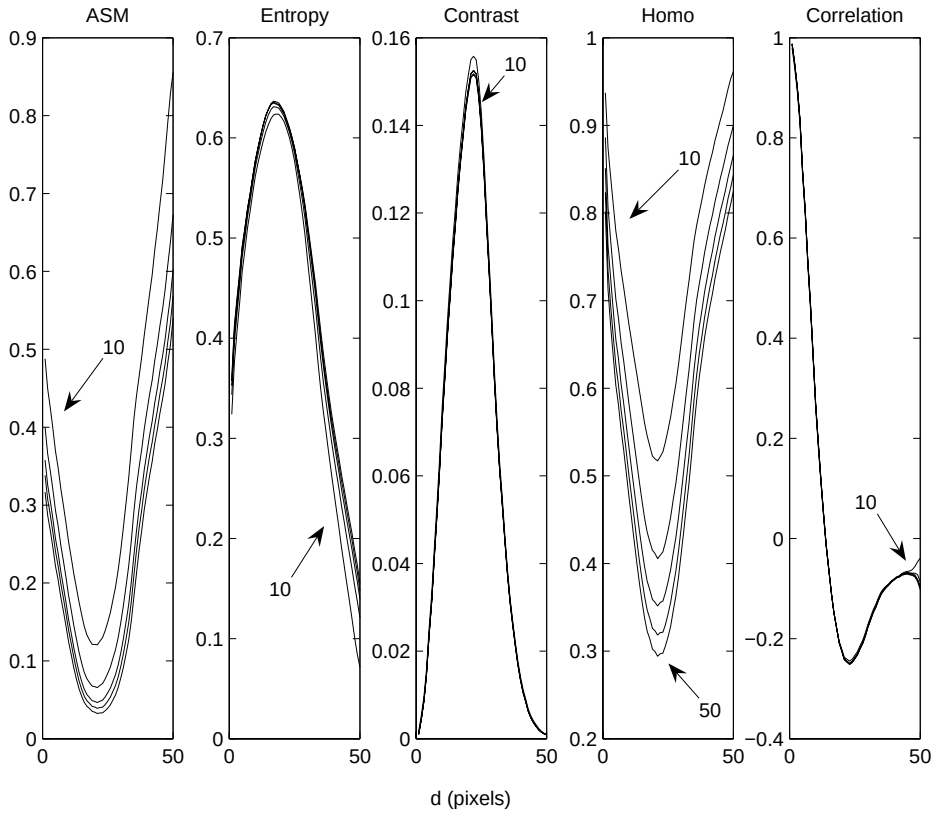


Figure 3.15: Co-occurrence statistics of the object in Figure 3.13 when using different numbers of bins.

Next, the real cloud images of the mid-latitude cyclone in Figure 3.5 are used to study how well the co-occurrence statistics can depict the texture of these meteorological events. First, all images are normalised so that the value of each pixel falls within [0,1]. Then, the co-occurrence matrices of the four events are computed. Figure 3.16 shows the statistics of these matrices when 16×16 bins are used. The legends a, b, c and d refer to convective clouds, stratiform clouds, clouds at the vortex centre and frontal clouds. These legends will be used throughout this section.

For the satellite images, their entropy reveals that the convective cloud is the most disorganised—there is no obvious relation between any two pixels in the image as its entropy is higher than that of other events. However, this is different from the statistics of the rain radar images which show that stratiform is the most disorganised than other events. On the contrary, the sharp gradient of the frontal cloud can be well captured by contrast of the satellite images. Its relationship to the distance between pixels also resembles that of the rain radar images. Such a similarity can also be found in the case of the convective cloud. Likewise, homogeneous and organised clouds at the centre of the cyclone also corresponds to the observed rain rates. This is reflected by homogeneity and entropy. From the results, ASM shows no differences between each event in contrast to contrast which is the best feature to distinguish the four events. More statistics that can well separate the four events are entropy of the IR images and homogeneity and correlation of the VIS images.

For the rain radar images, all statistics except correlation can well distinguish the four events. This is noticeable for the entropy which the four events are well-separated, similar to the entropy of the IR images although their arrangement of the four events are not the same. In addition, their correlation values resembles those of the VIS images. That is, there exists the similarity between the texture of cloud images and rain radar images.

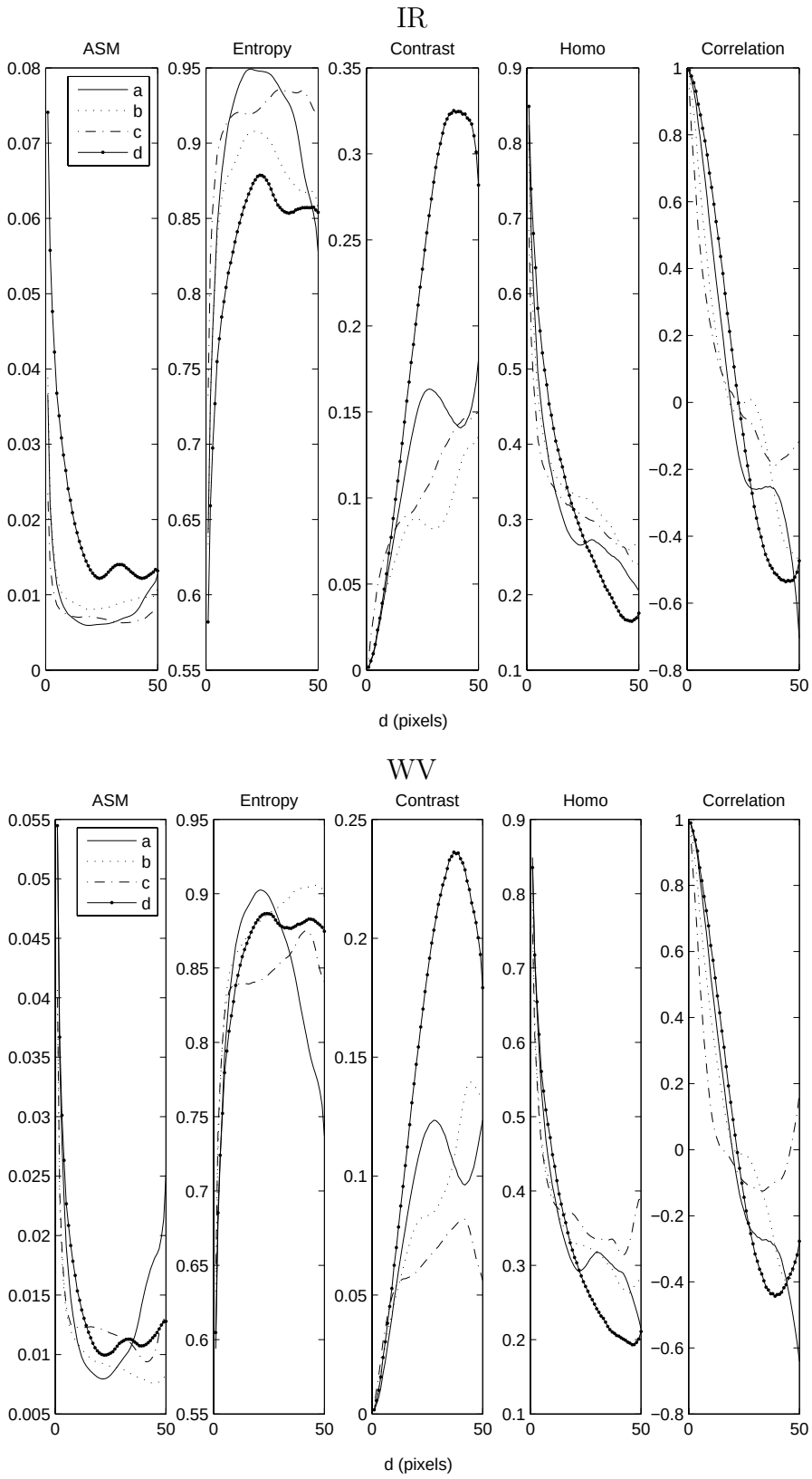


Figure 3.16: Co-occurrence statistics of IR, WV, VIS and rain radar test images. The legends a, b, c and d refer to convective clouds, stratiform clouds, clouds at the vortex centre and frontal clouds.

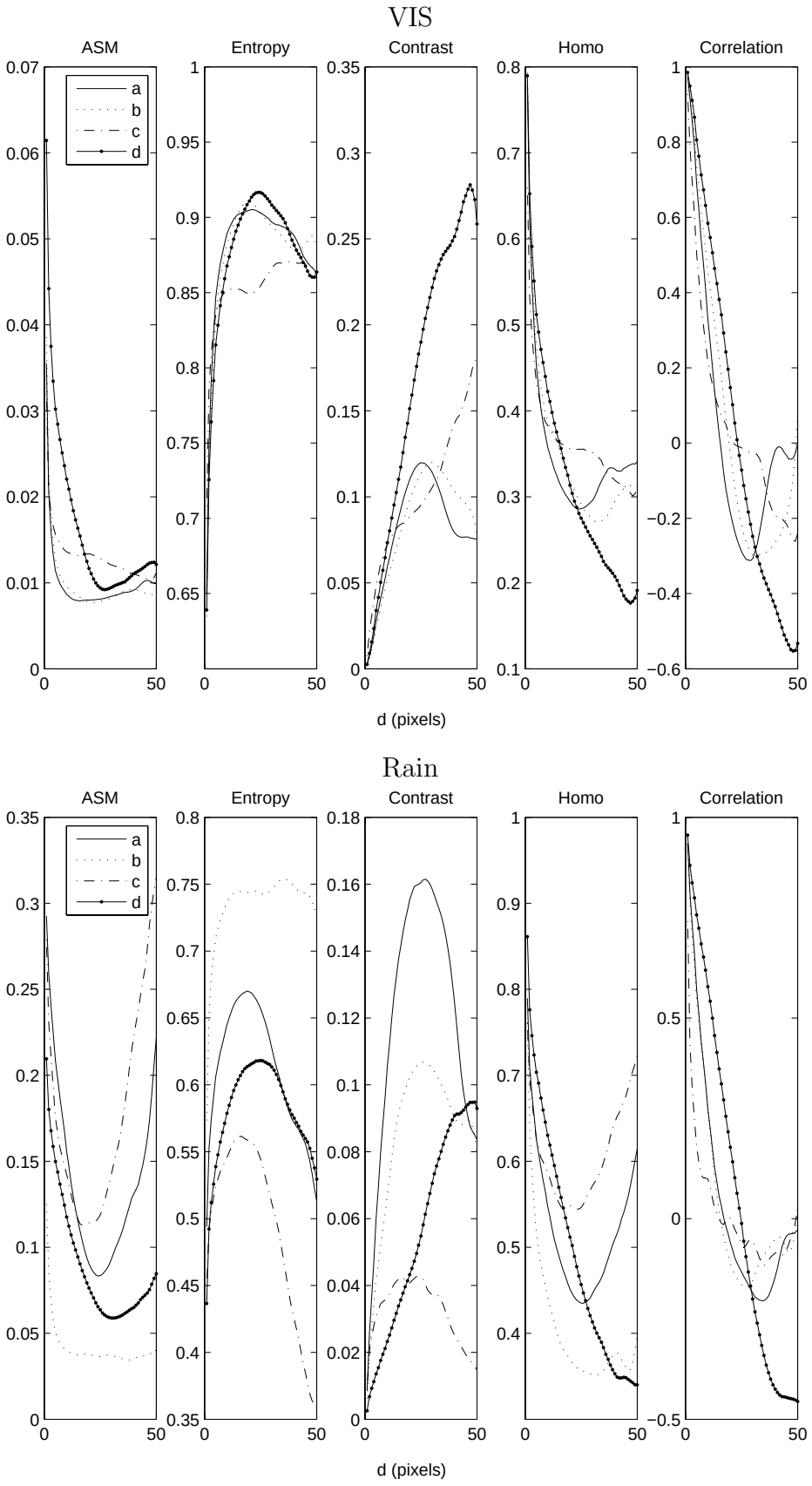


Figure 3.16 continued.

As discussed earlier, the co-occurrence matrix depends on the number of grey levels in the entire image. In addition, the matrix frequently results in zero entries of the whole rows and/or columns in the matrix, wasting memory in the computational process (Svolos and Todd-Pokropek, 1998). This is the major drawback of the co-occurrence matrix. However, the number of grey levels can be reduced by grouping into bins so that there are less zeros in the matrix. However, it should be large enough to be able to distinguish details of the texture. That is, an appropriate number of bins is required.

Grouping the grey levels can be done by using the same process as computing an image histogram. Several methods to define the optimal number of bins for a histogram have been proposed.

Some methods select the bin number for the calculation of a histogram by considering only from the number of samples (Cellucci et al., 2005) such as $1 + 3.3\log(n)$ (Sturgis' Rule, 1926), $1.87(n-1)^{0.4}$ (Bendat-Piersol's Rule, 1966), $\sqrt{n}$ (Mosteller-Turkey's Rule, 1977), and $(2n)^{\frac{1}{3}}$ (Terrell's Rule, 1983) where n is the number of data samples.

The distribution of the samples can also be used as a guidance to select the bin number such as $3.49\sigma n^{-1/3}$ (Scott's Rule, 1979) where σ is the standard deviation. It is suitable for Gaussian samples but might lead to oversizing bin widths. Another rule of thumb is to choose $2(IQR)n^{-1/3}$ (Freedman-Diaconis's Rule, 1981), where IQR is the interquartile range of the data, as a number of bins. It gives a narrower bin widths but often lead to the same choice of the bin width (Izenman, 1991).

Figure 3.17 compares these criteria in terms of the number of samples from a Gaussian distribution with a zero mean and a unit standard deviation. These numbers can be used as a guidance for selecting a bin size so that it is neither too coarse to capture the details of a texture nor so fine that there are many zero entries in a co-occurrence matrix. For example, according to Figure 3.17, an image of size 21×21 pixels needs 9.59-21.34 bins to represent its histogram.

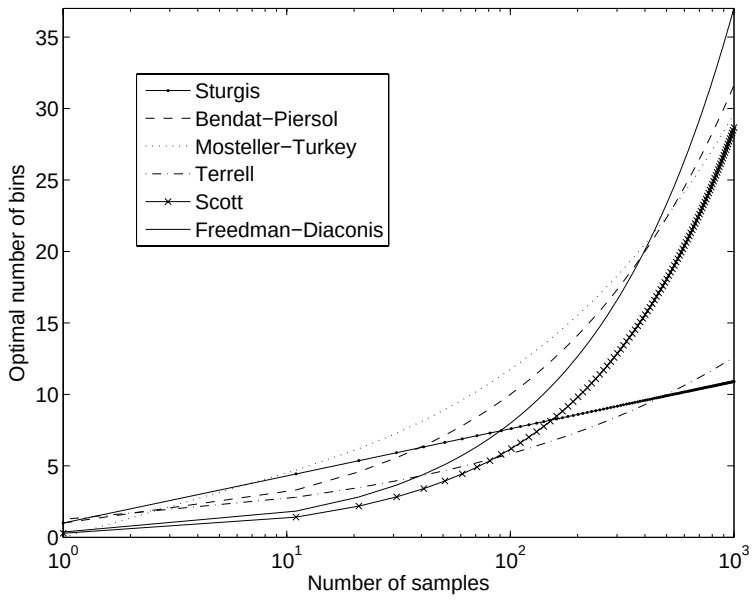


Figure 3.17: Optimal number of bins as a function of sample sizes, assuming the data is normally distributed with a variance of 1.

Multi-colour co-occurrence matrix

The co-occurrence matrix can also be applied to multi-spectral images. For example, the intensities of IR, VIS and WV images can be normalised to 1 and then mapped onto the RGB colour model. The model defines a *colour* by an addition of three primary colours which are red, green and blue. In this way, the colour of a pixel is a vector resulting from the combination of three independent vectors as in Figure 3.18. Figure 3.19 shows a colour image pixel as a vector in three dimensional space. The angle of the vector in the red-blue plane to the blue axis (ϕ) implies a proportion of cloud top temperatures and water contents, which can be used as an identification of possibly-precipitable clouds whereas the angle between the vector and the red-blue plane (θ) accounts for amounts of the rainfall through the cloud thickness properties (Suvichakorn and Tatnall, 2005).

The conclusion that the three channels convey different information of mid-latitude cyclones (Krennert et al., 1999) also supports the idea of using the multi-colour model for the cloud analysis.

The false colour images of the four sample events are shown in the upper panel of Figure 3.20. Based on an observation that a convective cloud with the lower IR and WV temperatures is usually thicker (brighter in VIS), the IR and WV temperatures are inverted so that the colder temperature is presented by the brighter shade of the RGB cube. As shown in the middle panel of Figure 3.20, the convective cloud has a bright white colours, which means that it has low IR and WV temperatures and high VIS reflectance. It can be noticed that the heavy rain corresponds to bright white shades. This is also applied to the frontal cloud where the rain areas corresponds to the white-pink shades. On the other hand, the rain areas at the centre of the storm and the light-to-moderate stratiform rain correlate to light green shades where the VIS intensities dominate the other two channels. The false colour images of the inverted IR and WV values also give more texture than the ones with non-inverted values and the shadow of the cloud is also preserved.

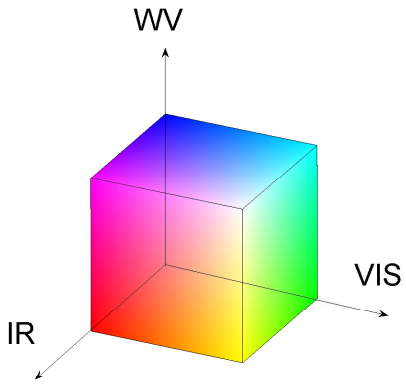


Figure 3.18: RGB cube model for IR, WV and VIS images.

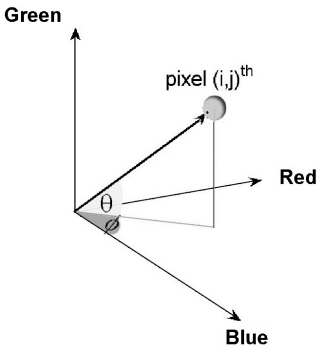


Figure 3.19: Colour image axis of a pixel.

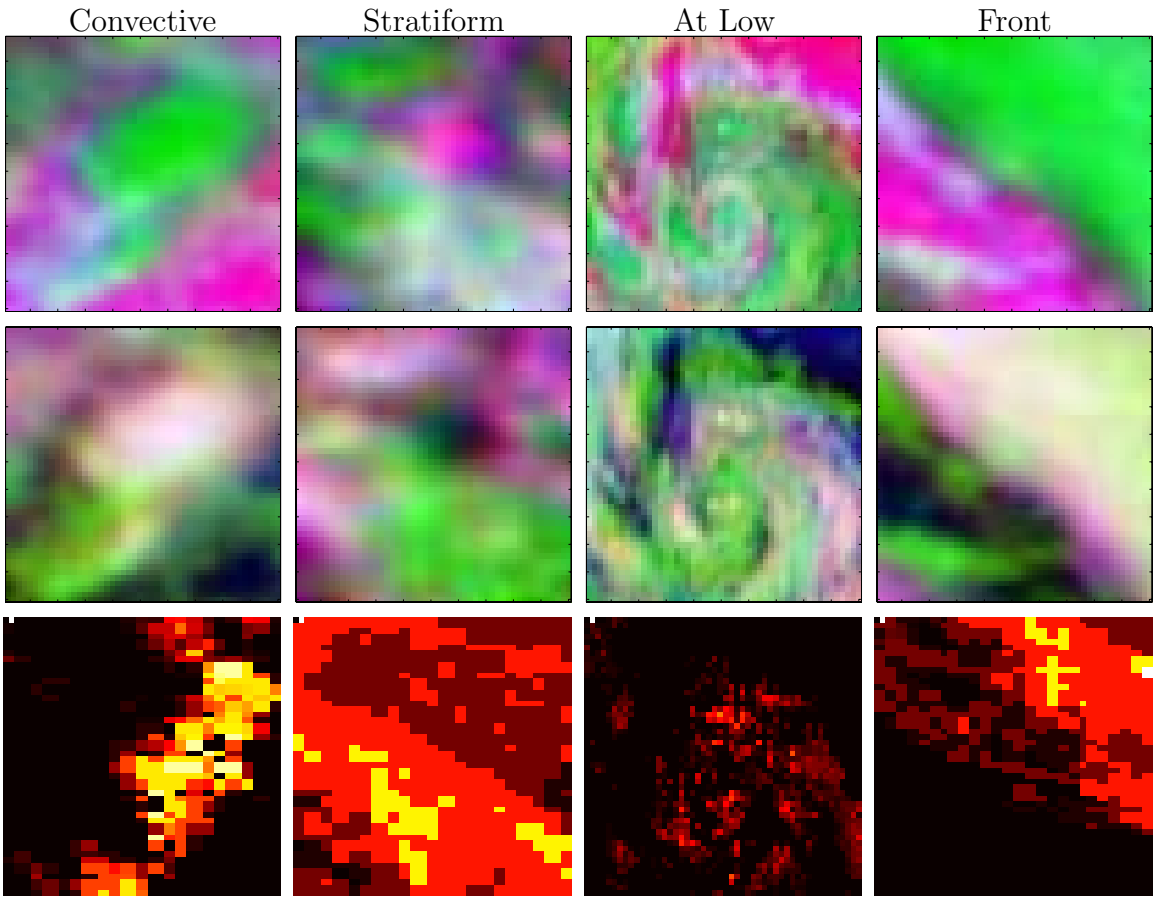


Figure 3.20: False colour images of the non-inverted (upper) and inverted (middle) IR and WV values of the four events in Figure 3.5. The corresponding rain radar images are in the lower panerl.

To derive textural features from the false colour image, the methods for monochromic images can still be applied. Given a displacement $d = ||(\Delta i, \Delta j)||$ between any two pixels, a co-occurrence matrix for an RGB space can be defined by

$$C_d(\vec{c}_1, \vec{c}_2) = \Pr\{I(i,j) = \vec{c}_1 \text{ and } I(i+\Delta i, j+\Delta j) = \vec{c}_2\}, \tag{3.2}$$

where $I(i,j) = \vec{c}_1$ refers to a pixel at the position (i,j) that has a colour vector $\vec{c}_1$ (Hauta-Kasari et al., 1996). The 3-dimensional co-occurrence matrices of the inverted and non-inverted false images are exactly the same. Their ASM and entropy are shown in Figure 3.21. Obviously, the two statistics can better distinguish the four events than those of the individual channels. This is similar to the two statistics of the rain radar images, although the arrangement of the statistic values of the two data is not the same. However, the computation of the 3-dimensional matrix requires a large storage to process. For example, if each colour channel has L values, the matrix will be $L^3 \times L^3$ in size.

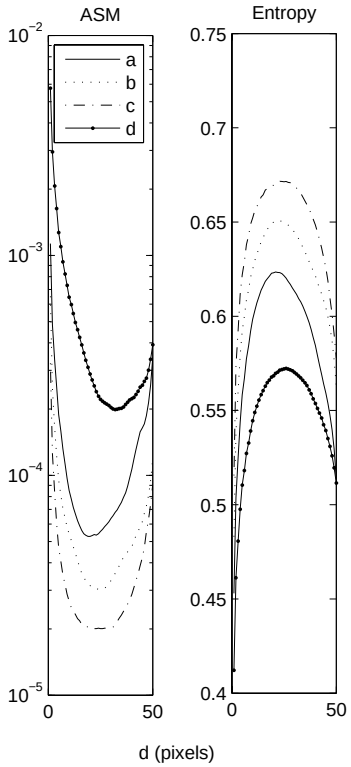


Figure 3.21: Statistics of co-occurrence matrices using the RGB images in Figure 3.20. The legends a, b, c and d refer to convective clouds, stratiform clouds, clouds at the vortex centre and frontal clouds.

Alternatively, the HSV (Hue, Saturation, Value) model, also called HSB (Hue, Saturation, Brightness), can be used to define a colour space. Hue represents the wavelength of a colour, analogous to position around a colour wheel. It ranges from 0-360° when colours goes from red to green, blue and back to red again. Saturation represents an richness of colour. A zero saturation means that the colour is colourless whereas a unit saturation means the maximum *colourfulness*. On the other hand, value or brightness means the intensity of the image and does not depend on colours. Figure 3.22 shows an illustration of the HSV cone model. Given a colour defined by $(R, G, B) \in [0, 1]$, an equivalent HSV colour can be determined by

$$\begin{aligned} V &= \max(R, G, B), \\ S &= 1 - \frac{\min(R, G, B)}{\max(R, G, B)}, \\ H &= \begin{cases} \arccos\left(\frac{G-B}{6S}\right) & \text{if } V = R, \\ \arccos\left(\frac{1}{3} + \frac{B-R}{6S}\right) & \text{if } V = G, \\ \arccos\left(\frac{2}{3} + \frac{R-G}{6S}\right) & \text{if } V = B. \end{cases} \end{aligned} \tag{3.3}$$

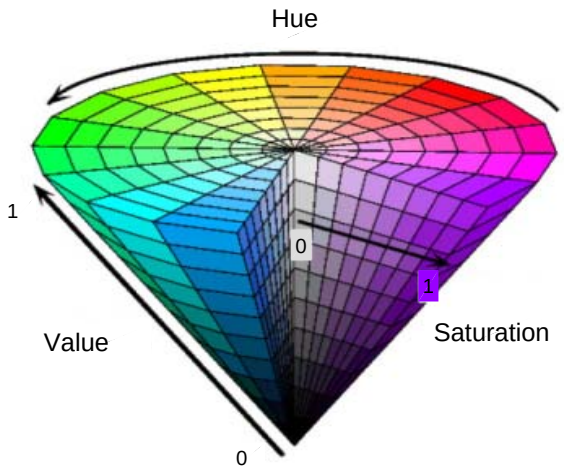


Figure 3.22: HSV cone model. Source: Matlab user's guide.

Value should be equal to the size of a colour vector in the RGB space. However, it is noticed that V in the equation is not the size of a vector in 3D. In fact, the HSV model treats the three primary colours as dependent variables. This can be beneficial to our rain estimation problem as each channel is not fully independent from others.

Arvis et al. (2004) characterised the colour texture by using V to compute the co-occurrence matrix. The mean and variance of hue and saturation were also included to specify the position of the vector in the colour space. However, the colour distribution of each event as presented in Figure 3.20 is characteristic and suggests hue for the representation of the cloud texture.

Figure 3.23 shows the statistics when using H, S and V to compute the co-occurrence matrices of the false image. The statistics of the non-inverted images are shown in the upper half of Figure 3.23 and those of the inverted images are in the lower half. The ASM and entropy of both inverted and non-inverted images are good at discriminating each meteorological event, especially when H and V are used. This is similar to those of the RGB model but much less memory is required. Additionally, the relation of homogeneity and correlation of (non-inverted) hue co-occurrence matrix as a function of displacement between two pixels (d) is also similar to those of the rain radar images. For instance, the correlation coefficients between them are 0.9294 and 0.9995 for the stratiform and the frontal cloud. The numbers are higher than the statistics of the individual channels which are presented in Table 3.1. Therefore, it is preferable to use the statistics especially ASM and entropy of the HSV model of the false images as the textural features if all the three channels are available.

At night, a combination a IR and WV images can be used instead of the VIS channel. Details of this method can be found in Suvichakorn and Tatnall (2005) (in Appendix D). The new false image resembles the false image when VIS image is used but it has less texture and different hues.

Table 3.1: Correlation coefficients of the co-occurrence statistics of the false image in comparison to those of the rain radar images.

Case	Homogeneity				Correlation			
	IR	VIS	WV	Hue	IR	VIS	WV	Hue
Convective	0.7856	**0.8980**	0.7867	0.8477	0.8794	**0.9889**	0.9150	0.9549
Stratiform	0.9442	**0.9905**	0.9483	0.9836	0.8756	0.8928	0.8830	**0.9294**
At Centre	0.1561	0.1733	**0.3888**	0.1331	0.9358	0.9010	**0.9477**	0.9315
Frontal	0.9877	0.9866	0.9770	**0.9985**	0.9970	0.9992	0.9856	**0.9995**

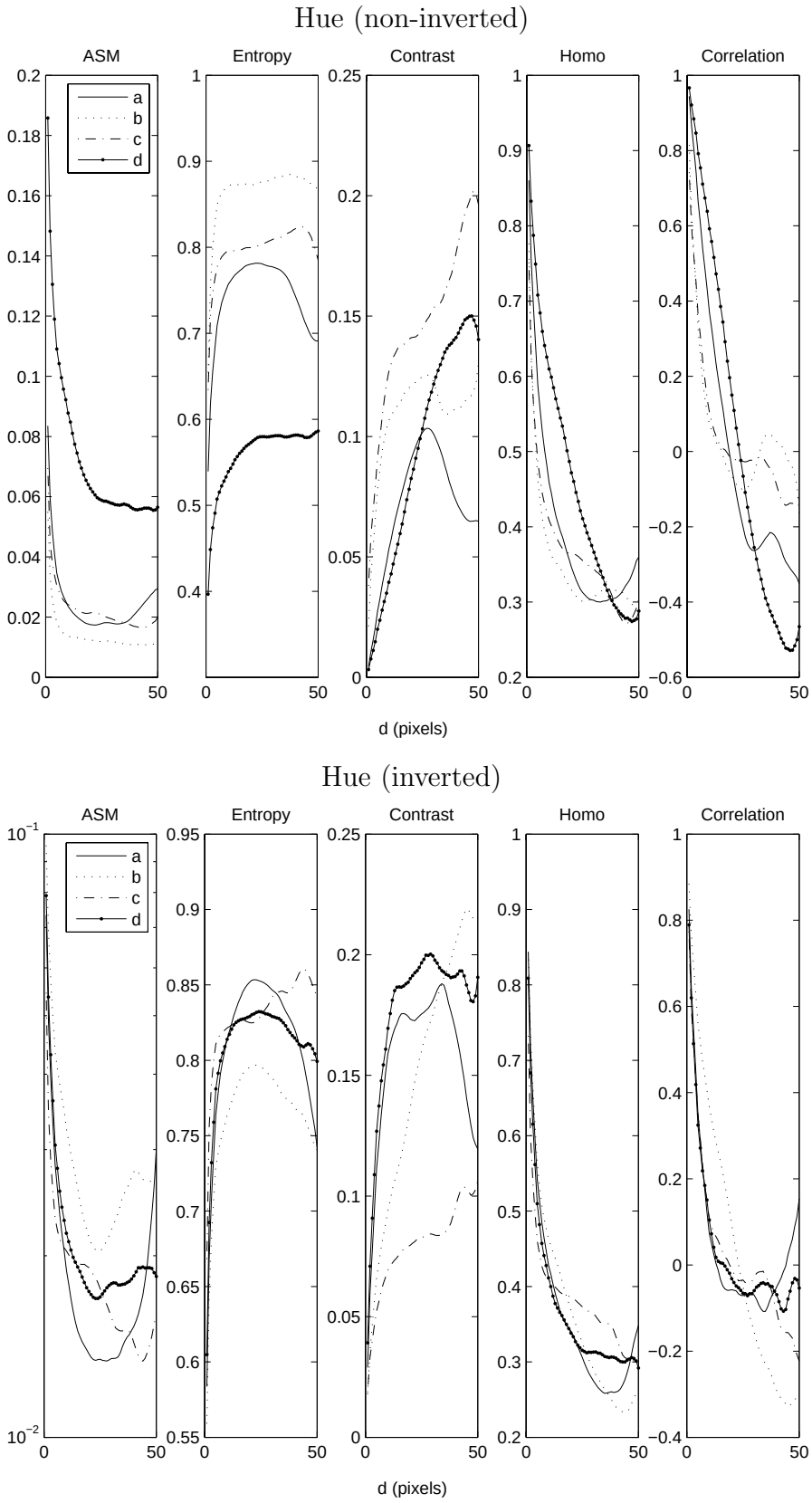


Figure 3.23: Statistics of co-occurrence matrices using of hue, saturation and Value of the HSV model. The legends a, b, c and d refer to convective clouds, stratiform clouds, clouds at the vortex centre and frontal clouds.

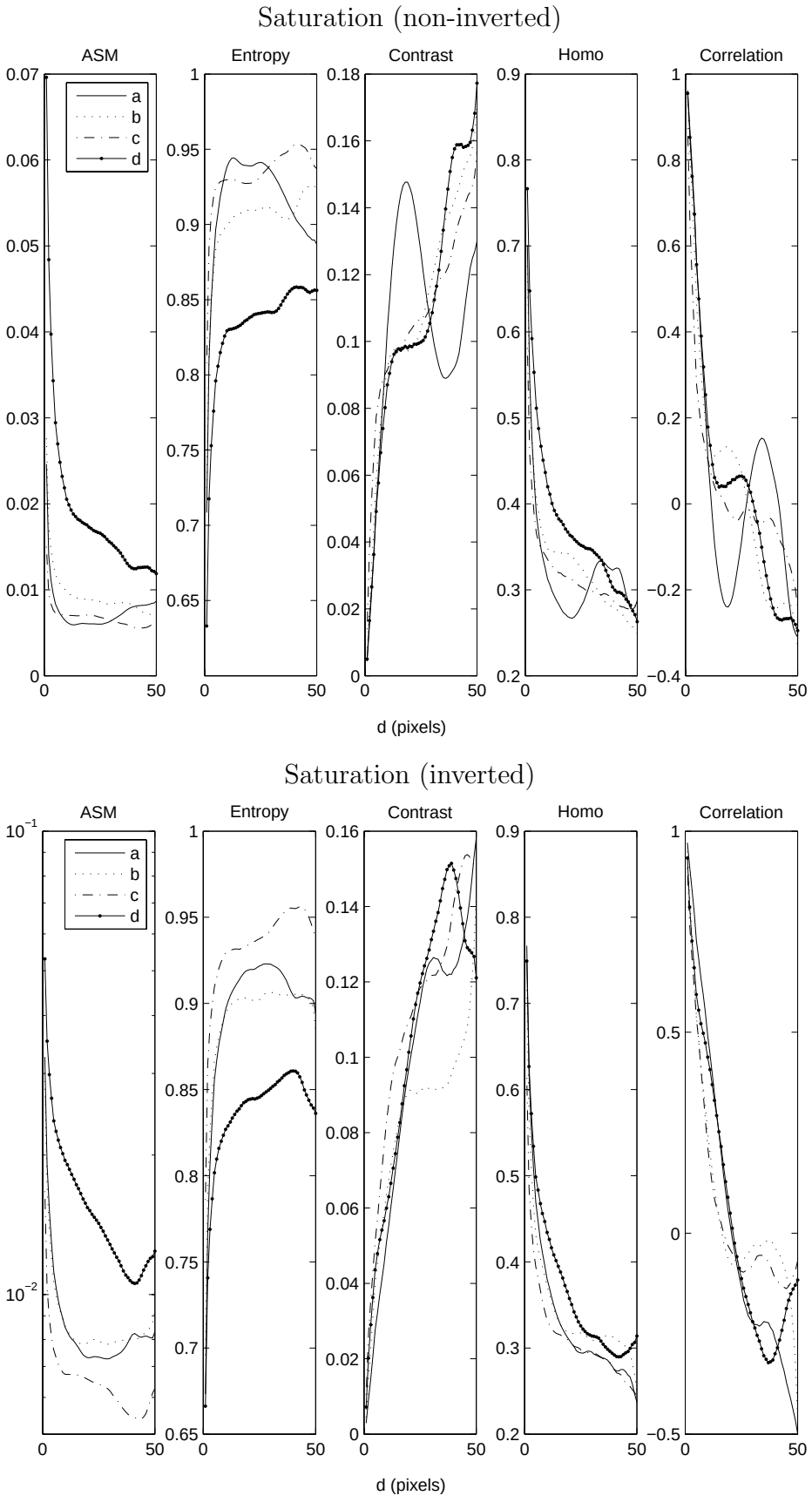


Figure 3.23 (continued)

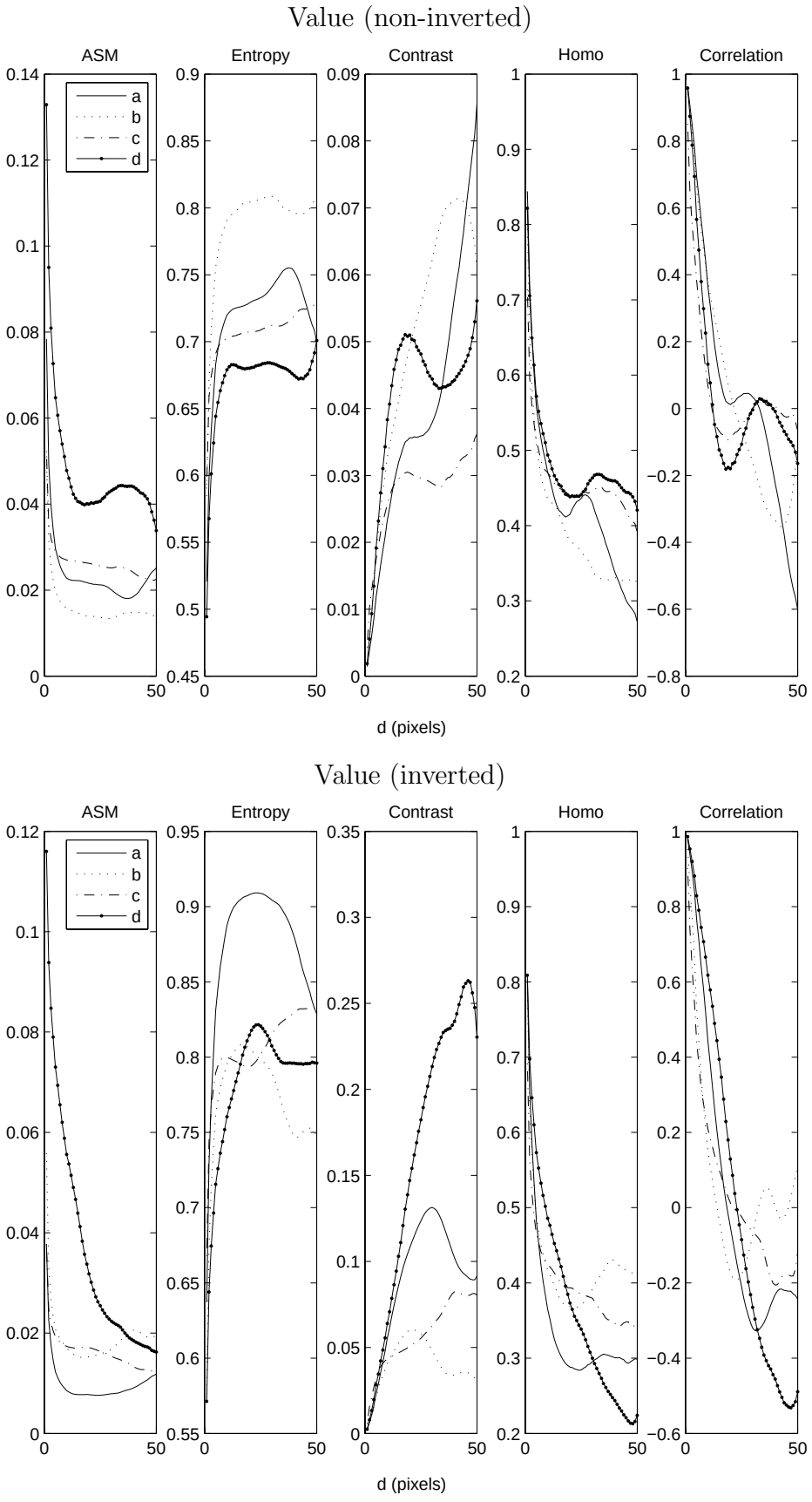


Figure 3.23 (continued)

Grey-level difference

Unlike the co-occurrence matrix that uses co-occurrences of intensities, the grey level difference method uses differences between intensities to describe texture. It is based on absolute grey level differences between pairs of pixels. As shown in Equation (2.2), the difference matrix D is defined as

$$D_{d,\theta}(i,j) = |I(i,j) - I(i+\Delta i, j+\Delta j)|,$$

where $d = ||(\Delta i, \Delta j)||$ is a distance between any two pixels and $\theta = \tan^{-1}(\Delta i/\Delta j)$, a probability density function can be defined as

$$f_{d,\theta}(k) = \Pr(D_{d,\theta}(i,j) = k).$$

Similar to the difference matrix, $f_{d,\theta}(k)$ is a function of distance (d) and direction (θ) between any two pixels as shown in Figure 3.24. If the texture is coarse relative to the distance d, the differences could be expected to be small and the probability density function large for small k. Accordingly, measures of the texture in an image may be computed from estimates of the spread in $f(k)$, e.g. by using its mean, standard deviation (STD) or entropy, at different displacements and angles as in the co-occurrence matrix.

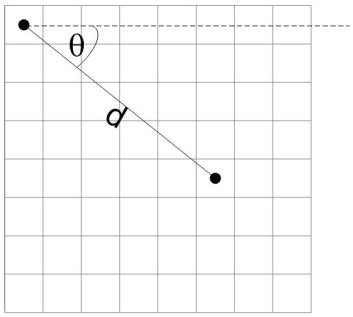


Figure 3.24: A diagram illustrates the distance (d) and the direction (θ) between any two pixels.

Figure 3.25 presents the statistics of the four types of images. Each tile of image represents the statistics as a function of direction from $0-180°$ in x-axis and distance between a pair of pixels in y-axis. The four rows correspond to the four meteorological events. The values of mean in the figure is in natural logarithmic scale to emphasise those with small values. For the images with apparent gradients such as the frontal cloud, there always exists the optimal value, either minimum or maximum, in a certain

direction. This is consistent in all types of images and is applicable to all kinds of statistics. There is also a resemblance between the standard deviation of the VIS and the rain radar images of the convective clouds. The similarity is illustrated in Table 3.2 by the correlation coefficient between the difference statistics of all the three channels and the rain radar images. However, if there are M values of d and N directions to be considered, the statistics at each position would produce a matrix of size $M \times N$. In practice, it is necessary to limit the value of M and N to reduce computation.

Table 3.2: Correlation coefficients of the grey-level difference statistics of the satellite images in comparison to those of the rain radar images. The maximum correlation coefficient of each statistic for each case is presented in boldface.

Case	Mean			STD			Entropy		
	IR	VIS	WV	IR	VIS	WV	IR	VIS	WV
Convective	0.50	**0.52**	0.44	0.64	0.60	**0.72**	0.72	0.69	**0.76**
Stratiform	0.60	**0.68**	0.46	0.73	**0.85**	0.73	0.75	**0.88**	0.74
At Centre	0.17	0.20	**0.36**	0.28	-0.02	**0.52**	0.33	0.16	**0.56**
Frontal	0.56	**0.62**	0.48	0.84	0.85	**0.85**	0.87	**0.89**	0.85

Several authors have used fixed numbers of the displacement and directions e.g. 1 pixel in the directions of 0°, 45°, 90° (Welch et al., 1988; Gu et al., 1991). However, during the experiments, it is found that the co-occurrence and difference statistics give the highest correlation at a particular value of the displacement d. Then, after this value, the correlation falls or becomes steady as illustrated in Figure 3.26. It is also noticed that at $d = 1$, which is commonly used for cloud classification, the correlation to rain rates is quite low (40-50% lower than the highest correlation coefficients). However, the value of d that gives the best correlation cannot be known a priori. In practice, the maximum or minimum of the statistics and their corresponding displacement might be used. On the other hand, fixing the directions can make the statistics of the difference matrix sensitive to the rotation of an object. For the case of a mid-latitude cyclone, in order to make the statistics invariant to the rotation of the cyclone, an angle with respect to the centre of the cyclone has been used. Figure 3.27 illustrates θ_R and θ_T which are the angles of the radius and its tangent with respect to x axis.

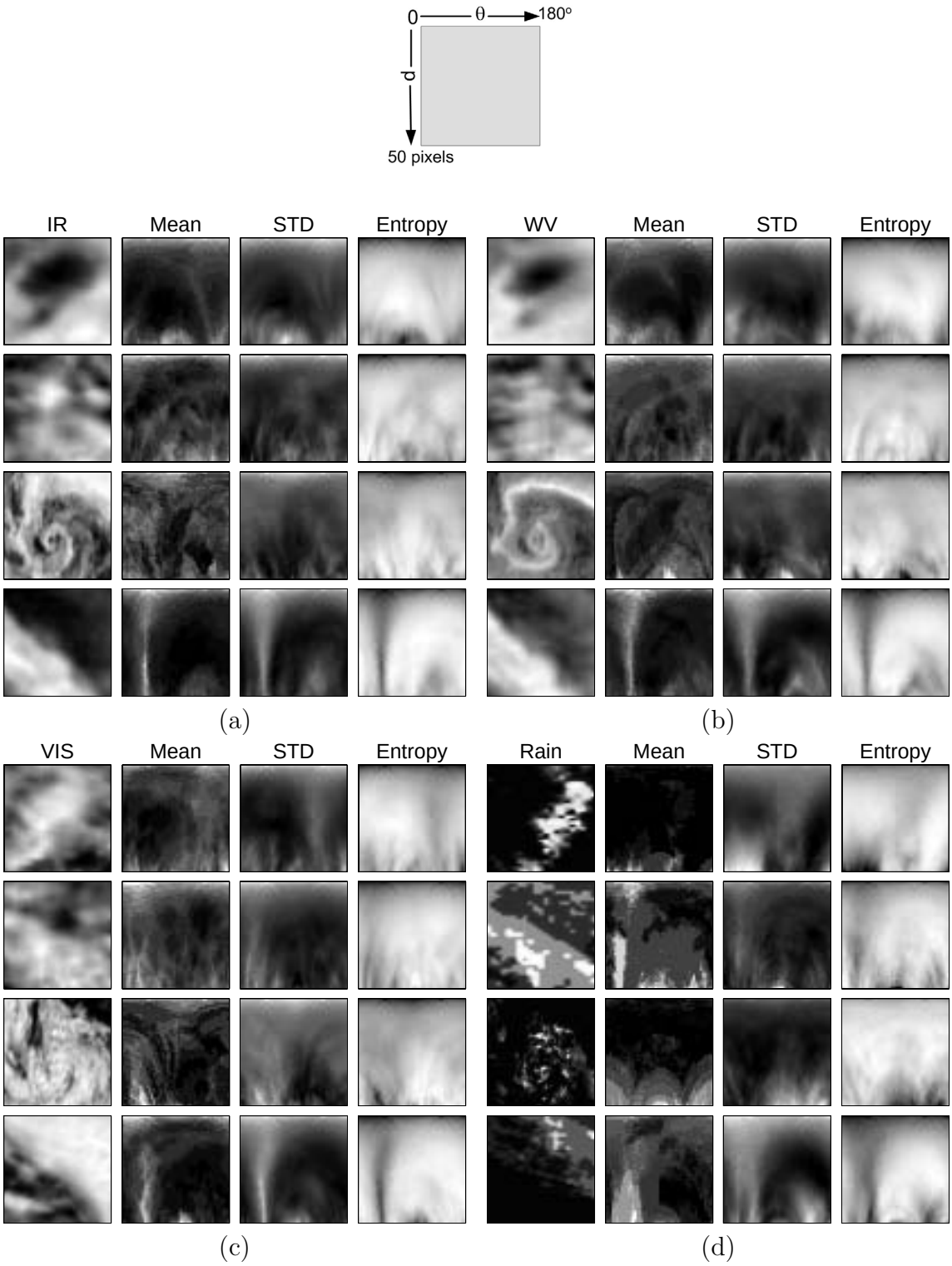


Figure 3.25: Grey-level difference statistics of (a) IR, (b) WV, (c) VIS images and (d) rain radar images in Figure 3.5.

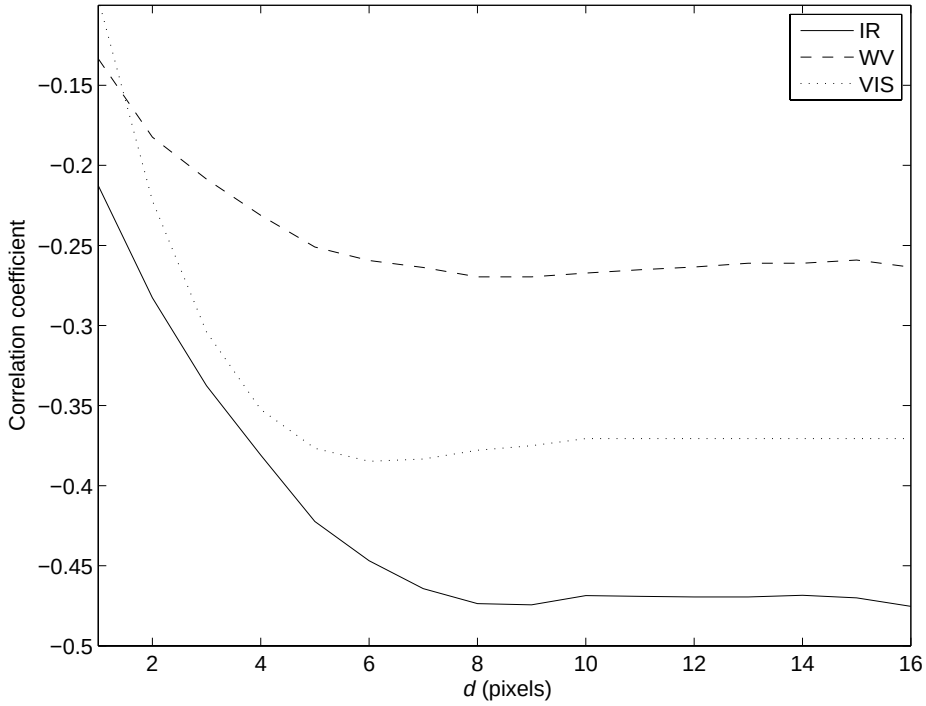


Figure 3.26: Correlation coefficients of ASM of a co-occurrence matrix as a function of displacements between two pixels (d).

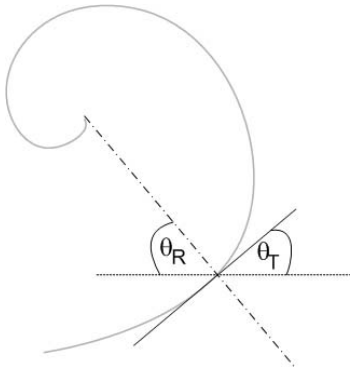


Figure 3.27: Definition of θ_R and θ_T.

3.2.2 Structural texture

The structural approach to textural analysis assumes that texture is composed of textural primitives e.g. lines and edges. Thus, it seeks to describe these dominant structural features in the image and their spatial organisations (Tuceryan and Jain, 1998), instead of intensity transitions as in the co-occurrence matrix or the grey-level difference statistics. Generally, edges[4] in VIS and IR images correlate to noticeable changes in cloud depth and height. The changes can be a result of evolution within the same cloud, which is related to the atmospheric instability. They can also be a result of a different height of two clouds such as a high cloud and a low cloud, which normally relate to different rain amounts. Either pieces of information are helpful to rain estimation. However, analysing edges in all possible directions is very computationally expensive. The dimensionality of the problem can be reduced by projecting[5] the image intensities onto a principal axis and then extracting edges along the axis. The technique to project a continuous function $f(x, y)$ onto any directions of $\theta \in [0, \pi)$ is called the Radon transform, which was first introduced in 1971 (Warrick and Delaney, 2001). The Radon transform is defined as

$$\mathscr{R}_f(\rho, \theta) = \int \int f(x, y)\delta(x \cos\theta + y \sin\theta - \rho)dxdy, \tag{3.4}$$

where $\delta(x)$ is the Dirac's delta function—it gives the value of infinity for $x = 0$ and zero elsewhere. The Radon transform is a function of the distance from the centre of the images (ρ) along the x axis (negative values refer to the distance to the left of the centre) and the angles between the projection line and the x axis (θ).

The Radon transforms of the four events of the rain radar images are presented in Figure 3.28. The maximum of the transformed values shows the direction of the significant edges. The characteristics such as gradients of a projection along the significant direction can be further extracted to use as texture signatures of the clouds.

Unlike the rain radar images, the transformed values of the same events of the infrared images shown in Figure 3.29 do not show any important differences between the events. The transformation of the IR and WV images also gives the same results. The main reason is due to the characteristics of the satellite images which is too smooth to show any importance along the projection lines, although our eyes can detect edges of the cloud objects. Therefore, Sobel's filter (Gonzalez et al., 2003) is applied to the images prior to the transformation to enhance the edges of the cloud

[4]Edges are defined as apparent transitions of image intensities

[5]Projection is done by integrating the image intensity onto a line

objects. The Radon transforms of the filtered IR images are presented in Figure 3.30. The transforms correlate to those of the radar images with the correlation coefficients of 0.71 and 0.66 for the case of the convective cloud and at the centre of the vortex. Figure 3.28 and 3.30 also reveal that there is a particular direction of the projection that gives the maximum transformed value, which can be seen as bright yellow to white colours. Such a direction refers to the most significant orientation of the texture, i.e. most of the object(s) in the texture lie in the direction of $90^o - \theta_{R_{\max}}$. The correlation coefficients between the Radon transforms of the satellite and the rain radar images and the differences between their $\theta_{R_{\max}}$ are presented in Table 3.3. It can be noticed that the least differences between $\theta_{R_{\max}}$ of both data lie within $\pm 25^\circ$ for each event. In addition, these angular errors do not exceed $\pm 60^\circ$ and the correct match (zero error) can also be found. Therefore, the direction of a highest Radon transform of the satellite images has been used as another textural feature for the rain estimation due to its correlation to that of the rain texture.

Figure 3.31 illustrates the correlation of $\theta_{R_{\max}}$ to the rain rates. The feature shows more relevance to the rain rates when compared to that of the spectral information—the number of pixels of the new features that have the correlation coefficients $> \pm 0.4$ is 16.81% whereas those of the spectral information is only 14.31 %. This encourages the use of Sobel's filter to enhance edges in the satellite images prior to applying the Radon transform.

Table 3.3: Correlation coefficients and $\theta_{R_{\max}}$ angular errors of the Radon-transforms of the satellite images in comparison to those of the rain radar images. The maximum correlation coefficient of each statistic for each case is presented in boldface.

Case	Correlation Coeff			Errors (degrees)		
	IR	VIS	WV	IR	VIS	WV
Convective	**0.71**	0.45	0.69	**25**	45	37
Stratiform	0.47	0.63	**0.72**	22	58	**0**
At Centre	**0.66**	0.43	0.34	**8**	-10	-10
Frontal	-0.25	-0.13	**-0.39**	**-12**	-13	-43

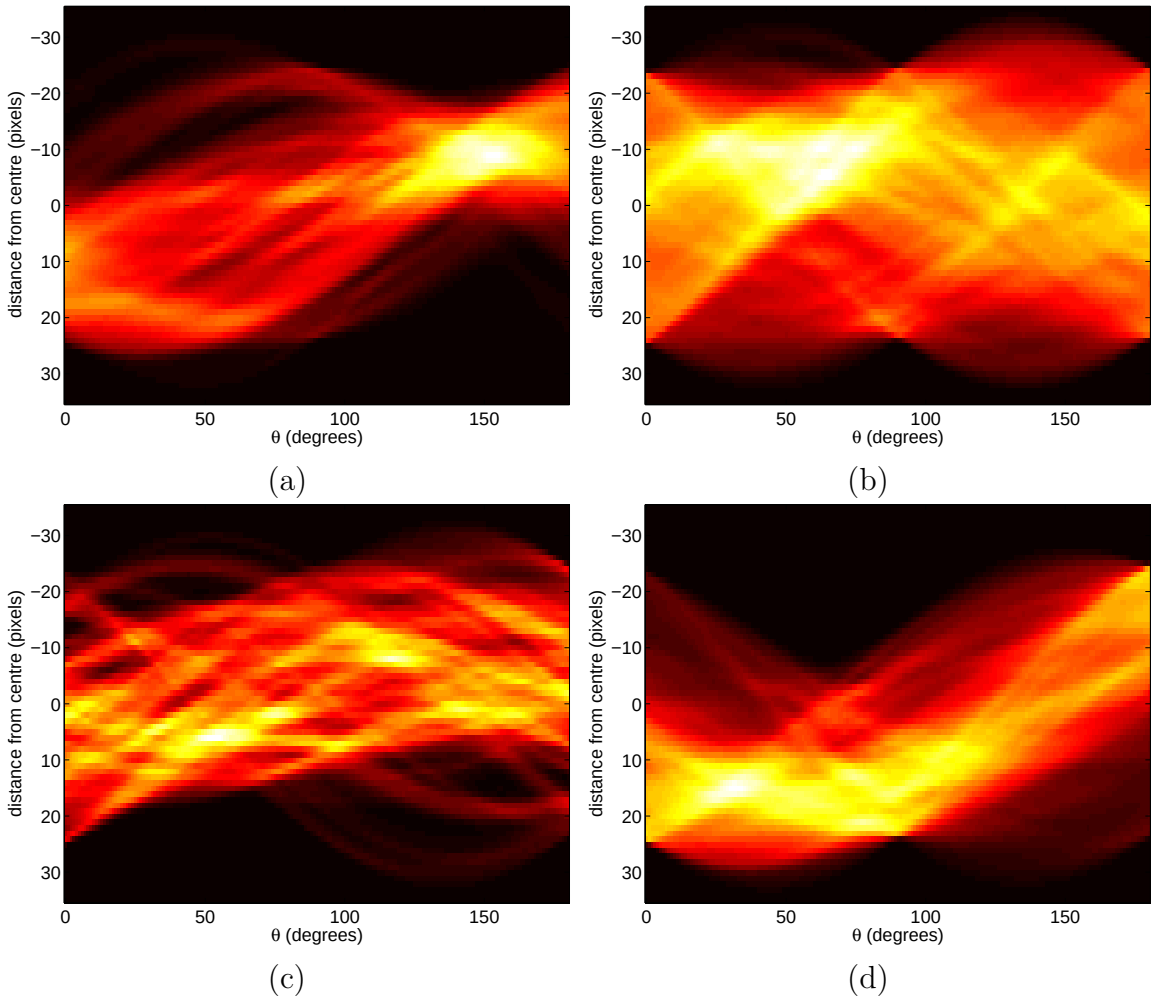


Figure 3.28: The Radon transformed values of rain radar images of (a) convective (b) stratiform (c) centre of the cyclone and (d) front. The values are shown in colours ranging from maximum in white, then yellow, red and black for zero values.

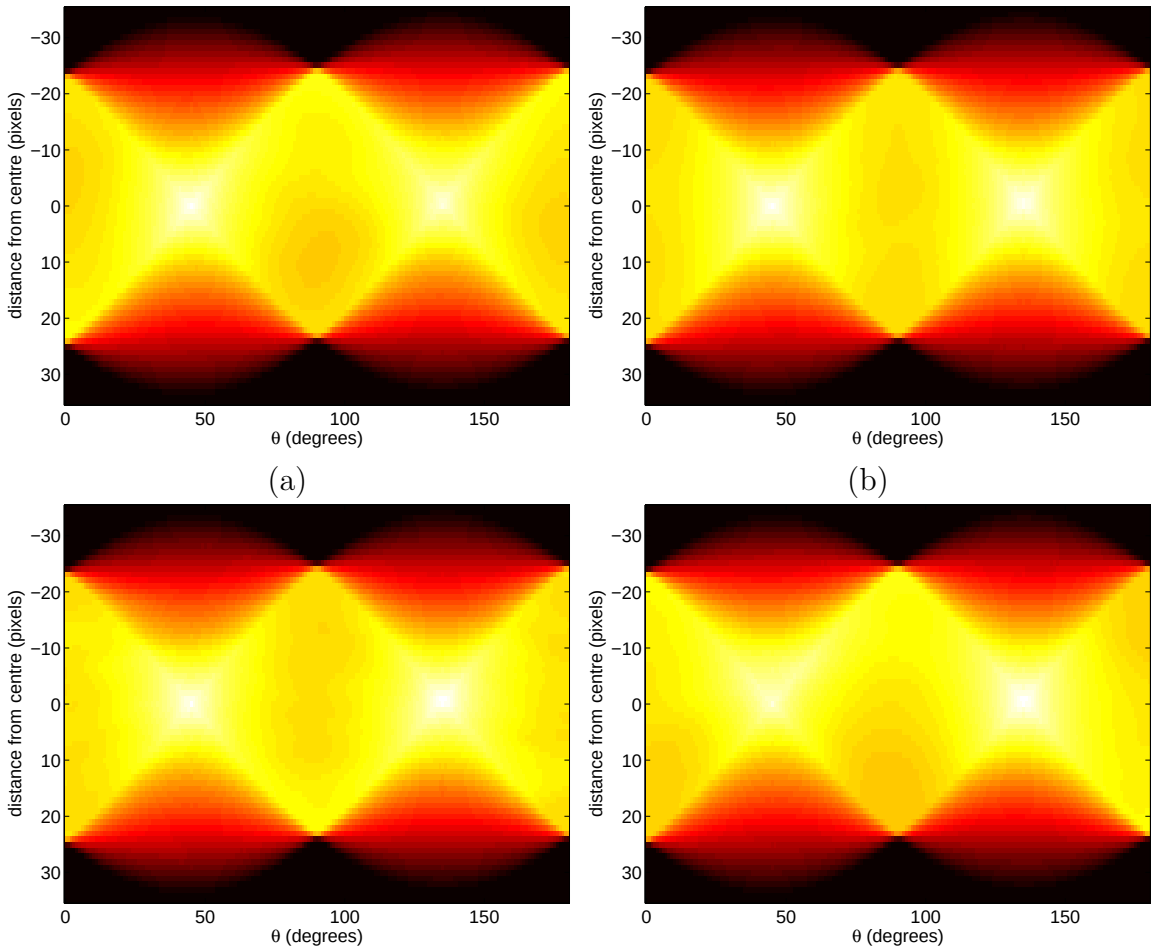


Figure 3.29: The Radon transformed values of infrared images of (a) convective (b) stratiform (c) centre of the cyclone and (d) front.

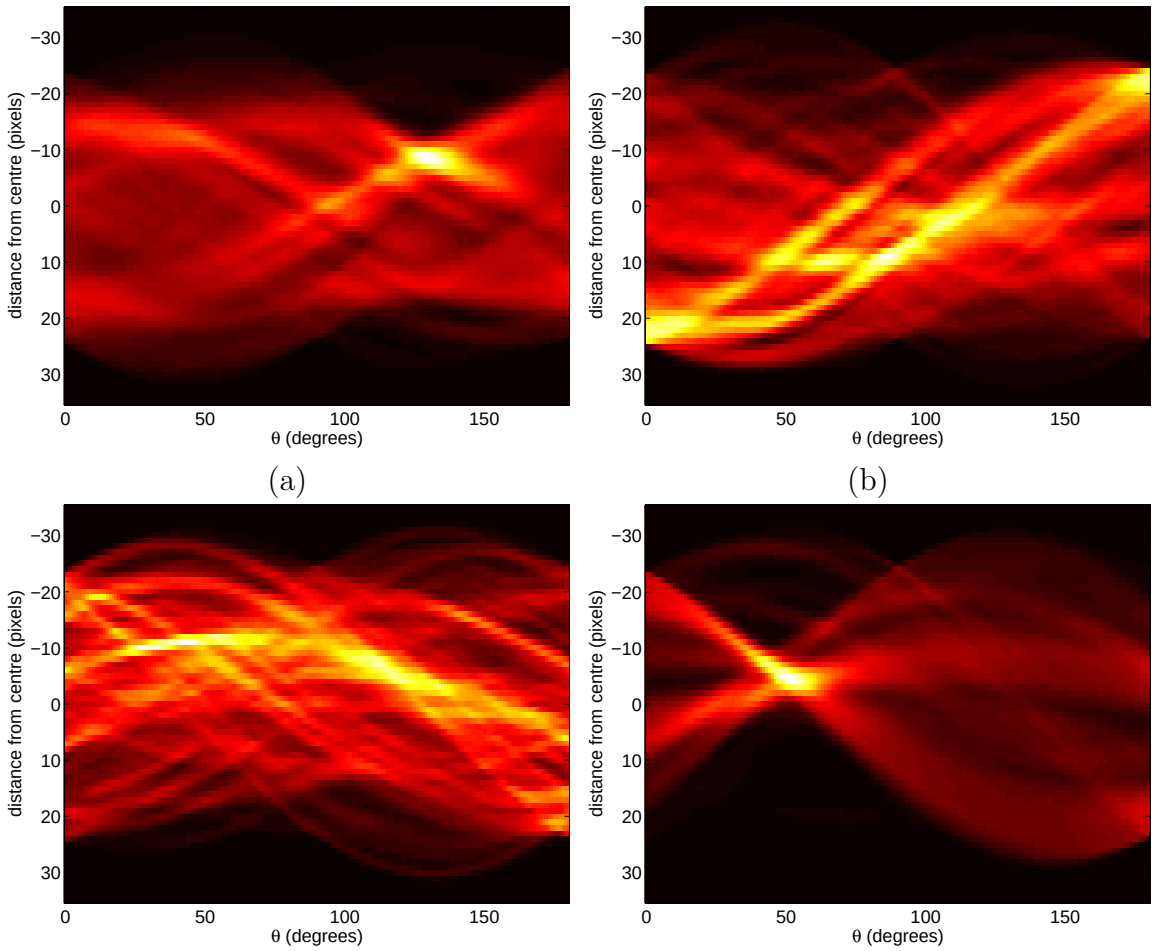
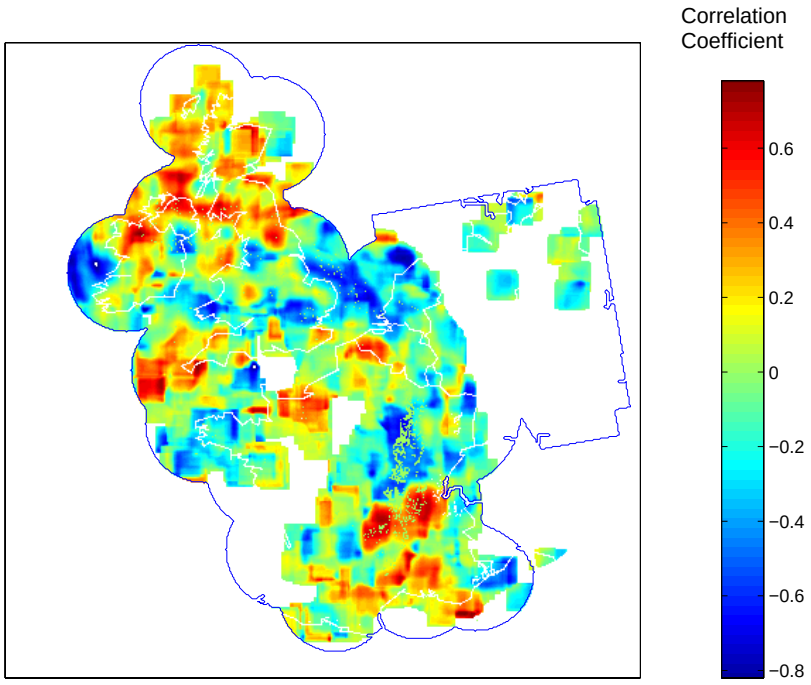
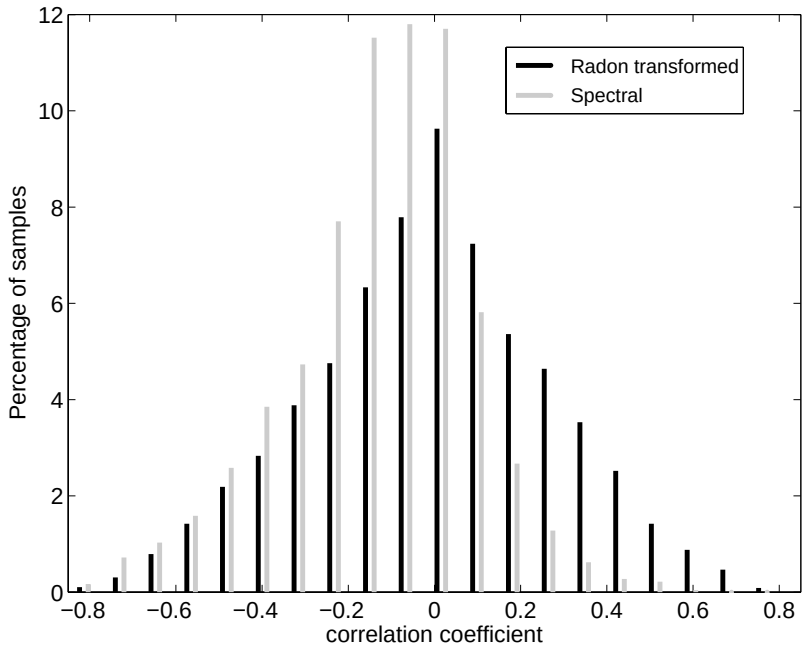


Figure 3.30: The Radon transformed values of gradients of the infrared images of (a) convective (b) stratiform (c) centre of the cyclone and (d) front.



(a)



(b)

Figure 3.31: (a) Correlation coefficient of the Radon transforms of the rain radar images and the infrared images of the mid-latitude cyclone on 23 June 2004 at 0030 UTC, calculated using a window of 21×21 pixels, and (b) its histogram when compared with that of Figure 3.2.

3.2.3 Frequency analysis

Some characteristics of texture in an image are not easily identifiable from the spatial domain. This includes the major repetitive components within the image. The need to distinguish these components leads to the mathematical representation of the image in terms of their orthogonal basis functions. The Fourier transform, for example, uses an infinite series of trigonometric functions at different frequencies. In the Fourier space image, each point represents a particular spatial frequency contained in the real domain image. These spatial frequency contents can reflect the texture of an image. For example, high spatial frequencies correspond to the presence of the sharp transitions in the intensity whereas low frequencies correspond to the absence of edges or regions of approximately uniform grey levels. The orientation of a spatial frequency corresponds to the orientation of an edge in the image. That is, the frequency contents of the image can be a measure of the coarseness and the directionality of textures. The Fourier transform of a signal $f(t)$ is defined as

$$\mathscr{F}(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-\hat{\jmath}\omega t) dt, \tag{3.5}$$

where $\hat{\jmath} = \sqrt{-1}$ and ω is the frequency of the transform.

The Discrete Fourier Transform (DFT) is the sampled Fourier Transform and therefore does not contain all frequencies forming an image, but only a set of samples which is large enough to fully describe the real domain image. The number of frequencies corresponds to the number of pixels in the real domain image; the image in the real and Fourier space are of the same size. For a square image of size $N \times N$, the two-dimensional DFT is given by

$$\mathscr{F}(k,l) = \frac{1}{N^2} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} I(x,y) \exp\left[-\hat{\jmath} 2\pi \left(\frac{kx}{N} + \frac{ly}{N}\right)\right], \tag{3.6}$$

where $I(x,y)$ is the real domain image and the exponential term is the basis function to decompose the image into its sine and cosine components. The Fourier transform of each spatial frequency $\omega = (k,l)$ is obtained by a process of multiplying the image with the corresponding basis function and summing the result. By using the same basis function, the real domain image can be derived from the Fourier image. The

inverse Fourier Transform is given by

$$I(x,y) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \mathscr{F}(k,l) \exp\left[\hat{\jmath} 2\pi \left(\frac{kx}{N} + \frac{ly}{N}\right)\right]. \tag{3.7}$$

The Fourier transform has the advantage of shift invariance. Strictly speaking, the magnitude of the Fourier transform is the same no matter what portion of the texture is considered. Changing the position only changes the phase of the transform. Figure 3.32 shows the Fourier image of the rain radar and the satellite images for the four type events in logarithmic scale. The Fourier images are shifted in such a way that the $\mathscr{F}(0,0)$ is displayed in the centre of the images. From the figure, it is noticeable that there are cross-shaped ripples and rectangular grids that overshadow other frequency responses.

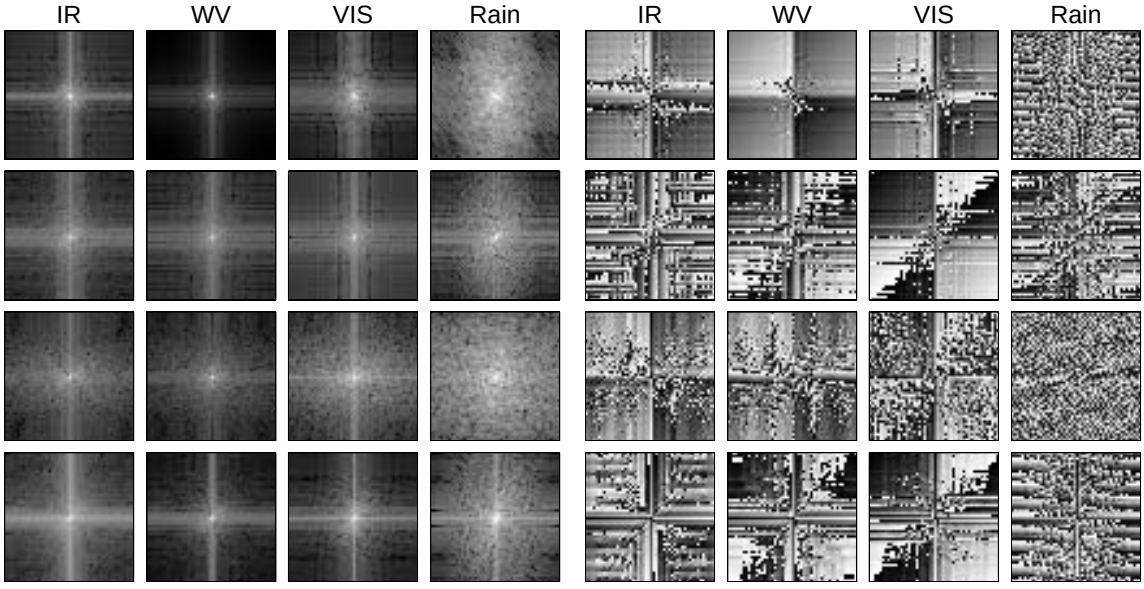


Figure 3.32: (a) Amplitudes and (b) phases of the Fourier transforms of the 4 events. Each row represents each event: (from top to bottom) convective, stratiform, at vortex's centre and frontal.

The cross-shaped ringing artifacts, known as the Gibbs phenomenon, are a consequence of treating an image as if it is part of a periodically replicated array of identical images extending horizontally and vertically to infinity. The discontinuity at edges of each virtually-tiling image aggregates frequency responses along the x and y axes of the Fourier image and therefore governs other features (Duhamel and Vetterli, 1990). These edge effects can be significantly reduced by windowing the image with a function that slowly tapers off at the edge such as a Gaussian function shown in Figure

3.33. In Figure 3.34, the Fourier transform of the windowed image can significantly reduce the white cross in the middle of the images.

On the other hand, the rectangular grids in the transformed images are a consequence of interpolating the satellite images to align with the finer resolution of the rain radar images. The effect is illustrated in a sequence of the images in Figure 3.34. From left to right, the images becomes smoother and their Fourier transforms produce more black rectangular grids. The grids are quite obvious in Figure 3.34 (c) when the white crosses that dominate the transforms are removed by the Gaussian windows.

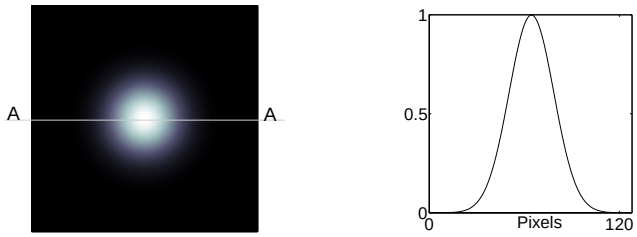
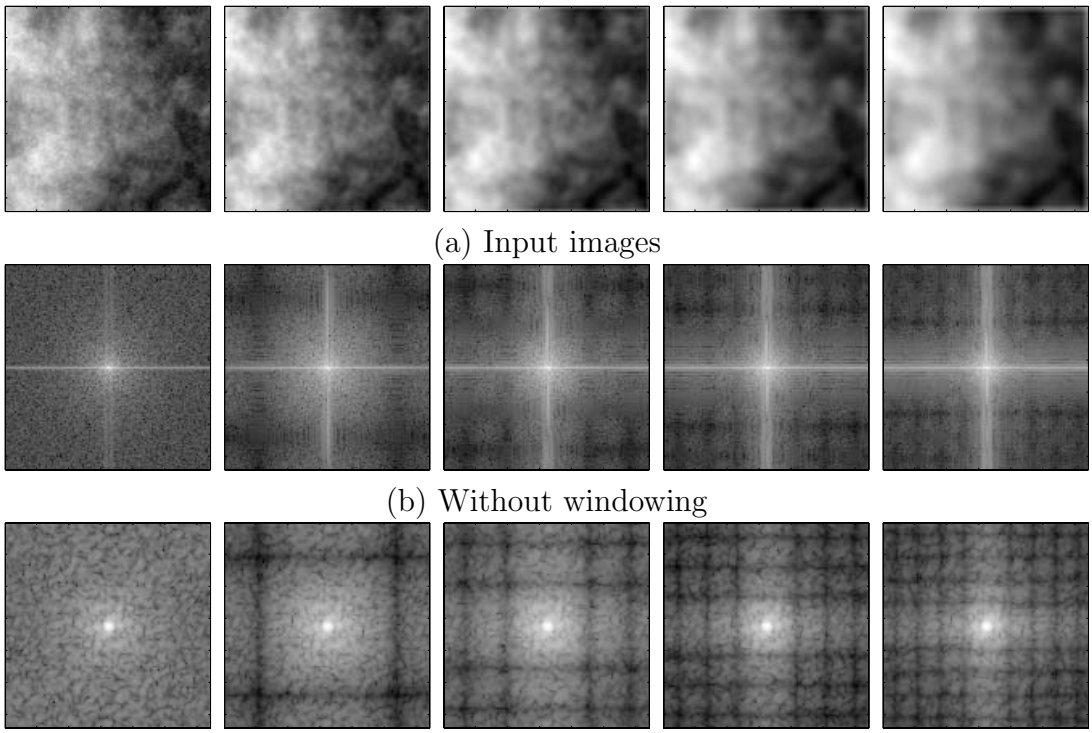


Figure 3.33: A Gaussian window for masking and image and its cross-section.



(a) Input images

(b) Without windowing

(c) With windowing

Figure 3.34: Fourier transforms of the smoothed images before and after using the Gaussian window.

The Fourier transforms when a Gaussian window is applied to the satellite and rain radar images are shown in Figure 3.35. The value at zero frequency $\mathscr{F}(0,0)$ at the centre of the images are by far the largest components. Using the Gaussian window can recover the direction of edges in the rain radar images. However, such a direction could not be observed from the satellite images. Therefore, Sobel's filter previously introduced is applied to the frontal cloud image prior to the Fourier transform to enhance structural information. The Fourier transforms of the filtered images and their corresponding principal components are shown in Figure 3.36. The principal components are defined by 20 pixels with the highest Fourier transforms. These pixels are represented as white in the lower panel of the figure. The illustration shows that the IR and VIS images can well extract the principal directions of the cloud.

In addition to the derived directions of the principal components, energy and entropy of the distribution in the Fourier space can be used as a rotation-invariant signature of texture characteristics.

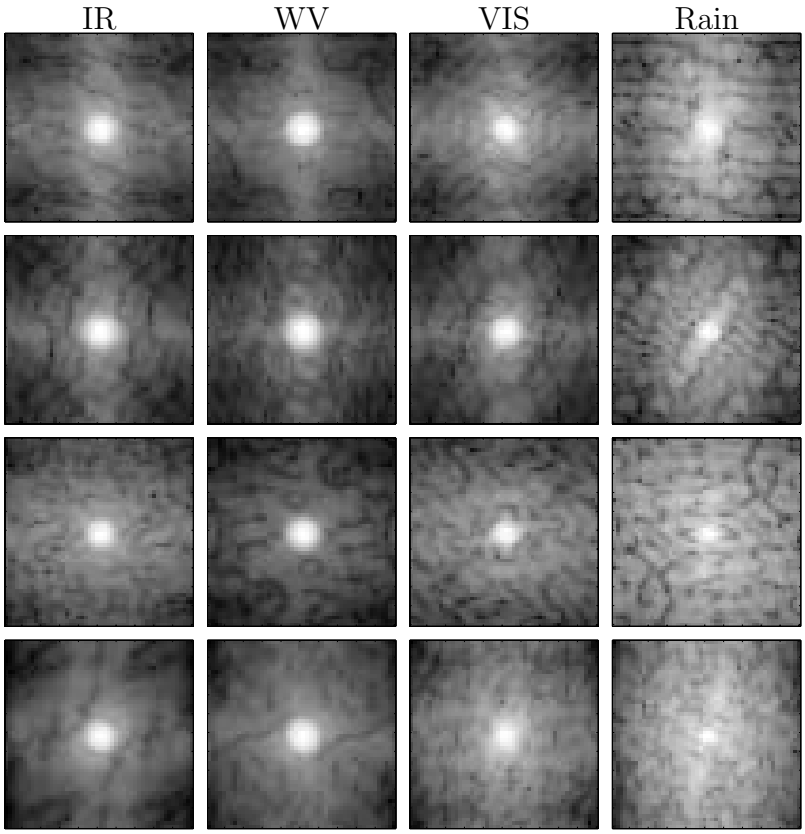


Figure 3.35: Amplitude of the windowed Fourier transform of the 4 events. Each row represents each event: (from top to bottom) convective, stratiform, at vortex's centre and frontal.

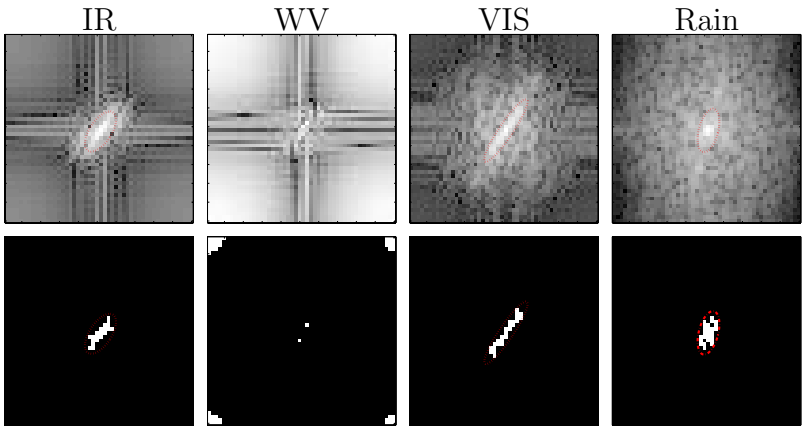


Figure 3.36: Amplitude (upper) and principal directions (lower) of the windowed Fourier transform after applying Sobel's filter to the frontal cloud.

3.2.4 Spatio-temporal analysis

Instead of fixing the Gaussian window at the centre of an image, the window can be shifted over the spatial domain to discover how the frequency contents evolve in space and time. This technique is known as the short-time Fourier transform (STFT). For a function $f(t)$, the technique can be mathematically represented by

$$\mathscr{F}_w(\tau,\omega) = \int_{-\infty}^{\infty} f(t)\exp(-\hat{\jmath}\omega t)w(t-\tau)dt. \tag{3.8}$$

The windowing function $w(t-\tau)$ allows the Fourier transform to calculate the frequency responses at each particular location τ. Therefore, the transform has become time-dependent as well as frequency-dependent.

Although achieving both time and frequency localisation, the short time Fourier transform still has a major disadvantage due to a fixed resolution—the same window size of the basis function is used for each frequency. In order to analyse the texture of an image, a short basis function (narrow window, wide band) is preferred to examine sharp gradients whereas a longer basis function (wide window, narrow band) is desirable to observe uniform details. The wavelet transform has been developed to overcome such resolution problems. The transform is done by applying a basis function that is localised in both time and frequency known as a mother wavelet and its family to a signal to acquire the pattern the signal is correlated with. The family of the basis function can be derived by translating and scaling the mother wavelet so that the wavelet transform becomes a function of these translating and scaling parameters

(Daubechies, 1990), i.e.

$$\Psi(\tau, s) = \int_{-\infty}^{\infty} f(t) \left[\frac{1}{\sqrt{|s|}} \psi^* \left(\frac{t-\tau}{s} \right) \right] dt, \tag{3.9}$$

where $\psi(t)$ is the mother wavelet function and $\psi_{s,\tau}(t)$ denotes variants of the function with a scale and translation parameter s and τ, and $(^*)$ denote the complex conjugate of the function. The meaning of the translation is similar to that in the STFT, i.e. it relates to the location of the window of a function that is shifted through the signal. However, instead of frequency ω, the scale parameter s is used. The scale parameter is a multiplicative inverse of the frequency; low scales (high frequencies) correspond to detailed information that usually lasts a short time such as edges in an image, and vice versa. In practice, both s and τ are sampled and discretised in order to reduce redundancy, otherwise the wavelet transform of an image would result in 4 dimensional space.

The wavelet transform can be applied to two-dimensional applications by using a simple hierarchical framework (Mallat, 1999). Its principle is to represent the image by its approximation, i.e. a down-sampled image at a coarser scale, and the difference between the approximation and its original image. The difference, which contains the high frequency content of the original image, can be precisely represented by a set of orthonormal wavelet basis functions that span over the frequency domain. The down-sampling is then successively applied to the approximated image so that the image becomes smaller and smaller. For the two dimensional analysis, if the downsampling of the wavelet transform is performed by a factor of two, the transform may be computed by filtering in each dimension separately. That is, the approximation of an image $f(x, y)$ can be characterised by

$$\begin{aligned} A_f(x, y) &= < f(x, y), \phi(x, y) > \\ &= < f(x, y), \phi(x)\phi(y) >, \end{aligned} \tag{3.10}$$

where $A_f(x, y)$ is the approximation of $f(x, y)$ and $\phi(x, y) = \phi(x)\phi(y)$ is a scaling function. The symbol $< , >$ denotes the inner product such that

$$< f(x), g(x) >= \int_{x=-\infty}^{\infty} f(x)g(x)dx,$$

Mallat also proved that, given $\phi(x, y)$ be a separable scaling function and $\psi(x)$ be the one dimensional wavelet associated with the scaling function, there exists three

wavelets

$$\begin{aligned} \psi_1(x,y) &= \phi(x)\psi(y) \\ \psi_2(x,y) &= \psi(x)\phi(y) \\ \psi_3(x,y) &= \psi(x)\psi(y) \end{aligned} \tag{3.11}$$

that form orthonormal basis functions of the image $f(x,y)$. Thus, the differences between $A_f(x,y)$ and $f(x,y)$ can now be described in terms of these functions. That is, the image $f(x,y)$ can be completely represented by 4 images which includes $A_f(x,y)$ and the three images which are the inner products between the orthonormal basis functions and the original image, i.e.

$$\begin{aligned} D1f(x,y) &= < f(x,y), \phi(x)\psi(y) > \\ D2f(x,y) &= < f(x,y), \psi(x)\phi(y) > \\ D3f(x,y) &= < f(x,y), \psi(x)\psi(y) > . \end{aligned} \tag{3.12}$$

If $\phi(x)$ and $\psi(x)$ are a perfect low pass filter and a perfect band pass filter, the image A_f will represent the low-frequency content of the original image. The image $D1f$ and $D2f$ will correspond to vertical high frequencies (horizontal edges) and horizontal frequencies (vertical edges). The image $D3f$ will indicate high frequencies in both directions such as corners. Accordingly, the wavelet transform becomes only an operation of cascaded discrete filters. That is, a decomposition of an image can be done by first performing a one-dimensional transformation on each row followed by a one dimensional transformation of each column. The decomposition can then be applied to A_f so that details at coarser resolution (higher level) of the image are extracted. The image can be reconstructed from the four sub-images by reversing the decomposition. The decomposition and reconstruction of an image at level j are shown in Figure 3.37. In the diagrams, $h(x)$ and $g(x)$ referring to the impulse response of $\phi(x)$ and $\psi(x)$ and $\tilde{h}(x) = h(-x)$.

Examples of $h(k)$ and $g(k)$ are the Daubechies and Coiflets, shown in Figure 3.38. The performance of the wavelets can be compared by using the number of their vanishing moments, which refer to the wavelets ability to represent polynomial behaviour or information in a signal. For example, the Daubechies orthogonal wavelets has a number of vanishing moments equal to half the number of coefficients. That is, the Daubechies wavelet with 6 coefficients (db6) has 3 vanishing moments, therefore it has power to represent constant, linear and quadratic signal components. There

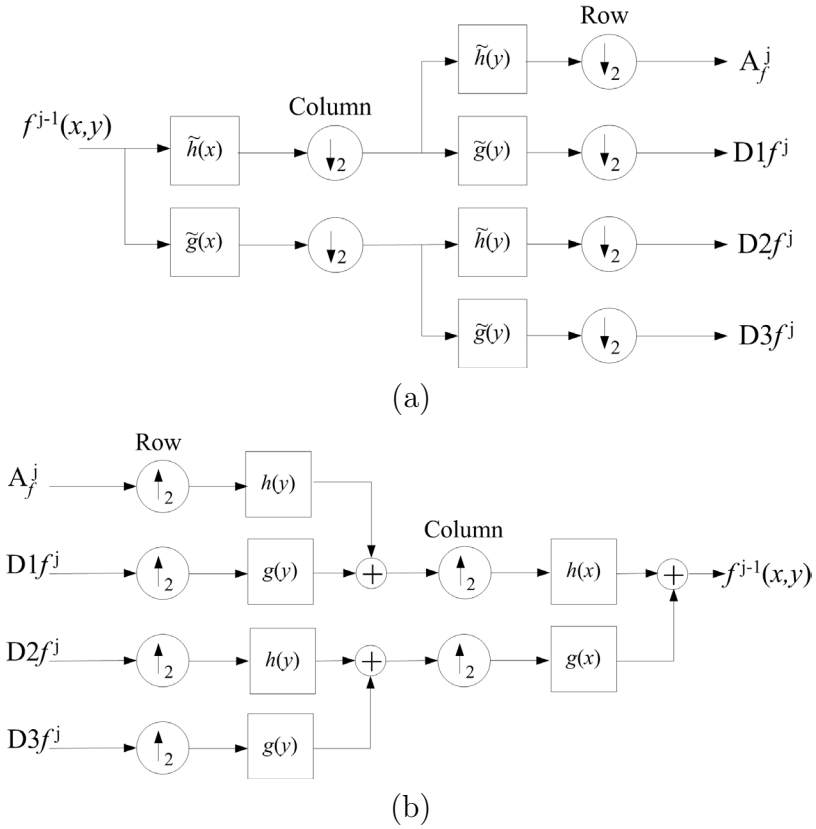


Figure 3.37: Block diagram showing the (a) decomposition and (b) reconstruction of an image using the wavelet transform (Mojsilović et al., 1997).

has been research using the Daubechies wavelet for cloud detection and classification (Tian et al., 1999; Amato et al., 2001) that has yielded a satisfactory result. However, because the scaling and wavelet functions are not symmetric, phases of the Daubechies are not linear. On the contrary, Coiflets are more symmetric compared to the Daubechies wavelets. This leads to the near linear phase characteristics in the Coiflets, though the number of the vanishing moments is equal to one-third of the number of the coefficients, which is less than that of the Daubechies wavelet. By experimental research, Calderbank et al. (1997) suggested that the number of vanishing moments required depends heavily on the nature of the image. In other words, which mother wavelets should be selected depend on the nature of texture in applications. For example, wavelets with more vanishing moments generally perform well with natural and smooth images but not for images with a lot of high frequency components.

Other wavelets with different properties are also available. For example, Symlet is a modified version of Daubechies wavelets that yields an increase in symmetry,

while retaining their relative simplicity (Daubechies, 1990). Instead of the same single wavelet for decomposition and reconstruction, a biorthogonal wavelet designed by Cohen et al. (1992) uses two different wavelets which results in a linear phase filter, which is a necessity in image compression. Therefore, it is the most widely used in the field of wavelet transform coding, and has been chosen as the only candidate in the FBI's fingerprint image compression standards.

Textural features that can be extracted from the wavelet decomposition includes the energy contained in each sub-image. Other statistics are co-occurrence statistics of the detailed sub-images (de Wouwer et al., 1999) and parameters representing the histogram of the wavelet coefficients at each level such as moments of the histogram. For example, The first moment is the mean of the coefficients decomposed by the wavelet. The second central moment is their variance, which describes how similar the coefficients are. The third central moment describes how symmetric the distribution of the coefficients is about the mean. In addition, Mallat (1989) experimentally found that histograms of the detailed sub-images of natural images can be modelled by

$$h(u) = K \exp\left[-\left(\frac{|u|}{\alpha}\right)^{\beta}\right], \tag{3.13}$$

where u is the wavelet coefficients. The parameter α models the width of the histogram peak and β modifies the decreasing rate of the peak. The method to derive both parameters is supplemented in Appendix C.

The results when a db2 wavelet was applied to the IR images of the four events are presented in Figure 3.39. The probability function of A_f coefficients are similar to the histogram of the original images but are slightly smoother. There is an obvious difference between the distribution of cloud at the centre of the cyclone in the four decomposed images and the other types of cloud. It has the most components (wide width of the distribution) in edges along x and y axes, and corners. This corresponds to what we can see from the test images that at the centre of the cyclone, clouds are distributed evenly in both x and y directions. The $D1f$ and $D2f$ distributions also show that the convective cloud has slightly more edges parallel to the x axis than to the y axis. On the contrary, the distributions of stratiform and frontal clouds do not give any preference. This also applies to the wavelet transform of the rain radar images presented in Figure 3.40, in which most of the components are nearly zero. Figure 3.41 compares an effect of using the Daubechies and Coiflets at the same length of filters. The non-zero $D3f$ coefficients when using the Daubechies are higher than those of the Coiflets, which means that the Daubechies wavelet can capture the transitions

in x and y directions better than the Coiflets. This corresponds to prior knowledge that the Daubechies wavelet has more vanishing moments and therefore can better approximate the functions of higher polynomial degrees. That is, the wavelet features can be used to measure the nonlinearity of the cloud texture in conjunction with other types of textural features for rain estimation.

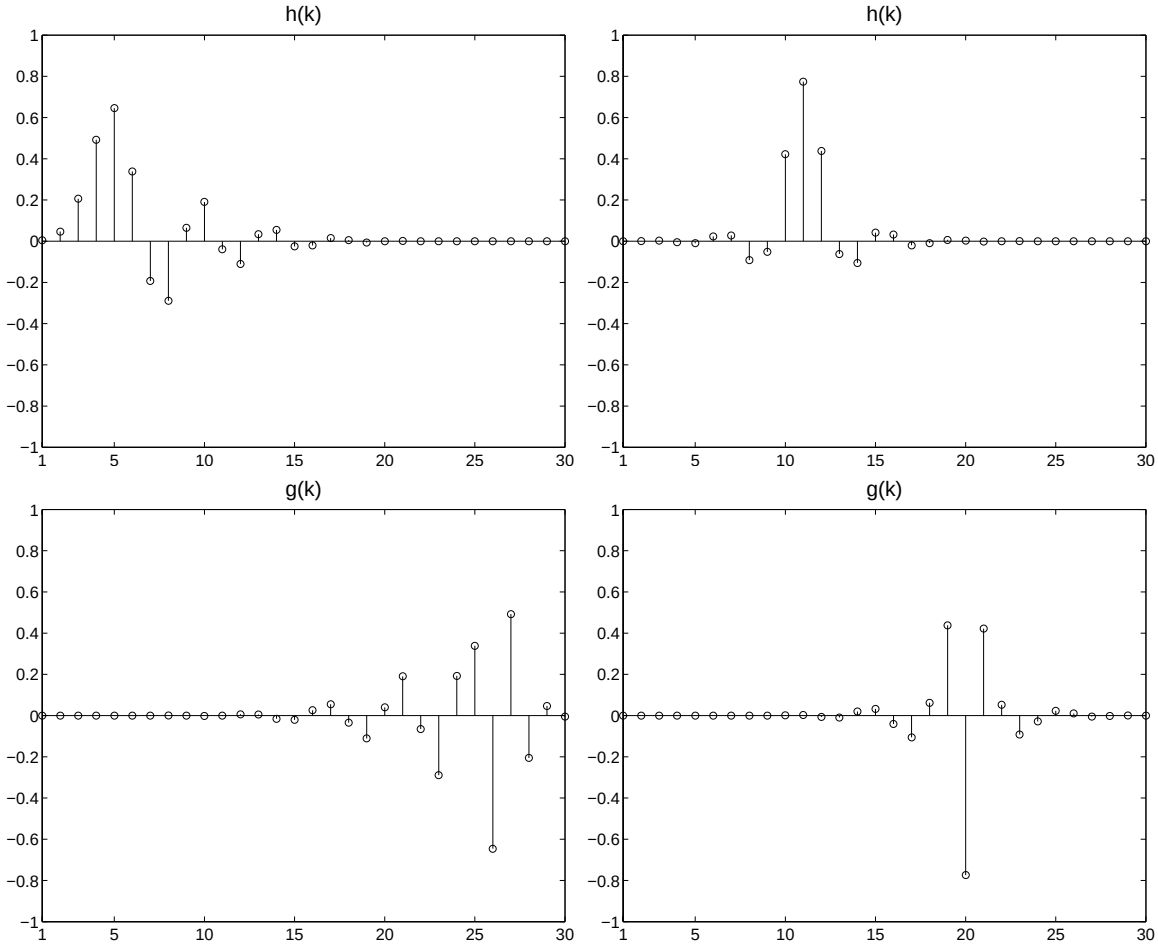


Figure 3.38: Coefficients of filters for (a) Daubechies and (b) Coiflets.

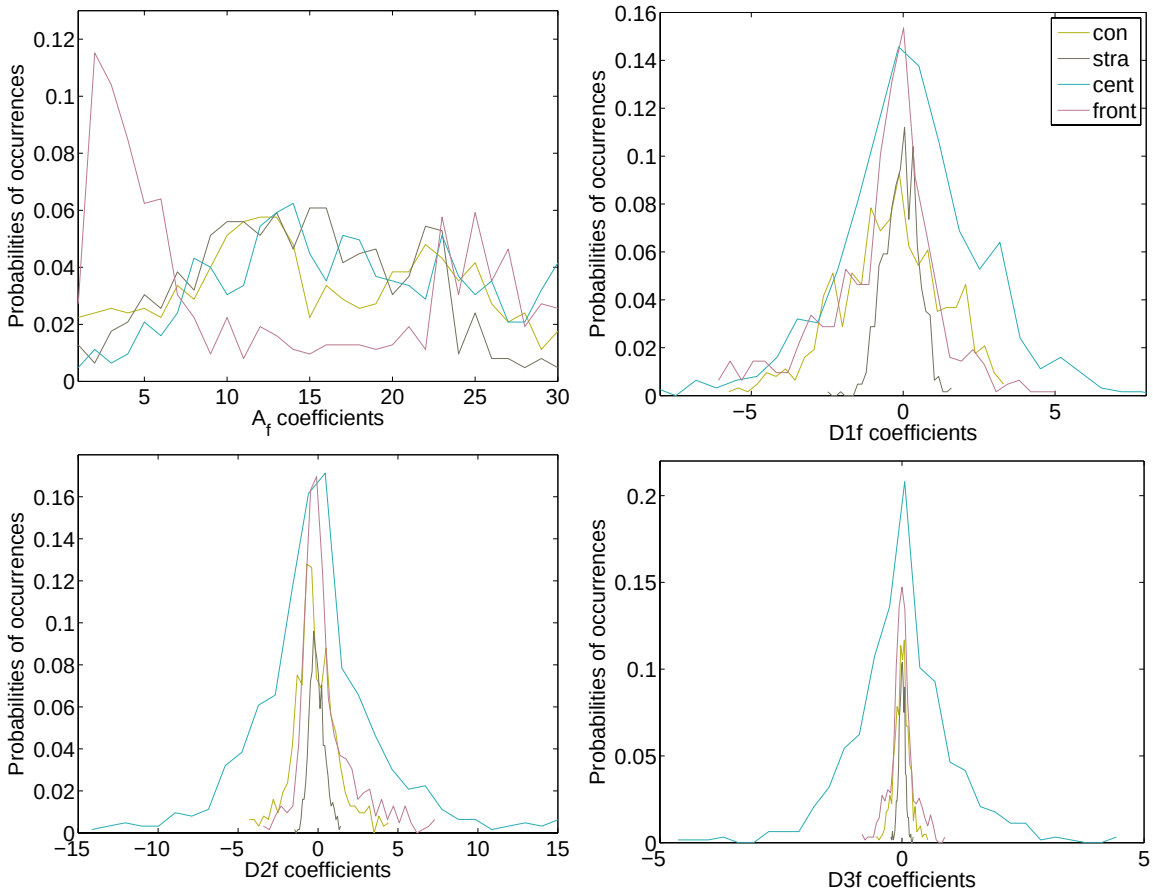


Figure 3.39: Probability functions of db2 coefficients of the first level decomposition of IR images for the four events.

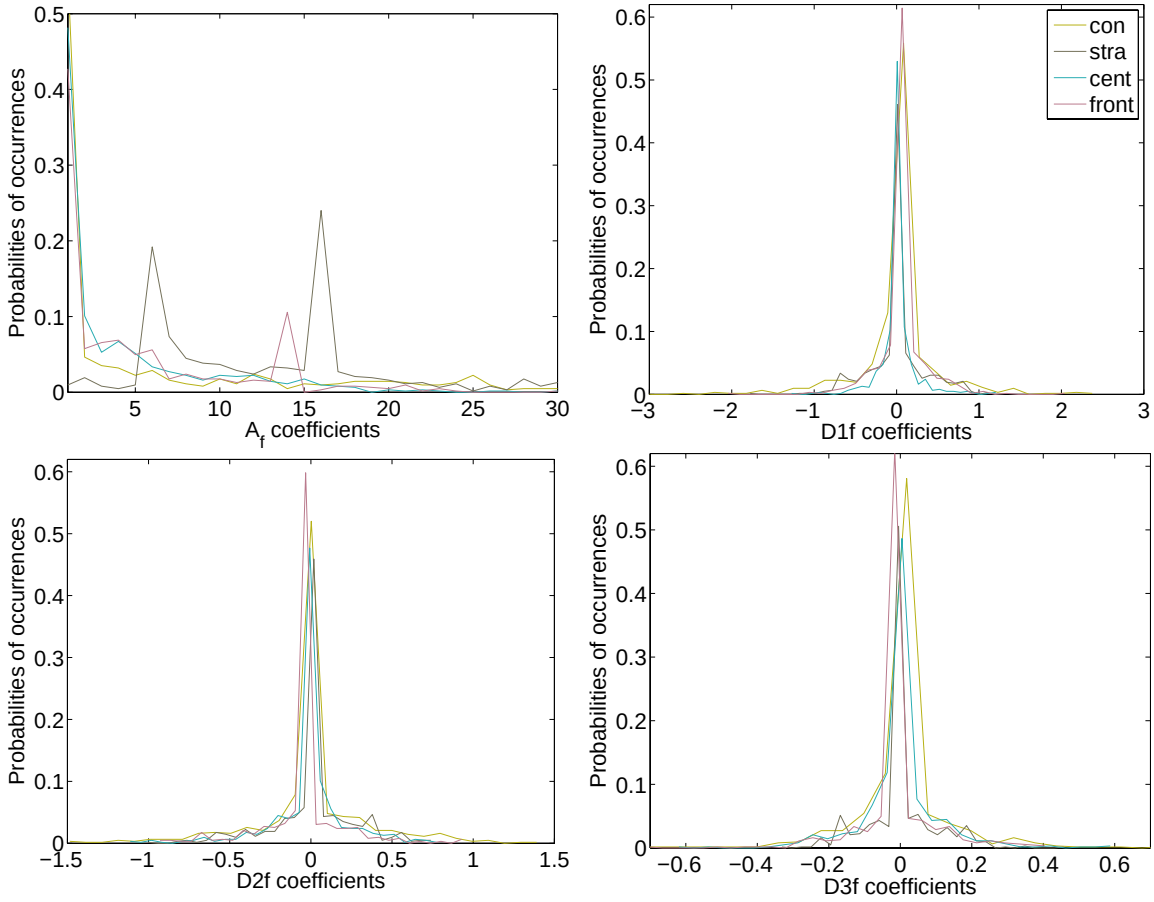
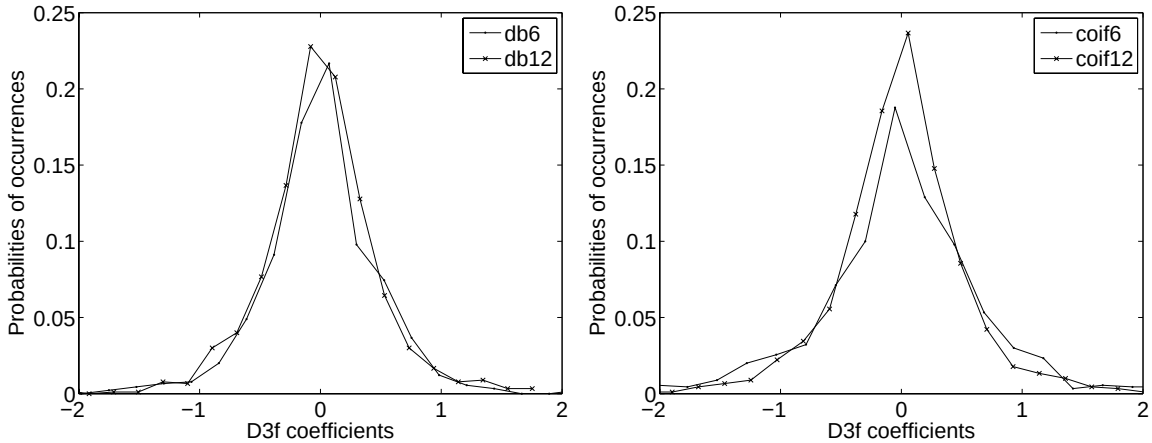


Figure 3.40: Probability functions of db2 coefficients of the first level decomposition of rain radar images for the four events.



(a) (b)

Figure 3.41: Comparison of probability functions of $D3f$ coefficients when (a) db and (b) coif of the same length are applied to the convective IR image.

3.3 The use of shadow

We are able to perceive the three dimensional nature of clouds by observing shadowing effects. For example, at the same solar zenith angle the longer length of shadows implies the higher, and probably thicker, clouds. Cloud shadows can be a useful indicator of the thickness of clouds and therefore provide information on rain. Consequently, cloud shadows may be taken into consideration in rain estimation.

For a flat earth approximation, the cloud base height h can be calculated in terms of the solar zenith angle Z and the length of the cloud shadow L_s by

$$h = \frac{L_s}{\tan Z}. \tag{3.14}$$

However, substantial errors can occur at a position off the satellite nadir (Berendes et al., 1998). For example as shown in Figure 3.42, the satellite sees the edge of the cloud at a distance $d = h \tan\theta$, where θ is the satellite zenith angle at the cloud edge. Therefore, the distance between cloud and shadow is $d + L_s$. As a result, the cloud base height that is seen by the satellite actually is

$$\begin{aligned} \acute{h} &= \frac{d + L_s}{\tan Z}, \\ h + \Delta h &= \frac{h \tan\theta + h \tan Z}{\tan Z}, \end{aligned}$$

which is higher than the actual cloud base height. The error $\Delta h/h$ is given by $\frac{\tan\theta}{\tan Z}$. It increases with increasing observation angle θ and decreasing solar zenith angle Z. The relative error of the derived cloud base height along the central meridian (0° longitude) as a function of latitudes for Meteosat7 is presented in Figure 3.43. It can be noticed that the error becomes worse at the higher latitude and the series of error lines also demonstrates that the worst error occurs at noon. Thus, using Equation (3.14) to derive the cloud base height directly can lead to substantial errors.

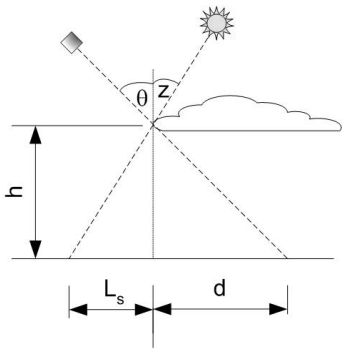


Figure 3.42: The geometry of a cloud, a satellite and the sun.

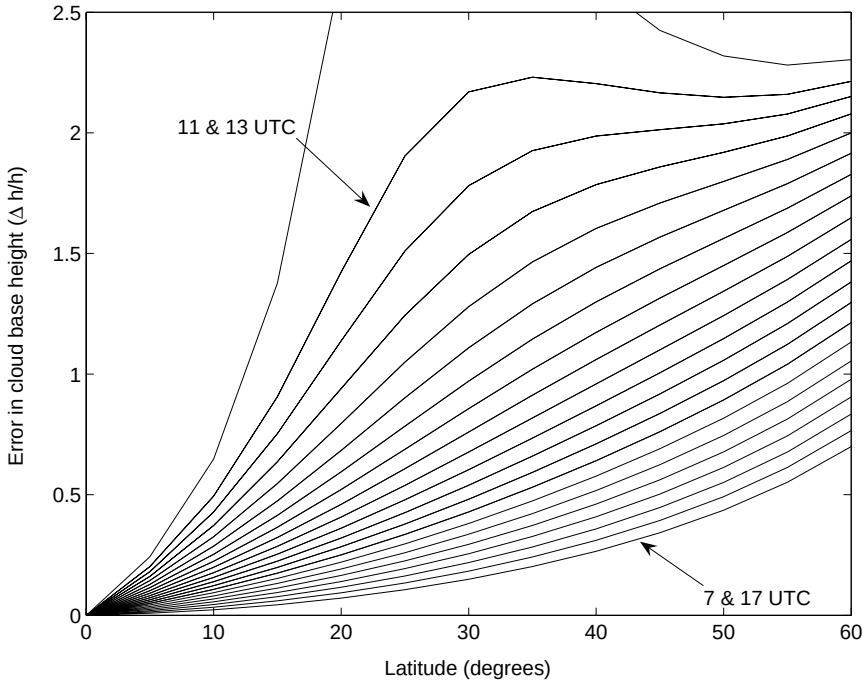


Figure 3.43: The relative error of cloud base heights derived from Equation 3.14 on 23 June 2004 from 0700 to 1700 UTC. Each line is fifteen minutes apart.

Alternatively, the shape from shading (SFS) technique can be used to reconstruct the surface shape from gradual variations of shading in the image (Ikeuchi and Horn, 1981). It attempts to simulate the mechanisms which human's eyes and brains actually use. The recovered shape might be expressed either in depth $Z(x, y)$, surface gradient $(p, q) = (\partial Z/\partial x, \partial Z/\partial y)$ which is the rate of change of depth in the x and y directions, or surface normal $\vec{N} = (p, q, 1)^T$.

To understand the SFS algorithm, the first step is to study the image formation process. A Lambertian model is the simplest explanation of the process. It assumes that the grey level at each pixel depends only on the light source direction and surface normal. Assuming that the surface is illuminated by a distant point source, the intensity of a pixel in an image is defined by

$$I(x, y) = \varrho \vec{N} \cdot \vec{L} = \varrho \frac{pl_1 + ql_2 + l_3}{\sqrt{1 + p^2 + q^2}} \tag{3.15}$$

where ϱ is the surface albedo which is a fraction of the incident light that is reflected and $\vec{L} = (l_1, l_2, l_3)$ is the light source direction in x, y, z coordinates with z axis normal to the image plane, illustrated in Figure 3.44.

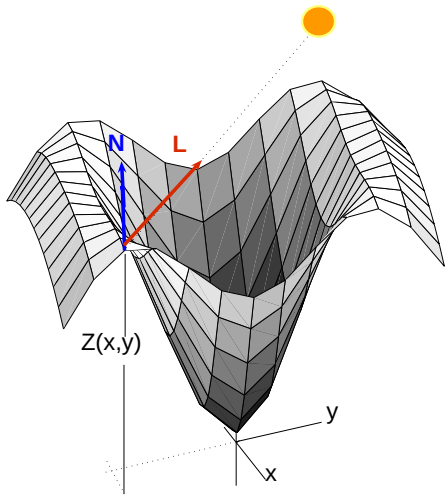


Figure 3.44: Diagram of an object and a light source of an image plane.

Assuming that ϱ and $\vec{L}$ are known, the surface gradient (p, q) are the two unknowns that the algorithm tries to determine from each pixel intensity $I(x, y)$. Therefore, SFS is an underdetermined problem in nature. In order to get a unique solution, if there is any at all, some constraints (Ikeuchi and Horn, 1981) are needed. These include

- brightness constraint

$$\iint (I - R)^2 dxdy, \tag{3.16}$$

- smoothness constraint

$$\iint (p_x^2 + p_y^2 + q_x^2 + q_y^2) dxdy, \tag{3.17}$$

- integrability constraint

$$\iint (R_x - p)^2 + (R_y - q)^2 dxdy, \tag{3.18}$$

where the subscript x and y denote the partial derivative in x and y directions. The brightness constraint ensures that the reconstructed surface (R) produces the same brightness as the input image. The smoothness constraint, in terms of the second order surface gradients, helps in reconstruction of a smooth surface. The integrability constraint ensures the recovered surfaces are valid, i.e. $Z_{x,y} = Z_{y,x}$, and correspond to the input image. Thus, by minimising

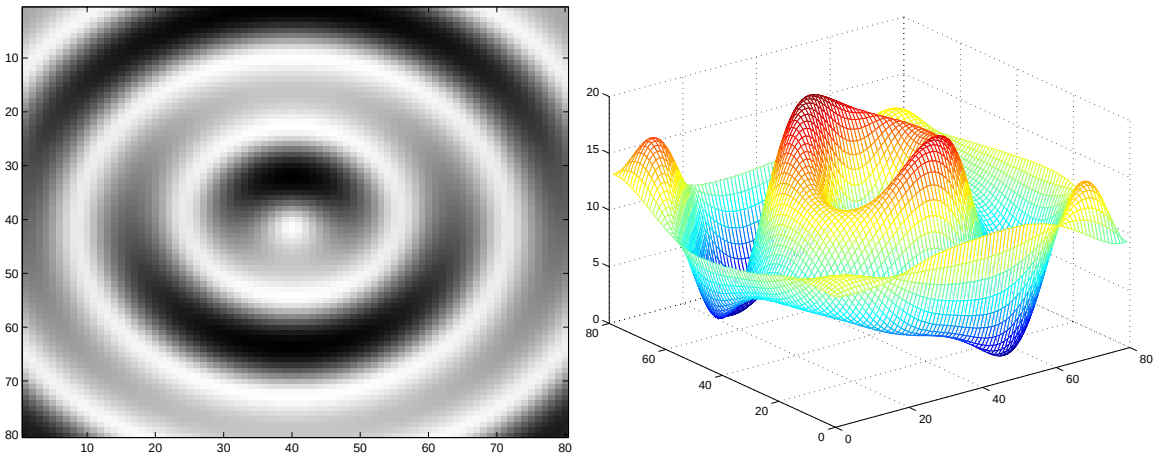
$$\iint \left\{ (I - R)^2 + \lambda_1 (p_x^2 + p_y^2 + q_x^2 + q_y^2) + \lambda_2 \left[(R_x - p)^2 + (R_y - q)^2 \right] \right\} dxdy, \tag{3.19}$$

where λ_1 and λ_2 are Lagrange multipliers, the shape R of the object can be recovered.

The assumptions made for the SFS problems are: 1) the surface is Lambertian, i.e. it reflects light in all directions, 2) the surface is illuminated by a distant point source, and 3) the position of the light source is known (Rajabi and Blais, 2001). Dealing with VIS satellite images, the light source is the sun which is not only far away from the scene, but also has a known direction. Therefore, the first assumption is the only one which is questionable. To this point, the intensity in the VIS channels is the reflectance of light from cloud particles, which can be assumed to have a spheroid shape and therefore reflect the light uniformly in all directions. In this regard, the Lambertian surface assumption can be acceptable when considering clouds.

Next, the performance of the algorithm is studied by the application to a synthetic image of an object with known-shape. Its light source direction $L = (0, 0.5, 0.5)$ is in the north direction of the object. The image and its recovered shape are shown in Figure 3.45. The object, which is in fact a circular sinusoidal wave, can be well

recovered by the algorithm with slightly distorted shape on top. The error is an effect of long shadows that occur at the first ridge. When tested with a VIS image of a mesoscale convective system in Figure 3.46, the cloud depth can be recovered but is distorted in shape. Similar to the result of the synthetic image that some shaded areas cannot be resolved, the recovered cloud height in some places are incorrectly determined to be lower than land. A reason for this is that a single $\vec{L}$ is used, contrary to the fact that each point receives sun light differently due to the earth's curvature. Moreover, the method derives the height relative to the background but clouds do not usually attach to each other. Special constraints must be included for such cases. In addition, in order to minimise Equation (3.19), an appropriate initial depth and the step sizes of the minimisation process are required.

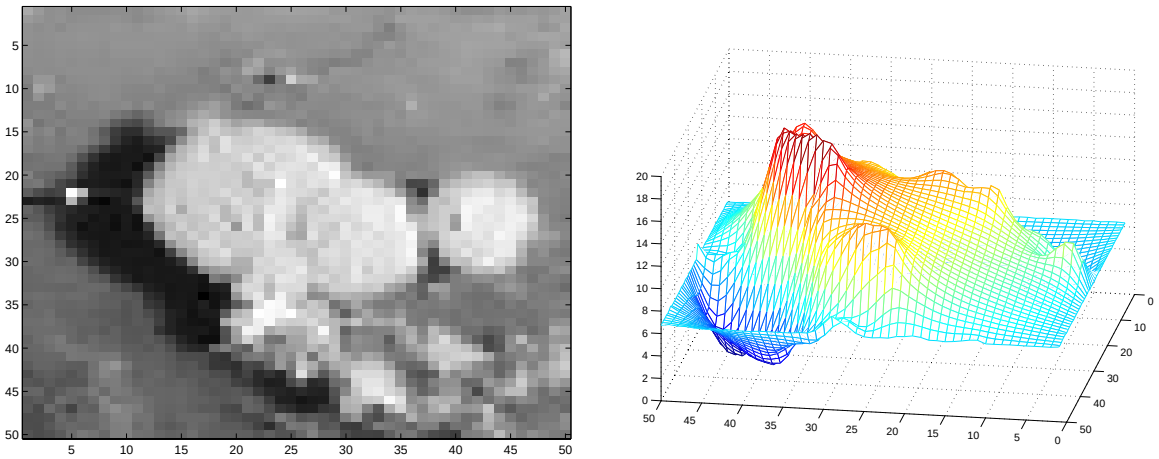


(a) (b)

Figure 3.45: A result after 200 iterations when the shape from shading approach (Ikeuchi and Horn, 1981) is applied to a synthesis image (a).

3.4 Conclusion

The relationship between the distribution of rain rates and the corresponding satellite images is complex and the use of the pixels alone leads to considerable error in the estimation of rainfall. Other pieces of information that can be extracted from the satellite images are textural features. These features are expected to enhance the spectral information for rain estimation. In this chapter, an analysis of textural features has been done in order to find out the features that uniquely describe cloud textures and how these features relate either to the texture of the corresponding rain rates or to the rain rates themselves. Four meteorological events which commonly relate to a



(a) (b)

Figure 3.46: The result after 200 iterations when the shape from shading approach (Ikeuchi and Horn, 1981) is applied to an image of a thunderstorm in (a).

mid-latitude cyclone are selected for the analysis. These events are a convective cloud, stratiform cloud, cloud at the centre of the cyclone and a frontal cloud.

The first feature that has been studied is a co-occurrence matrix which indicates how frequently two pixels with a specific grey level and being separated by a given distance appear in an image. The statistics of the matrix can refer to an arrangement of cloud texture. For example, entropy reveals that the convective cloud is the most disorganised—there is no obvious relation between any two pixels in the image as its entropy is higher than that of other events. The sharp gradient of the frontal cloud can also be captured by contrast of the satellite images. It is the best feature to distinguish the four events and its relationship to the distance between pixels also resembles that of the rain radar images. These statistics are also invariant to rotation and robust to noise. However, the major drawback of the co-occurrence matrix is zero entries within the matrix, wasting memories in the computational process. The problem can be alleviated by selecting an appropriate number of bins for the intensities to be grouped in. The number should be neither too coarse to capture the details of a texture nor so fine that there are many zero entries in the matrix. Several rules of thumb can be used as a guidance for the selection.

Instead of using individual satellite channels, the co-occurrence matrix can also be applied to multi-spectral images. The combination of IR, VIS and WV images into the RGB colour model is based on an observation that all the three channels give different pieces of information and contribute to the relation to rain rates differently. The values of ASM and entropy of the RGB co-occurrence matrix can better distinguish

the four events than those of the individual channels. However, the computation of the 3-dimensional matrix requires a large storage to process. Alternatively, the HSV (Hue, Saturation, Value) model can be used. Hue represents the wavelength of a colour whereas saturation represents an intensity of colour. Value refers to the intensity of the image and does not depend on colours. The HSV model treats the three primary colours as dependent variables. This can be beneficial to the rain estimation problem as each channel is not fully independent from others. The ASM and entropy of the HSV co-occurrence matrix are good at discriminating each meteorological event, similar to those of the RGB model, but much less memory is required.

The next texture descriptor is by statistics of a difference matrix, which is based on absolute grey level differences between pairs of pixels. The statistics are a function of the displacement between the pair (d) and the direction of a line crossing the pair. Similarities between the cloud texture and the rain texture of these statistics can be observed. For example, the frontal cloud which possesses apparent gradients has the highest/lowest statistics in a certain direction, resembling those of the rain image. However, the experiments found that the co-occurrence and difference statistics that give the highest correlation to the rain rates occur at a particular value of d. The maximum or minimum of the statistics and their corresponding displacement might be used in practice in order to avoid selecting an appropriate d. In addition, to make the statistics invariant to the rotation of the cyclones, a direction with respect to the centre of the cyclone can also be used.

Next, the VIS and IR intensity transitions which may be referred to cloud evolution or different cloud height of two clouds are extracted by means of the Radon transform. The Radon transform reduces the dimensionality of analysing the transitions in all directions by projecting the image intensities onto a principal axis and then extracting edges along the axis. However, applying the Radon transform directly to satellite images gives no significant results along the projection lines due to their smoothness. Therefore, Sobel's filter (Gonzalez et al., 2003) is applied to the images prior to the transformation to enhance the edges of the cloud objects. The Radon transform of the filtered images reveals the similarities of orientation of texture to those in the radar images with the correlation coefficients of 0.71 and 0.66 for the case of the convective cloud and at the centre of the vortex. The texture orientation extracted from the IR images is also more relevant to the rain rates than the spectral information. This encourages the use of Sobel's filter and the Radon transform as another textural feature for rain estimation.

The next features are energy and entropy of the distribution in the Fourier space which represents spatial frequencies of an image by a use of an infinite series of trigonometric functions. However, windowing and the use of Sobel's filter are required prior to the Fourier transform to get rid of discontinuities at edges of the image and smoothing effects which can result in cross-shaped ripples or rectangular grids.

Alternatively, the texture can be described by how its frequency contents evolve in time by the use of the wavelet transform. This spatio-temporal analysis decomposes an image into four sub-images that correspond to the low-frequency content of the original image, vertical high frequencies (horizontal edges), and horizontal frequencies (vertical edges) and high frequencies in both directions such as corners. There is a number of mother wavelets available but which to select depends heavily on the nature of texture appearing in an image.

Another piece of information is cloud shadows in VIS images which implies the thickness and height of clouds. Using this information can possibly identify structure and organisation of the cloud and relate them to rainfall. Two methods of shadow extraction were analysed in this chapter: (1) by assuming a flat earth surface and (2) shape from shading. The first technique calculates the cloud base height in terms of the solar zenith angle and the length of the cloud shadow. However, substantial errors can occur at a position off the satellite nadir. The second technique assumes that the reconstructed surface is Lambertian and is illuminated by a known-position distant point source. All of the assumptions are acceptable when applying to VIS images. The technique was applied to a VIS image of a mesoscale convective system. The cloud depth can be recovered but is distorted in shape. A reason for this is that a single light source is used, contrary to the fact that each point receives sun light differently due to the earth's curvature. Special constraints must be included for such cases. In addition, an appropriate initial depth and the step sizes of the minimisation process are required.

In summary, the textural features have some similarities to either the texture of the corresponding rain rates or the rain rates themselves. However, in this chapter these features were analysed separately. In the next chapter, a systematic analysis will be done to rank textural features according to their correlation to rain rates. Then, the best ranked features will be used for rain estimation.

Chapter 4

Rain Estimation

In this chapter, the textural features that have been analysed in Chapter 3 are related to rain rates. The features used for rain estimation are co-occurrence statistics, difference statistics, Radon transform, Fourier transform, wavelet transform to include statistical, structural, frequency and spatio-frequency characteristics of the cloud texture. In addition, the combinations of the three channels: IR, WV and VIS are also used. The combination is performed through a false image by associating the IR, VIS and WV to the red, green and blue colour bands, respectively. The hue, saturation and value of the false image are then used to derive the statistical features in the same way as individual channels. In Section 4.1, details of a mid-latitude cyclone that has been selected for the study are described. Then, the best features are selected regarding their relation to rain rates of the cyclone. The feature selection is presented in Section 4.2. The selected features are then used for rain estimation which is a combination of classification and function approximation to gain robust but continuous rain estimates. Details of the estimation are presented in Section 4.3.

4.1 Data

A set of Meteosat satellite images corresponding to a summer mid-latitude cyclone has been used for the study of rain estimation using cloud texture. The event occurred in mid-June 2004, starting on the 22nd and dissipating on the 24th. On the 22nd, the cloud and rain spread north-easterly during the day and by midnight were affecting all of England, Wales and Ireland, and parts off southern Scotland. The rain was associated with an unusually low depression for the summer with the mean sea level pressure down to 983 *mb* just of southern Ireland by midnight. The rain was heavy in places with gusts up to 40 knots in exposed places. The low pressure continued

to dominate the weather of the British Isles on the 23rd — heavy rain spread into northern Ireland and southern Scotland overnight and widespread falls occurred over southern parts of England. As the low pressure became slow-moving off the north-eastern coast of England, showers followed the rain. During this time, there were reports of landslides in Truro, floods in parts of Plymouth, power break-downs in Kent and Essex. Around the coasts of England, there were gusts of $60mph$ recorded, causing further damage to Brighton's West Pier, the Severn Crossing closed and ferry services across the English Channel suspended. On the 24th, the low pressure was moving over the North Sea. Showers developed in western areas which eventually died down in the evening.

This summer mid-latitude cyclone has been selected for the study due to the availability of its rain information. It was captured well by the Radarnet which has a coverage over the British Isles and western Europe, especially from the midnight of the 22nd to the late morning of the 23rd when the depression was growing to the mature stage. In addition, the cyclone also contains all the four meteorological processes investigated in the previous chapter. During the day of the 23rd, the visible images were also available for testing the multi-spectral technique which has been developed in the previous chapter. As shown in the pressure charts presented in Figure 4.1, the depression was a single cold-front dominated system with no other depressions nearby at the time. Its life cycle followed the theoretical description of the Norwegian model[1] with no disturbance from other systems that may make the analysis too complex. The weather analysis from SatRep[2] on 23 June 2004 at 0000 and 1200 UTC is presented in Figure 4.2.

[1] Details in Appendix A.

[2] a written report containing information of the IR Meteosat image in combination with NWP output and observations in terms of Conceptual Models, contributed by Austrian Meteorological Institute (ZAMG), Royal Netherlands Meteorological Institute (KNMI), Finnish Meteorological Institute (FMI) and European Organisation for the Exploitation of Meteorological Satellites (Eumetsat).

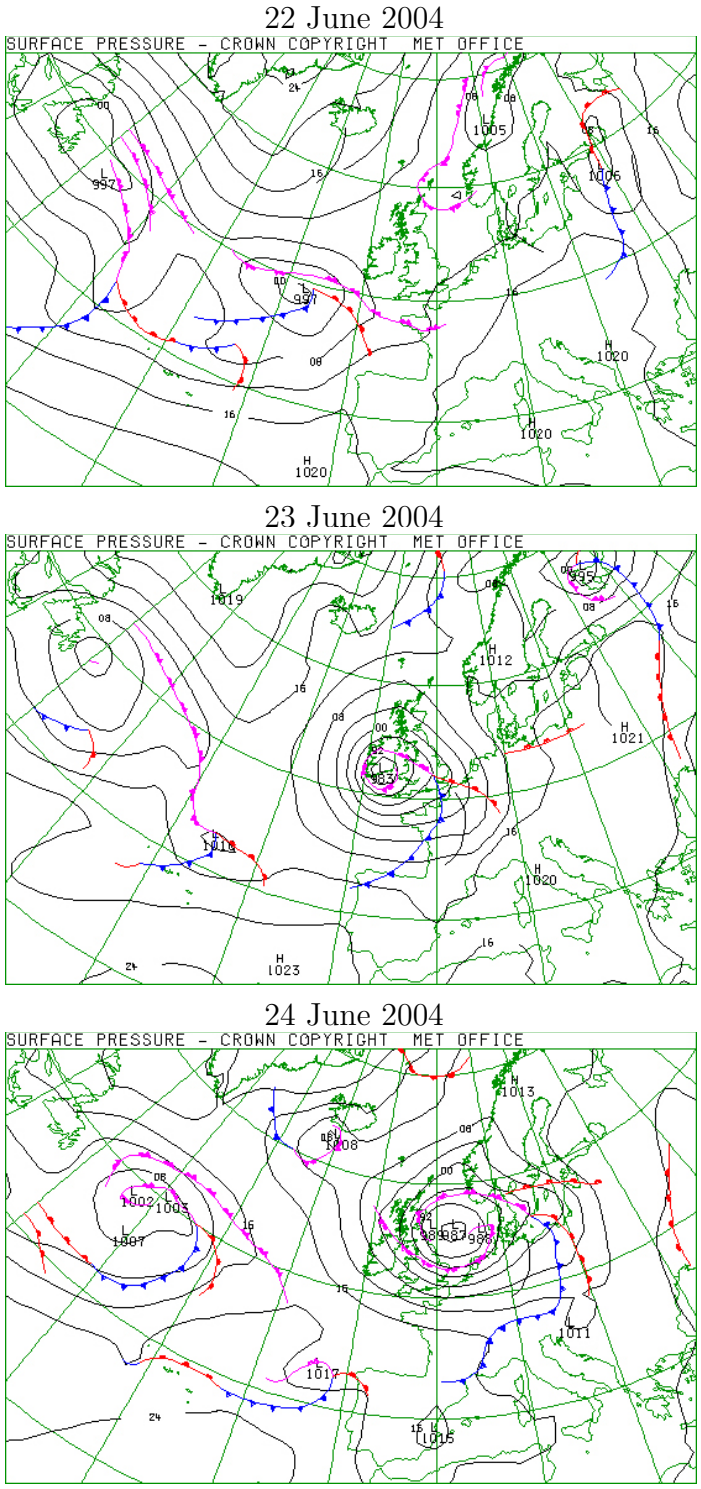
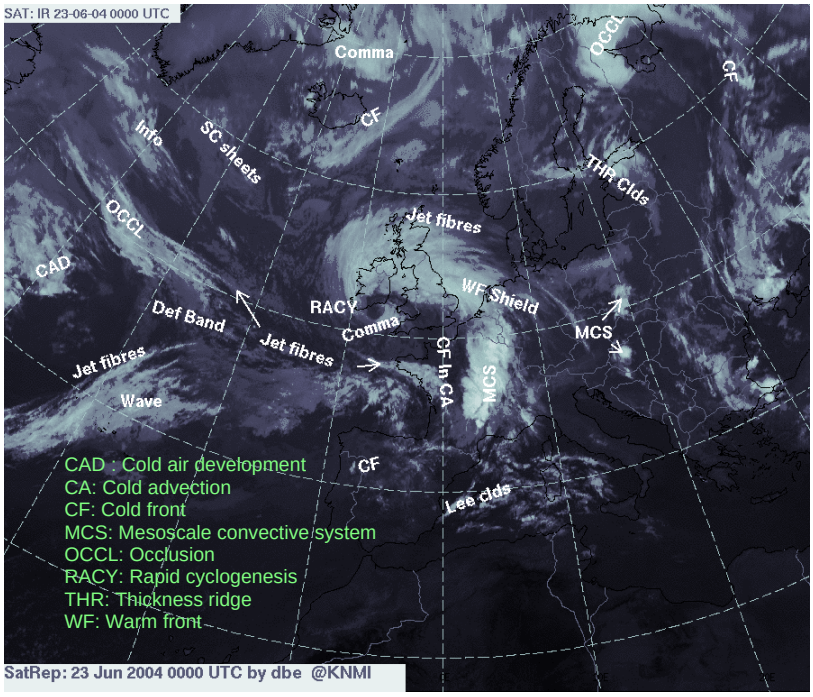
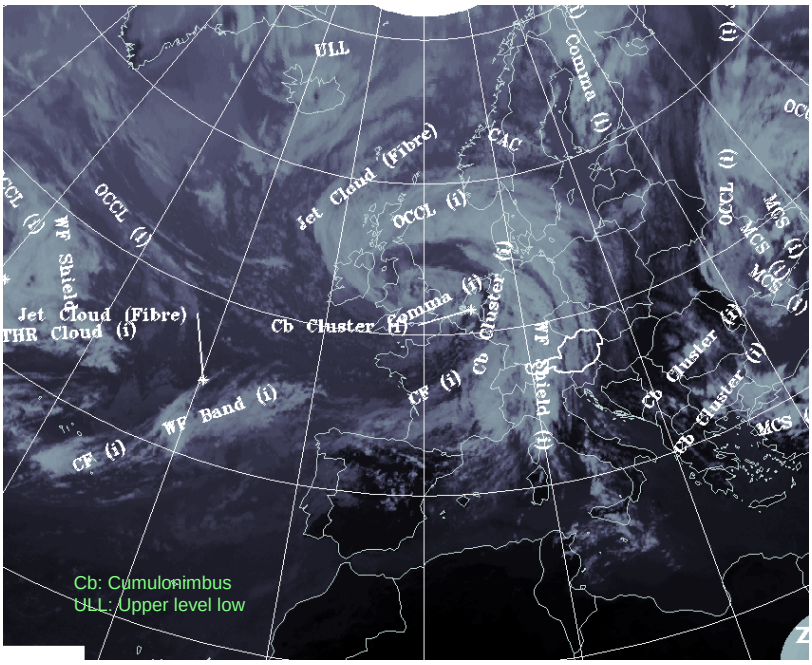


Figure 4.1: Pressure charts of the mid-latitude cyclone on 22-24 June 2004. Source: British Meteorological Office.



(a)



(b)

Figure 4.2: Weather analysis over Europe on 23 June 2004 at (a) 0000 UTC and (b) 1200 UTC. Source: Satrep `http://knmi.nl/satrep`.

The satellite observations were reprojected to polar stereographic projection to match that of the Radarnet at the resolution of 5×5 km per pixel. Next, the texture information of these selected pixels was extracted into features. A standardisation has been applied to the features to avoid biases between them. Each feature is rescaled by using the following equation:

$$\mathbf{z} = \frac{\mathbf{x} - \mu_x}{\sigma_x}, \tag{4.1}$$

where $\mathbf{z}$ is the standardised feature, μ_x and σ_x^2 are the mean and variance vectors of a feature vector $\mathbf{x}$. After standardisation, each feature has a mean of 0 and variance of 1. In order to find the relationship to rain rates, these features can be put into a classifier directly. However, large dimensionality of the input features and limited data samples can affect the classifier's performance and the computational complexity. A careful feature selection can alleviate the severity of the problem (Liu et al., 2005). This process is introduced in the next section. The selected features are then used for the rain estimation. The described process to prepare the data for the rain estimation is summarised in Figure 4.3.

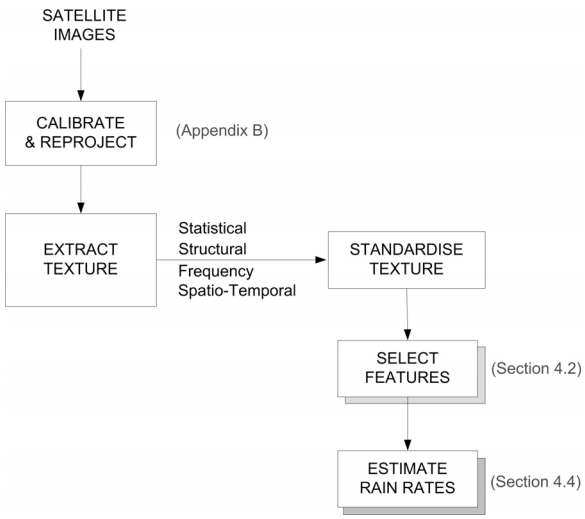


Figure 4.3: Diagram of a data pre-processing for rain estimation.

4.2 Feature selection

The computational requirements of classification are generally positively correlated with the number of input features. Accordingly, the feature selection is used to reduce the data and amount of computation by assessing how well the features are separated into classes or ranking their relevance. The selection can remove irrelevant and redundant features and reduce the amount of data needed for learning and thus increase the estimator's comprehensibility. That is, the feature selection is expected to reduce computational cost, focus on a relevant subset of feature, hence provide helpful meteorological knowledge. Two approaches can be used as a feature selection: dimensionality reduction and subset selection.

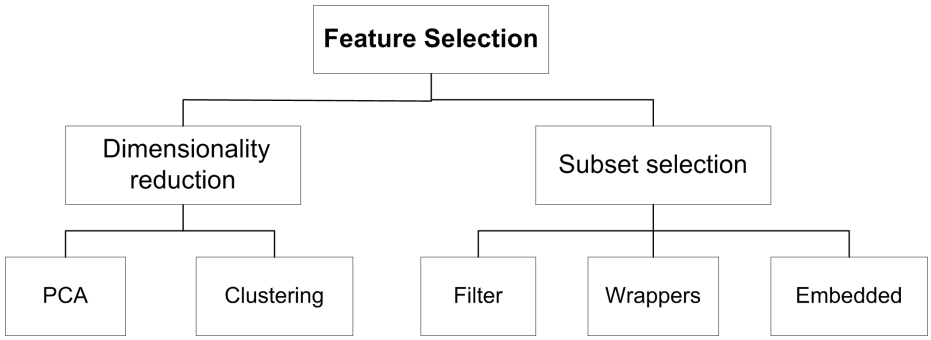


Figure 4.4: The two approaches of the feature selection mentioned in this section.

The first approach is to project the original feature vectors onto a space of smaller dimensionality on account of a covariance of the features or correlation among them. This approach is sometimes called feature extraction or feature reconstruction; it creates new features which are a combination of the original features. The most commonly used feature extraction among remote sensing and pattern recognition communities is the Principal Components Analysis (PCA) or the Karhunen-Loéve expansion (Richards and Jia, 1999). It projects the input features onto the eigenvectors of a covariance matrix (which is the variances between each feature dimension). The new data set spanning these eigenvectors have no dependencies between dimensions. Then, the least significant dimensions (least eigenvalue/least variances) are discarded. Therefore, the dimensionality of the input space can be reduced. However, dealing with a large dimension space especially with excessive samples in each dimension can be impractical. In addition, a great emphasis is usually placed on those features with the greatest variability. This is in contrast to the requirement of the classification, which is more concerned with maximising the separability between classes.

Another method is to exploit the clustering techniques by using similarity/distance measures in the input feature space to select a sub-feature dimension that gives maximum separability. An example of such method is the k-means algorithm which tries to cluster features into k separated Gaussian distributions. Alternatively, the partitioning can also be performed by maximising/minimising a criterion function that measures clustering quality of all possible partitions of the feature space. Examples of the function are the similarity of the features and their distribution within a cluster. However, its brute-force search to match a cluster of features with one of the classes may cause errors if the cluster corresponds to more than one class. Such cases, e.g. a cloud with a particular temperature can give either rain or no rain, have already been discussed in the previous chapter.

The second approach selects the most significance feature subsets and therefore retains the original meaning of the input features. This can be done in three ways (Guyon and Elisseeff, 2003):

- Filter method
- Wrapper method
- Embedded method

The latter two methods aim at selecting the optimal sets of features. The wrapper method, for example, utilises the learning machine of interest as a black box to score subsets of features according to their discriminant abilities. On the contrary, the embedded method performs the selection in the process of training and is usually specific to a given learning machine. In practice, the features are not ordered and the best feature set must be found from among all possible feature subsets. On the other hand, the filter method evaluates a measure of relevance of all the features and ranks them on the basis of that measure. The selection is not correlated to any specific classifiers and therefore the selected features have better generalisation. This method is widely used as a principal selection mechanism because of its simplicity, scalability and, most importantly, a good empirical success (Guyon and Elisseeff, 2003).

Computationally, the method requires only the computation and the sorting of scores. Statistically, it produces less variances than other methods and therefore can be used for any numbers of the sample sets. The simplest scoring function is the correlation criteria, which has been already introduced in Section 3.1. Referring to Equation (3.1), the correlation coefficient $r_{\mathbf{x},\mathbf{y}}$ describes the *linear* relation between

$\mathbf{x}$ and $\mathbf{y}$. Therefore, using $r^2_{\mathbf{x},\mathbf{y}}$ or $|r_{\mathbf{x},\mathbf{y}}|$ as a scoring function can sort the features according to the goodness of linear fit of the individual features. To lift the linear relationship restriction, the non-linear preprocessing can be used before calculating the correlation coefficient.

Next, the absolute value of a correlation coefficient $|r|$ is used as a scoring function to sort each textural features by its relevant to rain rates. First, all the samples are classified into three ranges of their corresponding rain rates: (0,2), [2,8), [8,∞) mm/h. The ranges are defined by following a guideline from NOAA's National Weather Service[3]. The correlation coefficients are calculated by comparing the features and the rain rates within a 21×21 pixel window. Figure 4.5 shows $|r|$ between homogeneity of an IR co-occurrence matrix and rain rates of 50 samples as a function of displacements between a pair of pixels. Samples that show no relevance to the rain rates such as the 5th, 7th, 9th sample are removed before computing the averaged correlation coefficients. The zero correlation is due to the homogeneity of the feature within the image window. These samples can be seen as blue strips in Figure 4.5 where $|r|$ of these samples becomes zero at all displacements.

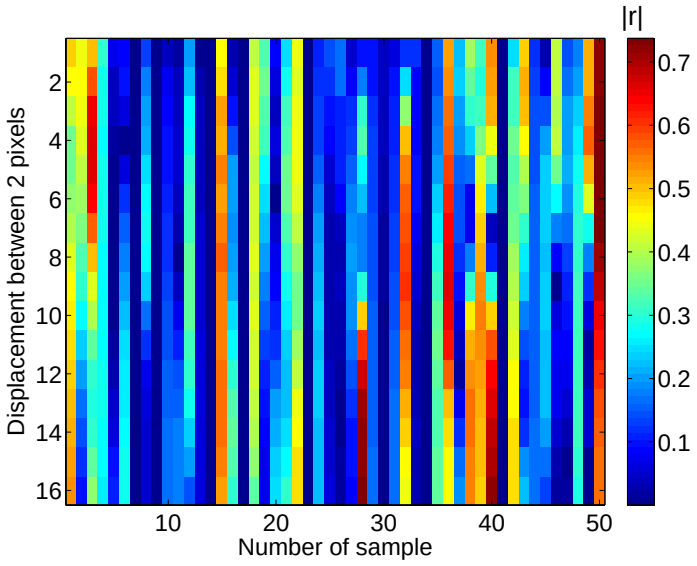


Figure 4.5: Absolute values of correlation coefficients between homogeneity of an IR co-occurrence matrix and moderate rain rates.

There are two ways of considering how the textural features correlate to the rain rates: (1) by the highest averaged r and (2) the highest percentage of samples that have

[3]NOAA's National Weather Service defines a rain rate ≤ 0.1 inches per hour as light rain, $0.11-0.3$ inches per hour as moderate rain and > 0.3 inches per hour as heavy rain.

high correlation to rain rates. These two parameters need to be used in combination to guarantee that a high r_{av} is not a result of outliers within the feature space. Therefore, the following scoring function has been used to combine both effects, i.e.

$$\varsigma = \frac{1}{1+k}(|r_{av}| + kN), \tag{4.2}$$

where N is a proportion of samples that have $|r| > 0.4$, which suggests that the feature and its corresponding rain rate within the window size of 21x21 correlates to each other according to Haldar and Mahadevan (2000), and $k \in (0, 1)$ is a constant to weigh the significance of N. The function ranges from 0 to 1. Figure 4.6 illustrates the scoring function when $k = 1$. The figure shows that the feature with a higher r_{av} might not be selected if N is too low, which can happen when the samples have a large deviation in r values.

Next, the scoring function of each textural feature is calculated. In the following feature selections, the value of k is chosen to be 1 to balance the significance of r_{av} and the number of high rain-correlated samples. The parameters can be adjusted to suit conditions in other applications. The scores are then used to rank the features of four types: statistical, structural, frequency and spatio-temporal.

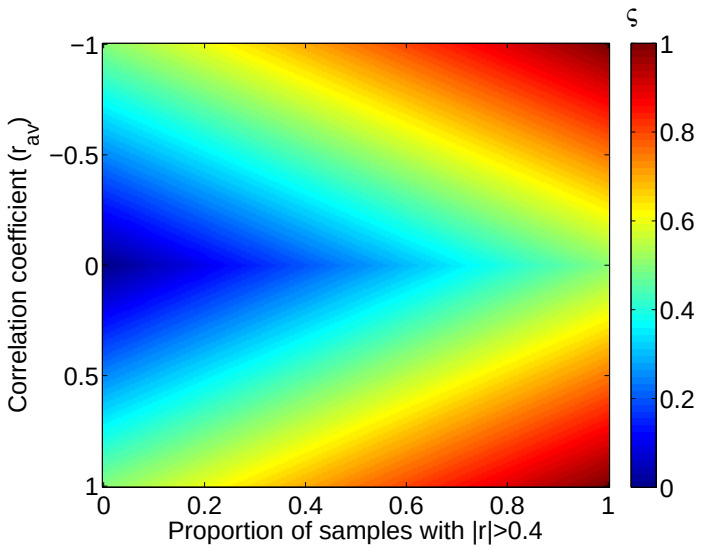


Figure 4.6: Scoring function as a function of r_{av} and N.

Statistical analysis

The statistical features used in this analysis are as follows.

Methods	**I**mage Type	**S**tatistics	**P**arameter
Co-occurrence (C_d)	IR, WV, VIS H, S, V	ASM Entropy Contrast Homogeneity Correlation	d=1-16 pixels $d_{\max}$, $d_{\min}$ $\max C_d$, $\min C_d$
Grey-level difference ($f_{d,\theta}$)	IR, WV, VIS H, S, V	Mean STD Entropy	d=1-16 pixels θ_d=0^o, 45^o, 90^o θ_T, θ_R

- θ_R and θ_T are angles between a line crossing the pixel and the centre of cyclone and x axis and its tangent line.
- θ_d is an angle between two pixels in the image window.

The parameters for considering the correlation of the co-occurrence features (C_d) and the rain rates are the distance between any two pixels in the considered 21×21 window (d) from 1-16 pixels, the highest and lowest values of these statistics ($\max C_d$, $\min C_d$) and the distance corresponding to these extreme values ($d_{\max}$ and $d_{\min}$). Likewise, the same distance parameter d = 1-16 pixels are used to consider the correlation of the difference ($f_{d,\theta}$) statistics. The statistics are derived in five directions i.e. 0, 45,90 degrees from the x axis and in the directions that correspond to the radius from the pixel to the centre of cyclone and its tangent line (θ_R and θ_T). The scoring function is calculated using these parameters for IR, VIS and WV images and for the HSV colour model constructed from these images.

Table 4.1 shows the statistical features that give the highest scoring function when $k = 1$ for light, moderate and heavy rain rates together with their r_{av} and N. For monochrome images, the textural features derived from VIS images are more correlated to light rain rates than other kinds of images. In addition, it can be observed that the difference statistics in θ_R and θ_T directions (see Figure 3.27) are in the top ranks. The statistics of these directions are derived in relation to the centre of the cyclone and therefore are invariant to the rotation of the cyclone and its location. However, the best scores belong to the statistics of a specific direction and displacement. For moderate rain rates, the statistics from a co-occurrence matrix dominate the top ranks. The statistics with the highest scoring function also have the number of pixels that have $|r| > 0.4$ as high as 46% of the total samples. That is, the relation between these features and their corresponding rain rates are likely to be linear.

On the other hand, the statistics derived from the HSV colour model, especially hue, generally give higher scores than the statistics of monochromic images.This is obvious for the case of moderate rain rates. The number of pixels with $|r| > 0.4$ are on average 58% of the total samples. The difference statistics also outrank those of the co-occurrence matrix.

It is also noticed that the maximum/minimum value of the statistics or the displacement that gives these optimal values do not show up in the table. These parameters have been introduced in the previous chapter in order to avoid fixing the value of displacement d. Therefore, an attention should be paid to select an appropriate d when these statistics are used.

Structural analysis

The structural features used in this analysis are as follows.

Methods	**I**mage Type	**P**arameter
Sobel's filter & Radon Transform ($R_{x,\theta}$)	IR, WV, VIS	ASM Ent $R_{\max}$ $x_{R_{\max}}$ $\theta_{R_{\max}}$

- x_R is the distance from centre of the image window along the projection line. It is negative for the pixel on the left side of the window.
- θ_R is an angle of the projection line with respect to the x axis.

Sobel's filter and the Radon transform are applied to a 21×21 pixel window surrounding each pixel. The statistics derived from the transformed values chosen for the consideration are angular second moment (ASM), entropy (Ent), the maximum of the transformed values ($R_{\max}$), and the radius length along the projection line and the angle between the projection line and the x axis that gives the maximum value ($x_{R_{\max}}$ and $\theta_{R_{\max}}$).

Table 4.1: Statistical features that are most correlated to rain rates.

Light rain rates

Rank	ς	Image	Method	Stat	para	r_{av}	$N(\%)$
1	0.3370	VIS	C_d	Ent	16px	-0.3558	31.8
2	0.3250	VIS	$f_{d,\theta}$	Mean	90°, 1px	-0.2864	36.4
3	0.3243	VIS	C_d	ASM	min	0.3077	34.1
4	0.3151	VIS	$f_{d,\theta}$	STD	90°, 1px	-0.2665	36.4
5	0.2913	VIS	C_d	Con	5px	-0.3099	27.3
6	0.2804	VIS	$f_{d,\theta}$	Ent	θ_R, 5px	-0.2199	34.1
7	0.2786	VIS	$f_{d,\theta}$	Ent	90°, 1px	0.2617	29.6
8	0.2782	VIS	$f_{d,\theta}$	Mean	θ_T, 1px	-0.1927	36.4
9	0.2717	VIS	$f_{d,\theta}$	Ent	θ_T, 1px	0.1798	36.4
10	0.2698	VIS	C_d	Homo	5px	0.2897	25.0

Moderate rain rates

Rank	ς	Image	Method	Stat	para	r_{av}	$N(\%)$
1	0.4283	IR	C_d	ASM	6px	-0.4165	44.0
2	0.4002	WV	C_d	Con	1px	-0.4004	40.0
3	0.3910	IR	C_d	Cor	9px	0.4219	36.0
4	0.3867	IR	C_d	Homo	8px	-0.3334	44.0
5	0.3831	VIS	C_d	Cor	7px	0.3462	42.0
6	0.3753	WV	C_d	Homo	1px	0.3526	40.0
7	0.3726	VIS	$f_{d,\theta}$	STD	θ_T, 10px	0.2853	46.0
8	0.3650	VIS	$f_{d,\theta}$	STD	90°, 10px	0.2699	46.0
9	0.3606	IR	C_d	Ent	5px	0.3012	42.0
10	0.3341	VIS	$f_{d,\theta}$	Ent	θ_T, 12px	-0.2283	44.0

Heavy rain rates

Rank	ς	Image	Method	Stat	para	r_{av}	$N(\%)$
1	0.4125	WV	C_d	Homo	1px	0.5271	29.8
2	0.3739	WV	C_d	Con	1px	-0.4499	29.8
3	0.3355	IR	C_d	Homo	7px	0.3943	27.7
4	0.3323	IR	C_d	ASM	4px	-0.3881	27.7
5	0.3292	IR	C_d	Cor	5px	-0.3817	27.7
6	0.3229	IR	C_d	Con	4px	0.4331	21.3
7	0.3166	IR	C_d	Ent	4px	0.3566	27.7
8	0.2797	IR	$f_{d,\theta}$	Ent	θ_T, 1px	0.1977	36.2
9	0.2794	IR	$f_{d,\theta}$	STD	90°, 1px	-0.1972	36.2
10	0.2750	IR	C_d	Ent	10px	0.1884	36.2

Table 4.2: HSV statistical features that are most correlated to rain rates.

Light rain rates

Rank	ς	Image	Method	Stat	para	r_{av}	$N(\%)$
1	0.3734	S	C_d	ASM	12px	0.4691	27.8
2	0.3679	S	C_d	Ent	1px	-0.4858	25.0
3	0.3061	H	C_d	Ent	15px	0.3623	25.0
4	0.3003	S	$f_{d,\theta}$	Ent	0°, 8px	-0.2143	38.6
5	0.2942	S	$f_{d,\theta}$	STD	0°, 8px	0.2020	38.6
6	0.2821	H	C_d	Con	14px	-0.3419	22.2
7	0.2804	S	$f_{d,\theta}$	Mean	0°, 4px	-0.1972	36.6
8	0.2755	S	C_d	Homo	16px	0.3010	25.0
9	0.2604	S	$f_{d,\theta}$	STD	θ_R, 2px	0.2254	29.6
10	0.2561	S	$f_{d,\theta}$	Ent	θ_T, 3px	-0.1485	36.4

Moderate rain rates

Rank	ς	Image	Method	Stat	para	r_{av}	$N(\%)$
1	0.4383	H	$f_{d,\theta}$	Ent	90°,5px	0.2766	60.0
2	0.4342	H	$f_{d,\theta}$	STD	90°,6px	-0.2653	60.0
3	0.4262	H	C_d	Cor	8px	0.2872	56.5
4	0.4200	H	$f_{d,\theta}$	Ent	θ_T, 3px	0.2599	58.0
5	0.4149	H	$f_{d,\theta}$	Mean	90°, 10px	0.2499	58.0
6	0.4147	H	$f_{d,\theta}$	Mean	θ_R, 16px	0.2494	58.0
7	0.4082	H	$f_{d,\theta}$	STD	θ_T, 4px	-0.2565	56.0
8	0.4065	H	$f_{d,\theta}$	STD	0°, 16px	-0.2531	56.0
9	0.4034	H	$f_{d,\theta}$	Mean	θ_T, 11px	0.2469	56.0
10	0.4022	H	$f_{d,\theta}$	STD	θ_R, 3px	-0.2444	56.0

Heavy rain rates

Rank	ς	Image	Method	Stat	para	r_{av}	$N(\%)$
1	0.3366	H	C_d	Homo	1px	0.4857	18.8
2	0.3318	H	C_d	Ent	16px	-0.4762	18.8
3	0.3129	S	C_d	Con	9px	-0.4383	18.8
4	0.3107	S	C_d	Homo	4px	0.4339	18.8
5	0.3092	S	C_d	Ent	1px	-0.4309	18.8
6	0.3037	H	C_d	Con	4px	-0.4199	18.8
7	0.2795	H	$f_{d,\theta}$	Ent	0°, 16px	0.1973	36.2
8	0.2772	H	$f_{d,\theta}$	STD	θ_R, 16px	-0.1926	36.2
9	0.2752	V	C_d	Homo	1px	0.4253	12.5
10	0.2730	H	$f_{d,\theta}$	Ent	θ_R, 15px	0.1843	36.2

Table 4.3 shows the structural features that give the highest scoring function when $k = 1$ for each rain class together with their r_{av} and N. Although these statistics show a high correlation to rain rates when tested with the four meteorological events in the previous chapter, there is no significant correlation between the features and the rain rates when tested on the whole image of the cyclone. The highest number of pixels with $|r| > 0.4$ can be found in each rain category ranges from 14-20% of the total samples only, though the number is still higher than the 14.31% when the spectral information is considered alone. In addition, there are features that have high r_{av} but low N which occur when there are many feature samples that do not show any relevance to rain rates (shown as blue strips in Figure 4.5, for example) being removed prior the calculation of r_{av}. This can be a result of no significant structural information within the window in which these samples locate.

Table 4.3: Structural features that are most correlated to rain rates.

Rain Category	Rank	ς	Image	para	r_{av}	$N(\%)$
Light	1	0.2331	WV	$x_{R_{\max}}$	-0.4235	4.27
	2	0.2298	VIS	Ent	-0.4142	4.55
	3	0.1604	IR	$R_{\max}$	-0.1617	15.91
Moderate	1	0.2113	IR	$R_{\max}$	-0.2826	14.00
	2	0.1987	VIS	$R_{\max}$	-0.2573	14.00
	3	0.1893	IR	ASM	-0.2786	10.00
Heavy	1	0.3097	IR	$R_{\max}$	0.4417	17.78
	2	0.3193	WV	$R_{\max}$	0.4385	20.00
	3	0.2420	WV	$\theta_{R_{\max}}$	0.4174	6.67

Frequency analysis

The frequency features used in this analysis are as follows.

Methods	**I**mage Type	**P**arameter
Windowed FFT (F)	IR, WV, VIS	Energy Entropy Principal direction

The windowed fast Fourier transform is applied to a 21x21 pixel window surrounding each rain pixel. The statistics derived from the transform are energy and entropy of the frequency contents. In addition, the principal direction which derived from the transformed values that contribute 10% of all frequency contents is also used to find the scoring function.

Table 4.4 shows the frequency features that give the highest scoring function when $k = 1$ for light, moderate and heavy rain rates together with their r_{av} and N. Similar to the scoring functions of the Radon transform statistics, the Fourier statistics do not give any significant correlation. For example, although $|r_{av}|$ of the moderate rain class are as high as 0.44, the number of samples that highly correlate to the rain rates are quite low. It is probably due to the fact that most of samples have $|r|$ near zero with some outliers that give extreme values of the correlation coefficients. The correlation to rain rates of the light rain class is also poorer than the Radon transform.

Table 4.4: Frequency features that are most correlated to rain rates.

Rain Category	Rank	ς	Image	para	r_{av}	$N(\%)$
Light	1	0.1050	IR	Ent	0.1401	6.98
	2	0.0638	IR	Ener	-0.0811	4.65
	3	0.0523	VIS	Ener	-0.0581	4.65
Moderate	1	0.2486	VIS	Ent	-0.4361	6.12
	2	0.2437	VIS	Ener	-0.4059	8.16
	3	0.2192	WV	direc	0.3365	10.20
Heavy	1	0.2744	WV	Ener	0.4513	9.76
	2	0.1580	VIS	Ener	-0.2916	2.44
	3	0.1530	VIS	direc	-0.3060	0.00

Spatio-temporal analysis

The spatio-temporal features used in this analysis are as follows.

Methods	**I**mage Type	**S**tatistics	**P**arameter
Daubechies (db2,db6, db10, db14, db18)	IR	$Af^{1,2}$	
Coiflets (coif6, coif18, coif30)	WV	$D1f^{1,2}$	μ_1, μ_2, μ_3
Symlets (sym2, sym6, sym10)	VIS	$D2f^{1,2}$	α, β (Mallat, 1989)
Biorthogonal (bior10, bior12, bior18)		$D3f^{1,2}$	

- The parameters: μ_1, μ_2, μ_3 (1st, 2nd, 3rd moment), α and β are computed from the histograms of the decomposed images Af, $D1f$, $D2f$ and $D3f$ at level 1 and 2.

The wavelet filters that are selected for the spatio-temporal analysis are Daubechies (db), Coiflets (coif), Symlets (sym) and Biorthogonal (bior) wavelets at various filter lengths. The Daubechies wavelet is selected because it provided promising results in cloud classification. Other wavelets are developed to gain a better property than the conventional Daubechies wavelet. For example, Symlets are modified by Daubechies to gain more symmetry while retaining the simplicity of the Daubechies wavelet. The vanishing moment of these two wavelets are equal to half of the filter lengths. Therefore, the vanishing moments in this analysis are 1, 3,, 9 for the Daubechies wavelet (db2,db6, , db18) and 1, 3, 5 for the Symlets (sym2, sym6, sym10). Coiflets are much more symmetrical than the Daubechies wavelet whereas the Biorthogonal wavelet consists of two wavelets, one for the decomposition and one for the reconstruction to gain an exact the reconstruction of the input image. These wavelets are selected as alternatives to the Daubechies wavelet. Although the symmetry of the filters are more significant in image coding, which wavelet has the best correlation to rain rates depends on its response to the nature of cloud texture.

Table 4.5 shows the spatio-temporal features that give the highest scoring function when $k = 1$ for each rain class. The spatio-temporal features are able to describe the cloud features better than frequency and structural features. The number of samples that highly correlate to rain rates is also higher. However, the averaged correlation coefficient and therefore the scoring functions of these features are quite low. This might be a consequence of using only a few parameters to represent the four sub-images—a lot of information is lost during the decoding process. However, it can be noticed that the wavelet that show the highest correlation is the Coiflets with a filter length of 30 coefficients. This implies that the rain rates correlate better with higher degree polynomials embedded in the texture.

Table 4.5: Spatio-temporal features that are most correlated to rain rates.

Rain Category	Rank	ς	Image	wavelet	para	r_{av}	$N(\%)$
Light	1	0.1142	IR	coif30	A_f^1, μ_3	0.1140	11.43
	2	0.1057	IR	db2	A_f^1, μ_3	0.0972	11.43
	3	0.1057	IR	sym2	A_f^1, μ_3	0.0972	11.43
	4	0.1003	IR	bior18	A_f^1, μ_3	0.1149	8.57
	5	0.0973	IR	coif18	A_f^1, μ_3	0.1089	8.57
Moderate	1	0.1400	VIS	coif30	A_f^1, μ_3	0.1238	15.63
	2	0.1398	VIS	db18	A_f^1, μ_3	0.1234	15.63
	3	0.1375	VIS	db14	A_f^1, μ_3	0.1188	15.63
	4	0.1370	VIS	db14	A_f^2, μ_3	0.1177	15.63
	5	0.1364	VIS	bior18	A_f^1, μ_3	0.1165	15.63
Heavy	1	0.1145	VIS	coif30	A_f^2, β	0.0910	13.79
	2	0.1033	WV	db10	A_f^1, μ_3	0.1032	10.34
	3	0.1025	WV	db14	A_f^1, μ_3	0.1016	10.34
	4	0.0999	WV	db10	A_f^2, μ_3	0.0964	10.34
	5	0.0992	WV	db14	A_f^2, μ_3	0.0950	10.34

In summary, the features that can be used for rain estimation can be classified into two sets: (1) the features that have the highest scores of all textural feature types and (2) the features that gives the highest scores of each textural feature type. These features are listed in Table 4.6 and 4.7.

Table 4.6: Features that have the highest scores of all textural feature types.

Rain Category	Rank	ς	Image	Method	para	r_{av}	$N(\%)$
Light	1	0.3734	S	C_d	ASM, 12 px	0.4691	27.8
	2	0.3679	S	C_d	Ent, 1 px	-0.4858	25.0
	3	0.3370	VIS	C_d	Ent, 16px	-0.3558	31.8
	4	0.3250	VIS	$f_{d,\theta}$	Mean, 90°, 1px	-0.2864	36.4
	5	0.3243	VIS	C_d	ASM, min	0.3077	34.1
Moderate	1	0.4383	H	$f_{d,\theta}$	Ent, 90°,5px	0.2766	60.0
	2	0.4342	H	$f_{d,\theta}$	STD, 90°,6px	-0.2653	60.0
	3	0.4283	IR	C_d	ASM, 6px	-0.4165	44.0
	4	0.4262	H	C_d	Cor, 8px	0.2872	56.5
	5	0.4200	H	$f_{d,\theta}$	Ent, θ_T, 3px	0.2599	58.0
Heavy	1	0.4125	WV	C_d	Homo, 1px	0.5271	29.8
	2	0.3739	WV	C_d	Con, 1px	-0.4499	29.8
	3	0.3366	H	C_d	Homo, 1px	0.4857	18.8
	4	0.3355	IR	C_d	Homo, 7px	0.3943	27.7
	5	0.3323	IR	C_d	ASM, 4px	-0.3881	27.7

Table 4.7: Features that gives the highest scores of each textural feature type.

Rain Category	Rank	ς	Image	Method	para	r_{av}	$N(\%)$
Light	1	0.3734	S	C_d	ASM, 12 px	0.4691	27.8
	2	0.2331	WV	$R_{x,\theta}$	$x_{R_{\max}}$	-0.4238	4.3
	3	0.1142	IR	coif30	A_f^1, μ_3	0.1140	11.4
	4	0.1050	IR	F	Ent	0.1404	7.0
Moderate	1	0.4383	H	$f_{d,\theta}$	Ent, 90°,5px	0.2766	60.0
	2	0.2486	VIS	F	Ent	-0.4361	6.1
	3	0.2113	IR	$R_{x,\theta}$	$R_{\max}$	-0.2826	14.0
	4	0.1400	VIS	coif30	A_f^1, μ_3	0.1238	15.6
Heavy	1	0.4125	WV	C_d	Homo, 1px	0.5271	29.8
	2	0.2744	WV	F	Ener	0.4513	9.8
	3	0.3097	IR	$R_{x,\theta}$	$R_{\max}$	-0.4417	17.8
	4	0.1145	VIS	coif30	A_f^2, β	0.0910	13.8

4.3 Relation to rain rates

Estimating rain rates by means of pattern recognition algorithms can be done by mainly 2 approaches—either by function approximation or by classification. The first approach is preferable in rain estimation as it provides an explicit and continuous value of rain rates, comparing to a class label from a classifier. Figure 4.7 shows a result of fitting a generalised linear model (GLM)[4] to find the best *linear* relation between IR temperature (T_{IR}) and rain rates (RR). The following relation gives the lowest sum of squared error when pixels with zero rain rates are excluded, i.e.

$$RR = 40.015865 - 16.344961 \log(T_{IR}). \tag{4.3}$$

The distribution of samples in Figure 4.7 reveals that most of the samples have very light rain rates. Therefore, heavy rain rates are almost negligible when fitting the linear model to the data. Although the ranges of the estimated rain rates are from 0 to 2 mm/h only, its root mean squared (RMS) error is 0.13 mm/h per sample only.

Alternatively, if both ranges of the rain estimates and the RMS error are required to be optimised, the (cumulative) histogram matching method can be used. The method can be implemented by, firstly, creating cumulative histograms of two data that need to be matched. Then, a look-up table to link the two data, e.g. temperature v.s. rain rates, that gives the same value of the cumulative histogram is derived. Hence, in theory, the rain rates derived by the histogram matching method should have the same histogram as the observed rain rates (that the look-up table is derived from). However, in practical, the look-up table, which is derived for a specific time and space, is intended to use with unknown distributions of the observed rain rates. The method has been commonly used to find a relationship between the observed rain rates and the radar signals. It has also been used to map IR temperatures to SSM/I passive microwave data in Eumetsat's Multi-sensor Precipitation Estimate (Heinemann et al., 2002).

The histogram matching technique is applied to map IR temperatures of the mid-latitude cyclone to rain rates by means of the look-up table. The relation is shown as a line marked with dots in Figure 4.7. It is capable of producing wider range of rain estimates than the linear model. The rain estimates by this method has an RMS error of 0.15 mm/h per pixel, which is slightly worse than the GLM method. However, the visualisation in Figure 4.8 shows that using this method produces more agreeable

[4]using GLMLab: A Generalized Linear Model Package for MATLAB by Peter Dunn at the Department of Mathematics and Computing, University of Southern Queensland (Australia).

results, although it fails to capture heavy rain areas near the centre of the cyclone and overestimates rain rates for the cold cloud bands. The histogram matching method is suitable to find a relationship between an output and *one* input. Embedding more textural features to rain estimation, for example, needs a large size of the look-up table, which can make the method impossible to be implemented. In addition, an investigation on sample by sample basis reveals that there is very low correlation (r=0.03) between the estimates and the actual rain rates when the histogram method is used. This can also be seen in the distribution of the two data presented in Figure 4.9 which reveals that there is no obvious relation between them. This is because, by using the histogram matching, the information of each point which should contribute to the rain rates is neglected. Such an effect is not obvious when the method is applied to a coarse resolution image ($5° \times 5°$) of convective clouds in Heinemann and Kerényi (2003) where rain rates (from SSM/I) and measured IR brightness temperatures have a good negative correlation.

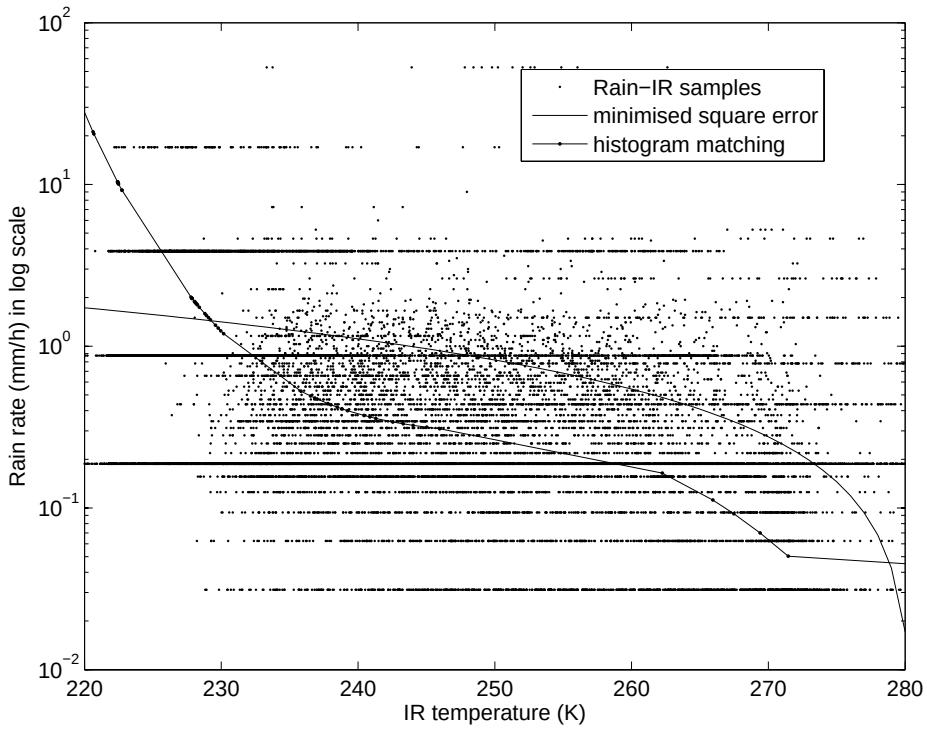
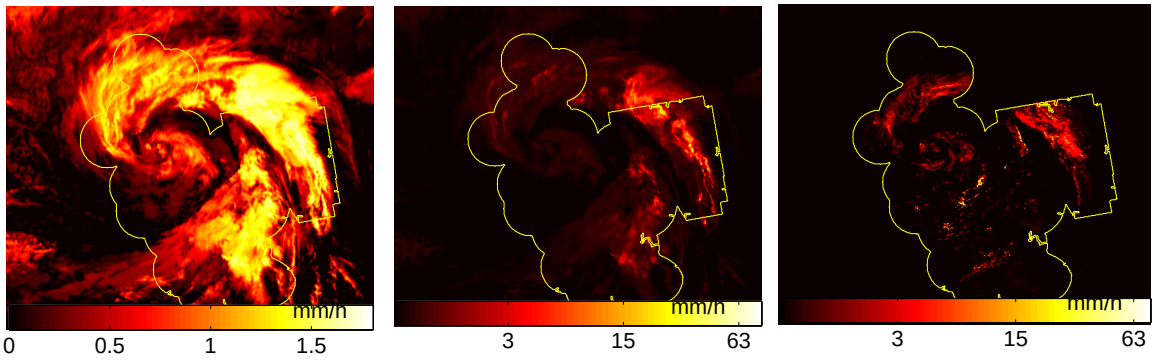


Figure 4.7: Distribution of rain rates and IR temperature.



(a) (b) (c)

Figure 4.8: Estimation of rain rates when (a) GLM and (b) histogram matching methods are used. The observed rain rates are in (c).

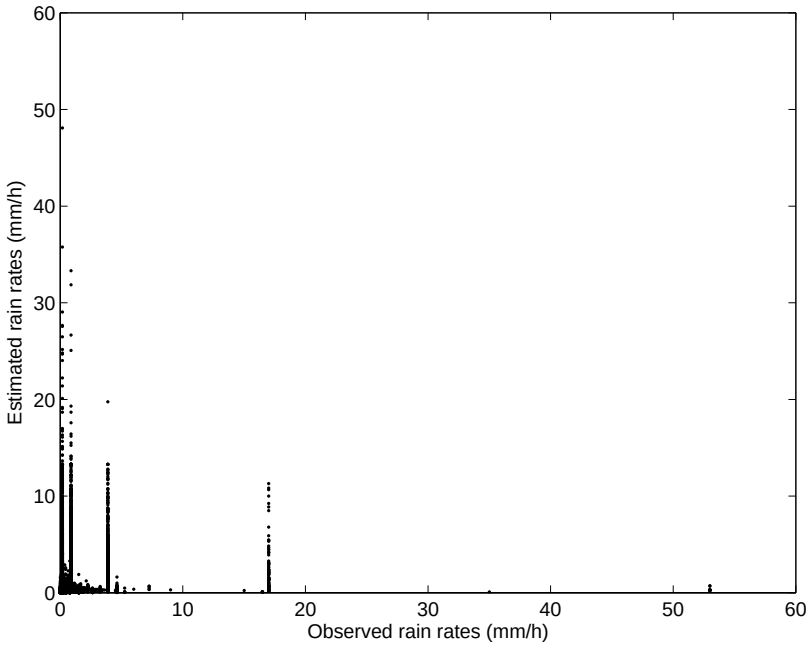


Figure 4.9: Rain estimates using the histogram matching method versus the observed rain rates.

The results illustrate two problems that occur when estimating rain rates:

- imbalanced distribution of rain rates
- assessment of accuracy.

First, the imbalance distribution of rain rates biases the result to the majority of samples, which are light rain rates in the case of mid-latitude cyclones. This effect is obvious when the GLM is used. Even if no heavy rain estimates are made, the GLM method still produces the RMS error as low as 0.13 mm/h.

Second is the use of the RMS error as an assessment of accuracy. The parameter is commonly used for the synoptic scale rainfall estimation where the quantity of interest is an averaged rain amount. It is useful to provide information on average, assuming that the errors on short time and space scales are not significant or can be cancelled out over large space and time scales. However, to diagnose the performance of the nowcasting rain rate estimation, the assessment should give no bias and provide useful information to understand the accuracy and limitations of the estimation.

For high resolution rain estimation, some standard assessment of accuracy used for quantitative precipitation forecasting can be applied in combination with the RMS error (Ebert, 2002). For instance, a two-way table presented in Figure 4.3 known as a yes/no contingency table is a useful way to see what types of errors are being made. A perfect estimation would produce only hits and correct negatives, and no misses or false alarms. The table can be used in the rain detection with a meaning of yes/no as rain/no-rain.

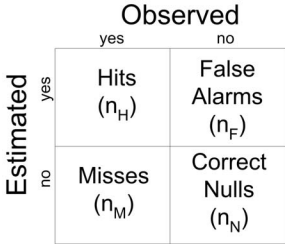


Figure 4.10: A contingency table for an estimation.

The statistics derived from the table are accuracy, bias, hit rate, false alarm rate, critical success index (CSI) and odds ratio. These statistics[5] are listed in Table 4.8.

Accuracy is a probability of correct classification whereas a hit rate is the number of hits relative to the total number of events observed. It is always used in combination

[5]Source: Forecast Verification - Issues, Methods and FAQ. `http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html`

Table 4.8: Statistics of a yes/no contingency table.

Statistics	Definition	Range	Perfect score
Accuracy	$\frac{n_H+n_N}{n_H+n_M+n_F+n_N}$	$[0, 1]$	1
Bias	$\frac{n_H+n_F}{n_H+n_M}$	$[0, \infty)$	1
Hit rate	$\frac{n_H}{n_H+n_M}$	[0,1]	1
False alarm rate	$\frac{n_F}{n_N+n_F}$	[0,1]	0
Critical success index	$\frac{n_H}{n_H+n_M+n_F}$	[0,1]	1
Odds ratio	$\log \frac{n_H n_N}{n_F n_M}$	$(-\infty, \infty)$	$>> 0$

with the false alarm rate which is defined as the number of false alarms relative to the non-events. If the ratio of a hit rate to a false alarm rate is unity, the classification can be regarded as unskilled as it is likely to produce a hit as it is likely to produce a false alarm. On the contrary, the odds ratio is the ratio of the probability of an event occurring to the probability of the event not occurring. It is equal to the ratio of the odds of making a hit to the odds of issuing a false alarm, i.e.

$$\text{Odds ratio} = \log \left(\frac{H}{1-H} \right) \left(\frac{F}{1-F} \right)^{-1}, \tag{4.4}$$

where H and F denote a hit rate and a false alarm rate. The odds ratio is not dependent on the frequency of the events as other statistics. On the other hand, Bias is defined as the ratio of events estimated to those observed. It tells whether the estimates are as many as the events observed e.g. whether the classification tends to say yes more then no to the events. It is unity for perfectly calibrated classification. The critical success index (CSI) is the ratio for successful events classified to the total number of event being classified. It assumes that instances when an event was neither expected nor observed are of no consequence. CSI is dependent upon the frequency of the events being classified. Schaefer (1990) proposed a modified CSI for removing frequency biases. The new statistic, namely the Gilbert skill score (GS), is defined as

$$\text{GS} = \frac{M}{(n_M + n_F)N + M}, \tag{4.5}$$

where $M = n_H n_N - n_M n_F$ and N is the total number of events, i.e. $N = n_H + n_M + n_F + n_N$. The more negative scores indicates the worse the classification is. It is always lower than the CSI and approaches the CSI when the event to forecast becomes rarer. The score of 0 indicates no skill, which can occurs when $n_H n_N = n_M n_F$, i.e. the hit rate is equal to the false alarm rate.

Instead of using only yes/no, the contingency table can be extended to $K \times K$ cells for K ranges of rain rates. By this way, the table provides more details about the estimation errors. A frequently used statistic is the Heidke skill score (HSS) which is defined as

$$\mathrm{HSS} = \frac{\sum_{k=1}^{K} \Pr(Y_k, O_k) - \sum_{k=1}^{K} \Pr(Y_k) \Pr(O_k)}{1 - \sum_{k=1}^{K} \Pr(Y_k) \Pr(O_k)}, \tag{4.6}$$

where $\Pr(Y_k, O_k)$ is the probability that the estimate and the observation falling in the k^{th} range and $\Pr(Y_k)$ and $\Pr(O_k)$ are their marginal probabilities. The perfect score is 1 whereas 0 indicates no skill. The Hanssen and Kuipers (HK) discriminant can also be used for the multi-category table. It is defined as

$$\mathrm{HK} = \frac{\sum_{k=1}^{K} \Pr(Y_k, O_k) - \sum_{k=1}^{K} \Pr(Y_k) \Pr(O_k)}{1 - \sum_{k=1}^{K} \Pr(Y_k)^2}. \tag{4.7}$$

The value ranges from -1 to 1 where the perfect score is 1 and 0 is no skill. The Hanssen-Kuipers discriminant and the Heidke skill score measures the fraction of correct estimates after eliminating those estimates that would have been correct due purely to random chance. However, the two statistics can overestimate skill of the estimators if they are heavily influenced by the most common category (Ebert, 2002). Therefore, a combination of these statistics, the RMS error and the correlation coefficient between the estimated and the actual rain rates have been used in this thesis to give the whole picture of the performance of the rain estimation.

To overcome the class imbalance problem, three methods can be applied to the feature space: over-sampling, under-sampling and cost-modifying. In principle, the first method randomly oversamples the minority class until it contains as many samples as others whereas the second method eliminates the over-sized class at random until it matches the size of the other classes. On the other hand, the cost-modifying method associates different costs with the misclassification errors so that the imbalance between classes is compensated. The under-sampling method is preferable if the minority class is the class of interest and the majority class includes a lot of irrelevant data (Domingos, 1999). The over-sampling and the cost-modifying methods are more

suitable for an application that the roles of both classes are equally important, though the over-sampling method also increases the size of the training samples (Japkowicz and Stephen, 2002). For rain estimation, the majority classes are light to moderate rain rates whereas the heavy rain which should be paid attention to fall in the minority class. Therefore, the under-sampling method seems to be more appropriate than the other two. By under-sampling the feature space so that the light, moderate and heavy rain classes have the same numbers of samples, the relation between the IR temperatures and rain rates in Equation (4.3) is modified to

$$RR = 114.030297 - 44.631970 \log(T_{IR}). \tag{4.8}$$

However, Equation (4.8) results in a sum of squared error of 914.99, 3 times higher than the original relation. The histogram of the estimates in Figure 4.11 also shows that the error of applying the GLM to the under-sampled data is a result of too high rain rate estimation, comparing to other methods that the estimates concentrate in low rain rates.

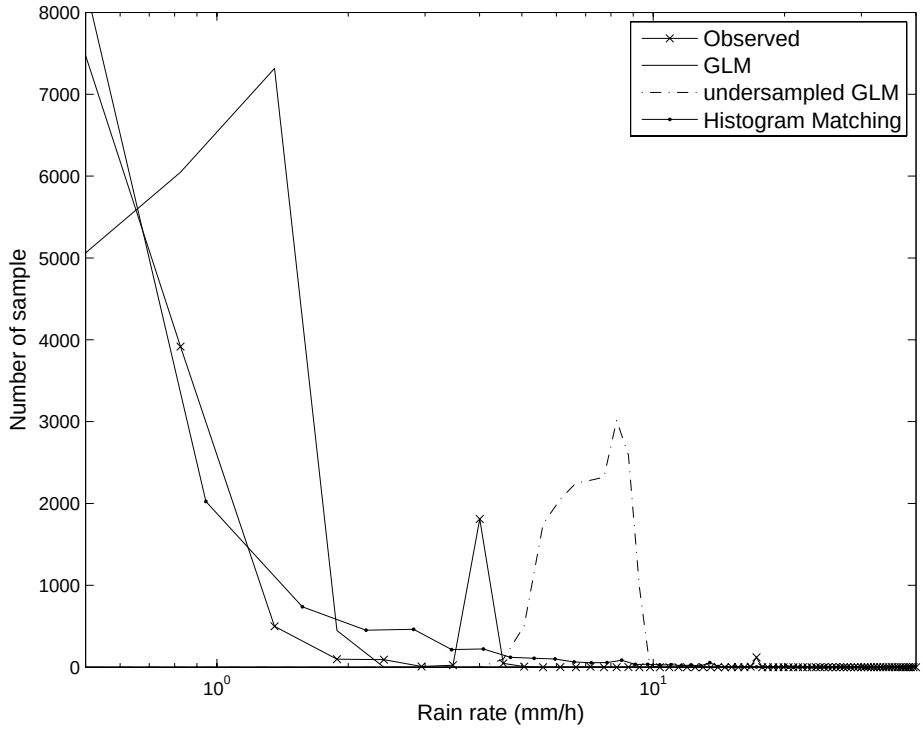
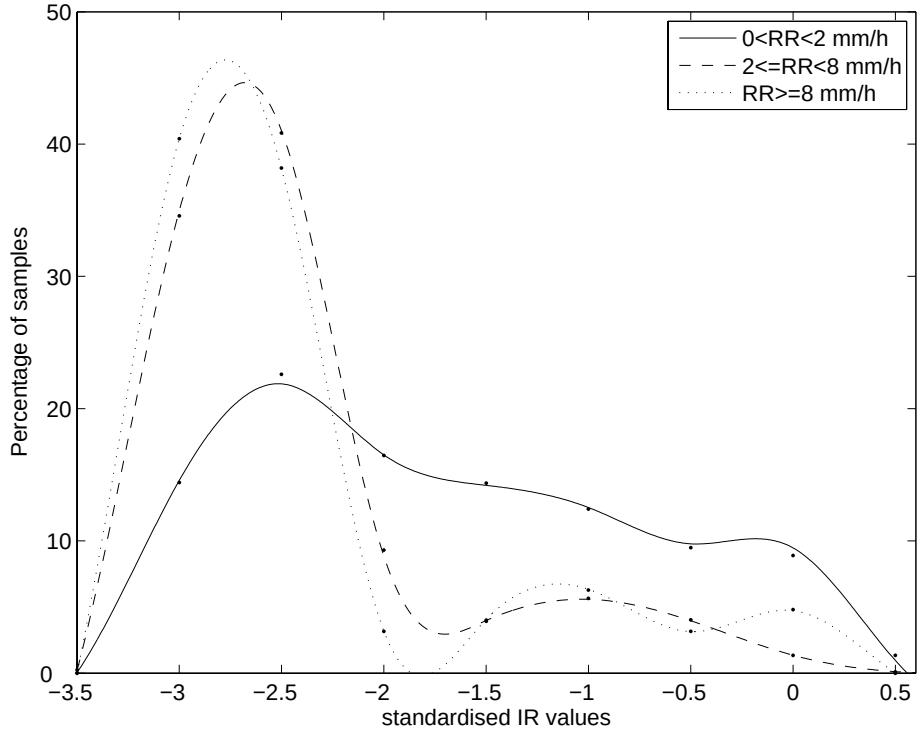


Figure 4.11: Histogram of rain estimates using GLM, undersampled GLM and Histogram Matching.

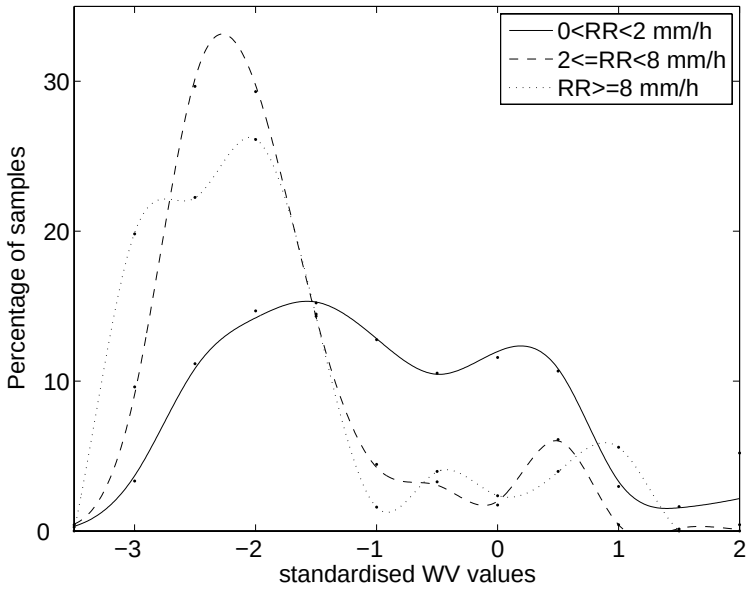
Alternatively, to preserve the numbers of high, moderate and light rain rates such

that the estimates have the same distribution as the actual rain rates, the cloud samples may be first classified into these categories. Classifying clouds by using the rain rate interval as a cloud label has an advantage of no variability in the class labelling which can happen even it is done by experts. In addition, this approach is also compatible with the feature selection that has been done previously according to light, moderate and heavy rain. In Figure 4.12, the distribution of satellite measurements which are standardised using Equation (4.1) of each rain classes shows that the spectral information can only separate light rain rates from others. For example, if the samples with the standardised IR values of more than -2 are regarded as light rain, more than 50 % of the light rain samples are correctly classified. However, such thresholds to separate moderate from heavy rain events could not be found. Therefore, the feature selection is revisited in search of features that are suitable for classification.

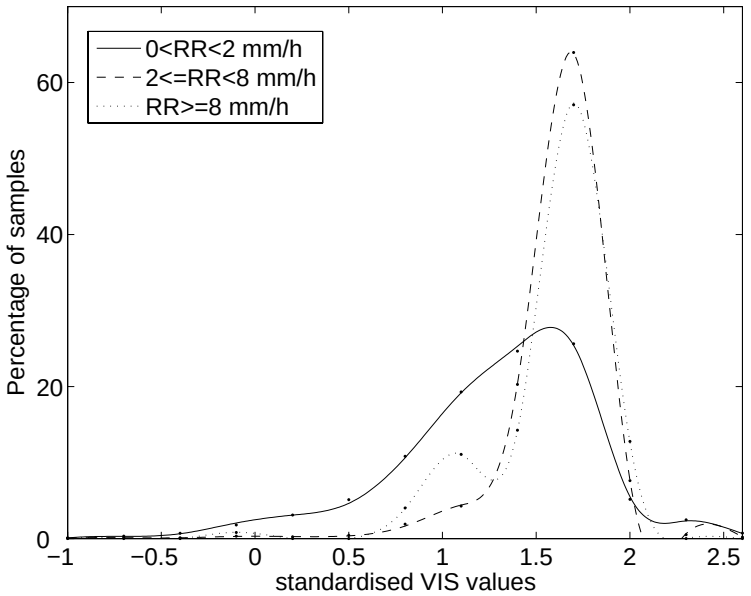


(a)

Figure 4.12: Distribution of standardised (a) IR, (b) WV and (c) VIS values.



(b)



(c)

Figure 4.12 continued.

Feature selection revisited

The textural features in Section 4.2 are ranked by their *linear* relation to rain rates in an expectation that the features would lead to an easier estimation of rain rates. By the same approach, the correlation coefficients of each rain rate category can also be used to rank each feature of a classification task. For example, if a feature has a high correlation score for light rain rates but falls in low ranks in the other two classes, it can be implied that the feature is highly linear to light rain rates but randomly (or nonlinearly) to the others. Therefore, the difference between the scoring functions of each feature within each class can be used as a scoring function for the classification. If the scoring functions of a feature $\vec{x}$ is defined as ς_L, ς_M and ς_H for light, moderate and high rain classes, the scoring functions of $\vec{x}$ for each class can be defined as

$$\begin{aligned} \varsigma_L^C(\vec{x}) &= (\varsigma_L - \varsigma_M) + (\varsigma_L - \varsigma_H), \\ \varsigma_M^C(\vec{x}) &= (\varsigma_M - \varsigma_L) + (\varsigma_M - \varsigma_H), \\ \varsigma_H^C(\vec{x}) &= (\varsigma_H - \varsigma_L) + (\varsigma_H - \varsigma_M), \end{aligned} \tag{4.9}$$

The feature that gives the highest scores of each class will be selected for the classification. In fact, a feature that can give the highest correlation to rain rates for every class is the best feature for classification, if such a feature could be found.

Five features with the highest scores for each class are listed in Table 4.9. Most of the listed features are the statistics of a difference matrix of hue at various directions and displacements. For the light rain, a large difference between ς_L and the other two scores indicates a good sign for using minimum ASM values of a co-occurrence matrix of VIS images to separate samples of light rain rates from th e other rain categories. In addition, angles between a line crossing a sample and the centre of cyclone and x axis (θ_R) seems to be the best direction to derive the difference statistics of hue images to distinguish heavy rain samples. However, the features in this class closely correlate to each other with the eigenvalues of their covariance matrix of [97.61, 1.21, 0.84, 0.18, 0.15]. This may affect the classification result.

Table 4.9: Textural feature candidates for classification.

Light rain rates

Rank	ς_L^C	Image	Method	Stat	para	ς_L	ς_M	ς_H
1	0.5059	VIS	C_d	ASM	min	**0.3243**	0.0820	0.0608
2	0.4445	H	$f_{d,\theta}$	STD	0°, 16px	**0.3734**	0.0841	0.2183
3	0.4250	H	$f_{d,\theta}$	Ent	0°, 1px	**0.3653**	0.0829	0.2227
4	0.4239	IR	$f_{d,\theta}$	Ent	0°, 4px	**0.2687**	0.0775	0.0359
5	0.4127	IR	$f_{d,\theta}$	STD	0°, 5px	**0.2550**	0.0619	0.0353

Moderate rain rates

Rank	ς_M^C	Image	Method	Stat	para	ς_L	ς_M	ς_H
1	0.5235	H	C_d	Cor	8px	0.0738	**0.3762**	0.1551
2	0.4313	V	$f_{d,\theta}$	Ent	90°, 3px	0.0752	**0.3448**	0.1831
3	0.4112	V	$f_{d,\theta}$	STD	θ_R, 16px	0.0677	**0.3464**	0.2139
4	0.3499	H	$f_{d,\theta}$	STD	45°, 2px	0.0860	**0.3220**	0.2081
5	0.3373	H	$f_{d,\theta}$	Mean	45°, 15px	0.0912	**0.3202**	0.2119

Heavy rain rates

Rank	ς_H^C	Image	Method	Stat	para	ς_L	ς_M	ς_H
1	0.3951	H	$f_{d,\theta}$	STD	θ_R, 9px	0.2004	0.1240	**0.3597**
2	0.3775	S	$f_{d,\theta}$	STD	90°, 5px	0.0471	0.0610	**0.2428**
3	0.3683	H	$f_{d,\theta}$	STD	θ_R, 14px	0.2219	0.1146	**0.3524**
4	0.3613	H	$f_{d,\theta}$	Ent	θ_R, 1px	0.2463	0.0939	**0.3507**
5	0.3386	H	$f_{d,\theta}$	Ent	θ_R, 2px	0.2529	0.0982	**0.3449**

4.4 Implementation

Next, the feature sets selected from the previous section will be used to estimate rain rates. Figure 4.13 summarises the implementation of rain estimation when the selected features are used. There are two sets of features for the estimation. First is a set of features for a classification task as listed in Table 4.9, labelled as Feature set 1 in Figure 4.13. Second is a set of features for rain rate approximation, labelled as Feature set 2 in Figure 4.13. These features are listed in Table 4.6 and 4.7. The process of rain estimation can be described as follows. First, the Feature set 1 is used to train a classifier, which will classify the samples into light, moderate or heavy rain. Details of the classification is presented in Section 4.4.1. Then, the Feature set 2 is used to find a regression model that relates these features to appropriate rain rates. Details of the regression is presented in Section 4.4.2.

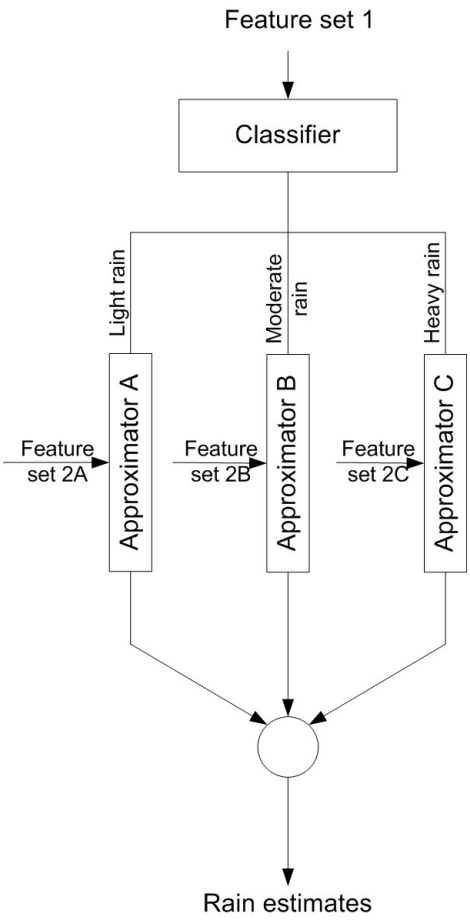


Figure 4.13: Implementation of rain estimation using the selected features.

For classification, a probabilistic neural network (PrNN) and a support vector

machine (SVM) which can perform effectively under the class imbalance environment (Tian et al., 1999; Japkowicz and Stephen, 2002) have been used.

PrNN is a three-layer feedforward network, consisting of an input layer, a pattern layer, and a summation layer. The input layer accepts feature vectors and supplies them to all neurons in the pattern layer. The pattern layer consists of K pools of pattern neurons where K is the number of classes. The pattern layer approximates probability density functions (PDFs) of each class for a feature vector by the Parzen window estimator, which states that a PDF can be defined as a summation of all kernel functions on each training sample. If a Gaussian kernel is used, the PDF of a feature vector $\mathbf{y}$ for each class c_i can then be defined as

$$\Pr(\mathbf{y}|c_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{(\sqrt{2\pi}\sigma)^d} \exp\left[-\frac{(\mathbf{y} - \mathbf{x}_j^{(i)})^T (\mathbf{y} - \mathbf{x}_j^{(i)})}{2\sigma^2} \right], \tag{4.10}$$

where $\mathbf{x}_j^{(i)}$ denotes the j^{th} training sample belonging to class c_i and acts as the centre of the Gaussian kernel, d is the dimension of the input vector, N_i is the number of samples in the training set and σ is a smoothing factor or the width of the Gaussian kernel. The output class is assigned to $\mathbf{y}$ by the maximum posterior probability, i.e.

$$C(\mathbf{y}) = \underset{c_i}{\operatorname{argmax}} \Pr(\mathbf{y}|c_i) \Pr(c_i), \tag{4.11}$$

where $\Pr(c_i)$ is prior class probabilities. The only control factor that needs to be selected for training a PrNN is the width of the Gaussian function. Too narrow σ causes a spiky approximation which cannot be generalised. Too wide σ smooths out detail of the PDF surface. An appropriate width can be chosen by selecting the width that produces the least errors. A PrNN can become a very large network since it contains the entire set of training samples. However, the greatest advantage of a PrNN is the training speed. Training a PNN actually consists mostly of copying training samples into the network, therefore close to instantaneous results can be expected.

The principle of the Gaussian basis function is used in a generalised regression neural network (GRNN) for function approximation. GRNNs have recently been employed in many signal processing applications, pattern recognition, and financial prediction (Ganchev et al., 2007). The network works in a similar fashion to PrNNs, but performs regression rather than classification tasks. The foundation of GRNNs is essentially based on the theory of nonlinear (kernel) regression, which was previously

proposed in 1964 by Nadaraya and Watson. As with the PrNN, Gaussian kernel functions are located at each training sample. Each sample can be regarded as evidence of a response surface with a given height at that sample point in the input space, which is progressively decaying in the immediate vicinity. A GRNN copies training samples into the network to be used to estimate the response on new points. The output is estimated using a weighted average of the outputs of the training samples. A GRNN train almost instantly though the network tends to be large, retarding the computation.

A support vector machine on the other hand is a supervised learning method that can be used for both classification and regression. For classification, SVMs attempt to find a hypersurface in the feature space from a subset of the training samples, namely support vectors. The hypersurface is aimed to be capable of separating samples of different classes from each other. The surface can be modified by applying a kernel function to the input space so that it forms a nonlinear separation of the feature space. Unlike PrNNs that give a posterior probability of a class given an input, the output of an SVM is a real value of its decision function. For a two-class problem, the class label is assigned to a feature vector according to the negative or positive value of the decision function. For a multi-class problem, the class that gives the highest value of the decision function is assigned to the feature vector.

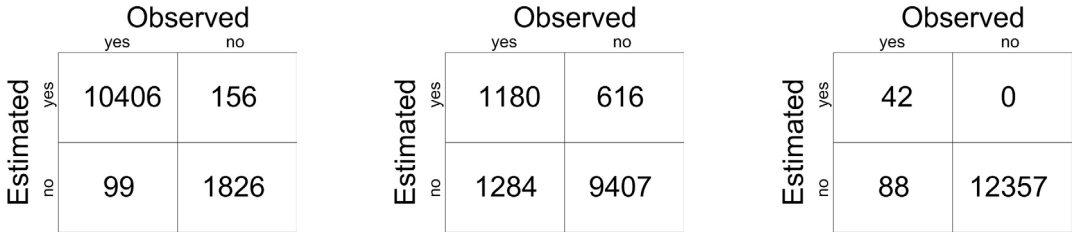
The technique can also be applied to the regression problem by using the support vectors to model the data. Its kernel function extends the capacity of SVMs beyond a linear regression model. The details of support vector machines are briefly introduced in Appendix C.2.

4.4.1 Classification

Next, the two classifiers, i.e. a PrNN and a Gaussian-kernel SVM[6], are used to evaluate the Feature set 1 listed in Table 4.9 on how well they can classify cloud pixels into their corresponding rain classes. The classification results of the two classifiers when the whole data are used as both training samples and test samples are presented in Figure 4.14 and 4.15 using contingency tables. The width of the Gaussian kernel is 0.1 for the PrNN and 0.4 for the SVM. These widths, by experiments, give the best classification results for each classifier.

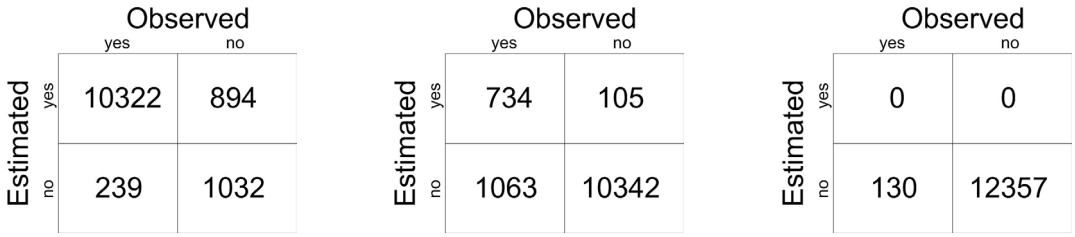
Table 4.10 shows the statistics derived from the contingency tables. Using the

[6]using SVM and Kernel methods Matlab toolbox (Canu et al., 2005)



(a) (b) (c)

Figure 4.14: Classification results using a probabilistic neural network for (a) light (b) moderate and (c) heavy rain rates.



(a) (b) (c)

Figure 4.15: Classification results using a support vector machine for (a) light (b) moderate and (c) heavy rain rates.

selected features, the PrNN can achieve 84.8-99.3 % accuracy whereas the SVM achieve 77.0-99.0 % accuracy. The high accuracy of the classification when using the PrNN shows that the feature space can be well separated according to rain classes. However, the low accuracy of the SVM for the moderate rain class indicates that the features of this class cannot be well separated using a Gaussian kernel. In addition, even though all samples of heavy rain are misclassified when using the SVM, the classification still yields a high accuracy due to the non-events. This is another example that underlines the class imbalance problem and the assessment of accuracy.

The ratio between the hit rate and the false alarm rate of more than one indicates some skills of the classifiers. By an inspection, the SVM produces less incorrect results and should be graded as a better classifier for the moderate rain class. The ratio can depict this circumstance by giving the ratio of 41 to the SVM and 8 to the PrNN. Similar result is found when using the odds ratio which gives a score of 1.83 to the SVM and only 1.15 to the PrNN.

However, as the false alarms for the heavy rain are zero, the odds ratio cannot be retrieved although the PrNN shows more skills of higher hit events.

Similar to accuracy, CSI shows that both classifiers are skilful for the case of light rain with the score of 0.98 for the PrNN and 0.90 for the SVM. Although it is obvious

that the PrNN can classify the feature space better than the SVM, i.e. its false alarms and missed events are lower than those of the SVM, CSI could not tell which classifier is better for the task as the CSIs of both classifiers are quite similar. The Gilbert skill score can identify such incorrect classification. It gives the score of 0.86 to the PrNN and only 0.43 to the SVM. In addition, the score of the heavy rain can still be retrieved. The Gilbert score of zero when the SVM is used indicates that all samples need to be support vectors to generate complex hypersurface in order to separate the heavy rain from other classes. In other words, the SVM could not find an appropriate boundary between classes. Other non-redundant feature set should be selected. Otherwise, the PrNN might be more appropriate for the classification task using the given feature set.

Table 4.10: Classification statistics using a contingency table.

Statistics	**L**ight rain		**M**oderate rain		**H**eavy rain	
	PrNN	SVM	PrNN	SVM	PrNN	SVM
Accuracy (%)	97.96	90.93	84.78	88.70	99.30	98.96
Hit rate (H)	0.99	0.98	0.48	0.41	0.32	0
False alarm rate (F)	0.08	0.46	0.06	0.01	0	0
H:F	12.4	7.1	8.0	41.0	N/A	N/A
Odds ratio	3.09	1.70	1.15	1.83	N/A	N/A
Bias	1.01	1.06	0.73	0.47	0.32	0
Critical success index	0.98	0.90	0.38	0.39	0.32	0
Gilbert skill score	0.86	0.43	0.30	0.34	0.32	0

The classfication results in Table 4.10 are derived from the case when all samples are used as a training set and a test set. To give an overview on the general performance of using the selected textural features for the classification, generalised scoring statistics of the PrNN are computed using the k-fold cross-validation. The k-fold method can be described as follows. First, the original feature space is partitioned into k sub-samples. Then, of the k sub-samples, a single sub-sample is retained for testing the classifiers and the remaining $(k-1)$ sub-samples are used as training data. The cross-validation process is repeated k times with each of the k sub-samples being used as the validation data. The k results from all *folds* is then averaged to produce the generalised statistics of the classification.

Figure 4.16 shows the generalised accuracy and Gilbert skill scores ($k = 10$) when different widths of a Gaussian function (σ) of the PrNN are used. Increasing the width of the Gaussian function of the PrNN increases the accuracy and the Gilbert skill score. After some values of σ, both statistics start to decrease. The accuracy

of the heavy rain class which is not shown in Figure 4.16 (a) becomes a constant of 0.9896 for $\sigma > 0.2$ whereas the Gilbert skill score becomes zero. The large σ makes the PrNN become too generalised for the feature space, leading to misclassification. It can also be noticed that the generalised statistics are lower than when the whole samples are used as a training set. For example, at $\sigma = 0.1$ the accuracy drops from 97.96 % to 91.28 % and the Gilbert skill score drops from 0.86 to 0.49 when the PrNN is used.

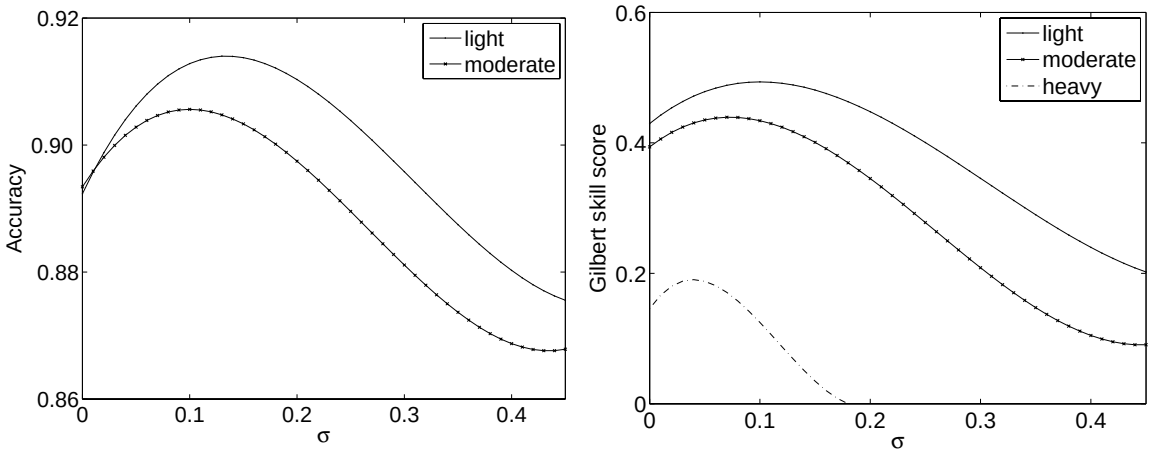


Figure 4.16: Generalised statistics: (a) accuracy and (b) Gilbert skill score using a probabilistic neural network with different σ.

When the PrNN is used as a classifier for multi-class problems, an appropriate class can be assigned to a sample by the following procedure. Due to the fact that the output of PrNNs is the PDF of a feature input vector being in each class, the sample can be assigned to a class c if its PDF of being in the class c is more than a random chance of 0.5 and higher than other classes. A sample that cannot be assigned to any classes using this rule can be re-assigned by majority classes of its neighbours. The result when three PrNN are used without any class re-assignment is presented in Figure 4.17. The σ of the three PrNNs used for light, moderate and heavy rain are 0.1, 0.07 and 0.04. These values give the highest Gilbert scores for each class when k-fold cross-validation is used. Comparing to the actual classes defined by the observed rain rates, most of the samples that do not fall into any classes belong to zero rain rates. Therefore, the class re-assignment may not be necessary, based on this observation. The Heidke skill score and Hanssen-Kuipers discriminant of the result when non-assigned samples are regarded as zero rain rates are 0.4784 and 0.5099. The scores show that the PrNNs can well classify the feature space into zeros, light, moderate and heavy rain even if the zero rain rates are not included for the training. The classification result has a number of false alarms of light rain class. The class is overestimated especially at the centre of the cyclone. In contrast, the moderate rain is underestimated. However, most of the heavy rain are well captured.

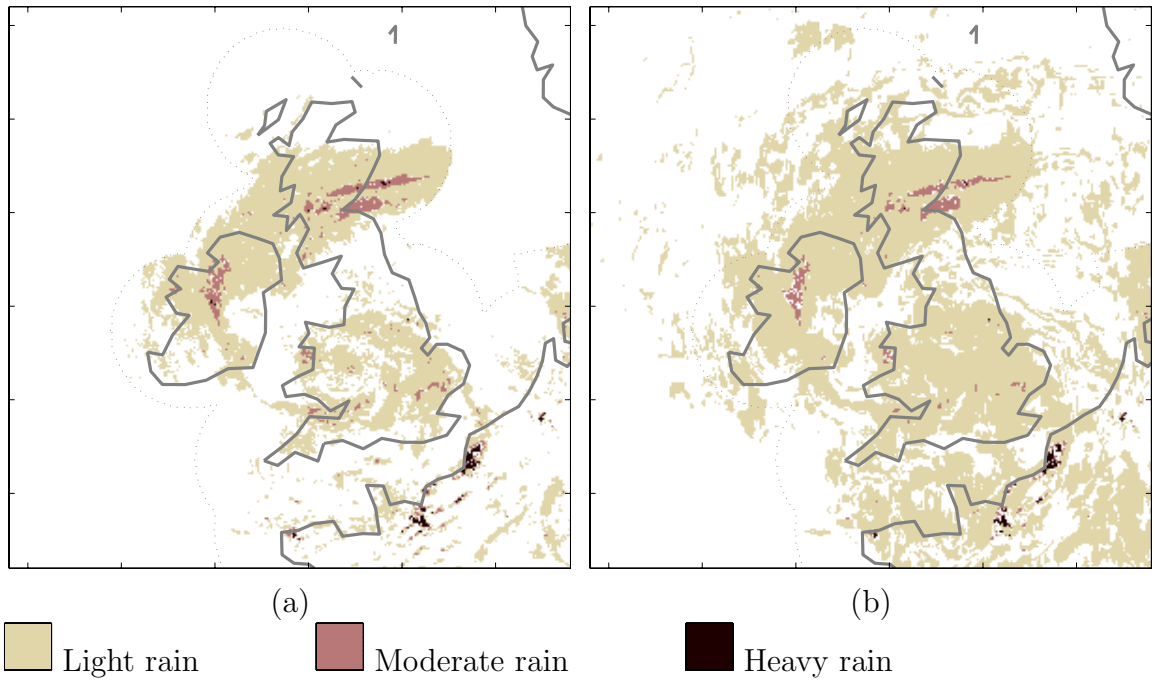


Figure 4.17: (a) Actual rain classes and (b) classification results using three PrNNs.

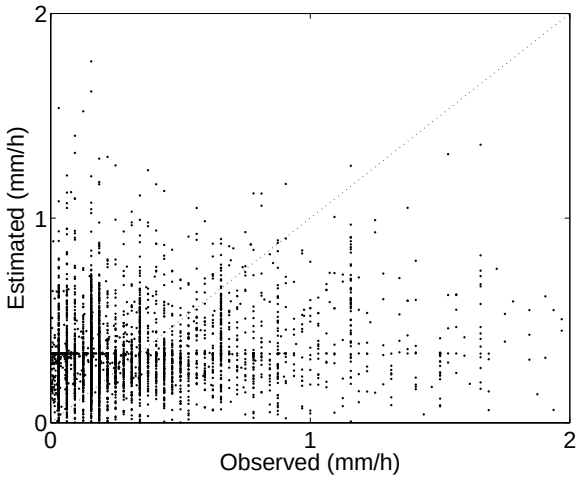
4.4.2 Approximation

Next, a generalised regression neural network (GRNN) and an SVM are applied to the Feature set 2 of each rain class to estimate rain rates. There are two types of feature sets: (a) features that have the highest scores of all textural feature types and (b) features that gives the highest scores of each textural feature type. The Feature set 2 (a) and Feature set 2 (b) are listed in Table 4.6 and 4.7. Table 4.11 shows the correlation coefficient (r) between the estimated and observed rain rates when the GRNN ($\sigma = 1$) and the SVM ($\sigma = 0.4$) are applied to both feature sets of each class. The rain class used for the regression is the actual rain class derived by the observed rain rates, assuming that each sample is correctly classified by the classifiers. This is to verify the approximation results. (The performance when both classifiers and approximators are used is presented later in this section.) The moderate and heavy rain estimates achieved a correlation coefficient to rain rates of 0.4-0.6 for both regression methods. For the case of light rain, r is relatively low.

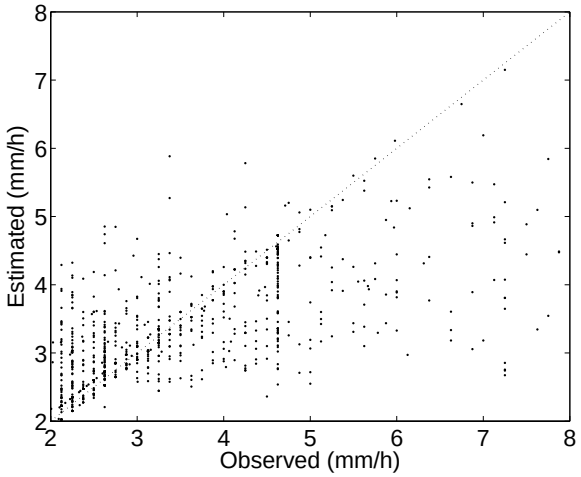
Figure 4.18 shows a scatter plot between the observed and estimated rain rates when the SVM is used with the Feature set 2 (b). The scatter plot reveals that there are cases that a sample of the same observed rain rates is assigned to different rain estimates. This indicates that there exists different cloud textures that give the same rain rates. Additional data might be needed to estimate rain rates for such cloud types. In addition, the SVM tends to underestimate the rain rates > 30 mm/h.

Table 4.11: Correlation coefficients between the observed and the estimated rain rates.

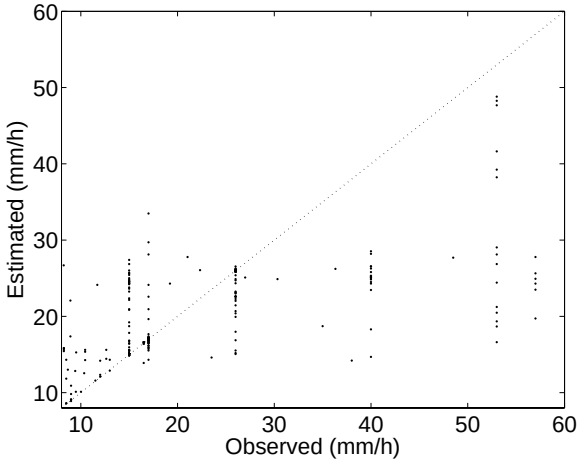
Rain Classes	**F**eature Set 2 (a)		**F**eature set 2 (b)	
	GRNN	SVM	GRNN	SVM
Light	0.12	0.17	0.18	0.15
Moderate	0.51	0.66	0.48	0.68
Heavy	0.44	0.60	0.43	0.62



(a)



(b)



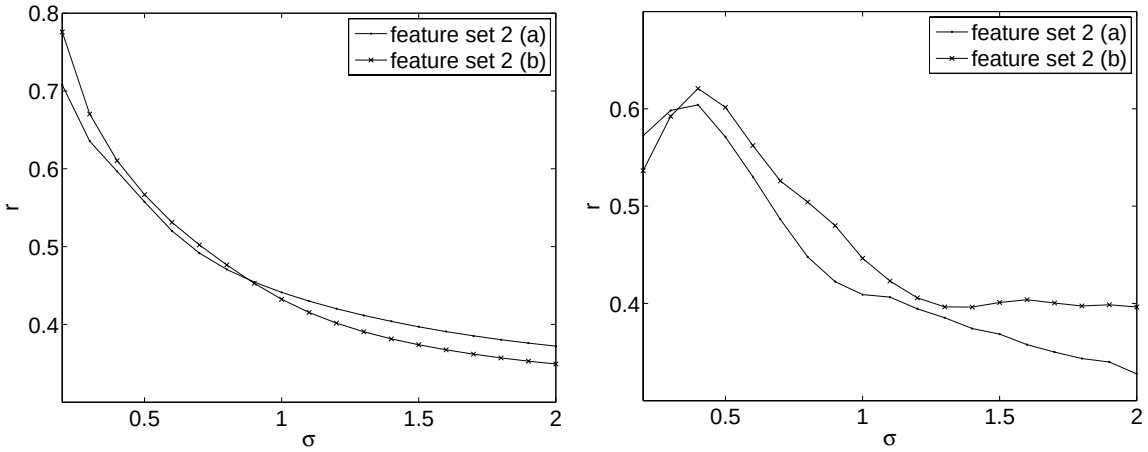
(c)

Figure 4.18: Estimation of rain rates using a support vector machine for (a) light (b) moderate and (c) heavy rain rates.

The performance of both estimators is controlled by the width of a Gaussian function (σ) in a similar way of the PrNN. Figure 4.19 shows the influences of σ when both estimators are applied to heavy rain. Although the same Gaussian function is used, σ in both estimators has different impacts on the classification. As shown in Equation (4.10), a radial basis neural network such as PrNNs attempt to approximate the probability of the feature space given a class ($\Pr(\mathbf{x}|c_i)$, where $\mathbf{x}$ is an input vector) or the surface response of the feature space given a desired output for the case of GRNNs. Its σ controls the width of an area in the input space to which each neuron responds. For example, a larger σ leads to a large area around the input vector where the radial basis neurons respond with significant outputs. Unlike the radial basis function network where the whole feature vectors must be used, the SVM uses a Gaussian function as a kernel between an input vector and the support vectors. Its σ controls the smoothness of the boundary, i.e. a small σ allows the SVM to fit any forms of the boundary whereas a large σ tends to smooth it.

For the GRNN, the correlation r between the observed and estimated rain rates decreases gradually when a larger σ is used. As σ increases, the Feature set 2 (a) tends to be more resistant to the changes of σ; its correlation becomes higher than that of the Feature set 2 (b) when $\sigma > 0.9$.

Similar to the classification task, the SVM has an optimum width of the Gaussian function ($\sigma \approx 0.4$) that gives the highest correlation to the estimated rain rates. Not reaching as high r as the GRNN, the SVM tends to be more robust to σ as its r changes only within ± 0.27 of its optimum value. That is, the SVM can generalise the regression problem while retaining the similar correlation coefficient. The Feature set 2 (b) is generally suitable for the SVM.



(a) (b)

Figure 4.19: Correlation coefficients between the observed rain rates and the rain estimates when (a) GRNN (b) SVM with different σ are used for heavy rain estimation.

Figure 4.20 shows the estimation of the whole system when both classification and approximation are used. Here, three PrNNs are used for classifying samples into zero, light, moderate and heavy rain. Then, a GRNN is applied to the Feature set 2 (a) of each class. The result when the PrNN is used is in Figure 4.20 (b). Meanwhile, an SVM is applied to the Feature set 2 (b) for another rain estimates. Its estimated rain rates is presented in Figure 4.20 (c).

By visual inspection, there is a good agreement between the observed and estimated rain rates. Some overestimated light rain rates occur at the centre of the cyclone as a consequence of misclassification result by the PrNN. The estimation using the SVM gives smoother texture for moderate and heavy rain. The RMS errors of the GRNN and SVM estimates are 0.80 and 0.69 mm/h. Although having higher RMS errors than the histogram matching methods, the estimates show higher correlation of 0.81 and 0.84.

The rule of thumb that thick, cold clouds should produce heavy rain cannot be applied to mid-latitude cyclones as it is seen that most of the cloud bands produce only light rain. However, cloud texture has been shown to have an important role in the estimation of rain rates.

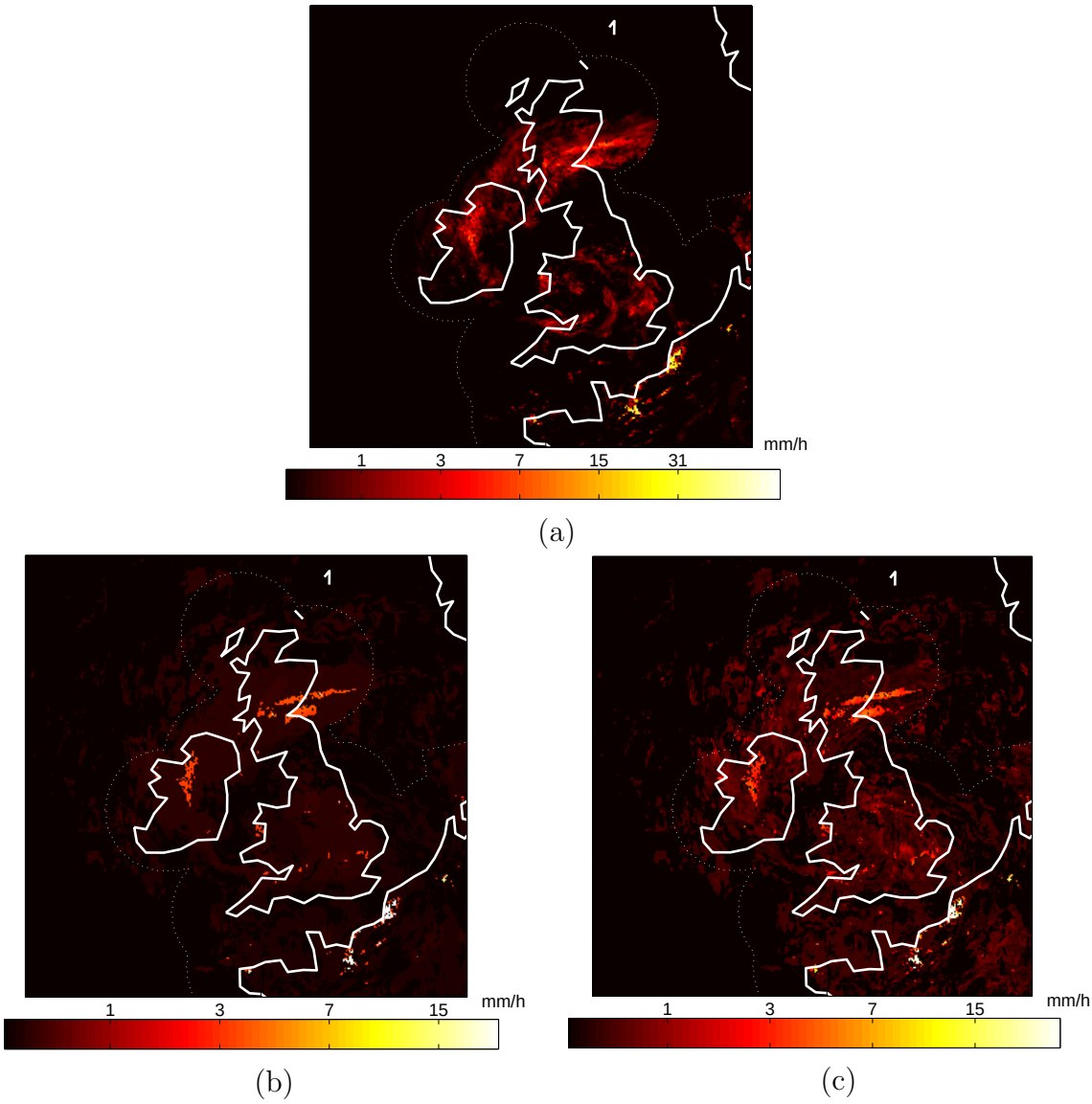


Figure 4.20: A comparison between (a) observed rain rates and estimated rain rates when (b) GRNN and (c) SVM methods are used for approximation and three PrNNs are used for classification.

4.5 Conclusion

In this chapter, the textural features derived from a mid-latitude cyclone on 22-24 June 2004 are related to rain rates. First, the features are ranked according to their linear relationship to the rain rates. Two parameters, i.e. the averaged correlation coefficient r_{av} and the number of samples that have high correlation, are used in combination as a scoring function. By means of the scoring functions, the features that best correlate to the actual rain rates are used for rain estimation. The rain estimation which has been used here consists of a classification and a function approximation to enable robust but continuous rain estimates.

For classifying samples into light, moderate and heavy rain, a PrNN and an SVM have been trained by the selected features. The Gilbert skill score of 0.40-0.49 for the case of light and moderate rain implied high skilfulness of the classifiers and a good separation between classes within the feature space. However, the high dependency feature set that is used for classifying heavy rain leads to a very low skill score. A class is assigned to each sample by considering the maximum probability function of its feature being in each class. The Heidke skill score and Hanssen-Kuipers discriminant of classifying the feature space into zeros, light, moderate and heavy rain are as high as 0.4784 and 0.5099 even if samples of zero rain rates have not been included in the training.

The classification result of each class is then used for approximating rain rates. A GRNN and an SVM are selected for this task. By applying both regression methods to the selected features, the estimation results visually agree with the observed rain rates with a correlation coefficient of 0.81 when the GRNN is applied to the feature set with the highest scores of all feature types and 0.84 when the SVM is applied to the feature set with the highest scores of each texture type. Four types of the textural features are statistical, structural, frequency and spatio-frequency. Though having slightly higher RMS errors, the correlation coefficients of both methods are higher than when only spectral information or the distribution of rain rates is used. This leads to the conclusion that textural information can be used for rain estimation when only geostationary satellite measurements are available. Three channels of the satellite measurements are used. These are infrared, water vapour and visible channels. Any other additional channels or information that may be available are merely unnecessary if higher accuracy cannot be achieved.

Chapter 5

Rain Prediction

The conventional nowcasting methods can be divided into three categories: persistence method, steady-state method[1] and functional method (Dell'Acqua and Gamba, 2003). The persistence method assumes that each pixel in the forecasted image is associated with the same value of the previous image. The second technique assumes that the structures in the original image are transported but unchanged in size and intensity, i.e. the object in the forecasted image is assumed to be equal to the last available image, translated by a suitable motion vector. On the contrary, the functional method uses either a linear function to linearly extrapolate the future value of a pixel or other nonlinear learning machines such as artificial neural networks to learn and predict nonlinear temporal relations. The decisive goal of this chapter is to exploit information derived from geostationary satellite images, temporal features in particular, to predict rain rates.

Figure 5.1 compares the movement of rain patterns in the radar images and clouds in the IR images of the four meteorological events in Chapter 3. These movements are manually tracked by comparing a sequence of images of the events. The lines in the figure indicate paths of rain and clouds as seen from space, starting from the original coordinate (0,0). The dimension of each image is in pixels which can be referred to the actual displacement that the rain and clouds move—a pixel has a resolution of approximately 5 km. The location of clouds in the sequence is marked by $\bullet$ and the location of rain is marked by $\times$ in each image. Each mark is separated by a period of half an hour. Therefore, the distance between each mark can be referred to as speed of the rain and clouds. The number at each mark indicates a correlation coefficient between the values at the present time and the values at the time the object was at the initial location. At the initial location, the coefficient correlation is 1 as the object

[1] also known as Lagrangian persistence or extrapolation method

at the present time is being compared to itself. From the illustration, it can be seen that clouds move in the same direction as their corresponding rain and at the same speed. All except frontal clouds have the differences in displacement between the rain and the cloud lower than 15 pixels, which is approximately 65 km. Therefore, the movement of cloud can be used to predict the location of rain in the future period, assuming that the rain information is not available at that time. This can be useful when the rain information can be acquired less frequently than the IR measurements. An example of such rain information is the microwave-based rain estimates such as SSM/I, which are available at least three times a day.

CMORPH (Joyce et al., 2004) produces global precipitation analyses at very high spatial and temporal resolution. The motion vectors derived, by means of correlation, from infrared images are used to transport the microwave-derived precipitation features during periods when the microwave data are not available. The method is flexible, i.e. any precipitation estimates from any microwave satellite source can be incorporated. However, it governs the movement of the precipitation features only. A quick drop in correlation coefficients as a function of time in Figure 5.1 indicates a low predictability of rain rates, i.e. rain evolves with time. It implies that the relation between clouds and rain rates that is derived at the present time could change considerably in the future period; even the correlation between the actual rain rates themselves drops almost 50% at half an hour later. Among the four events, the convective rain is found as the best to retain the correlation.

To predict the rain rates by movement of clouds, automatic derivation of clouds is needed. In Section 5.1, two motion estimation techniques that are commonly used to derive motion vectors from images are studied using cloud images. The two techniques are maximum cross-correlation and optical flow techniques. In Section 5.2, the derived cloud motions by both techniques will be used for predicting rain rates from the current estimates derived in the previous chapter.

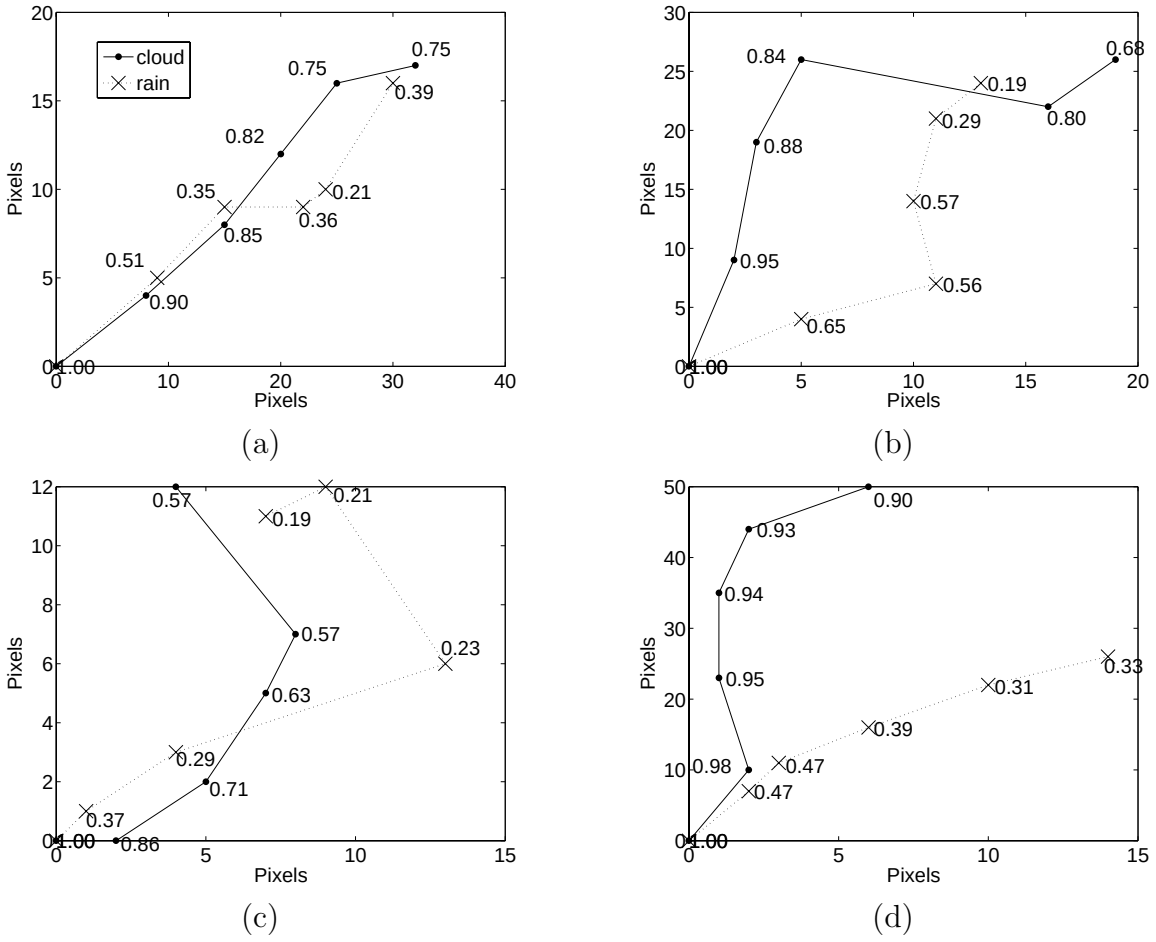


Figure 5.1: Predictability of rain and cloud clouds of (a) convective (b) stratiform (c) central of the vortex and (d) frontal clouds in terms of displacements and correlation coefficients. The locations of cloud and rain are marked by • and × in each image. Each mark is separated by a period of half an hour whereas the number indicates the correlation between the object at the original coordinate (0,0) and itself as time passes by.

5.1 The motion of clouds

5.1.1 Maximum cross correlation

First, the performance of the cross-correlation-based method to derive motions of clouds is studied. The cross-correlation technique is the most common statistical technique used for target matching because of its robustness (Vega-Riveros and Jabbour, 1989; Nillius and Eklundh, 2002). It searches for the same object in two images by finding an area with the maximum cross-correlation coefficient. The cross-correlation can be derived from a similarity measure e.g. the squared Euclidean distance:

$$\begin{aligned} d^2_{f,g}(\Delta\vec{x}) &= \sum_{\vec{x}}[f(\vec{x}) - g(\vec{x}+\Delta\vec{x})]^2 \\ &= \sum_{\vec{x}}[f^2(\vec{x}) - 2f(\vec{x})g(\vec{x}+\Delta\vec{x}) + g^2(\vec{x}+\Delta\vec{x})], \end{aligned}$$

where $f(\vec{x})$ is a target object in the first image and $g(\vec{x})$ is a search area within the second image, $\vec{x}$ denotes a position index of the object, and $\Delta\vec{x}$ is a displacement that the object moves during the time interval between the first image and the second image. The smaller d the more similarity there is between the object and the search area-of-interest. Assuming that $\sum f^2(\vec{x})$ and $\sum g^2(\vec{x}+\Delta\vec{x})$ are constant, the similarity measure can be reduced to the cross-correlation as a function of displacement, namely

$$r(\Delta\vec{x}) = \sum_{\vec{x}} f(\vec{x})g(\vec{x}+\Delta\vec{x}). \tag{5.1}$$

The maximum score of $r(\Delta\vec{x})$ represents the highest probability of being the correct match between the object and the search area. The displacements $\vec{x}$ can then be linked to the speed of the object when the time interval are taken into consideration. However, Equation (5.1) is sensitive to light conditions and size of the object. Moreover, if $\sum f^2(\vec{x})$ varies with position and time, the assumption is violated. Therefore, it is preferable to use the normalised cross-correlation (Lewis, 1995) which is defined by

$$c_n(\Delta\vec{x}) = \frac{\sum[f(\vec{x}) - \bar{f}(\vec{x})][g(\vec{x}+\Delta\vec{x}) - \bar{g}(\vec{x}+\Delta\vec{x})]}{\left(\sum[f(\vec{x}) - \bar{f}(\vec{x})]^2 \sum[g(\vec{x}+\Delta\vec{x}) - \bar{g}(\vec{x}+\Delta\vec{x})]^2\right)^{\frac{1}{2}}}, \tag{5.2}$$

where $\bar{f}(\vec{x})$ is an averaged intensity of the object $f(\vec{x})$. Next, the convective cloud image in Chapter 3 is used for motion analysis. Figure 5.2 shows the cloud on 23 June 2004 at 1030 UTC and its location at half an hour later. The box in Figure 5.2 (a)

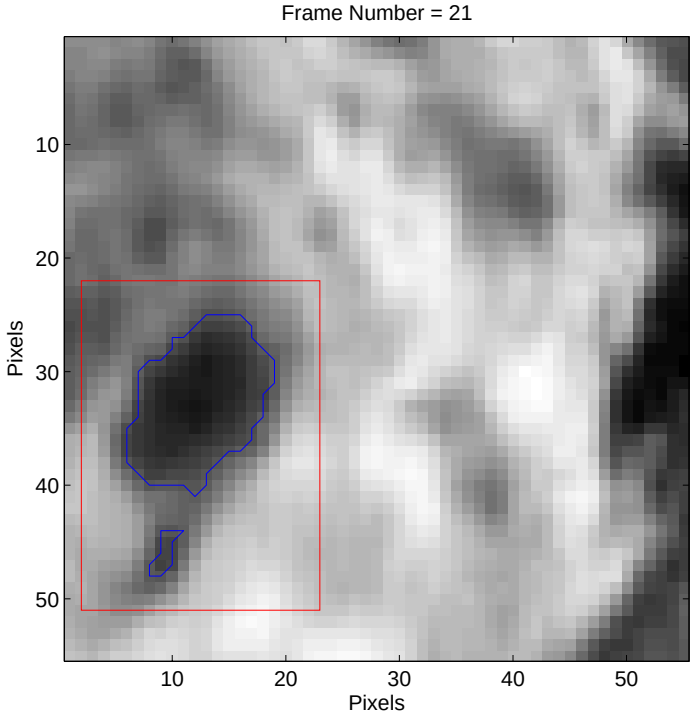
indicates an area that is used to find the cross-correlation in the second image. Figure 5.2 (b) shows that the cloud moves to the north-east of the original location at half an hour later. The cross correlation coefficient between this area and the target image is 0.91, more than the coefficients of other areas. The correlation is presented as contours in Figure 5.3. The contours also show that the surface of the coefficients lies many peaks but only the maximum of them indicates the best match of the target image. The maximum coefficient however does not reach the value of 1 due the fact that the cloud transforms itself with respect to time i.e. it expands and becomes colder half an hour later.

The cross-correlation technique is robust. That is, it can find the correct displacement between a target and a search area no matter how large the displacement is. Nevertheless, it does not support rotation of an object. As illustrated in Figure 5.4, the cross-correlation between the object in Figure 5.4 (a) and its rotated version decreases as a function of angles between the two. The decreasing coefficient results in a shallower peak in the correlation coefficient's surface and might lead the search for the maximum value into the wrong direction. However, the technique is robust to noise as the coefficient changes only slightly when a Gaussian noise is added to the image as shown in Figure 5.5.

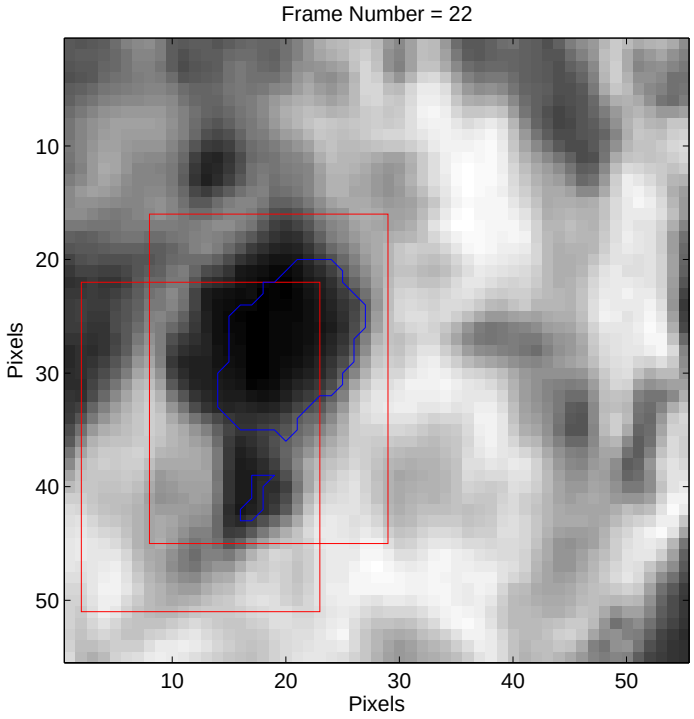
Next, to study the performance of the cross-correlation technique on expansion of an object, a Gaussian surface

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-0.5\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \tag{5.3}$$

is created. The symbols σ_x and σ_y denote standard deviations of the intensity in x and y direction. The object is expanded by increasing σ_x and σ_y by 5%. The result of motions derived by the cross-correlation technique from an expanding object is presented in Figure 5.6. The intensity of the expanded object is shown in colour contours with its peak at the centre of the image in white colour. By increasing σ_x and σ_y, the object is expanding out from the centre. However, the cross-correlation is being trapped to the pattern that is simiar to the original image and produces the motion vectors pointing to north-east, where the first peak of the correlation coefficients is found.



(a)



(b)

Figure 5.2: Two consecutive images showing a movement of a convective cloud on 23 June 2004 over the Netherlands at (a) 1030 UTC and (b) 1100 UTC.

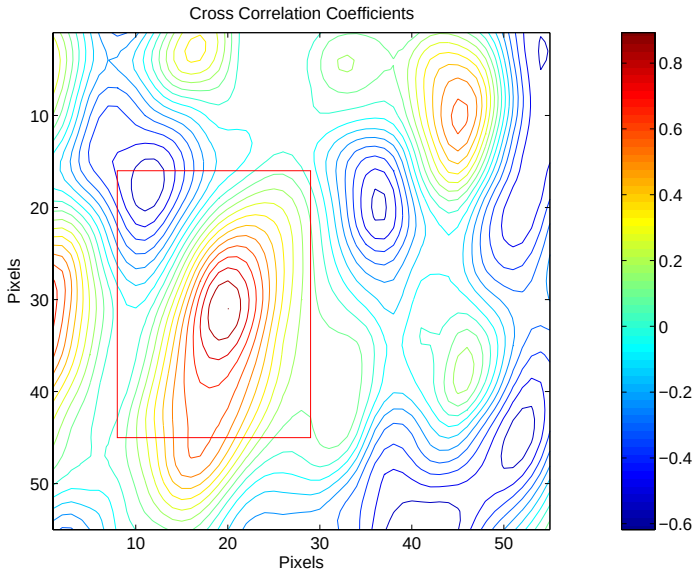


Figure 5.3: Contours of cross-correlation coefficients of the convective cloud in Figure 5.2

To estimate motion on a block-by-block basis by brute force requires expensive computation e.g. an image of size $m \times m$ within a search window $(-w, w)$ needs $m^2(2w+1)^2$ evaluations of the matching criterion. Assuming that the maximum speed of a mid-latitude cyclone is 150 km/h (approximately 15 pixels per half an hour), the search window needs to be $(-15, 15)$ pixels. This leads to 6×10^8 comparisons for an image of 800×800 pixels over the British Isles and western Europe.

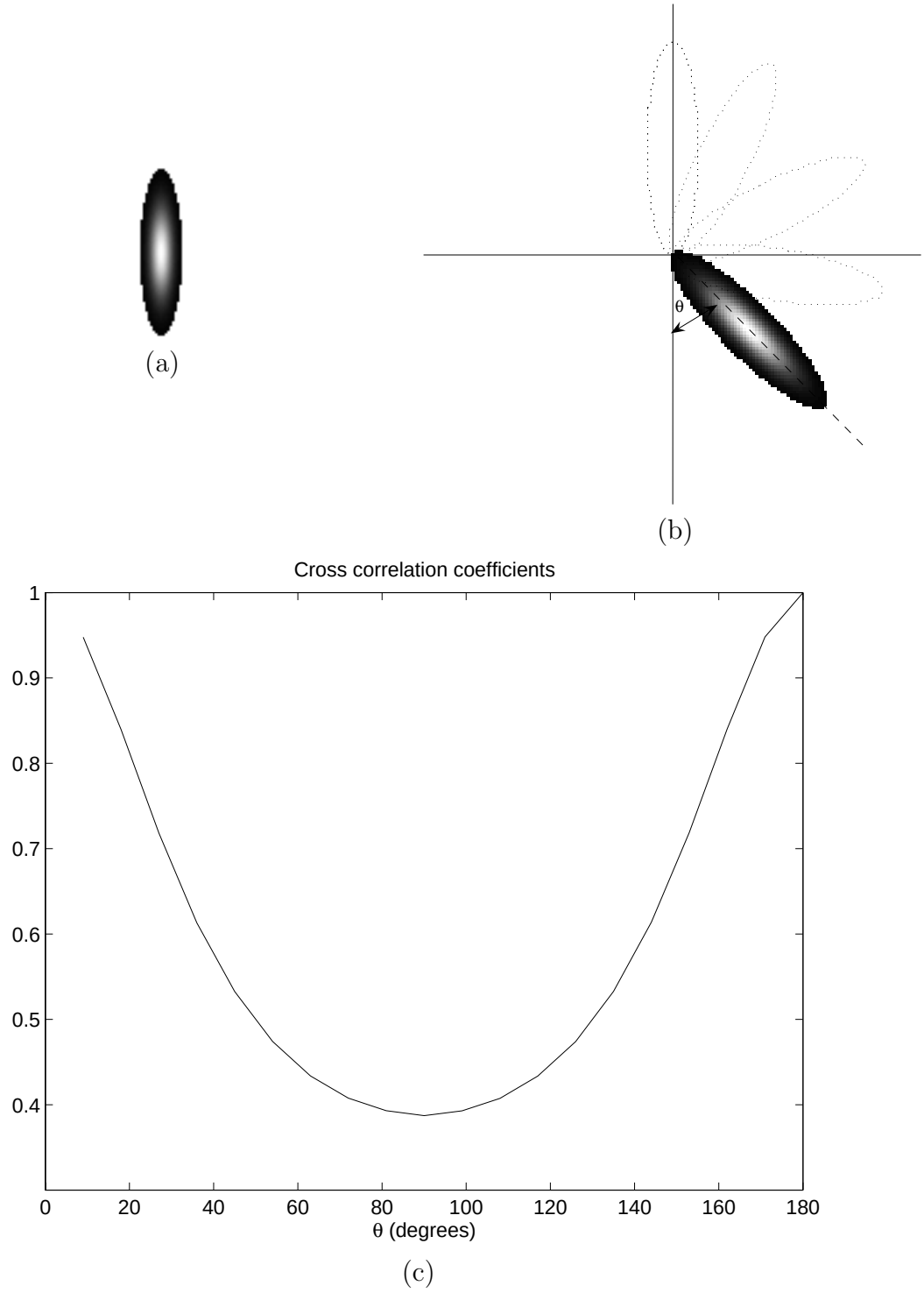


Figure 5.4: An effect of rotation on the cross-correlation technique. (a) a target object (b) rotation of the object and (c) correlation coefficients between a target object and the rotated images.

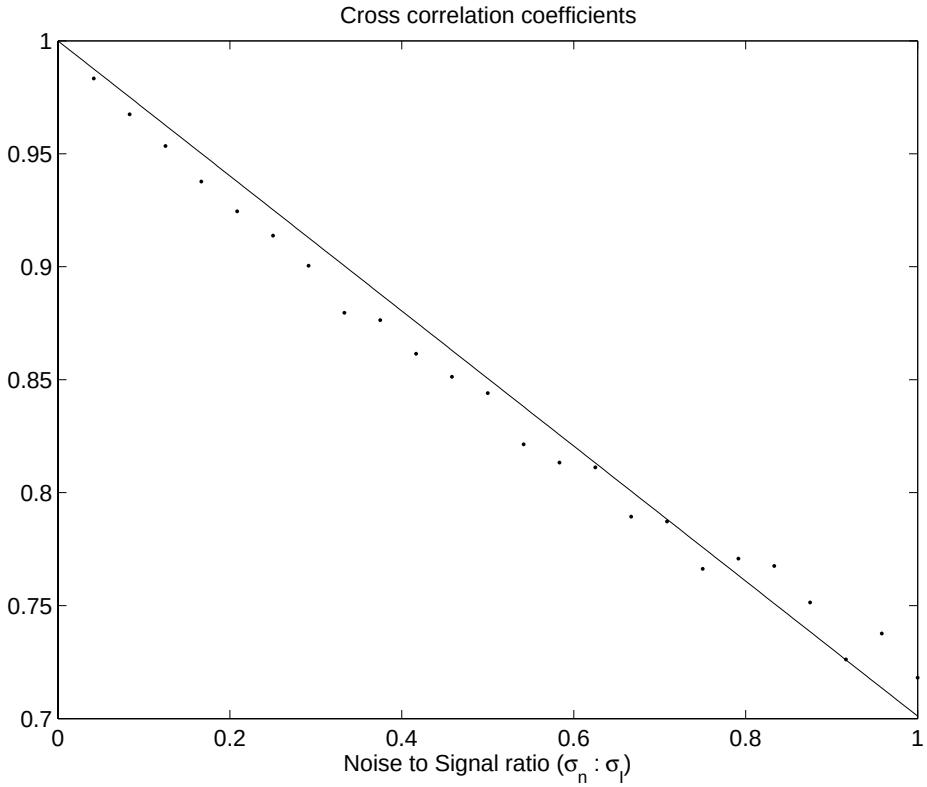


Figure 5.5: An effect of additive Gaussian noise to the cross-correlation technique.

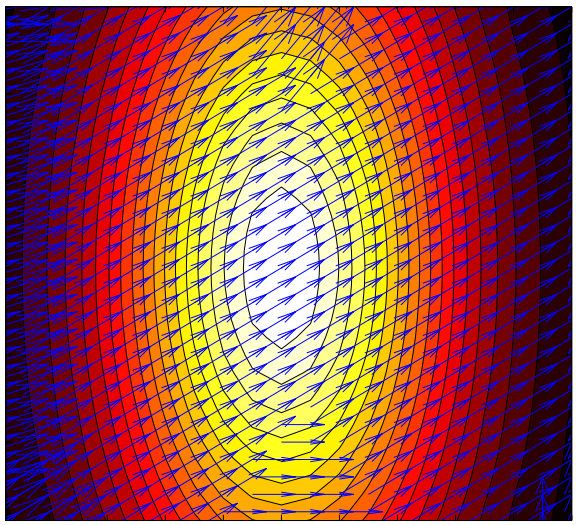


Figure 5.6: Derived velocity from expansion of an object by the cross-correlation technique.

5.1.2 Optical Flow

Another method that has been used to derive motions of clouds is the optical flow technique. Deriving motion by the optical flow technique is commonly based on two main assumptions: constant brightness in Equation (5.4) and smoothness of the velocity in Equation (5.5):

$$(\nabla I) \cdot \vec{v} + I_t = 0, \tag{5.4}$$

$$(\nabla \vec{v})^T \cdot (\nabla \vec{v}) = 0, \tag{5.5}$$

where $I = I(\vec{x}, t)$ denotes the image intensity as a function of space ($\vec{x}$) and time (t), I_t is the first order derivatives of I with respect to time and $\vec{v}$ is the image velocity vector.

Equation (5.4) assumes that the intensities of local image regions e.g. a cloud object are approximately constant over time:

$$I(\vec{x}, t) \cong I(\vec{x} + \delta\vec{x}, t + \delta t) = I(\vec{x}, t) + \nabla I \cdot \delta\vec{x} + I_t \delta t + ... \tag{5.6}$$

Assuming that the high order terms in Taylor series expansion are negligible, we get the brightness constraint, which resembles the law of mass conservation:

$$\begin{aligned} \nabla I \cdot \frac{\partial \vec{x}}{\partial t} + I_t &= 0, \text{ or} \\ \nabla I \cdot \vec{v} + I_t &= 0 \end{aligned} \tag{5.7}$$

Alongside, the smoothness constraint is used on a basis that neighbouring velocities should be practically identical, if corresponding to a solid object. Thus, $\vec{v}$ can be obtained by minimising

$$E(\vec{v}) = \int_{\mathscr{I}} (\nabla I \cdot \vec{v} + I_t)^2 dxdy + \lambda \int_{\mathscr{I}} ||\nabla \vec{v}||^2 dxdy, \tag{5.8}$$

defined over a domain $\mathscr{I}$ (Horn and Schunck, 1981). The Lagrange multiplier λ reflects the relative influence of the smoothness constraint. However, the smoothness assumption is insufficient when dealing with fluid motions because the neighbouring elements are usually free to move according to the underlying fluid dynamics (Zhou et al., 2000).

The conventional optical flow method can be an exact image motion if a number of conditions e.g. uniform illumination, Lambertian surface reflectance[2] and pure translation to the image plane are satisfied (Barron et al., 1994). However, the intensity of infrared images commonly changes from frame to frame due to temperature changes at tops of the clouds. This violates the first assumption. The brightness variability appears not only in infrared images of clouds but also in other applications such as magnetic resonance images (Prince and McVeigh, 1992). Accordingly, Schunck (1985) added a divergence of image velocity which expresses compressible property of the flow to Equation (5.4), namely

$$(\nabla I) \cdot \vec{v} + I(\nabla \cdot \vec{v}) + I_t = 0. \tag{5.9}$$

A number of equations of state suggest that there exists a relation between the temperature, the density and the pressure. For example, the ideal classical gas law

$$P = \rho RT, \tag{5.10}$$

where P, ρ and T is the pressure, the density and the temperature of the ideal gas and R is the ideal gas constant ($8.314472\ J(mol \cdot K)^{-1}$) is the simplest form of these equations of state. Referring the image intensity to the density, Equation (5.9) also matches the mass continuity of the compressible fluids e.g. gases and vapours, which are the main components of clouds. In addition, their density usually depends on the coordinates $\vec{x}$ in space and may depend on time as well.

For the compressible optical flow method, $\vec{v}$ can be found by minimising

$$E(\vec{v}) = \int_{\mathscr{I}} (\nabla I \cdot \vec{v} + I(\nabla \cdot \vec{v}) + I_t)^2 dxdy + \lambda \int_{\mathscr{I}} ||\nabla \vec{v}||^2 dxdy. \tag{5.11}$$

As E is convex, there is a unique minimal value. Defining U and V are the components of $I\vec{v}$ in x and y directions, the Euler-Lagrange equations associated to E are

$$-\frac{\partial}{\partial x}(U_x + V_y + I_t) - \lambda \Delta U = 0 \text{ and} \tag{5.12}$$

$$-\frac{\partial}{\partial y}(U_x + V_y + I_t) - \lambda \Delta V = 0, \tag{5.13}$$

which by using the finite difference method of iteration (Bérériat and Berroir, 2000),

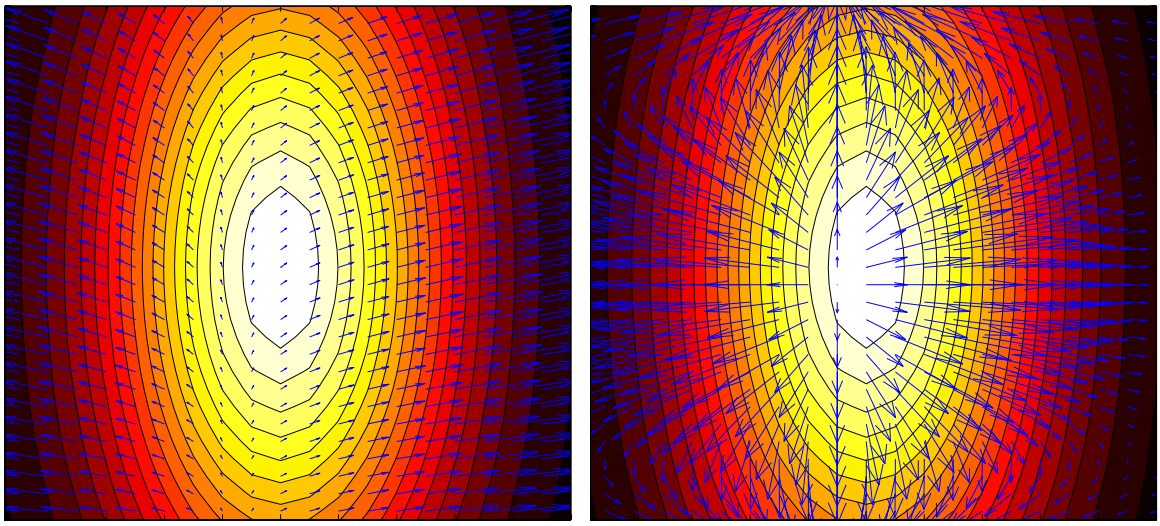
[2]The brightness of a Lambertian surface is constant regardless of viewing angles.

yield

$$U^{k+1} = U^k + \frac{\partial}{\partial x}(U_x^k + V_y^k + I_t) + \lambda \Delta U^k, \tag{5.14}$$

$$V^{k+1} = V^k + \frac{\partial}{\partial y}(U_x^k + V_y^k + I_t) + \lambda \Delta V^k. \tag{5.15}$$

Figure 5.7 shows results when the conventional and the compressible optical flow techniques are applied to the expanded object in Figure 5.6 that the cross-correlation technique fails to cope with. The optical flow methods can describe the expansion better than the cross-correlation technique, especially the compressible optical flow method which can depict the expansion from the centre of the image as shown in Figure 5.6 (b).



(a) Horn and Schunck (1981) (b) Bérériat and Berroir (2000)

Figure 5.7: Derived velocity from expansion of an object by the optical flow techniques.

The compressible optical flow approach tends to work well to recover cloud motion, since it utilises fluid dynamics constraints which are applicable to clouds. Bérériat and Berroir (2000) have adopted this approach to the whole cloud object and, owing to the relationship between cloud's temperature and its elevation, inferred that by their method the cloud volume is conserved. Suvichakorn and Tatnall (2004) (in Appendix D) assumed that a mid-latitude cyclone moves as an equiangular spiral[3] and used a constant angular velocity as another constraint for the compressible optical flow method. The method gave a better result than when the conventional method was

[3]An equiangular spiral is a spiral that forms a constant angle between a line from the origin to any point on the curve.

used.

However, owing to the fact that the optical flow technique is based on the approximation in intensities as in Equation (5.6), the comparison is made between the intensity of a local image region and itself at the new location. This means that the new location of *itself* must have been known approximately. Therefore, if the motion is significant i.e. ≥ 1 pixel/frame, some initial estimates of the displacement is required before calculating the optical flow method. This induces a circular problem as the displacement is also the answer that is desired when we use the optical flow technique. Bergen et al. (1993) proposed using global and local motion models in a hierarchical order to improve the conventional optical flow technique. That is, the image was firstly viewed at a coarse resolution so that the motion can be roughly estimated. Later, the method was applied to finer and finer resolution to produce final motion vectors.

A large displacement between two consecutive images is a result of too low sampling rate that the continuous movement cannot be traced, leading to an aliasing in the frequency domain. Therefore, temporal oversampling (Lim et al., 2004) and an appropriate post-sampling spatial filter (Christmas, 2000) can be used to solve the circular problem. However, working in the frequency domain requires the whole image transform and, consequently, fails to give a good result at a local scale.

Alternatively, we may combine the robustness of the maximum cross-correlation technique in conjunction with the fluid dynamics of the optical flow method as follows. First, an image is divided into non-overlapping windows. Next, the cross correlation technique is applied to each window to find the location of the present window in the consecutive images. Then, the optical flow techniques are applied to the aligned images to find the velocity of each pixel. By this method, there is no need to compute the cross correlation for every pixel, leading to a reduction in computational time. Meanwhile, the optical flow technique becomes more robust by using the initial displacement derived from the cross correlation. The final velocity can be computed from the initial displacement calculated by the cross-correlation technique and refined to a more accurate velocity by the optical flow method by means of vector addition as shown in Figure 5.8.

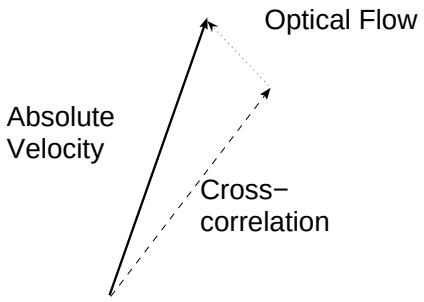
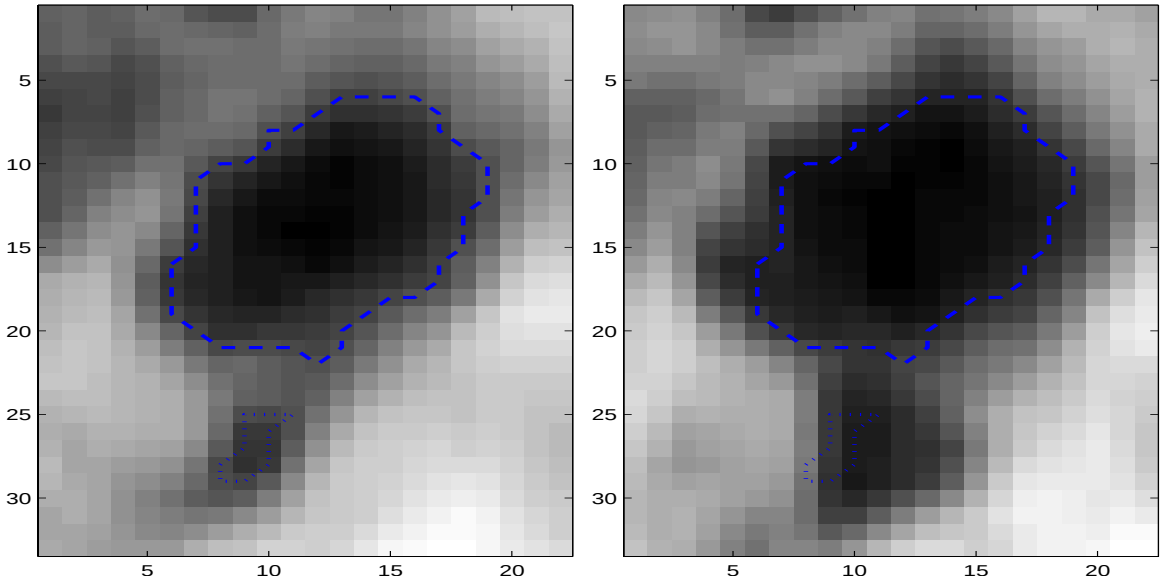


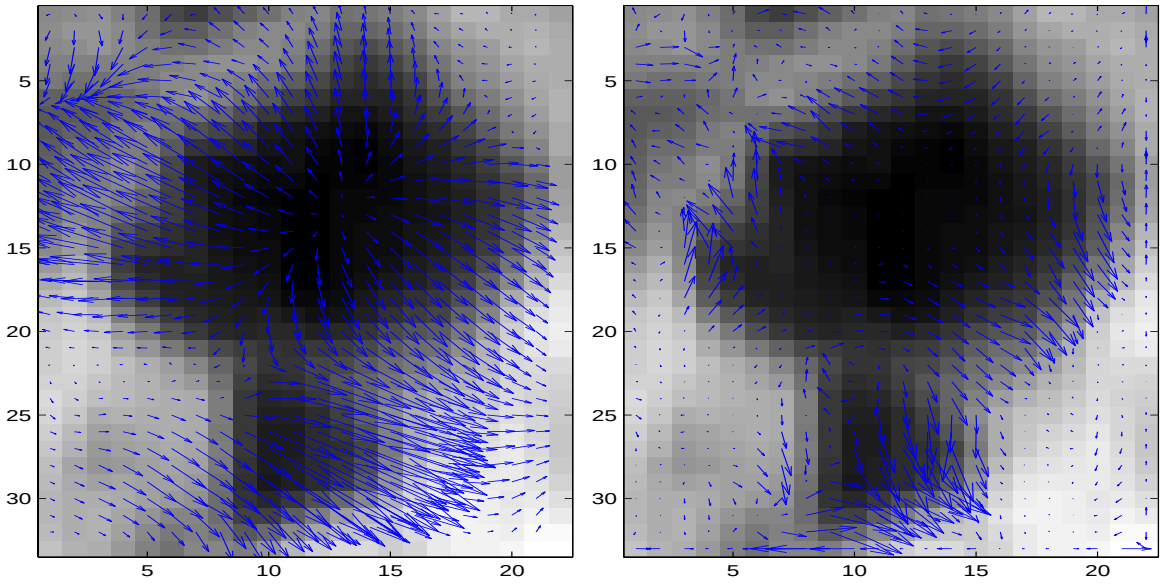
Figure 5.8: Vector addition to derive the velocity from the cross-correlation and the optical flow techniques.

The combination of both techniques can be illustrated in the following experiment. First, the two consecutive images of the convective cloud in Figure 5.2 are aligned by using the maximum cross-correlation. The cloud after alignment presented in Figure 5.9 (b) shows that at half an hour later the large cloud stretched out to the north-west and the south-east directions whereas the small cloud expands in the south-east direction. Figure 5.10 shows the results when the optical flow techniques are applied to the aligned image. From the illustration, both optical flow techniques, i.e. the conventional (Horn and Schunck, 1981) and the compressible (Béréziat and Berroir, 2000) methods, can well capture the expansion of the convective cloud. That is, they can capture the expansion of the cloud in the north-west and the south-east directions, shown as large motion vectors in the figure. However, the conventional technique gives smoother motion vectors than the compressible optical flow method.



(a) (b)

Figure 5.9: The convective cloud in Figure 5.2 after alignment by the maximum cross-correlation.



(a) Horn and Schunck (1981) (b) Bérézia t and Berroir (2000)

Figure 5.10: Derived velocity from expansion of a convective cloud in Figure 5.9 by the optical flow techniques.

5.2 Implementation

Information from geostationary satellite images can be used for rain prediction in two ways: (1) by using the rain estimators derived from textural information at present time and (2) by using cloud motion vectors derived in the previous section. Both methods are summarised in Figure 5.11.

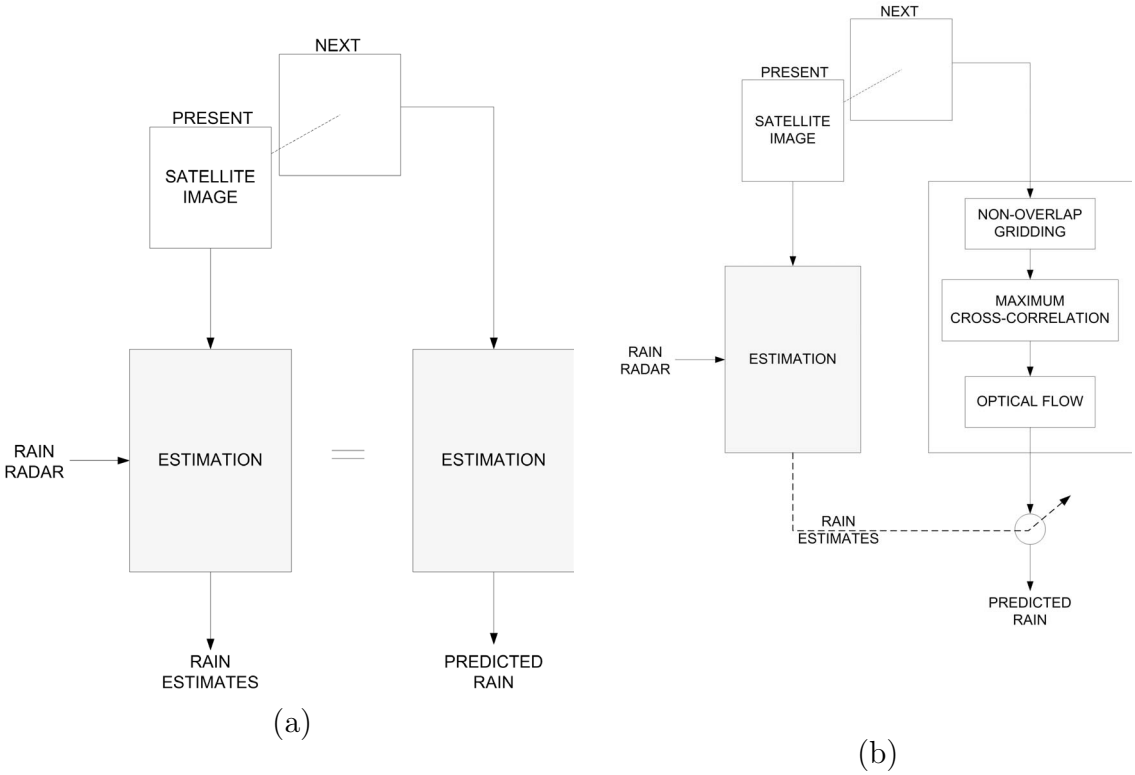


Figure 5.11: Methods that can be used for rain prediction: (a) by using the estimators of the present time and (b) by using motion vectors.

The first method is based on the assumption that the relation between the cloud's textural information and its corresponding rain rates stays consistently for some period of time before being changed by other atmospheric factors. This can be seen as one of the persistence now-casting techniques. Figure 5.12 (b) shows the classification result of using the three PrNNs trained by the Feature set 1 (Table 4.9) derived at 0830 UTC to predict rain rates at 1330 UTC. The Heidke skill score of the classification result drops from 0.48 at 0830 UTC to -0.01 at 1330 UTC whereas the Hanssen-Kuipers discriminant drops from 0.51 to -0.01. At this time, the correlation of the predicted rain rates to the observed rain rates is 0.003 and the RMS error is 4.63 mm/h per sample. On the contrary, the classification result of using the new three PrNNs trained by the same types of feature derived at 1330 UTC can achieve the Heidke skill score

and the Hanssen-Kuipers discriminant of 0.55 and 0.59. The scores are even higher than the result at 0830 UTC. Table 5.1 also shows that the PrNNs trained by the most recent information can achieve as high as 89.0-99.8 % accuracy. The Gilbert skill score of 0.40 on average for each class also indicates a good skill of the PrNNs for classification when the recent rain information is used. This implies that the textural features can perform well for classifying samples into zero, light, moderate and heavy rain but frequent updating the PrNNs is desirable if rain information is available for training.

Table 5.1: Performances of the PrNNs for classifying samples into each rain class.

Accuracy (%)

Methods	**L**ight rain	**M**oderate rain	**H**eavy rain
I	97.76	84.78	99.30
II	11.61	95.87	99.08
III	89.02	99.46	99.82

Gilbert skill score

Methods	**L**ight rain	**M**oderate rain	**H**eavy rain
I	0.86	0.32	0.32
II	0.004	0.005	0.004
III	0.44	0.54	0.23

- Method I: Rain estimation at 0830 UTC. Training features derived at 0830 UTC.
- Method II: Rain prediction at 1330 UTC. Training features derived at 0830 UTC.
- Method III: Rain estimation at 1330 UTC. Training features derived at 1330 UTC.

The second prediction method that has been implemented in the study assumes that clouds propagate at similar speeds to their corresponding rain. To apply the motion derived from the previous section for rain prediction, the IR image is divided into equal non-overlapped grids as shown in Figure 5.13. The maximum cross-correlation (MCC) technique is then applied to the grids to calculate a rough displacement of the rain estimates in the next half an hour (the Meteosat7's time interval). Subsequently, the optical flow method finds dense velocities between each grid at the present time and its matching grid in the next period that gives the highest correlation coefficient. Finally, the motion vectors are then applied to the rain estimates in order to predict how the rain estimates would be in the next half an hour by the continuity equation. If $R(x, y, t)$ denotes a rain rate/rain volume within each grid as a function of space

and time, the continuity equation states that

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\partial R}{\partial t} + \frac{\partial Ru}{\partial x} + \frac{\partial Rv}{\partial y} = 0, \tag{5.16}$$

where u and v are velocities in x and y directions (Germann and Zawadzki, 2002). Using the velocities derived by the optical flow method, the equation gives $\frac{\partial R}{\partial t}$ that can be used to update the rain rate at present time. If there is no update rain information available, this process can be continue to the next couple of hours.

Next, the proposed technique is applied to the mid-latitude cyclone as follows.

1. First, the MCC method is applied to the IR image at present time (0830 UTC) to approximate the location of rain in the next period. The derived displacement at 0900 UTC over the British Isles in illustrated in Figure 5.14 (c). The original location of each grid is displayed in grey colour whereas the new location derived by the MCC method is displayed in black colour. The result follows well with the cyclonic movement of the cyclone. There are some errors e.g. uncommon large displacements Northwest of Ireland due to the formation of a cooling cloud.

2. Second, the velocities of clouds derived by Bereziat's optical flow technique is applied to the rain rates via the continuity equation in Equation (5.16) to approximate how the rain rates would change in the next period. The initial input of the prediction is the *known* rain rates at present time. For a purpose of comparison, the observed rain rates at 0830 UTC and rain estimates at the same time are used as the initial input of the prediction. The rain estimates are derived by the three PrNNs in conjunction with the support vector machine as described in the previous chapter.

3. Finally, the approximated rain rates are mapped onto the distorted grids in Figure 5.14 (c) to give the predicted rain rates as a final result.

Figure 5.15 and 5.16 show the accuracy and correlation coefficients between the predicted and observed rain rates. Using a combination of both MCC and optical flow techniques can provide the correlation of 0.3 to the observed rain rates at an hour later. At the time, the RMS error is 1.3 mm/h. However, the prediction skill gradually decreases with time. The correlation becomes 0.013 and the RMS error becomes 2.72 mm/h three and a half hours later. The rain estimates derived by the SVM gives a slightly lower RMS error and a slightly higher correlation than when the observed rain rates are used. The 5-hour predicted results over the British Isles as shown in Figure

5.18 reveals a visual agreement with the observed rain rates, especially at the centre of the cyclone which moved from Wales to Norfolk. The movement of the centre matched that of the predicted rain rates. The result demonstrates the ability of the proposed motion estimation to handle with the movement of clouds and confirmed the relation between the movement of clouds and rain. The rain band in the North Atlantic Ocean above Ireland of the predicted rain rates when the actual rain rates at 0830 UTC are used as the initial input of the predictor also have the same vertically-arranged texture as that of the observed rain rates. However, it can be noticed that the heavy rain rates at the coastlines of France e.g. in Normandie and Belgium, stay almost the same place at 5 hours later in contrast to those of the predicted rain rates which move eastwardly. The rain could possibly be influenced by other factors, probably spurious effects of the rain radar images, not by the cyclone itself. Such an error can be recognised by considering the previous sequence of the rain radar images. The same method can be applied to correct the unrealistic distorted grid in some areas as seen in Figure 5.14 (c). That is, a set of previous sequence of IR images can be used to correct the derived speed of clouds which may result from generating or decaying clouds.

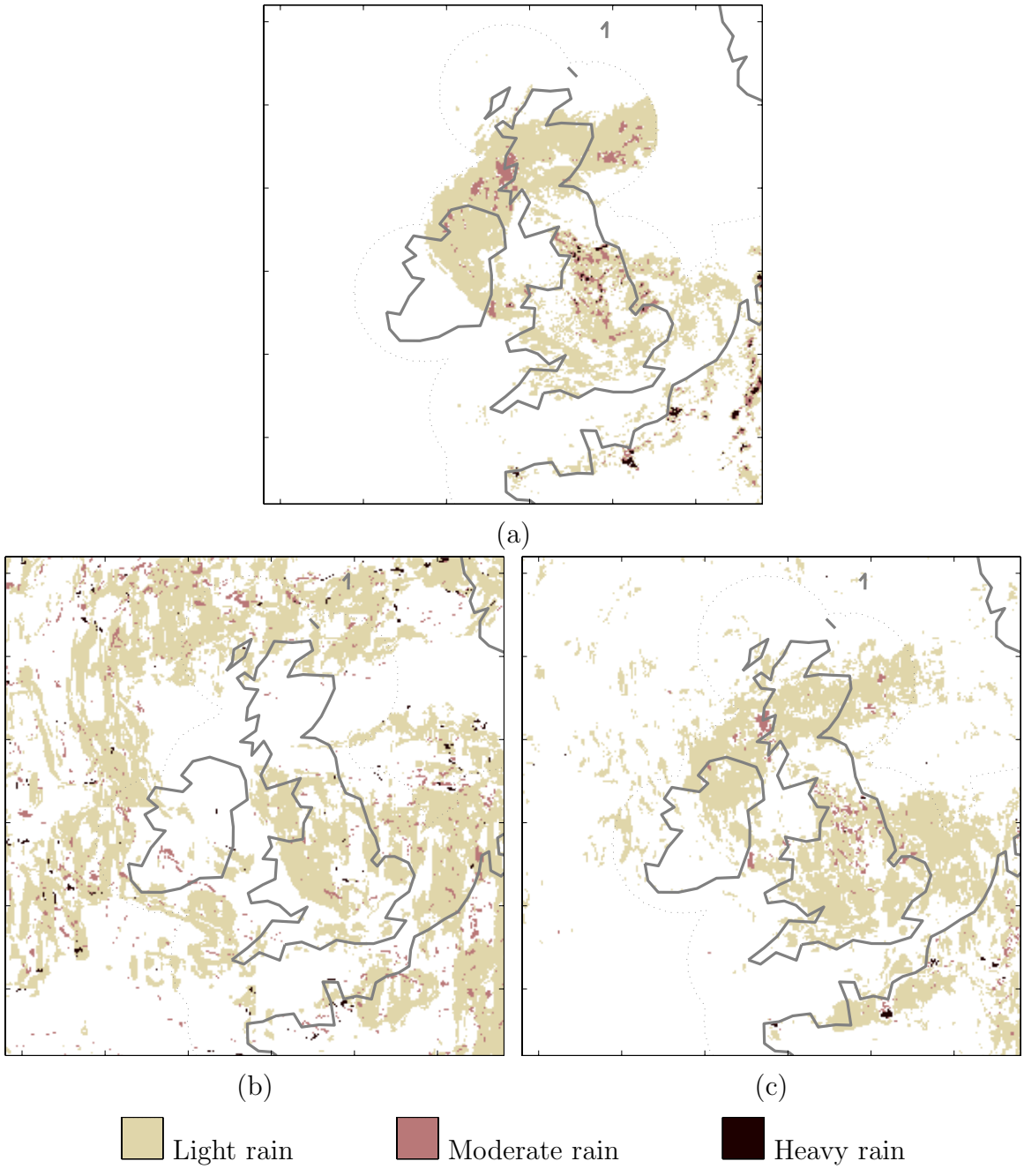
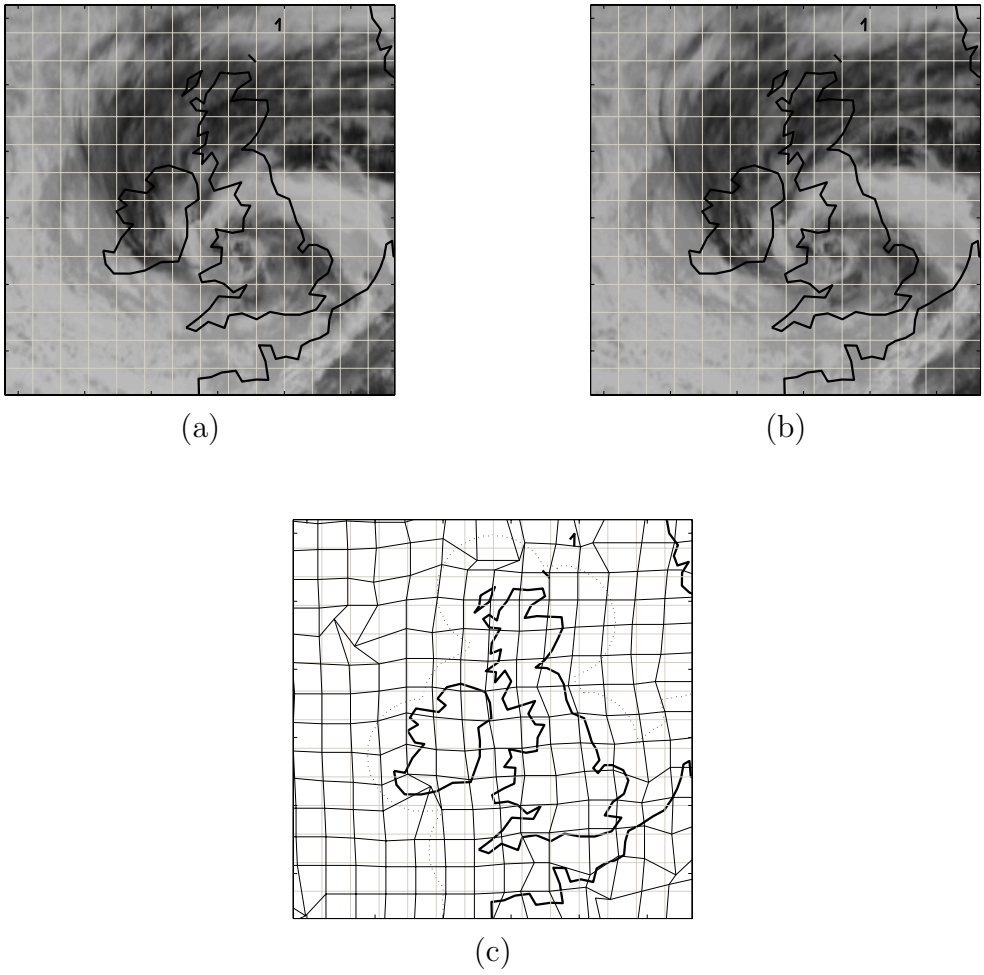


Figure 5.12: A comparison between (a) observed rain rates and predicted rain rates at 1330 UTC when (b) three PrNNs are trained by the features of 0830 UTC and (c) three PrNNs are trained by the features of 1330 UTC.

Figure 5.13: Non-overlapped grid for the maximum cross-correlation technique.



(a) (b)

(c)

Figure 5.14: The mid-latitude cyclone on 23 June 2004 at (a) 0830 UTC and (b) 0900 UTC over the British Isles and the displacement calculated by the MCC method in (c).

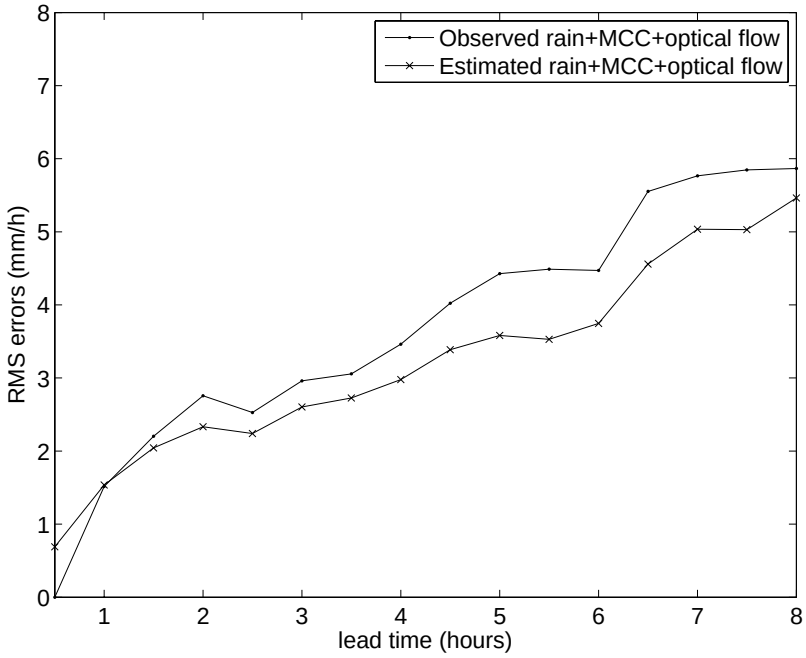


Figure 5.15: RMS errors of the predicted rain rates.

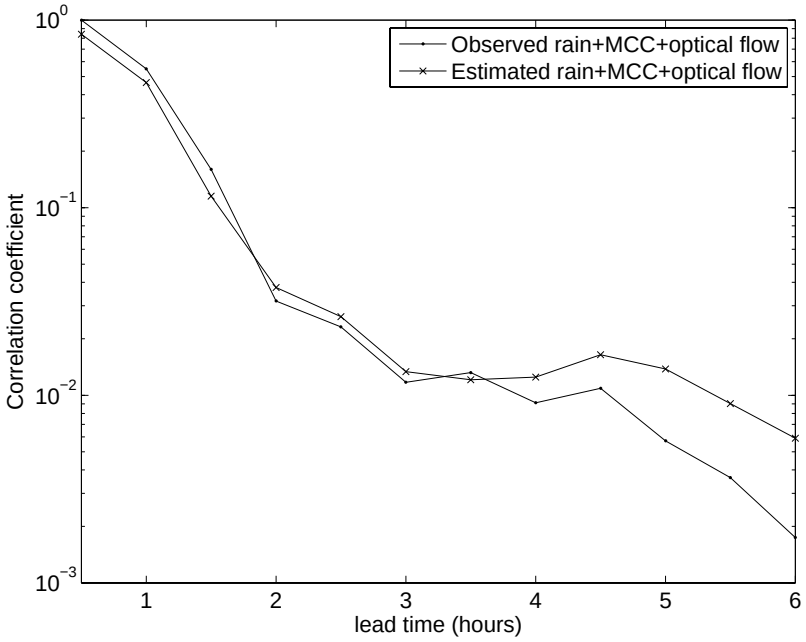


Figure 5.16: Correlation coefficients between observed and predicted rain rates.

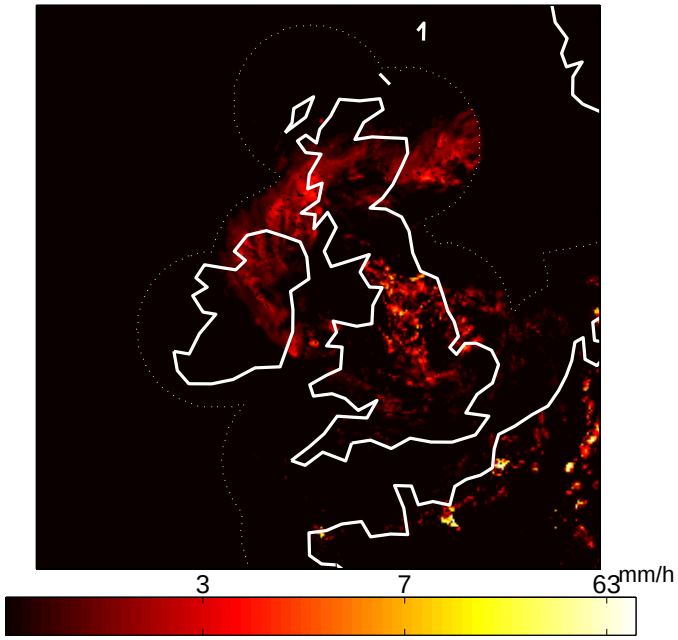


Figure 5.17: Observed rain rates at 1330 UTC.

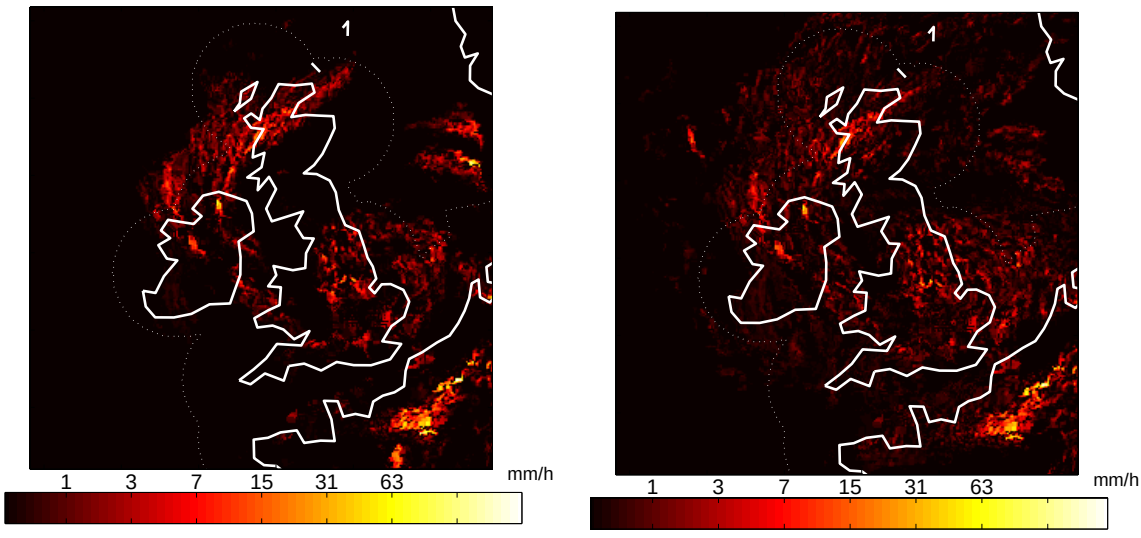


Figure 5.18: Predicted rain rates at 1330 UTC by using velocities from the optical flow method applied to (a) observed rain rates at 0830 UTC and (b) estimated rain rates at 0830 UTC.

5.3 Conclusion

The performance of the cross-correlation-based and optical flow methods to derive motions of clouds is studied. The cross-correlation technique is robust. However, using block-by-block basis makes the method require expensive computation. On the contrary, the optical flow method and its modified version can describe the expansion better than the cross-correlation technique by means of fluid dynamics constraints. However, some initial estimates of the displacement are required before calculating the optical flow method. Alternatively, we may combine the robustness of the maximum cross-correlation technique in conjunction with the fluid dynamics of the optical flow method. The final velocity can be computed from the initial displacement calculated by the cross-correlation technique and refined to a more accurate velocity by the optical flow method by vector addition. By this method, there is no need to compute the cross correlation for every pixel, thus reducing computational time.

The velocities derived from cloud images can be used for prediction. By assuming a steady state of the rain rates, the cloud motion can be used to predict the location of the rain rates in the next period. The displacement can be found by the MCC method. In addition, velocities from the optical flow method can be used to derive how the rain intensities change by using the rule of rain mass continuity. The predicted rain rates from a combination of the MCC and optical flow techniques show a visual agreement with the observed rain rates even at 5 hours later than the present time. It can provide a correlation of 0.3 to the observed rain rates at an hour later. The RMS error is 1.3 mm/h at that time. However, the RMS error increases and the correlation coefficients between the predicted and observed rain rates decrease as a function of lead time.

Chapter 6

Conclusion

The work described in this thesis addresses the potential for estimating and predicting rain rates from satellite imagery. Initially the spectral information available at each pixel from satellite imagery was used to estimate rainfall. The analysis performed confirmed other research that showed that whilst there is a correlation between this spectral information and rainfall, it is not satisfactory for quantifying rainfall. Based on evidence that texture provides a more robust and characteristic property of the clouds, a comprehensive analysis of cloud texture for use in rain estimation which has not yet found in other previous research is done. A number of advanced techniques have been used to identify a relationship between the cloud imagery and the rain rates. In addition, the use of cloud motions derived from infrared images to predict rain rates has also been presented. The case study used was a mid-latitude cyclone over the British Isles and western Europe.

Rain estimation

The appearance and type of cloud are known to be related to rain rate but the motivation for this work was the potential for accurately quantifying the rain rate from the satellite images. Cloud texture from these satellite measurements is also less sensitive to atmospheric attenuation or instrumental noise. The types of textural information examined can be categorised as statistical, using a co-occurrence matrix and a difference matrix; structural using Sobel's filtered Radon transform; frequency using Sobel's filtered, windowed Fast Fourier Transform; and spatio-frequency using Daubechies'wavelet, Coiflet, Symlets and Biorthogonal wavelets. Features derived from the HSV model of the combination of IR, WV and VIS images were also proposed to represent cloud textures. An analysis of these textural features has been

done to find out how they describe cloud textures and how these features relate to rain rates.

The relation to rain rates was described via a correlation coefficient. It has been found that cloud texture features have some similarities to those of rain texture. In addition, some features such as structural features show a much better correlation with rain rates, compared with those of the spectral information. The correlation coefficient was also used to select the feature that best correlated with rain rate prior to the rain estimation. This was in order to reduce the dimensionality of the problem. The scoring function used in the study takes into account not only the averaged correlation coefficient but also the fraction of the samples that have a high correlation. The texture of hue of the HSV false images was found to correlate better with rain rates than the texture of monochrome images. The finding supports the use of this novel feature for rain estimation.

Selected features were then used for classifying clouds into light, moderate and heavy rain. A classifier that is appropriate for this task is a probabilistic neural network (PrNN). By using three PrNNs to separate each rain class and assigning a class label to each sample by its maximum probability density function as proposed in Chapter 4, the classifiers can achieve a Gilbert skill score as high as 0.86 for the case of light rain and 0.49 on average. This reveals that the feature space comprised of textural information can be well separated, compared to overlapped distribution of spectral information. The PrNN is selected to demonstrate the separable feature space due to its simplicity, i.e. only the width of a Gaussian function needed to be adjusted, and its fast computation. Next, another set of features was used for estimating rain rates within each class. Two approximation methods, i.e. a generalised regression neural network and a support vector machine, that have been selected for the task give a result with good visual agreement. Though having higher root-mean-square errors than a linear discriminant method or the histogram matching applied to IR spectral information, the rain estimates of both methods have a significant relationship with the observed rain rates with a correlation coefficient of 0.81 and 0.84.

This method of rain estimation, in which textural features are classified into rain classes and rain rates estimated by a support vector machine, would seem to resolve some of the problems associated with rain-rate analysis from mid-latitude satellite data. The estimation results were robust and continuous rain estimates were achieved. Unlike other rain estimation techniques in which additional information has always been added, the results showed that textural information alone can be used to estimate rain rates. The proposed method is preferable when only satellite measurements are

available. However, frequent updating of the observed rain rates is desirable to retain the accuracy of the estimation. In addition, the three channels of Meteosat7 are the standard channels, i.e. they are common to all geostationary satellites. Therefore, the the proposed textural features and the rain estimation can be applied to other, or newer, geostationary satellites.

Rain prediction

Information from geostationary satellite images has been used for rain prediction in this research in two ways: by using the rain estimators derived from textural information for the present time and by using cloud motion vectors. The first approach is based on the assumption that the present relation between the cloud's textural information and its corresponding rain rates stays the same for some period of time before being changed whereas the second approach assumes that the present rain rates are transported but unchanged in size and intensity.

Two motion estimation techniques were applied to derive the motion vectors from the cloud images for the study of the second prediction approach. First, the cross-correlation technique which calculates the maximum correlation between the initial target and subsequent scene was used to estimate the displacement. This technique is known for its robustness. The second technique exploits a fluid dynamics constraint in order to derive a real velocity rather than the displacement calculated in the cross-correlation technique. This technique, known as the optical flow technique, can be modified to support the compressible property of the image flow, i.e. the intensity of the same object that may change according to time. This makes the method appropriate for deriving the cloud motions. However, this method makes a comparison between the intensity of a local image region and itself at the new location and this requires that the new location of the object must be known approximately. Therefore, in this thesis, a novel method combining the robustness of the cross-correlation technique and the flexibility of the optical flow method was used. The combination was implemented by first computing initial displacements by the cross-correlation technique and then refining them to a more accurate velocity by the optical flow method. The final motion vector of each image pixel was achieved by means of vector addition. The result when the proposed combination was applied to a convective cloud showed that the cross-correlation method could successfully find the future location of the cloud in consecutive images while the optical flow method could describe the expansion of the cloud.

To implement the prediction using the two motion estimation techniques, the displacement of a cloud derived by the maximum cross-correlation technique was used to approximate the location of rain in the next period. Then, its velocity was derived by the optical flow technique and used to predict how the rain rates would change. The rain rates predicted by this novel method provided a correlation of 0.3 to the observed rain rates an hour later. However, as expected, the prediction skill gradually decreased with time. Three and a half hours later the correlation had fallen to 0.013 and the RMS error increased to 2.72 mm/h.

Information from top-of-the cloud measurements has a limited indirect relation to rain rates. The objective of this work was not to remove that limitation but rather to develop a method that can use as much information as possible to aid rain estimation. Based on the analysis of textural features derived from a mid-latitude cyclone, the experimental results shown in this thesis reveal that there is good evidence that using selected textural information can enhance the rain estimation. In addition, the method does not depend on location or time as long as there is initial rain information for selecting appropriate features and training the classifiers. This rainfall rate information could be provided by microwave-based rain estimates such as SSM/I or TMI. Using the motion of rain areas, it has been found that it is possible to use the trained system to estimate rain rates at a later time, and then the trained system can be updated as new rain information becomes available. In conclusion, it has been found that the cloud texture seen in geostationary satellite images can provide useful rain estimation and therefore is a good candidate for the development of operational rain estimation.

Future improvements

There are many improvements that could be done in the future to improve the estimation and prediction results.

- The basic differences between the estimates and the measurements were largely due to evolving meteorological conditions. The UK is a particularly complex region with large differences in rainfall over short distances and highly dynamic weather patterns. The methods described in this thesis are complementary to forecasting techniques such as numerical weather prediction and a combined system could expect to greatly improve results.

- Continuous updating would provide more accurate rain estimation because it would compensate for the ever-changing relationship in time and space between

rain measurements.

- Different types of rain move with different velocities. Therefore, the cloud motion derived by the MCC and optical flow methods could be used as another feature for rain estimation. The implementation of this idea is presented in Suvichakorn and Tatnall (2005) (in Appendix D).

- A more complex assessment method than the correlation coefficient could be used to assess the correlation between textural feature and rain rates. Mutual information that takes into account the whole dependency structure can be applied to measure the dependency between them. The mutual information is based on information entropy first introduced in the communication theory as a quantity of uncertainty in a communication channel (Shannon, 1948). It is successfully used in the registration for medical images (Pluim et al., 2003) and remote sensing images (Chen et al., 2003), outperforming the cross-correlation technique.

- The source of rain information used in this study is rain estimates derived from radar signals and has many limitations, some of which are described in Chapter 3. Improvements in the derivation of rain rates could include calibration with weather station data or the use of other sources of rain information such as SSM/I where there is a scarcity of radar stations.

It has been shown that the use of textural features can be beneficial to rain rate estimation. Textural information combined with motion analysis provide the best potential for predicting the location and rate of rainfall. The simple yet effective estimation and prediction of rain rate can also be applied to future applications where more information at the other frequencies of the Meteosat Second Generation are available.

Bibliography

Adler, R. F., G. J. Huffman, and P. R. Keehn, 1994: Global tropical rain estimates from microwave adjusted geosynchronous IR data. Remote Sensing Reviews, **11**, 125–152.

Adler, R. F. and A. J. Negri, 1988: A satellite technique to estimate tropical convective and stratiform rainfall. Journal of Applied Meteorology, **27**(1), 30–51.

Aguado, E. and J. Burt, 2003: Understanding Weather and Climate. Prentice Hall, 592.

Amato, U., G. Masiello, C. Serio, and V. Cuomo, 2001: Cloud detection from satellite images by wavelets. In Technical proceedings on ERS/Envisat symposiums.

Arkin, P. A., 1979: The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B–scale array. Monthly Weather Review, **107**(10), 1382–1387.

Arvis, V., C. Debain, M. Berducat, and A. Nernassi, 2004: Generalization of the cooccurrence matrix for colour images: Application to colour texture classification. Image Analysis and Stereology, **23**, 63–72.

Azimi-Sadjadi, M. and S. Zekavat, 2000: Cloud classification using support vector machines. In Proceedings of the IEEE 2000 International Geoscience and Remote Sensing Symposium, volume 2, Hawaii, 669-671.

Bader, M., G. Forbes, J. Grant, R. Lilley, and A. Waters, 1995: Images in Weather Forecasting—A Practical Guide for Interpreting Satellite and Radar Imagery. University Press, Cambridge, 499.

Bankert, R., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. Journal of Applied Meteorology, **33**, 909–918.

Barrett, E. C., 1970: Forecasting daily rainfall from satellite data. Monthly Weather Review, **98**, 322–327.

Barrett, E. C. and D. W. Martin, 1981: The Use of Satellite Data in Rainfall Monitoring. Academic Press, New York, 340.

Barron, J. L., D. J. Fleet, and S. S. Beauchemin, 1994: Performance of optical flow techniques. International Journal of Computer Vision, **12**(1), 43–77.

Bellon, A., S. Lovejoy, and G. L. Austin, 1980: Combining satellite and radar data for the short-range forecasting of precipitation. Monthly Weather Review, **108**(10), 1554–1566.

Berendes, T., S. K. Sengupta, R. M. Welch, B. A. Wielicki, and M. Navar, 1998: Cumulus cloud base height estimation from high spatial resolution Landsat data: a Hough transfrom approach. IEEE Transactions on Geoscience and Remote Sensing, **30**(3), 430–443.

Bérézia t, D. and J.-P. Berroir, 2000: Motion estimation on meteorological infrared data using a total brightness invariance hypothesis. Environmental Modelling and Software, **15**(6-7), 513–519.

Bergen, J. R., P. Anandan, K. J. Hanna, and R. Hingorani, 1993: Hierarchical model-based motion estimation. In Motion Analysis and Image Sequence Processing, Sezan, M. I. and Lagendijk, R. L., editors. Kluwer, Norwell, 237-252.

Bonifacio, R., G. Dugdale, and J. Milford, 1993: Sahelian rangeland production in relation to rainfall estimates from Meteosat. International Journal of Remote Sensing, **14**(14), 2695–2711.

Boser, B. E., I. Guyon, and V. Vapnik, 1992: A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory (COLT 1992), Pittsburgh, 144-152.

Brown, O. B. and P. J. Minnett, 1999: MODIS infrared sea surface temperature algorithm. Algorithm Theoretical Basis document version 2.0. Technical report, University of Miami.

Brown, R. and P. Wickham, 1995: Basic principles. In Images in Weather Forecasting—A Practical Guide for Interpreting Satellite and Radar Imagery,

Bader, M., Forbes, G., Grant, J., Lilley, R., and Waters, A., editors. University Press, Cambridge, 7-16.

Calderbank, A., I. Daubechies, W. Sweldens, and B.-L. Yeo, 1997: Lossless image compression using integer to integer wavelet transforms. In Proceedings of the International Conference on Image Processing, volume 1, 596-599.

Canu, S., Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, 2005: SVM and kernel methods Matlab toolbox. Perception Systémes et Information, INSA de Rouen, Rouen, France.

Castro, S., W. Emery, and G. Wick, 2004: Skin and bulk sea surface temperature estimates from passive microwave and thermal infrared satellite imagery and their relationships to atmospheric forcing. In Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium, volume 2, 831-834.

Cellucci, C. J., A. M. Albano, and P. E. Rapp, 2005: Statistical validation of mutual information calculations: comparison of alternative numerical algorithms. Physical Research, **71**(066208), 1–14.

Chen, H.-M., P. K. Varshney, and M. K. Arora, 2003: Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, **41**(11), 2445–2454.

Cherkassky, V. and Y. Ma, 2004: Practical selection of svm parameters and noise estimation for svm regression. Neural Networks, **17**(1), 113–126.

Christmas, W. J., 2000: Filtering requirements for gradient-based optical flow measurement. IEEE Transactions on Image Processing, **9**(10), 1817–1820.

Christodoulou, C. I., S. C. Michaelides, and C. S. Pattichis, 2003: Multifeature texture analysis for the classification of clouds in satellite imagery. IEEE Transactions on Geoscience and Remote Sensing, **41**(11), 2662–2668.

Cohen, A., I. Daubechies, and J.-C. Feauveau, 1992: Biorthogonal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics, **45**(5), 485–560.

Conners, R. W. and C. A. Harlow, 1980: A theoretical comparison of texture algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, **2**(3), 204–222.

Cotton, W. and R. Anthes, 1989: Storm and Cloud Dynamics. Academic Press, 880.

Cristianini, N. and J. Shawe-Taylor, 2000: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, Cambridge, 189.

Daubechies, I., 1990: The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Information Theory, **36**(5), 961–1005.

de Wouwer, G. V., P. Scheunders, and D. V. Dyck, 1999: Statistical texture characterization from discrete wavelet representations. IEEE Transactions on Image Processing, **8**(4), 592–598.

Dell'Acqua, F. and P. Gamba, 1998: Rainfield tracking using radial basis functions. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, volume IV, Seattle, 2068-2070.

Dell'Acqua, F. and P. Gamba, 2003: Pyramidal rain field decomposition using radial basis function neural networks for tracking and forecasting purposes. IEEE Transactions on Geoscience and Remote Sensing, **41**(4), 853–862.

Derrien, M. and H. Gléau, 2005: MSG/SEVIRI cloud mask and type from SAFNWC. International Journal of Remote Sensing, **26**(21), 47074732.

Derrien, M., H. L. Gléau, J.-F. Daloze, and M. Haeffelin, 2005: Validation of SAFNWC/MSG cloud products with one year of SEVIRI data. In EUMETSAT Meteorological Satellite Data Users' conference.

Domingos, P., 1999: Metacost: a general method for making classifiers cost-sensitive. In Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 155-164.

Duhamel, P. and M. Vetterli, 1990: Fast Fourier transforms: a tutorial review and a state of the art. Signal Processing, **19**, 259–299.

Ebert, E. E., 1987: A pattern recognition algorithm for distinguishing surface and cloud types in the polar regions. Journal of Climate and Applied Meteorology, **26**, 1412–1427.

Ebert, E. E., 1989: Analysis of polar clouds from satellite imagery using pattern recognition and a statistical cloud analysis scheme. Journal of Applied Meteorology, **28**(5), 382–399.

Ebert, E. E., 2002: Verifying satellite precipitation estimates for weather and hydrological applications. In 1st International Precipitation Working Group (IPWG) Workshop, Madrid, Spain.

Endlich, R., D. Wolf, D. Hall, and A. E. Brain, 1971: Use of a pattern recognition technique for determining cloud motions from sequences of satellite photographs. Journal of Applied Meteorology, **10**, 105–117.

Ferraro, R. R., 1997: Special sensor microwave imager derived global rainfall estimates for climatological applications. Journal of Geophysical Research, **102**(D14), 16715–736.

Fisher, R., 1936: The use of multiple measurements in taxonomic problems. Annals of Eugenics, **7**, 179–188.

Follansbee, W. A., 1973: Estimation of average daily rainfall from satellite cloud photographs. Technical report, NOAA Technical Memorandum. NESS 44, Washington DC, 39.

Ganchev, T., D. Tasoulis, M. Vrahatis, and N. Fakotakis, 2007: Generalized locally recurrent probabilistic neural networks with application to text-independent speaker verification. Neurocomputing, **70**, 1424–1438.

Garand, L., 1988: Automated recognition of oceanic cloud patterns. part I: Methodology and application to cloud climatology. Journal of Climate, **1**(1), 2039.

Germann, U. and I. Zawadzki, 2002: Scale-dependence of the predictability of precipitation from continental radar images part I: description and methodology. Monthly Weather Review, **130**, 2859–2873.

Gléau, H. and M. Derrien, 2006: Validation report for the PGE01-02-03 of the SAFNWC/MSG. Technical report, MeteoFrance.

Gonzalez, R., R. Woods, and S. Eddins, 2003: Digital Image Processing Using Matlab. Prentice Hall, New York, 609.

Govaerts, Y. M. and M. C. N. Clerbaux, 2004: Operational calibration of the Meteosat radiometer VIS band. IEEE Transactions on Geoscience and Remote Sensing, **42**(9), 1900–1914.

Grazzini, J., A. Turiel, H. Yahia, and I. Herlin, 2005: A multifractal approach for extracting relevant textural areas in satellite meteorological images. Environmental Modelling and Software.

Grazzini, J., H. Yahia, I. Herlin, and A. Turiel, 2003: Analysis and comparison of functional dependencies of multiscale textural features on monospectral infrared images. In Proceedings of 2003 IEEE International Geoscience and Remote Sensing Symposium, volume 3, 764-767.

Griffith, C. G., W. L. Woodley, P. G. Grube, D. W. Martin, J. Stout, and D. N. Sikdar, 1978: Rain estimation from geosynchronous satellite imagery-visible and infrared studies. Monthly Weather Review, **106**(8), 1153–1171.

Grimes, D. I. F., E. Coppola, M. Verdecchia, and G. Visconti, 2003: A neural network approach to real-time rainfall estimation for Africa using satellite data. Journal of Hydrometeorology, **4**(6), 1119–1133.

Gu, Z. Q., C. N. Duncan, P. M. Grant, C. F. N. Cowan, E. Renshaw, and M. Mugglestone, 1991: Textural and spectral features as an aid to cloud classification. International Journal in Remote Sensing, **12**, 953–968.

Gu, Z. Q., C. N. Duncan, E. Renshaw, M. A. Mugglestone, C. F. N. Cowan, and P. M. Grant, 1989: Comparison of techniques for measuring cloud texture in remotely sensed satellite meteorological image data. In IEE Proceedings on Radar and Signal Processing, volume 136(5), 236-248.

Guyon, I. and A. Elisseeff, 2003: An introduction to variable and feature selection. Journal of Machine Learning Research, **3**, 1157–1182.

Haldar, A. and S. Mahadevan, 2000: Probability, Reliability and Statistical Methods in Engineering Design. John Wiley and Sons, Inc., New York, 304.

Harrison, D. L., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. Meteorological Applications, **7**, 135–144.

Hauta-Kasari, M., J. Parkkinen, T. Jaaskelainen, and R. Lenz, 1996: Generalized cooccurrence matrix for multispectral texture analysis. In Proceedings of the 13th International Conference on Pattern Recognition, volume 2, Vienna, 785-789.

Heinemann, T. and J. Kerényi, 2003: The eumetsat multi-sensor precipitation estimate (MPE) : Concept and validation. In Proceedings of the EUMETSAT users conference, Germany.

Heinemann, T., A. Lattanzio, and F. Roveda, 2002: The Eumetsat multi-sensor precipitation estimate (MPE). In Proceedings of the International Precipitation Working Group, Madrid.

Holmlund, K., 2000: The atmospheric motion vector retrieval scheme for Meteosat Second Generation. In Proceedings of the 5th International Winds Workshop, Lorne, Australia.

Horn, B. K. P. and B. G. Schunck, 1981: Determining optical flow. Artificial Intelligence, **17**(1-3), 185–203.

Houze Jr., R., 1993: Cloud Dynamics. Academic Press, 573.

Huang, W., K. K. Lai, Y. Nakamori, and S. Wang, 2004: Forecasting foreign exchange rates with artificial neural networks: A review. International Journal of Information Technology & Decision Making, **3**(1), 145–165.

Ikeuchi, K. and B. K. P. Horn, 1981: Numerical shape from shading and occluding boundaries. Artificial Intelligence, **17**(1-3), 141–184.

Izenman, A. J., 1991: Recent developments in nonparametric density estimation. Journal of the American Statistical Association, **86**(413), 205–224.

Japkowicz, N. and S. Stephen, 2002: The class imbalance problem: a systematic study. Intelligent Data Analysis, **6**(5), 429–450.

Joachims, T., 1999: Making large-scale SVM learning practical. In Advances in Kernel Methods - Support Vector Learning, Schölkopf, B., Burges, C., and Smola, A., editors. MIT Press, Cambridge, MA, 169-184.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of Hydrometeorology, **5**, 487–503.

Khazenie, N. and K. A. Richardson, 1991: Classification of cloud types based on spatial textural measures using NOAA-AVHRR data. In Proceedings of 1991 IEEE International Geoscience and Remote Sensing Symposium, volume 3, 1701-1705.

Kidd, C., 2001: Satellite rainfall climatology: a review. International Journal of Climatology, **21**, 1041–1066.

Kostinski, 2000: On the spatial distribution of cloud particles. Journal of the Atmospheric Sciences, **57**(7), 901–915.

Krennert, T., R. Winkler, B. Zeiner, and V. Zwatz-Meise, 1999: Manual of Synoptic Satellite Meteorology Conceptual Models. Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Austria.

Kurz, M. and M. V. Young, 1995: Occlusion and mature depressions. In Images in Weather Forecasting—A Practical Guide for Interpreting Satellite and Radar Imagery, Bader, M., Forbes, G., Grant, J., Lilley, R., and Waters, A., editors. University Press, Cambridge, 213-286.

Lamei, N., K. D. Hutchison, M. M. Crawford, and N. Khazenie, 1994: Cloud-type discrimination via multispectral textural analysis. Optical Engineering, **33**(04), 1303–1313.

Larsen, R., K. Conradsen, and B. Ersboll, 1998: Estimation of dense image flow fields in fluids. IEEE Transactions on Geoscience and Remote Sensing, **36**(1), 256–264.

Lee, J., R. C. Weger, S. K. Sengupta, and R. M. Welch, 1990: A neural network approach to cloud classification. IEEE Transactions on Geoscience and Remote Sensing, **28**(5), 846–855.

Lensky, I. M. and D. Rosenfeld, 1997: Estimation of precipitation area and rain intensity based on the microphysical properties retrieved from NOAA AVHRR data. Journal of Applied Meteorology, **36**, 234–242.

Lensky, I. M. and D. Rosenfeld, 2006: The time-space exchangeability of satellite retrieved relations between cloud top temperature and particle effective radius. Atmospheric Chemistry and Physics, **6**, 28872894.

Lewis, H. G., 1998: The Use of Shape, Appearance and the Dynamics of Clouds for Satellite Image Interpretation. PhD thesis, University of Southampton.

Lewis, J., 1995: Fast normalized cross-correlation. Vision Interface.

Lim, S. H., J. Apostolopoulos, and A. E. Gamal, 2004: Benefits of temporal oversampling in optical flow estimation. In Proceedings of the 2004 International Conference on Image Processing, 2567-2570.

Liu, H., E. R. Dougherty, J. G. Dy, K. Torkkola, E. Tuv, H. Peng, C. Ding, F. Long, M. Berens, L. Parsons, Z. Zhao, L. Yu, and G. Forman, 2005: Evolving feature selection. IEEE Intelligent Systems, **20**(6), 64–76.

Love, T., V. Kumar, P. Xie, and W. Thiaw, 2004: A 20-year Africa precipitation climatology using satellite and gauge data. In AMS Conference on Applied Climatology, P 5.4.

Lovejoy, S. and G. L. Austin, 1979a: The delineation of rain areas from visible and IR satellite data for GATE and mid-latutudes. Atmosphere-Ocean, **17**, 77–92.

Lovejoy, S. and G. L. Austin, 1979b: The sources of error in rain amount estimating schemes from GOES visible and IR satellite data. Monthly Weather Review, **107**(8), 1048–1054.

Lovejoy, S. and D. Schertzer, 1986: Scale invariance, symmetries, fractals, and stochastic simulations of atmospheric phenomena. Bulletin of the American Meteorological Society, **67**(1), 21–32.

Machado, L. A. T., W. Rossow, R. Guedes, and A. Walker, 1998: Life cycle variations of mesoscale convective systems over the Americans. Monthly Weather Review, **120**, 392–406.

Mallat, S., 1999: A Wavelet Tour of Signal Processing. Academic Press, California, 620, 2 edition.

Mallat, S. G., 1989: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, **11**(7), 674–693.

Mandelbrot, B. B., 1982: The Fractal Geometry of Nature. Freman, San Francisco.

Mandelbrot, B. B., 2001: Gaussian, Self-affinity and Fractals. Springer-Verlag, New York, 654, 3 edition.

Minnis, P., L. Nguyen, D. F. Young, D. R. Doelling, M. L. Nordeen, D. A. Spangenberg, R. Palikonda, G. D. Nowicki, and M. Haeffelin, 2004: Comparison of cloud properties from Meteosat-8 and surface observations. In Proceedings of the Meteosat Second Generation Research Announcement of Opportunity Workshop, 91-96.

Mojsilović, A., M. V. Popović, A. N. Nešković, and A. D. Popović, 1997: Wavelet image extension for analysis and classification of infarcted myocardial tissue. IEEE Transactions on Biomedical Engineering, **44**(9), 856–866.

Negri, A. J., R. F. Adler, and P. J. Wetzel, 1984: Rain estimation from satellites: an examination fo the Griffith-Woodley techniques. Journal of Climate and Applied Meteorology, **23**, 102–116.

New, M., M. Todd, M. Hulme, and P. Jones, 2001: Precipitation measurements and trends in the twentieth century. International Journal of Climatology, **21**, 1899–1922.

Nieman, S. J., W. P. Menzel, C. M. Hayden, D. Gray, S. T. Wanzong, C. S. Velden, and J. Daniels, 1997: Fully automated cloud-drift winds in NESDIS operations. Bulletin of the American Meteorological Society, **78**(6), 1121–1133.

Nillius, P. and J.-O. Eklundh, 2002: Fast block matching with normalized cross-correlation using walsh transforms. Technical Report TRITA-NA-P02/11, ISRN KTH/NA/P–02/11–SE, CVAP-268, Royal Institute of Technology, Stockholm, 10.

O'Sullivan, F., C. H. Wash, M. Stewart, and C. E. Motell, 1990: Rain estimation from infrared and visible GOES satellite data. Journal of Applied Meteorology, **29**(3), 209–223.

Pankiewicz, G. S., 1995: Pattern recognition techniques for the identification of cloud and cloud systems. Meteorological Applications, **2**(3), 257–272.

Papin, C., P. Bouthemy, and G. Rochard, 2002: Unsupervised segmentation of low clouds from infrared Meteosat images based on a contextual spatio-temporal labeling approach. IEEE Transactions on Geoscience and Remote Sensing, **40**(1), 104–114.

Pidwirny, M., 2004: Fundamentals of Physical Geography. Okanagan University College.

Pluim, J. P. W., J. B. A. Maintz, and M. A. Viergever, 2003: Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging, **22**(8), 986–1004.

Prince, J. L. and E. R. McVeigh, 1992: Motion estimation from tagged MR image sequences. IEEE Transactions on Medical Images, **11**(2), 238–249.

Rajabi, M. A. and J. A. R. Blais, 2001: Densification of digital terrain elevations using shape from shading with single satellite imagery. In Proceedings of International Conference on Computational Science, volume 2, 3-12.

Reudenbach, C. and J. Bendix, 2002: Satellite based rainfall retrieval with Meteosat, GOES and MSG in the mid-latitudes and the tropics. In Workshop on Precipitation.

Richards, F. and P. Arkin, 1981: On the relationship between satellite-observed cloud cover and precipitation. Monthly Weather Review, **109**, 1081–1093.

Richards, J. A. and X. Jia, 1999: Remote Sensing Digital Image Analysis: An Introduction. New York Springer, New York, 3 edition.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. Weather and Forecasting, **5**(4), 570–575.

Schmetz, J., K. Holmlund, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, and L. V. de Berg, 1993: Operational cloud-motion winds from Meteosat infrared images. Journal of Applied Meteorology, **32**, 1206–1224.

Schultz, D. M., D. Keyser, and L. F. Bosart, 1998: The effect of large-scale flow on low-level frontal structure and evolution in midlatitude cyclones. Monthly Weather Review, **126**, 1767–1791.

Schunck, B. G., 1985: Image flow fundamentals and future research. In IEEE Proceedings of Computer Vision and Pattern Recognition, San Francisco, USA, 560-571.

Scofield, R. A. and L. E. Spayd Jr., 1984: A technique that uses satellite, RADAR, and conventional data for analyzing and short-range forecasting of precipitation from extratropical cyclones. NOAA Technical Memorandum NESDIS 8, Satellite Application Laboratory, U.S. National Environmental Satellite Data and Information Service, Washington D.C., 51.

Seddon, A. M. and G. E. Hunt, 1985: Segmentation of clouds using cluster analysis. International Journal of Remote Sensing,, **10**, 717–731.

Seze, G. and M. Desbois, 1987: Cloud cover analysis from satellite imagery using spatial and temporal characteristics of the data. Journal of Climate and Applied Meteorology, **26**, 287–303.

Shannon, C. E., 1948: A mathematical theory of communication. The Bell System Technical Journal, **27**, 379–423, 623–656.

Shapiro, M. and S. Gronas, 1999: The Life Cycles of Extratropical Cyclones. American Meteorological Society, 359.

Smith, K. T. and G. L. Austin, 2000: Nowcasting precipitation—a proposal for a way forward. Journal of Hydrology, **239**, 34–45.

Snyder, J., 1987: Map Projections—A Working Manual. U.S. Government Printing Office., Washington, D.C.

Specht, D. F., 1990: Probabilistic neural network. Neural networks, **3**, 109–118.

Specht, D. F., 1991: A general regression neural network. IEEE Transactions on Neural Networks, **2**(6), 568–576.

Stout, J. E., D. W. Martin, and D. N. Sikdar, 1979: Estimating GATE rainfall with geosynchronous satellite images. Monthly Weather Review, **107**, 585–598.

Suvichakorn, A. and A. Tatnall, 2004: Flow analysis of cloud images from geostationary satellites. One Day BMVA symposium at the Royal Statistical Society.

Suvichakorn, A. and A. Tatnall, 2005: The application of cloud texture and motion derived from geostationary satellite images in rain estimation - a study on mid-latitude depressions. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, volume 3, Seoul, 1682-1685.

Svolos, A. E. and A. Todd-Pokropek, 1998: Time and space results of dynamic texture feature extraction in MR and CT image analysis. IEEE Transactions on Information Technology In Biomedicine, **2**(2), 48 – 54.

Tapiador, F. J., C. Kidd, V. Levizzani, and F. S. Marzano, 2004: A neural networks-based fusion technique to esimate half-hourly rainfall estimates at 0.1° resolution from satellite passive microwave andinfrared data. Journal of Applied Meteorology, **43**, 576–592.

Thomson, M. G. A. and D. H. Foster, 1997: Role of second- and third-order statistics in the discriminability of natural images. Journal of the Optical Society of America A: Optics, Image Science, and Vision, **14**(9), 2081–2089.

Tian, B., M. A. Shaikh, M. R. Azimi-Sadjadi, T. H. V. Haar, and D. L. Reinke, 1999: A study of cloud classification with neural networks using spectral and textural features. IEEE Transactions on Neural Networks., **10**(1), 138–151.

Todd, C., C. Kidd, D. Kniveton, and T. J. Bellerby, 2001: A combined satellite infrared and passive microwave technique for estimation of small-scale rainfall. Journal of Atmospheric and Oceanic Technology, **18**(5), 742–755.

Tso, B. and P. M. Mather, 2001: Classification methods for remotely sensed data. Taylor and Francis Inc., New York.

Tsonis, A. A. and G. A. Isaac, 1985: On a new approach for instantaneous rain area delineation in the midlatitudes using GOES data. Journal of Climate and Applied Meteorology, **24**, 1208–1218.

Tuceryan, M. and A. K. Jain, 1998: Texture analysis. In The Handbook of Pattern Recognition and Computer Vision, Chen, C. H., Pau, L. F., and Wang, P. S. P., editors. World Scientific Publishing Co., 2 edition, 207-248.

Turk, F. J., G. Rohaly, J. D. Hawkins, E. A. Smith, A. Grose, F. S. Marzano, A. Mugnai, and V. Levizzani, 2000: Analysis and assimilation of rainfall from blended SSM/I, TRMM and geostationary satellite data. In Proceedings of the 10th Conference on Satellite Meteorology and Oceanography, AMS, Long Beach, CA, 66-69.

Uddstrom, M. J. and W. R. Gray, 1996: Satellite cloud classification and rain-rate estimation using multispectral radiances and measures of spatial texture. Journal of Applied Meteorology, **35**, 839–858.

Uddstrom, M. J., J. A. McGregor, W. R. Gray, and J. W. Kidson, 2001: A high-resolution analysis of cloud amount and type over complex topography. Journal of Applied Meteorology, **40**(1), 16–33.

Vega-Riveros, J. F. and K. Jabbour, 1989: Review of motion analysis techniques. In Proceedings of IEE Communications, Speech and Vision, volume 136, South Carolina, 397-404.

Warrick, A. L. and P. Delaney, 2001: Signals, systems and computers, 1996. In Conference Record of the Thirtieth Asilomar Conference, volume 2, IEEE, 1245-1249.

Welch, R., S.K.Sengupta, A. Goroch, P. Rabindra, N. Rangaraj, and M. Navar, 1992: Polar cloud and surface classification using avhrr imagery: an intercomparison of methods. Journal of Applied Meteorology, **31**, 402–430.

Welch, R. M., K. Kuo, and S. K. Sengupta, 1990: Cloud and surface textural features in polar regions. IEEE Transactions on Geoscience and Remote Sensing, **28**(4), 520–528.

Welch, R. M., S. K. Sengupta, and D. W. Chen, 1988: Cloud field classification based upon high spatial resolution textural features 1. gray level co-occurrence matrix approach. Journal of Geophysical Research, **93**(D10), 12 663–12 681.

Welch, R. M., S. K. Sengupta, and D. W. Chen, 1989: The effect of spatial resolution upon texture-based cloud field classifications. Journal of Geophysical Research, **94**(D12), 14 767–14 781.

Wu, R., J. A. Weinman, and R. T. Chin, 1985: Determination of rainfall rates from GOES satellite images by a pattern recognition technique. Journal of Atmospheric and Oceanic Technology, **2**(3), 314–330.

Xu, L., X. Gao, S. Sorooshian, P. A. Arkin, and B. Imam, 1999: A microwave infrared threshold technique to improve the GOES precipitation index. Journal of Applied Meteorology, **38**(5), 569–579.

Young, M. V., G. S. Forbes, and L. Neil, 1995: Types of cyclogenesis. In Images in Weather Forecasting—A Practical Guide for Interpreting Satellite and Radar Imagery, Bader, M., Forbes, G., Grant, J., Lilley, R., and Waters, A., editors. University Press, Cambridge, 213-286.

Zhou, L., C. Kambhamettu, and D. B. Goldgof, 2000: Fluid structure and motion analysis from multi-spectrum 2D cloud image sequences. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, South Carolina, 744 -751.

Appendix A

Mid-latitude Cyclones

Mid-latitude cyclones[1] are described here to give a better understanding of cyclogenesis or how the cyclones generates, moves and dissipates and how these events can relate to rainfall.

- All cyclones begin as a small-amplitude disturbance along the stationary polar front[2] where cold air from polar regions meets warm air from the south. The collision of these two air masses results in the uplift of the warm air into the upper atmosphere forming an inspiral[3] around a low pressure centre. Now, a storm is born.

- During the middle stages, the storm intensifies and the pressure at the storm's centre drops. The warm air south of the low's centre and between the two fronts–warm and cold, known as the warm sector, is narrowed since the cold front is observed to rotate around the system faster than the warm front. This is sometimes called the mature stage of travelling cyclones.

- As the cold front advances, it will catch up to the warm front causing the air in the warm sector to be lifted into the upper atmosphere. The resulting boundary between the cold and cool air masses is called an occluded front.

- Eventually, a day or two after the occlusion, the cyclone dissipates, winds subside, and the stationary front forms on the surface of the Earth again.

[1] also known as frontal cyclones, travelling cyclones, wave cyclones and extra-tropical cyclones

[2] Zone of transition between polar and tropical air masses . Its average position during the winter is at about 30°latitude and during the summer at about 60°latitute.

[3] An inward-moving spiral pattern. In the Northern Hemisphere, the inspiral will be counter-clockwise since the Coriolis deflection is to the right. In the Southern Hemisphere the inspiral is clockwise.

These cold-front dominated cyclones, known as Norwegian model cyclones, are seen to be meridionally elongated from satellites[4]. However, some recent studies (Shapiro and Gronas, 1999; Young et al., 1995) found that there are also warm-front dominated cyclones where the weak[5] cold front moves nearly perpendicular to the warm front and then air parcels in the warm frontal zone transport their baroclinicity[6] west of the cyclone centre. Unlike the Norwegian model, the catch-up of the warm front never occur. This type of cyclones, named Shapiro-Keyser, tends to be zonally elongated[7] and can be observed over the North Pacific Ocean. Both types of cyclones are presented as conceptual models in Figure A.2. The symbols used in the diagram and charts is in Figure A.9.

The mid-latitude cyclone is rarely motionless and commonly travels about 1200 km in one day. Its direction of movement is generally eastward. Precise movement of this weather system is controlled by the orientation of the polar jet stream[8] in the upper troposphere. An estimate of future movement of the mid-latitude cyclone can be determined by the winds directly behind the cold front (Pidwirny, 2004). Mid-latitude cyclones can produce a wide variety of precipitation types, depending on types of clouds embedded in the fronts. For example, cumulonimbus clouds along cold fronts produce heavy precipitation and can develop into severe thunderstorms if conditions are right, whereas stratiform clouds generate less intense, wide-spread precipitation ahead of the warm front. On the contrary, the warm sector may produce none to only a brief, convective shower if it has enough moisture. Next, details about cold fronts, warm sectors and warm fronts, together with their characteristics on satellite images are reviewed.

Cold front

Cold fronts vary in associated temperature contrast, wind shift, and precipitation regimes. Two basic groups of cold fronts have been suggested: Ana and Kata cold front. The conceptual models of both groups are shown in Figure A.3.

[4]with much north/south component and migrate slowly from west to east

[5]This is defined by a weakness in horizontal temperature gradient amplitude along the poleward portion near the low centre.

[6]Baroclinicity is a cold air advection/warm air advection couplet that increases atmospheric instability. On analysis and forecast charts it is the isotherms crossing the height contours.

[7]migrate rapidly from west to east with little north/south component

[8]Polar jet stream is a relatively fast uniform winds concentrated within the upper atmosphere in a narrow band. It exists in the mid-latitudes at an altitude of approximately 10 km.

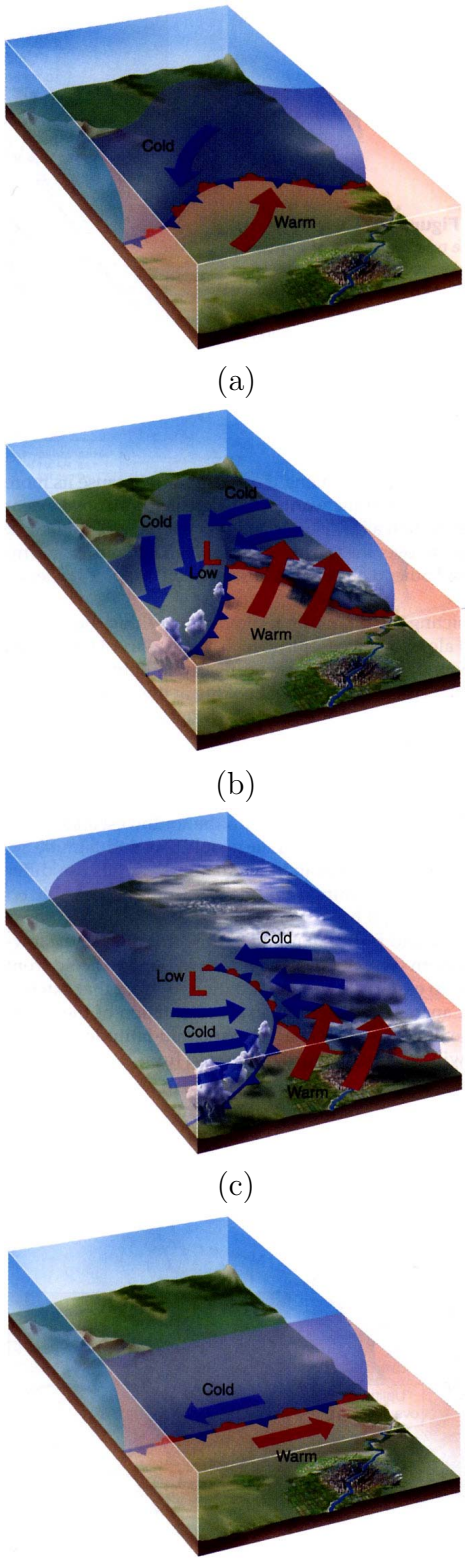
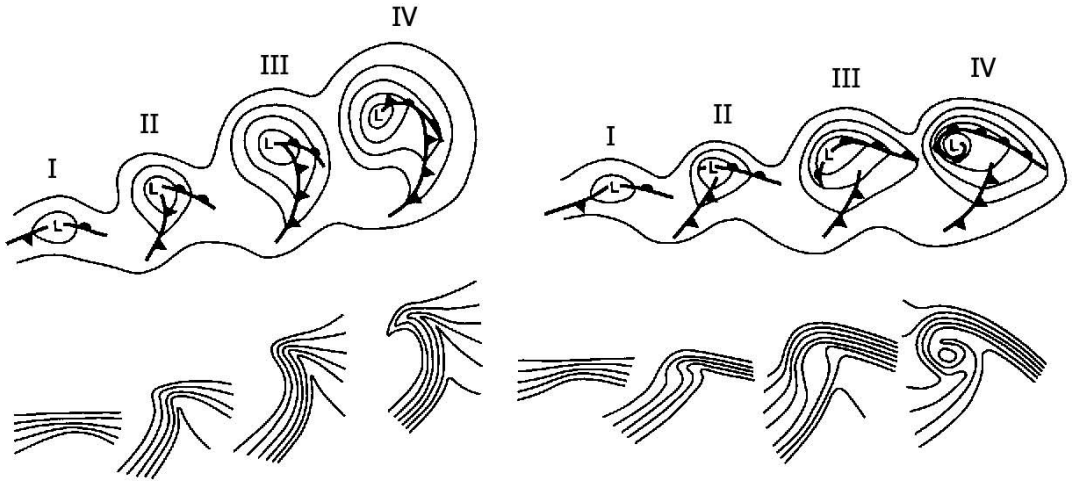


Figure A.1: Illustration of a Norwegian cyclone life cycle. Source: Aguado and Burt (2003).



(a) Norwegian cyclone model (b) Shapiro and Keyser model

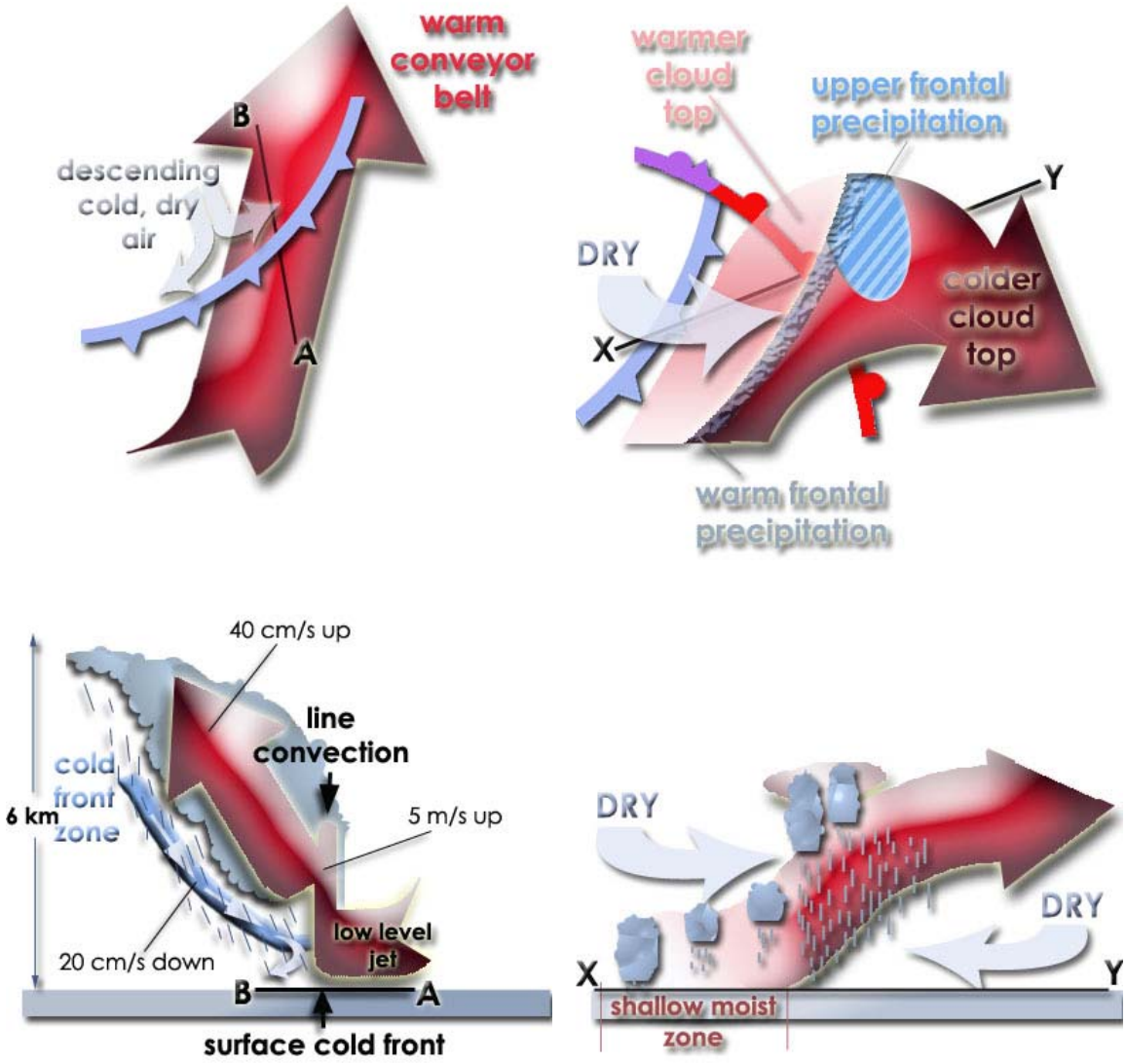
Figure A.2: Conceptual models of cyclone evolution (Schultz et al., 1998) showing lower-tropospheric (e.g. 850-hPa) geopotential height and fronts (top), and lower-tropospheric potential temperature (bottom). (a) Norwegian cyclone model: (I) incipient frontal cyclone, (II) and (III) narrowing warm sector, (IV) occlusion; (b) ShapiroKeyser cyclone model: (I) incipient frontal cyclone, (II) frontal fracture,(III) frontal T-bone and bent-back front, (IV) frontal T-bone and warm seclusion. The characteristic scale of the cyclones based on the distance from the geopotential height minimum, denoted by L, to the outermost geopotential height contour in stage IV is 1000 km.

Ana-cold fronts are associated with relatively deep/sharp troughs aloft, i.e. the upper flow is roughly parallel to the surface front. In these cases, the surface fronts often possess sharp temperature changes and wind shifts, and significant vertical motion, resulting in a line of showers or thunderstorms along the front and extensive post-frontal precipitation. Most of this classical type of cold frontal bands have cold tops, representing as bright in IR images. Such bands can sometimes have a distinct forward edge. When this edge is discernible, the surface front is located close to it. Sample images of Ana cold front as seen from Meteosat are shown in Figure A.4. In VIS images, the cloud area is similar to that in the IR images, although the brightest (thickest) cloud lies towards the leading edges of the frontal cloud band. In WV images, there is a broad band of moisture, with a sharp edge on the cold side of the front.

Kata-cold fronts are not as well-defined as Ana fronts, and usually possess weak-to-moderate temperature gradients and scattered showery precipitation along the front. Kata fronts occur with a less amplified flow pattern (i.e. more westerly momentum) aloft, which results in a system-relative forward sloping ascent and a band or area of more significant precipitation ahead of the surface front. Precipitation patterns associated with Kata fronts have been explained via a split cold front or cold front aloft (CFA) model. The split cold front is associated with the leading edge of cold advection aloft, as well as with frontogenesis and convergence between warm, moist air ahead of the CFA and cooler, drier air behind it. This typically produces significant lift and a band or area of stratiform and/or convective precipitation along and ahead of the CFA within the surface warm sector. In IR images, the front appears as two adjacent bands with distinctly different cloud-top temperature, clearly seen in Figure A.5 (a) over the Oceania (eastern side of New Caledonia). The leading band has cold tops and a sharp rear edge in both IR and WV images, marking a discontinuity in humidity aloft. In VIS images, the edge may often be observed from a shadow cast onto the trailing band. The rear edge of the warmer cloud band can also be clearly seen. In contrast, a cloud in IR images is sometimes difficult to distinguish from the underlying surface. These bands are most common and vigorous near developing depressions. The TRMM Microwave Imager in Figure A.5 (c) shows that there are heavy rainfalls along the edge of colder clouds.

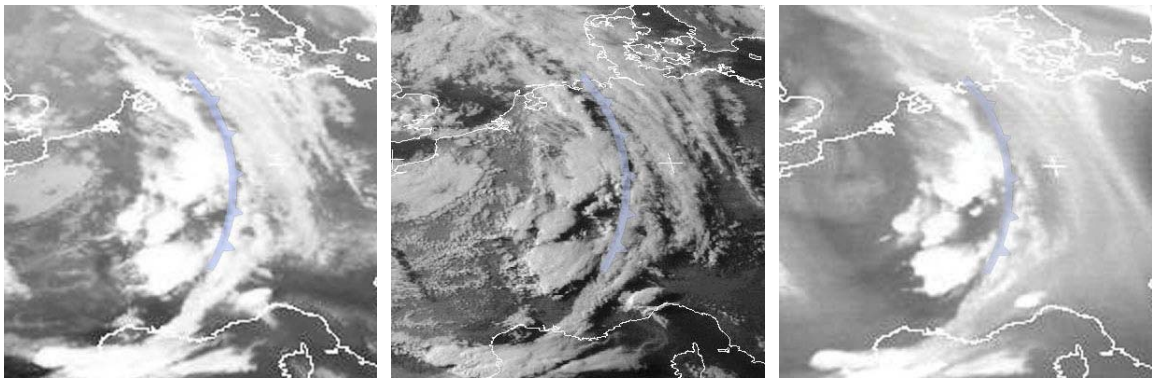
Warm front and warm sector

In satellite imagery warm fronts can have two main different shapes: they either appear as a *warm front shield* or as a *warm front band*, distinguished by cloudiness within the warm sector. For a warm front shield in a well-developed case, the grey shades in VIS images are generally white within the warm sector and near to the surface warm front. The grey shades become gradually more and more grey towards the forward cloud edge. In IR images the grey shades of a cloud shield are usually white indicating cold cloud tops. Its substructures may differ from case to case. The pixel values in WV images within the area of the cloud shield are high. At the leading edge of as cloud shield, there is a sharp gradient from white to black indicating dry air at the cyclonic jet side. For a warm front band, its ideal main characteristics in a VIS image of the cloud pattern associated with a warm frontal cloud band are the grey shades, generally bright at the rear edge and becoming darker towards the forward edge; whereas in the IR image the grey shades appear brightly towards the forward



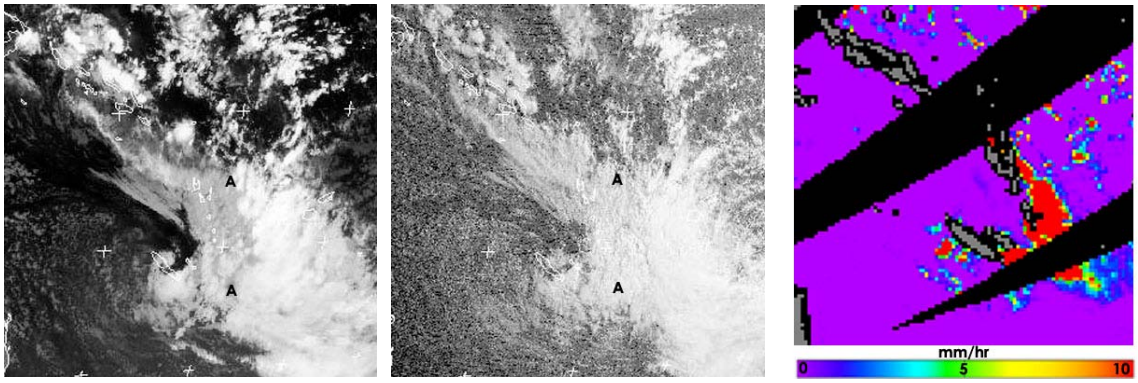
(a) (b)

Figure A.3: Top view (upper) and side view (lower) of (a) Ana cold front and (b) Kata cold front model.



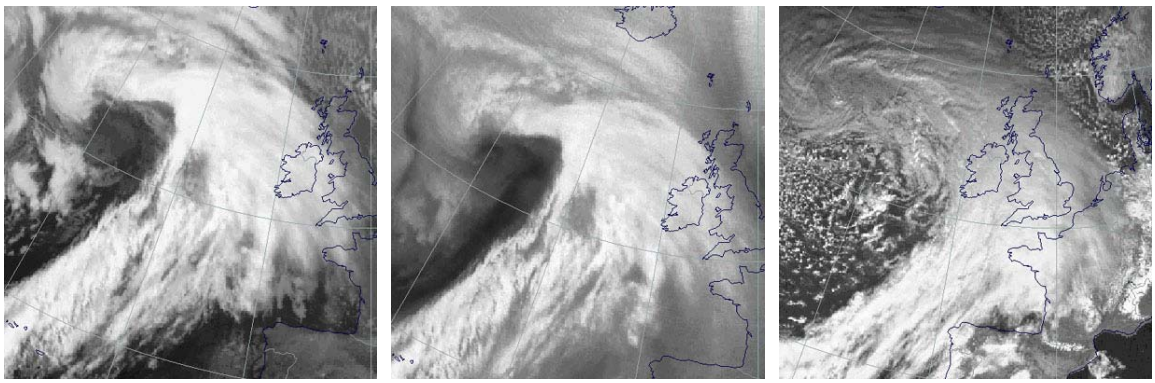
(a) (b) (c)

Figure A.4: (a) IR, (b) VIS and (c) WV images of a classical cold front, accompanied with single cell thunderstorms on the right side and a short band of warm frontal at the bottom. Source: Meteosat over Germany on 30 July 2003, 1800 UTC.



(a) (b) (c)

Figure A.5: (a) IR and (b) VIS images of a Kata cold front from GMS over Pacific ocean on 15 July 2003 and its corresponding rainfall rate from TMI.



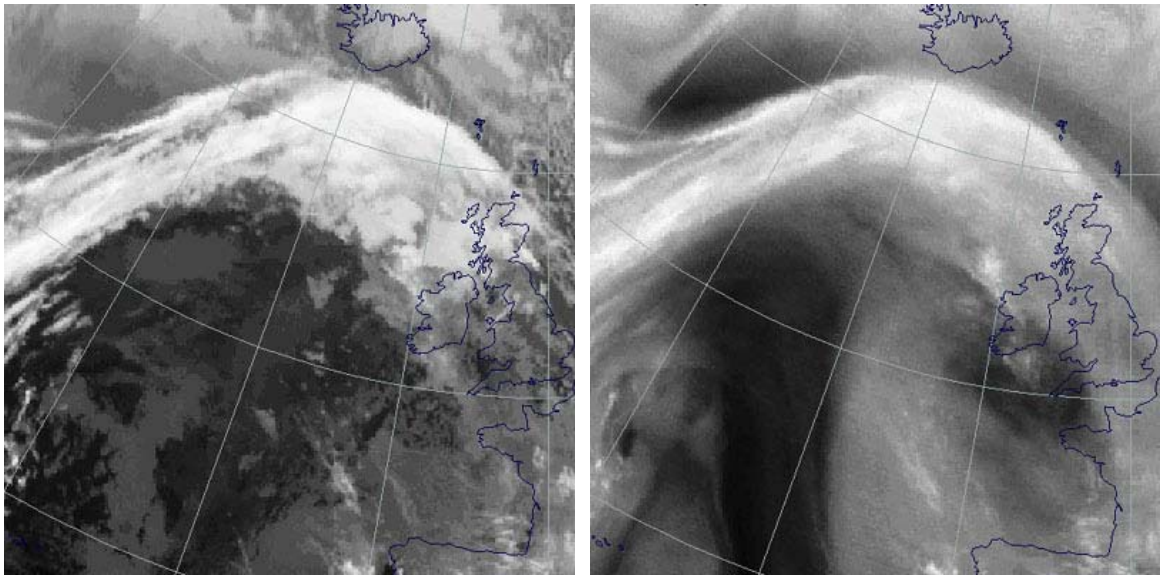
(a) (b) (c)

Figure A.6: (a) IR and (b) WV images of a warm front shield from the Atlantic (east of approximately 20°W) across Ireland to the Bay of Biscay on 27 January 1995 at 0030 UTC and (c) VIS image at 1200 UTC shows an eastward propagation of this shield as far as the Benelux countries and west Germany. Source: Krennert et al. (1999).

and downstream cloud edge. In WV images, the band has similar appearance to the warm front shield—bright within the area and a sharp gradient at the leading edge. Examples of warm front shield and warm front band are shown in Figure A.6 and A.7.

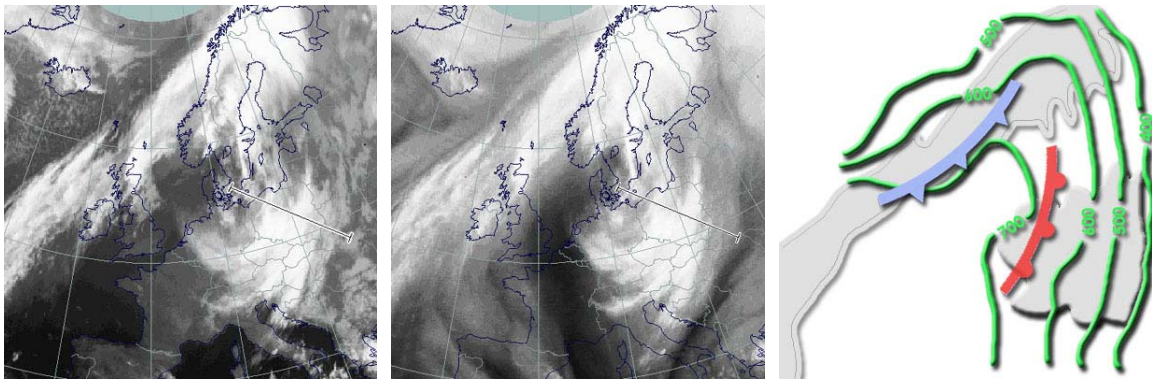
The areas of precipitation are embedded within the warm conveyor belt and move therefore faster than the surface front. As the rain bands move further ahead of the surface front, due to the still ascending warm conveyor belt, the precipitation dies out since there is less moisture. Directly at and behind the surface front it is often observed that the precipitation, in the form of rain, is changing into drizzle. For the warn front shield, the warm sector is overcast and precipitation occurs which might also be intensive; this is a consequence of the ascending warm conveyor belt within this region.

Sometimes, warm frontal cloud may appear completely detached from other areas of cloud and can no longer be explained by the frontal processes. Orography can also dominate the weather. For example, in Central Europe (the Alps), weather events are intense, but on the northern side of the Alpine region (the windward side) the cloud leads to heavy and enhanced precipitation. Cloudiness can lead to light to moderate rain or drizzle as well. Characteristics of a detached warm frontal cloud on satellite images are thick in VIS and cold in IR images. It appears as a wide broad anticyclonic curved area with high pixel values in WV images, indicating high moisture content. Figure A.8 shows a detached warm front seen in IR and WV images, together with its isobar and surface front.



(a) (b)

Figure A.7: (a) IR and (b) WV images of a warm front cloud band observed on the Atlantic (east of approximately 25°W) is extending to Scotland and Northern Ireland on 1 March 1995 at 0006 UTC. Source: Krennert et al. (1999).



(a) (b) (c)

Figure A.8: Images on (a) IR and (b) WV channel of detached warm frontal cloud on 22 March 1995 at 2330 UTC. Source: Krennert et al. (1999). Its relationship with isobar and surface front is in (c).

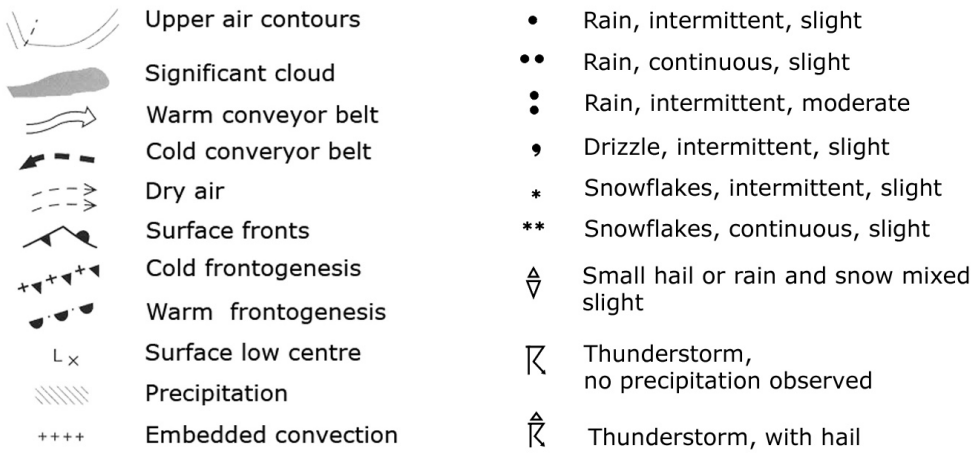


Figure A.9: Symbols used in the schematic diagrams and charts.

Appendix B

Data Preprocessing

B.1 Meteosat7 Radiance Calibration

The calibration is to ensure that any variation in time is due to changes of the signal coming from the observed target only, and not from variation of the observing instruments. It associates the raw measured units, expressed as digital counts received from the satellite's sensors, with an actual geophysical value of the object seen. First, the radiance is calculated from a digital count using an appropriate calibration coefficient.

- For IR images, sea surface temperatures from the National Centers for Environmental Prediction (NCEP) data sets are converted to radiances, taking the impact of the intervening atmosphere into account. These expected radiances are then correlated with observed counts of pixels that have been identified as sea surface scenes. The relation between the expected radiances and the observed counts are the instantaneous calibration coefficients.

- For WV images, the expected radiance at the top of the atmosphere is determined using temperature and humidity profiles from radiosonde observations. These radiances are then related to the mean WV count within a segment, if the radiosonde station located within the 550 great circle arc around the sub-satellite point and the segment is free of clouds above 700 hPa.

The calibration relation is given by

$$R = \alpha(DC - C_0), \tag{B.1}$$

where R is the radiance ($Wm^{-2}sr^{-1}$), α is the calibration coefficient, DC is an observed digital count and C_0 is a space count, which is the radiometric reading when

the satellite is viewing space.

Table B.1: Calibration parameters, applied to IR and WV images on 22-24 June 2004.

Date	Julian Day	Slot	**IR**		**WV**	
			α_{IR}	$C_{0,\mathrm{IR}}$	α_{WV}	$C_{0,\mathrm{WV}}$
22.06.04	174		0.107578		0.010427	
23.06.04	175	25	0.107950	5	0.010467	6
24.06.04	176		0.107803		0.010452	

Since 29th May 2000, the blackbody calibration is also operated on board. The black body mechanism consists of two black bodies located on opposite sides of the main optical path, one kept at ambient temperature of about 290 K as a cold reference and the other heated to about 50 degrees higher, as a warm reference. The temperatures of the two black bodies are telemetered to the Primary Ground Station (PGS) together with the corresponding black body imaging values. Thus, the observed counts can then be related to the known radiance. However, the main radiometer optics are not in the optical path when viewing the black bodies during calibration. Hence a correction model is needed to cope with these effects. In this work, the vicarious calibration is used because at present only their parameters are available from Eumetsat. Regardless of the calibration method used, the radiances are then converted into brightness temperatures using

$$T = \frac{B}{\ln(R) - A}, \tag{B.2}$$

where R is the radiance ($Wm^{-2}sr^{-1}$), T is a brightness temperature (K), A and B are regression coefficients of an exponential fit of the temperature-radiance conversion table.

Table B.2: Regression coefficients of the temperature-radiance conversion for Meteosat 7.

	IR2	WV1
A	6.9618	-1255.5465
B	9.2477	-2233.4882

For VIS channel, Eumetsat uses the Meteosat Visible and Infrared Imager (MVIRI) calibration algorithm (Govaerts and Clerbaux, 2004), which based upon radiative transfer simulation over bright desert and sea targets accounting for the spectral response of the calibrated channels. The radiation transfer includes total column water

vapour, total column ozone and aerosol optical thickness. The estimated calibration coefficient at a given date t is computed with the following formula:

$$\alpha_{\mathrm{VIS}}(t) = \alpha_{\mathrm{VIS}} + (D_f N_t \times 10^{-5}), \tag{B.3}$$

where α_{VIS} is the calibration coefficient at launch date, D_f is a daily drift, and N_t is the number of days since launch date. (α_{VIS} and D_f are 0.9147 $Wm^{-2}sr^{-1}$ per digital count and 5.7453 $Wm^{-2}sr^{-1}$per digital count per day$\times 10^5$ for Meteosat 7, which had been lauched since 2 September 1997.) The effective radiance R_{eff} can be retrieved with the following formula:

$$R_{eff} = \alpha_{\mathrm{VIS}}(t)(DC - DC_0), \tag{B.4}$$

where DC_0 is estimated as the mean offset of the two detectors of the VIS channel. However, DC_0 is available until 6 October 2003 only.

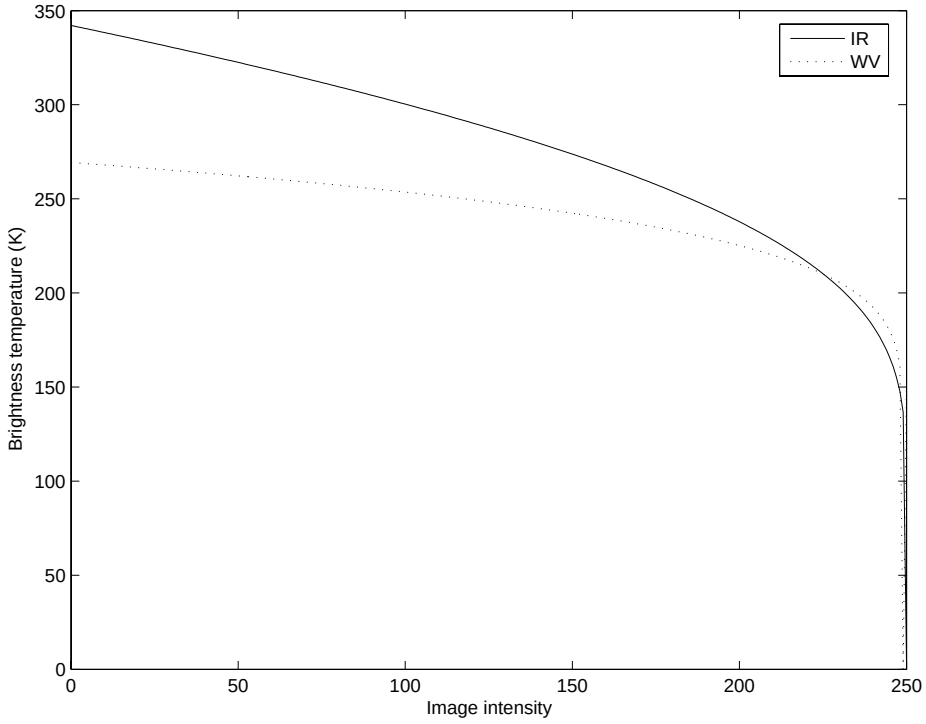


Figure B.1: The relationship between image intensity and brightness temperature on 22 June 2004.

The scaled radiance values are then normalised by the cosine of the solar zenith angle Z shown in Figure B.3 to account for different solar radiation effects at different

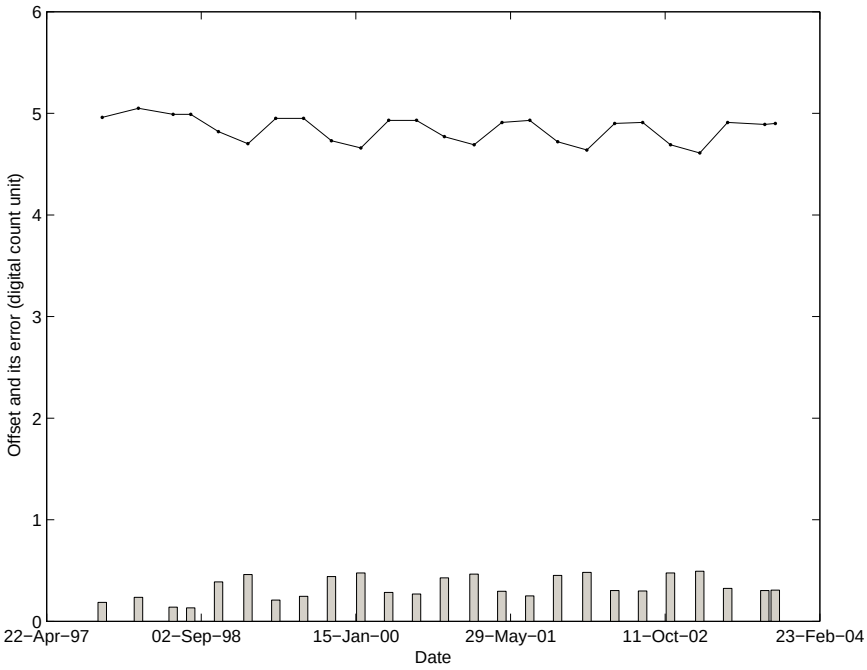


Figure B.2: The relationship between offset DC_0 and its error for Meteosat 7 VIS image calibration and date used.

time and different locations:

$$\tilde{R}_{eff} = \frac{R_{eff}}{\cos Z}. \tag{B.5}$$

The solar zenith angle is a function of date and time, and location:

$$\cos Z = \sin \Phi \sin \delta + \cos \Phi \cos \delta \cos H, \tag{B.6}$$

where Φ is the latitude of a pixel. H is the hour angle i.e. the angle that the earth has rotated since solar noon. At solar noon the hour angle = 0°, for example. Then, it measures time from solar noon in terms of fifteen degrees per hour. δ is the solar declination angle. It is an angle between the plane perpendicular to incoming solar radiational axis of the earth, which is tilted bewtween 23.5° on June 21/22 and -23.5°

on December 21/22. Changes in the solar declination angles contribute to seasons.

$$\begin{aligned} H &= \frac{15}{180}(\text{local time} - 12 - \text{daylight saving time}) \\ \delta &= 0.322003 - 22.971\cos(t) - 0.357898\cos(2t) \\ &\quad -0.14398\cos(3t) + 3.94638\sin(t) \\ &\quad +0.019334\sin(2t) + 0.05928\sin(3t), \end{aligned}$$

where t is the day angle and is calculated by $2\pi(\text{Julian date} - 1)/365$.

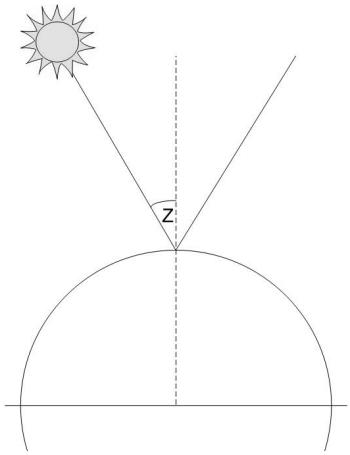


Figure B.3: Solar zenith angle.

B.2 Latitude and longitude of a pixel

The oblate spheroid shape of the Earth can be represented in spherical coordinates as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_{eq}\cos\theta\sin\phi \\ r_{eq}\sin\theta\sin\phi \\ r_p\cos\phi \end{bmatrix}, \tag{B.7}$$

where r_{eq} is the equatorial radius (6378 km), r_p is the polar radius (6356 km). An image of a geostationary satellite is taken by mapping these coordinates onto a plane as illustrated in Figure B.4. Therefore, the spherical coordinate of a point $\acute{A}$ on the image can be calculated from the intersection of $\overline{BE}$ piercing through the ellipsoid.

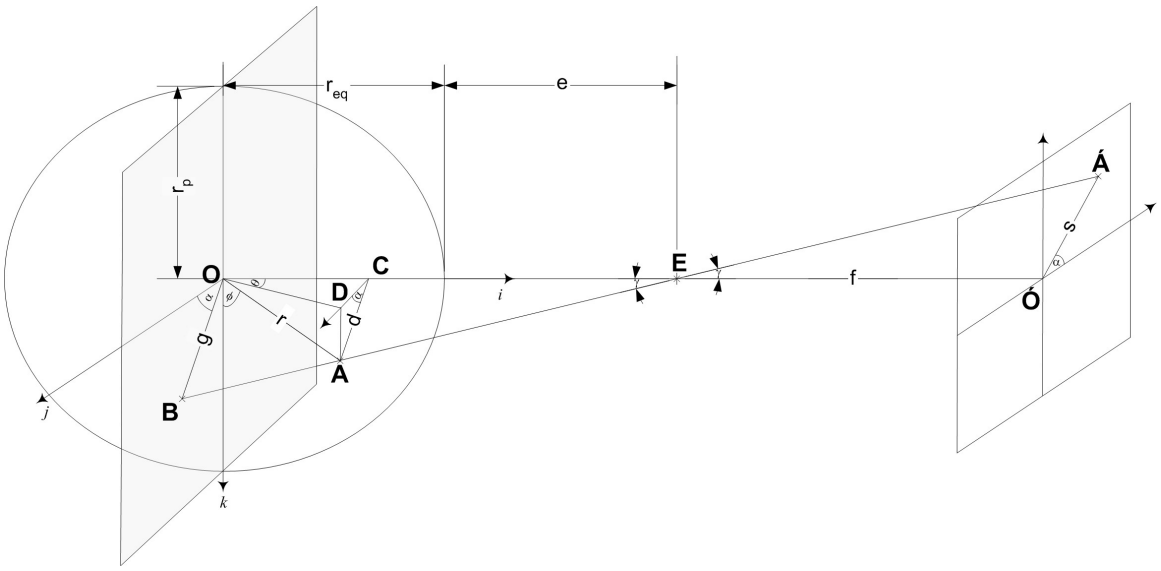


Figure B.4: Perspective projection of ellipsoid onto a plane.

That is,

$$\begin{bmatrix} r_{eq}\cos\theta\sin\phi \\ r_{eq}\sin\theta\sin\phi \\ r_p\cos\phi \end{bmatrix} = \begin{bmatrix} r_{eq}+e \\ 0 \\ 0 \end{bmatrix} + m \begin{bmatrix} -(r_{eq}+e) \\ g\cos\alpha \\ g\sin\alpha \end{bmatrix}, \tag{B.8}$$

where the left side of the equation is the parametric representation of $\overline{BE}$ in 3D-space, m is a parameter to control the line equation, α is the angle of the point from the x-axis of the image, g is a projected radius of the globe onto the $y-z$ plane, which is parallel to the image and $r_{eq}+e$ is the geostationary satellite's altitude from the origin, which is 42 164 km. The solution of the equation is

$$m = \frac{-c_2 \pm \sqrt{c_2^2 - 4c_1c_3}}{2c_1},\ 0 \le m \le r_{eq}+e \tag{B.9}$$

$$\begin{aligned} c_1 &= (r_{eq}+e)^2 + g^2 \\ c_2 &= -2(r_{eq}+e) \\ c_3 &= (r_{eq}+e)^2 - (r_{eq}^2\sin^2\phi + r_p^2\cos^2\phi). \end{aligned} \tag{B.10}$$

Consequently, the spherical coordinates (θ, ϕ) of the Earth are given by:

$$\begin{aligned} \phi &= \cos^{-1}\left(\frac{mg\sin\alpha}{r_p}\right) \\ \theta &= \sin^{-1}\left(\frac{mg\cos\alpha}{r_{eq}\sin\phi}\right), \end{aligned} \tag{B.11}$$

and g can be estimated from

$$\frac{g}{r_{eq}+e} = \frac{s}{f}. \tag{B.12}$$

where s is a distance of the pixel away from the origin of the image and f is the focal length of the projection and can be found from the known-location, known-projected point from the given image. However, the spinning Meteosat satellite acquires data on a line by line basis; each line of data is gathered with a tilt mirror fixed at a constant angle with respect to the spin axis of the satellite. Therefore, each line of an image corresponds to a cone swept out in space. This means that the image also has *depth* and is not flat as we assume. Thus, f becomes non-linear, as shown in Figure B.5.

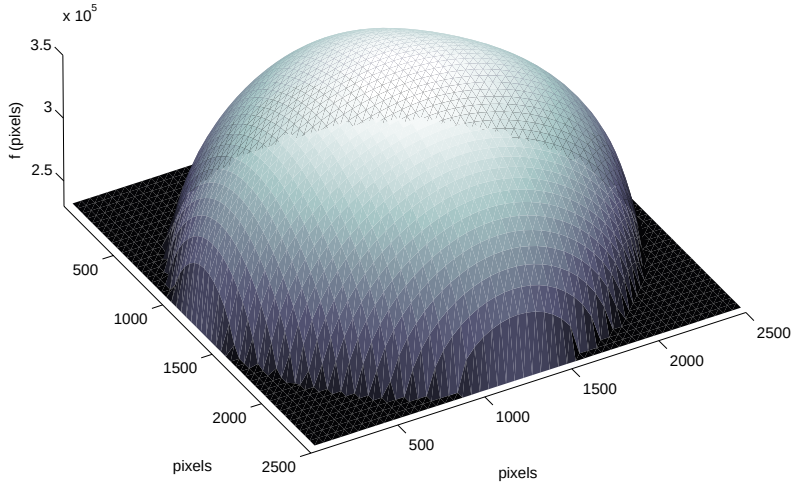


Figure B.5: Calculated f from known location.

However, the Meteosat image represents an $18° \times 18°$ view of the globe divided on an equi-angular basis. For example, an IR image of 2500×2500 pixels is assumed to to be $18/2500$ degrees per pixel. Therefore, another way to represent $\overline{BE}$ in Figure

B.4, as suggested in the Meteosat Archive User Handbook[1], is by using an angle of the image row, α_L, and that of the image column, α_S. Both angles can be defined as

$$\begin{aligned} \tan\alpha_L &= \frac{r}{\sqrt{p^2+q^2}} \\ \tan\alpha_S &= \frac{q}{p}, \end{aligned} \tag{B.13}$$

where $[pqr]^T$ is a nonzero vector lies on three points: a location on the earth's coordinate, the satellite and a point in an image. It is parallel to $\overline{BE}$ in Figure B.4. Repeating the previous process to find where the line $\overline{BE}$ piercing through the globe yields latitudes and longitudes of pixels of an image without using the non-linear f. The intersection between the line and the globe gives

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} R_{eq}+e \\ 0 \\ 0 \end{bmatrix} + m \begin{bmatrix} p \\ q \\ r \end{bmatrix}, \tag{B.14}$$

where

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix} = \begin{bmatrix} 1 \\ \tan(\alpha_S) \\ \frac{\tan(\alpha_L)}{\sec(\alpha_S)} \end{bmatrix}. \tag{B.15}$$

Note that the calculated latitude is the geocentric latitude, which is the angle between

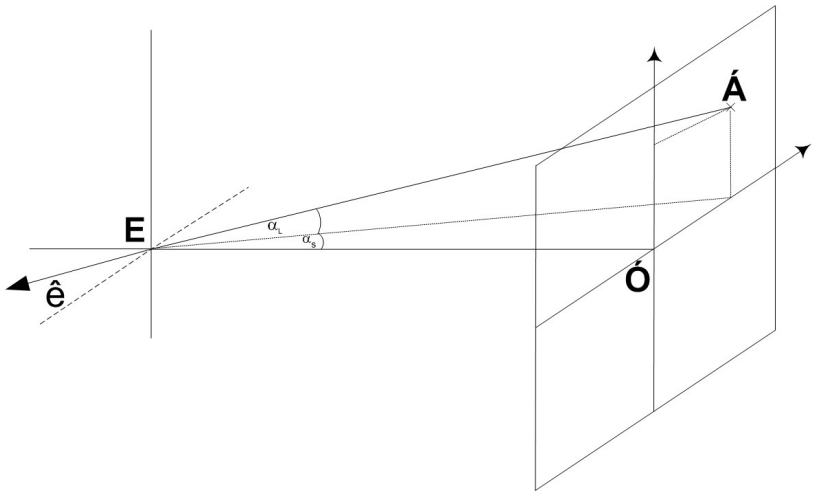


Figure B.6: Angles of pixels in an image.

[1]EUM TD 06 2.6 (November 2001), available at `http://www.eumetsat.de`

the equatorial plane and a line from the centre of the Earth. Then, it is converted to the geodetic (geographic) latitude, which is the angle at which the normal vector forms with the plane of the geodetic equator by

$$\phi_g = \arctan \frac{\tan(\phi)}{1-\varepsilon^2}, \tag{B.16}$$

where ε is the eccentricity of the globe defined as $\sqrt{1-(R_p/R_{eq})^2}$ and is 0.08227185.

Figure B.7 illustrates the resolution of pixels along the equatorial line, ranging from 5 km at the nadir to 30 km. The calculation approximates the *great circle* distance between any two of these pixels by $|\hat{r}|\cos^{-1}\delta$, where δ is an angle between two vectors e.g. $\hat{r}_1$ and $\hat{r}_2$, pointing to two geographical points and $|\hat{r}|$ defines the size of the vector. That is,

$$\cos\delta = \frac{\hat{r}_1 \cdot \hat{r}_2}{|\hat{r}_1||\hat{r}_2|}. \tag{B.17}$$

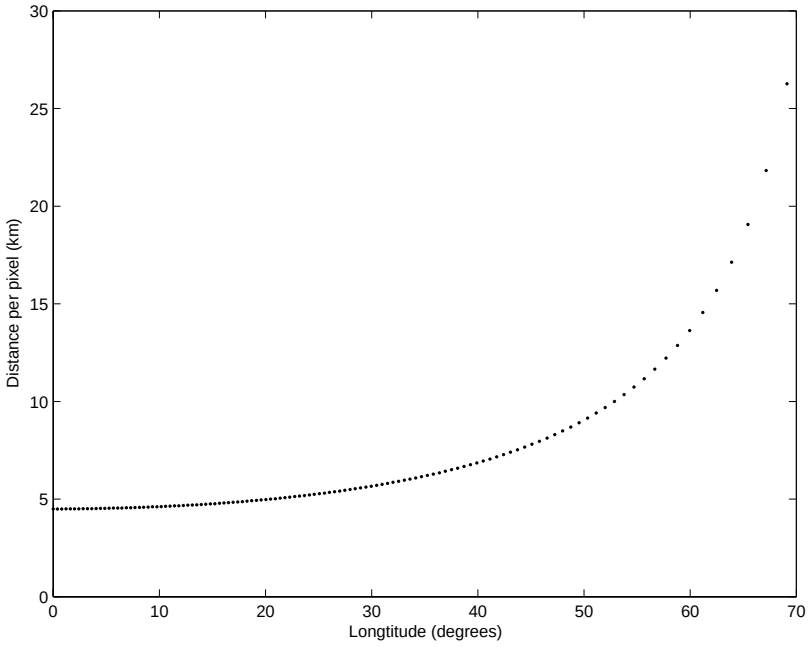


Figure B.7: Resolution per pixel (km) at different distance away from the centre of the image.

B.3 Orthographic to stereographic projection

Next, a Meteosat image is projected onto a plane in contact with the earth at the north pole, known as the polar stereographic projection, in order to collocate with a

rain radar image from the British Meteorological Office. Following Snyder (1987), the easting (E) and northing (N) position in a new image of a spherical coordinate (ϕ, θ) are

$$\begin{aligned} E &= E_f + \rho \sin(\theta) \\ N &= N_f - \rho \cos(\theta), \end{aligned} \tag{B.18}$$

where

$$\rho = \frac{2k_0 r_{eq} t}{\sqrt{(1+\varepsilon)^{(1+\varepsilon)} \cdot (1-\varepsilon)^{(1-\varepsilon)}}} \tag{B.19}$$

$$t = \frac{\tan(\frac{\pi}{4} - \frac{\phi}{2})}{[(1 - \varepsilon \sin\phi)/(1 + \varepsilon \sin\phi)]^{\varepsilon/2}}, \tag{B.20}$$

where k_0 is a scale on central meridian which is chosen to be 1 in the work and ε is the eccentricity of the globe.. The false-easting E_f and false-northing $N - f$ of the projection are set to zero. Next these coordinates are mapped to (x, y) coordinates of the rain radar images by

$$\begin{aligned} y &= -1.8695 \times 10^4 N - 428.7301 \\ x &= 1.8685 \times 10^4 E + 410.2852 \end{aligned} \tag{B.21}$$

The projected satellite images are finely matched with the Nimrod rain radar image with a displacement error not exceeding 1 pixel when considering coastlines and latitude-longitude grids.

B.4 Nimrod rain radar image to rain rates

The radar images from 15 C-band radars around the British Isles and Europe via the Radarnet reception system are received and corrected by the British Meteorological Offices Nimrod system, then distributed to users. The corrections include the identification and removal of spurious echoes resulting from anomalous propagation of the radar beam, errors resulting from variations in the vertical profile of reflectivity and radar sensitivity errors. The images are in Graphic Interchange Format (.gif), with its intensity corresponds to rain rates. Figure B.8 illustrates a function that is used to map the image intensities (0-255) to rain rates in the work, together with the colour of

each intensity that appears in the image. The intensity values of 206 and 207 are borders of countries and locations outside the radar coverage, respectively. The borders are replaced by averaged rain rates of their neighbourhoods. Figure B.9 illustrates rain radar and IR images before and after the pre-processing. The resolution of the reprojected image is presented in Figure 1.6.

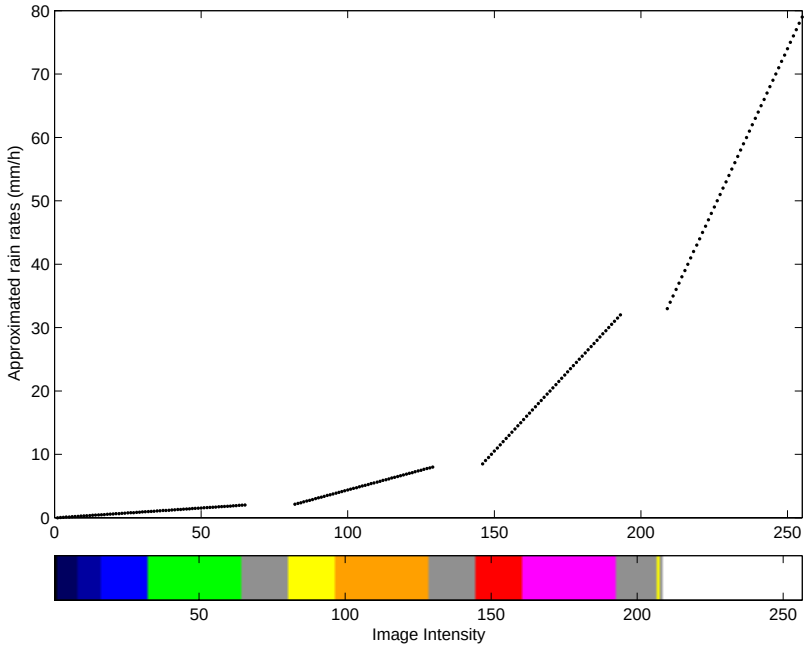


Figure B.8: Function for mapping radar image intensities to rain rates.

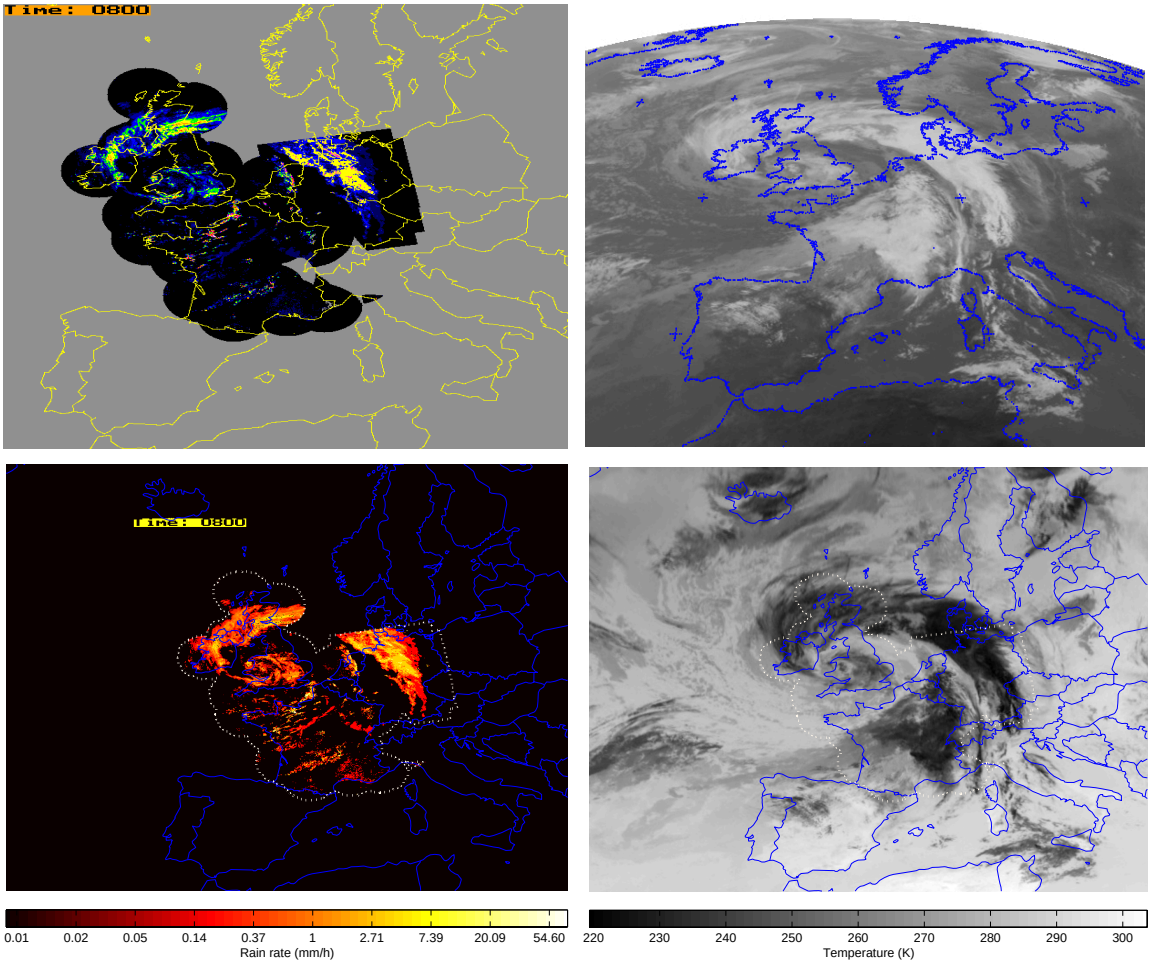


Figure B.9: A rain radar image and an infrared image on 23 June 2004 at 0800 UTC received from BADC are in the upper panel. The lower panel shows the images on the polar stereographic projection. The dotted line indicates the area that the rain information is available.

Appendix C

Miscellaneous

C.1 Histogram representation

Mallat (1989) suggested a method to derive the two parameters from the histograms of a variable u using these following equations:

$$h(u) = K \exp\left[-\left(\frac{|u|}{\alpha}\right)^{\beta}\right],$$

where

$$\begin{aligned} \alpha &= \frac{m_2 \Gamma\left(\frac{1}{\beta}\right)}{N\Gamma\left(\frac{3}{\beta}\right)}, \\ \beta &= F^{-1}\left(\frac{m_1^2}{m_2 N}\right), \end{aligned} \tag{C.1}$$

where

$$\begin{aligned} F(x) &= \frac{\Gamma\left(\frac{2}{x}\right)^2}{\Gamma\left(\frac{3}{x}\right)\Gamma\left(\frac{1}{x}\right)} \\ m_1 &= \int_{-\infty}^{\infty} |u| h(u) du \\ m_2 &= \int_{-\infty}^{\infty} u^2 h(u) du, \end{aligned}$$

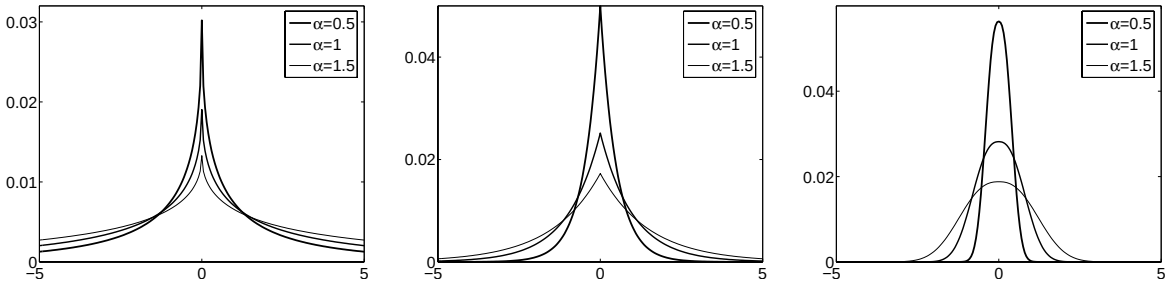
and the constant K is a normalisation parameter to keep

$$\int_{-\infty}^{\infty} h(u) du = N,$$

where N is the total number of pixels in the given detailed sub-image. The function $\Gamma(\cdot)$ is called a Gamma function and is defined by

$$\Gamma(x) = \int_0^\infty \exp(-u)u^{x-1}du. \tag{C.2}$$

Figure C.1 illustrates the histogram models of various α and β whereas Figure C.2 compares an effect of different values of β to the models . If $\beta = 2$, the model becomes a Gaussian function and as a consequence α is a standard deviation. The parameter m_1 and m_2 are the mean deviation and the normalised energy of a sub-image.



(a) $\beta = 0.5$ (b) $\beta = 1$ (c) $\beta = 2$

Figure C.1: Histogram model using Equation (3.13) with different values of α and β

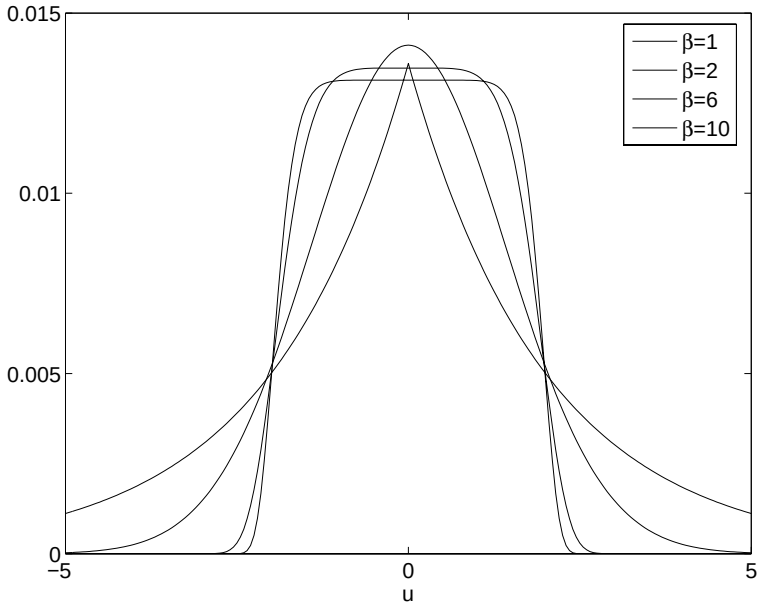


Figure C.2: Histogram model using Equation (3.13) with different values of β at $\alpha = 3$

C.2 Introduction to support vector machines

The simplest model of Support Vector Machine is the *maximal margin classifier*. It works for data which are linearly separable in the feature space. Given a set $S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)\}$ where $\mathbf{x} = (x_1, \ldots, x_n)^T$ is a sample in the vector space of n dimensions and y is its label, and a linear function $f(\mathbf{x})$

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{w} \cdot \mathbf{x} \rangle + b \\ &= \sum_{i=1}^{n} w_i x_i + b, \end{aligned}$$

where $\mathbf{w}$ and b are a weight vector and bias that control the function, the data are said to be linearly separable if there exists a hyperplane $(\mathbf{w}, b)$ that can correctly classifies the data, i.e. there is still a (positive) margin between the samples and a linear decision boundary that separates the samples into classes. The margin is usually referred to as a geometric margin. The hyperplane that realises the maximum value of the geometric margin is known as a maximal margin hyperplane. According to the Vapnik Chervonenkis theory, such a hyperplane also has a minimum probabilistic classification error bound. The SVM finds this optimal hyperplane by using a convex optimisation method as follows. First, let a functional margin of a sample with respect to the hyperplane $(\mathbf{w}, b)$ be defined as

$$\gamma_i = y_i(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b).$$

The geometric margin is a Euclidean distance of a sample point from a decision boundary and therefore equal to the functional margin of a normalised weight vector. If the functional margin is fixed to be equal to 1, that is

$$\begin{aligned} \left\langle \mathbf{w} \cdot \mathbf{x}^+ \right\rangle + b &= +1, \\ \left\langle \mathbf{w} \cdot \mathbf{x}^- \right\rangle + b &= -1, \end{aligned}$$

The geometric margin is then equal to

$$\begin{aligned} &= \frac{1}{2}\left(\left\langle \frac{\mathbf{w}}{||\mathbf{w}||} \cdot \mathbf{x}^+ \right\rangle - \left\langle \frac{\mathbf{w}}{||\mathbf{w}||} \cdot \mathbf{x}^- \right\rangle \right) \\ &= \frac{1}{||\mathbf{w}||} \frac{\langle \mathbf{w} \cdot \mathbf{x}^+ \rangle - \langle \mathbf{w} \cdot \mathbf{x}^- \rangle}{2} \\ &= \frac{1}{||\mathbf{w}||}. \end{aligned} \tag{C.3}$$

Hence, the hyperplane that solves the optimisation problem

$$\begin{aligned} \text{minimise} \quad & \langle \mathbf{w} \cdot \mathbf{w} \rangle & \text{(C.4)} \\ \text{subject to} \quad & y_i(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b) \geq 1, & \text{(C.5)} \\ & i = 1, \ldots, l, \end{aligned}$$

realised the maximal margin hyperplane with a geometric margin $= 1/||\mathbf{w}||_2$. The amount of loss associated with a particular choice of parameters $(\mathbf{w}, b)$ can be calculated using the primal Lagrangian:

$$L = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle - \sum_{i=1}^{l} \alpha_i [y_i(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b) - 1], \tag{C.6}$$

where $\alpha_i \geq 0$ are the Lagrange multipliers. By differentiating L with respect to $\mathbf{w}$ and b, the following relations *on* the decision boundary are obtained:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}_i = \mathbf{0} \\ \mathbf{w} &= \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}_i \end{aligned}$$

and

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l} y_i \alpha_i = 0$$

Substituting the relations into the primal, thus maximising

$$L = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle . \tag{C.7}$$

subject to $\alpha_i \geq 0$ and $\sum_i \alpha_i y_i = 0$ will give an optimal solution. The optimisation problem now becomes a linear combination of the training samples[1]. According to the Karush-Kuhn-Tucker (KKT) complementarity condition which states that the optimal solution $\mathbf{w}^*$, α_i^* and b^* must satisfy

$$\alpha_i^* [\langle \mathbf{w}^* \cdot \mathbf{x}_i \rangle + b] = 1, \ i = 1, \ldots, l, \tag{C.8}$$

[1]This is called the dual representation

and that all the other α_i are zero, only a number of the input $\mathbf{x}_i$ which the functional margin equal to one and lie closest to the hyperplane involve in the hyperplane expression:

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{w}^* \cdot \mathbf{x} \rangle + b^* \\ &= \sum_{i \in \text{sv}} y_i \alpha_i^* \langle \mathbf{x}_i \cdot \mathbf{x} \rangle + b^* \\ &= \sum_{i \in \text{sv}} y_i \alpha_i^* K(\mathbf{x}_i, \mathbf{x}) + b^*. \end{aligned} \tag{C.9}$$

These input vectors that correspond to the optimal Lagrange multipliers are commonly called support vectors. Other input vectors that are not support vectors have no influences and therefore do not affect the solution. The symbol $K(\mathbf{x}, \mathbf{z})$ denotes a kernel function that maps the input $\mathbf{x}$ and $\mathbf{z}$ to an inner product feature space. In Equation (C.9), K is a linear function of the inner product which might not be suitable for the data that are not linearly separable.

To extend the use of the support vector machine to the data that are not linearly separable, some errors are allowed in the optimisation problem Equation (C.4) to avoid over-fitting effect (Cristianini and Shawe-Taylor, 2000). This can be done by introducing slack variables $\xi_i \geq 0$ to the hard-margin constraint in Equation (C.5), i.e.

$$y_i(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b) \geq 1 - \xi_i, \tag{C.10}$$

such that

$$\text{minimise } \langle \mathbf{w} \cdot \mathbf{w} \rangle + C \sum_{i=1}^{l} \xi_i \tag{C.11}$$

where $C > 0$ is a regularisation constant yields an optimal solution. Its dual problem is

$$\begin{aligned} \max_{\alpha} \quad & \textstyle\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j), \\ \text{subject to} \quad & \textstyle\sum_i \alpha_i y_i = 0, \\ \text{and} \quad & 0 \leq \alpha_i \leq C, \end{aligned}$$

and the KKT conditions

$$\begin{aligned} \alpha_i = 0 &\Rightarrow y_i f(\mathbf{x}_i) \geq 1 \quad \text{and} \quad \xi_i = 0, \\ 0 \leq \alpha_i \leq C &\Rightarrow y_i f(\mathbf{x}_i) = 1 \quad \text{and} \quad \xi_i = 0, \\ \alpha_i = C &\Rightarrow y_i f(\mathbf{x}_i) \leq 1 \quad \text{and} \quad \xi_i \geq 0, \end{aligned} \tag{C.12}$$

reveal that outside the margin are the optimal α_i are zero and only α_i that corresponds to the edge of or inside the margin area are non-zero. This sparsity property allows large scale SVM practical. That is, the problem can be decomposed into a small number of optimisation variables that can be sequentially solved. Details of the decomposition can be found in Joachims (1999).

The support vector machine can be extended multiclass problems by using

- one against one method
- one against all method
- all at once method.

Given k as a number of classes, the one against one method separately solve the optimal decision of each class against others. Then, the class that give the maximum decision function $f(\mathbf{x})$ is assigned to the input vector. This method requires k classifiers to solve the problem. The second method solves the optimal boundary function between any two classes, then the class of the largest wins is selected. It requires $k(k-1)/2$ classifiers with less numbers of variables for each classifier. On the other hand, the all at once method solves only one optimisation problem:

$$\text{minimise} \sum_{m=1}^{k} \langle \mathbf{w}_m \cdot \mathbf{w}_m \rangle + C \sum_{m=1}^{k} \sum_{i=1}^{l} \xi_i^m. \tag{C.13}$$

Boser et al. (1992) suggested a way to create non-linear classifiers by applying a kernel to maximum-margin hyperplanes, i.e. every dot product is replaced by a non-linear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space. Thus, the non-linearity is induced in the kernel-transformed feature space. Maximum margin classifiers are well regularized, so the infinite dimension does not spoil but improve the results. Some common kernels include

- Polynomial (homogeneous): $K(\mathbf{x}, \mathbf{z}) = \langle \mathbf{x} \cdot \mathbf{z} \rangle^d$
- Polynomial (inhomogeneous): $K(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x} \cdot \mathbf{z} \rangle + 1)^d$
- Radial Basis Function (RBF): $K(\mathbf{x}, \mathbf{z}) = \exp(-\gamma ||\mathbf{x} - \mathbf{z}||^2)$, for $\gamma > 0$
- Gaussian RBF: $K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{||\mathbf{x}-\mathbf{z}||^2}{2\sigma^2}\right)$
- Sigmoid: $K(\mathbf{x}, \mathbf{z}) = \tanh(k \langle \mathbf{x} \cdot \mathbf{z} \rangle + c)$, $\exists k > 0$ and $c < 0$.

The support vector machine can also be used for a regression problem where the goal is to estimate an unknown continuous-valued function based on a finite number set of noisy samples. SVM regression performs linear regression in the high-dimension feature space. Using the previously-defined notation, the linear model is given by

$$f(\mathbf{x}) = \sum_{i=1}^{n} w_i g_i(x_i) + b, \tag{C.14}$$

where g_i denotes a set of nonlinear transformation. The optimal solution is gained by minimising ε-insensitive loss and the geometric margin:

$$\begin{aligned} \text{minimise} \quad & \langle \mathbf{w} \cdot \mathbf{w} \rangle + C \textstyle\sum_{i=1}^{l} (\xi_i^+ + \xi_i^-) && \text{(C.15)} \\ \text{subject to} \quad & y_i - f_i(\mathbf{x}) \leq \varepsilon + \xi_i^+, && \text{(C.16)} \\ & f_i(\mathbf{x}) - y_i \leq \varepsilon + \xi_i^-, && \text{(C.17)} \\ & && \text{(C.18)} \end{aligned}$$

where ξ_i^+ and ξ_i^- are slack variables for the optimisation problem. As in the classification, this problem can transformed into the dual problem with its solution is given by

$$\begin{aligned} f(\mathbf{x}) \quad &= \textstyle\sum_{i \in \text{sv}} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b, && \text{(C.19)} \\ & && \text{(C.20)} \end{aligned}$$

where $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function

$$K(\mathbf{x}_i, \mathbf{x}) = \sum_{j=1}^{l} g_j(\mathbf{x}) g_j(\mathbf{x}_i). \tag{C.21}$$

Parameter $C > 0$ determines the trade off between the model complexity (flatness) and the degree to which deviations larger than ϵ are tolerated in optimisation formulation. For example, if C is too large (infinity), then the objective is to minimize the empirical risk only, without regard to model complexity part in the optimisation formulation. Parameter ϵ controls the width of the ϵ-insensitive zone, used to fit the training data. The value of can affect the number of support vectors used to construct the regression function. The bigger ϵ , the fewer support vectors are selected. On the other hand, bigger ϵ results in more *flat* estimates. Hence, both C and ϵ affect model complexity and the SVM's generalisation performance. The empirical study in Cherkassky and

Ma (2004) showed that the following prescription for regularisation parameter:

$$C = \max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|) \tag{C.22}$$

$$\epsilon = 3\sigma\sqrt{\frac{\ln l}{l}}, \tag{C.23}$$

where $\bar{y}$ is the mean of the training responses (outputs), σ and σ_y is the standard deviation of the input and output, can effectively handle outliers in the training data.

Appendix D

Publications

1. A. Suvichakorn and A. Tatnall. Flow analysis of cloud images from geostationary satellites. One Day BMVA symposium at the Royal Statistical Society, 2004.

2. A. Suvichakorn and A. Tatnall. The application of cloud texture and motion derived from geostationary satellite images in rain estimation—a study on mid-latitude depressions. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, volume 3, pages 1682–1685, Seoul, 2005.