UNESCO-IHE
INSTITUTE FOR WATER EDUCATION

May 2007 Flooding England

Using a hybrid approach to improve rainfall prediction for water management

CARLOS ARTURO MARTINEZ CANO
MSc Thesis WSE-HI 09-14
September 2009

UNESCO-IHE
Institute for Water Education

Using a hybrid approach to improve rainfall prediction for water management

Master of Science Thesis

By

CARLOS A. MARTINEZ C.

Supervisors:

Dr. Ir. Yunqing Xuan, MSc. (UNESCO-IHE)
Dr. Ir. Gerald Corzo, MSc. (UNESCO-IHE)

Examination committee

Prof. Dr. D. P. Solomatine (UNESCO-IHE), Chairman
Dr. Ir. Shreedhar Maskey, MSc. (UNESCO-IHE)
Dr. Ir. Yunqing Xuan, MSc. (UNESCO-IHE)

This research is done for the partial fulfilment of requirements for the Master of Science degree at the UNESCO-IHE Institute for Water Education, Delft, the Netherlands

Delft, The Netherlands
September 2009

“Oh Mary, conceived without sin,
Pray for us who turn to you”.
Amen

Abstract

In the last decades the increased availability of hydrological spatial models and operational online systems have boosted the necessity for more accurate short term spatial rainfall forecast in water resources management. In short term forecast (hours or minutes), radar observations and numerical weather models are the most common sources of spatial rainfall information. Rainfall prediction based on weather radars presents better performance than numerical weather prediction (NWP) models over the first hours. This can be explained by the fact that radar does not use important physical information, like wind, temperature, pressure and others, to calculate the rainfall information. Moreover, the forecast information obtained from radar is commonly based on tracking spatial patterns of rain. Therefore, on extended forecast horizon the high variability of physical variables (e.g. wind, humidity and others) are contemplated by the NWP and not by the tracking of radar rainfall information. On the other hand, NWP models do not generally capture well the initial rainfall distribution. Nevertheless, over longer time scales, NWP models would perform better than weather radars as they resolve dynamically the large scale flow.

In this study the integration of radar tracking forecast and NWP models results have been explored. The NWP model used is the meso-scale model five (MM5), and was set-up for a region in Southeast England. Radar tracking forecast based on spatial correlation of the radar patterns over consecutive time steps was applied. The Nimrod radar data from the British Atmospheric Database Centre (BADC) is used as measured rainfall. Data driven modelling techniques such as model trees (M5), linear regression model (LR), artificial neural networks (ANN) and modular neural networks (MNN) have been used for the integration of both forecasting models. These techniques have been applied to three cases using different precipitation events.

- Case 1 using data sampling in the process of building the DDM.
- Case 2 using consecutive rainfall events selected on expert knowledge
- Case 3 using a rule based integration of models

The results showed an overall improvement in almost all data-driven modelling techniques applied. The techniques were compared with those obtained from weather radar and MM5 model. It is important to highlight that although an improvement was found in the overall region, radar tracking and NWP models still have low performance at highly retailed spatial resolution. The use of improved models or inclusion of weather atmospheric variables might bring an important alternative that needs to be tested. The use of modular models is probably the best alternative to integrate the NWP and radar tracking; however, we found limitations in its application due to the low number of rainfall events available in the short data set used.

Keywords:

Nowcasting, weather radar, radar tracking, Numerical Weather Prediction, MM5 model, data driven modelling.

Table of Contents

1 INTRODUCTION

1.1 Background

It is predicted that severe weather events such as snow, rain and waves will be even more extreme over the next few decades because of climate change. Looking at extreme rainfall events their intensity also may change in the future.

Flooding is a natural event that occurs when there is a heavy rainfall that fills rivers and streams above their normal capacity. The water excess that gathers cannot be controlled by normal boundaries and follows path of least resistance. Floods can also occur when rainwater collects on the ground and cannot find a source to drain into; a typical example of this is the surface water runoff.

Many decision making processes in water management require a clear understanding of the distribution of water volume and/or quality both spatially and temporarily. Over the years, increasingly sophisticated modelling systems have been developed and deployed at different levels in order to meet this demand in terms of employing hydraulics and hydrological models and their integration. These models, despite how complicated their structures may achieve, actually work as a sort of distributor able to re-distribute the inputs which in many cases are precipitations. Therefore, it is common practice to invest efforts in order to obtain precipitation from various observations and one of the important sources is ground- based rain gauges.

For some cases, decision making can be conducted sufficiently well with the rainfall observations and utilities provided by modelling tools. However, in many others cases additional rainfall information is needed to make a sensible decision. The following three cases indeed show the relevance of this extra information in decision making process:

1. Flash flood forecasting and/or warnings of a small (mountainous) region where the concentration time naturally is so short and gauge based observation cannot provide either proper or sufficient information to support decision making.

2. The water management over large river catchment needs to have effective coordination along the river. This coordinating process often requires rainfall information in a short (a couple of days) or medium (up to one week) period of time ahead. For instance, the regulation of reservoir(s) normally benefits from the accurate short-term rainfall forecast so that appropriate operation can be achieved with the help from good inflow prediction based on the precipitation forecast.

3. Damage and losses caused by increasingly frequent urban flooding have come into the vision of both city planners as well as modelers. In urban flooding detail information of the rainfall distribution over the area is crucial. This information, traditionally provided through ground-based observations like weather stations, nowadays is gradually replaced (or at least assisted) by remote-sensing technique like

weather radars. As such, attention also has been drawn to a have better rainfall estimation and/or prediction.

The term of *precipitation nowcasting* is used to refer the very short term nature of the precipitation forecast (i.e. 0-6 hoursWilson, 2003). Its capability has become an important specialized topic that affects daily forecasting and decision making processes in weather service and water management. Nowcasting relies on the use of up-to-date observations of weather and simple forecasting methodologies to produce precipitation forecasts with high spatial and temporal resolutions. It has been almost exclusively based on extrapolation of radar echoes, satellite imagery of clouds and/or lighting location data. Extrapolated based approaches to the short range forecasting (nowcasting) of precipitation using radar have proved quite successful as the motion of precipitation in the past is a good predictor to its motion in the near future (Pierce et al., 2005).

Precipitation nowcasting can be obtained by using information from weather radars and numerical weather prediction models (NWP). Weather radars are based on electromagnetic wave response interpretation: it has limited area and limited lead-times, forecast depends on analysis of 2D spatial pattern dynamics. NWP produces forecast for different lead times and is based upon the solutions to a set of governing equations that describe the atmospheric dynamics and physics. NWP also depends on data assimilation where a complementary model is used to identify errors and then correct them.

Numerical weather prediction (NWP) models are typically run over large spatial scales with coarse spatial and temporal resolutions. Compared with extrapolation forecasting using weather radars, NWP uses current weather conditions as input into mathematical models of the atmosphere to predict the weather. Numerical weather models have a disadvantage of the variable representation of rainfall in relation to the real precipitation field. Quantitative precipitation forecast (QPF) is the expected amount of rainfall accumulated over a specified time period over specified area. NWP tends to smooth the sub grid rainfall variability that is important to runoff generation within the grid, therefore a scaling error will be introduced to hydrological model as long as the QPF has not reached the real scale of the rainfall-runoff process. NWP models are still very difficult to be initialised for the very short range forecast whereas the weather radar information can predict well instantaneous rainfall. The combinations of the two methods arise in the hope of utilising the advantages from both in order to get better rainfall prediction (Xuan, 2007).

The precipitation information provided by weather radar can be used to predict instantaneous rainfall distributed but its accuracy remains as a major issue due to substantial amount of errors inherited in the observations as well as the methods employed for prediction. There are some nowcasting techniques such as high-resolution numerical modelling and advection forecasting that show problems with accuracy to describe the local weather condition and their expected evolution during the first hours.

Indeed, the precipitation prediction has been one of the most difficult parts for atmospheric models due largely to the fact the underlying physics and dynamics cannot be expressed explicitly as done with many other models. As to the pattern-extrapolation based radar nowcasting, it is even more so as the dynamics normally has nowhere to fit in using traditional approach. Nevertheless, radar-based short-term precipitation forecasting (less than 1 hour) and NWP-based forecast (few hours to days or weeks) are currently main sources of prediction available to many "downstream" services, e.g. water management and flood risk management.

It has been found that whilst the radar nowcasting performs reasonably better than NWP for very short lead-time forecast, its skill drops rapidly with the time horizon going beyond hours. On the other hand, the NWP forecasts need to have several hours for "spinning up" during which time their skill remains low and decreases quickly as well after certain time. A schematic diagram can reveal such a time-wise trend of the performance of both methods. (Figure 1-1)

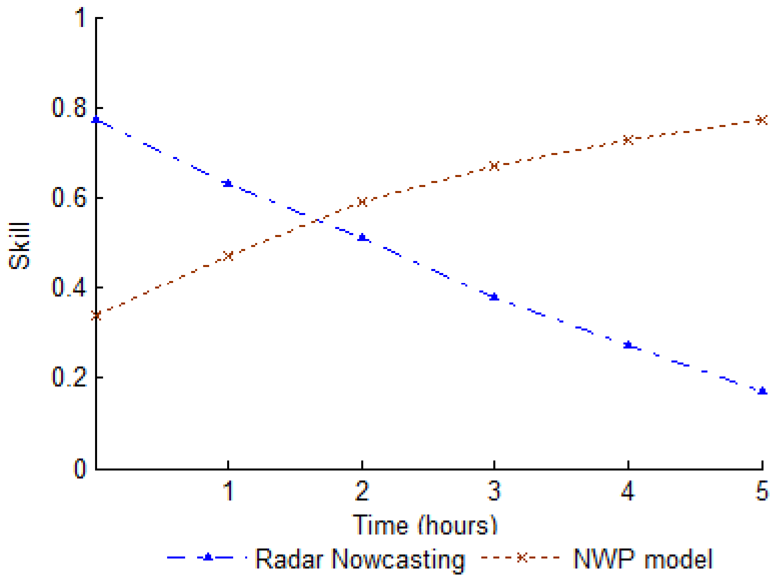


Figure 1-1 *Radar nowcasting and NWP model performance*

It is a very much natural thinking to utilise information from both methods to improve the skill of precipitation forecasting. The past decade has seen such an effort in merging NWP with radar nowcasting using various techniques, e.g., the NIMROD system (Golding, 1998) developed by the UK Met Office in 1990s using weighted averaging of both sources of information, and most recently the STEPS system that employs a scale-dependent merging technique able to take into account errors and uncertainties propagating at various scales (Bowler et al., 2006).

In view of the inherent difficulties in presenting precipitation process due to it high nonlinearity, intermittency, the data-driven modelling techniques may be a better choice in terms of merging the two kinds of forecasting techniques, without bothering too much the underlying, physics, since the complex relationship can be revealed by the data itself through proper modelling procedure. In fact, data-driven modelling techniques have been widely applied to a variety of applications in water sectors such as modelling, data classification, reservoir optimisation, building flood severity maps based on aerial or satellite photos and short-term forecasting (Solomatine et al., 2008). Also for the treatment of precipitation uncertainty in rainfall-runoff-routing models in the framework of fuzzy set theory assisted by genetic algorithms (Maskey et al., 2004). Those models do not represent the physics and knowledge about the details of the processes but data-driven models are built to obtain the relationship between inputs and outputs variables.

It is probable to get a better performance and accuracy levels combining NWP information and radar information to produce precipitation events by a hybrid approach. Hybrid refers to the combination of two or more different things (in this case precipitation data from different sources) aimed at achieving a particular objective or goal (rainfall prediction). This integration process using data driven modelling techniques has been very useful in modelling (Ganguly, 2002; See and Stan, 2000) obtaining results that show the importance to make efforts to achieve improved results.

1.2 Objectives and Scope

The general objective of this research is to explore the use of data-driven models (DDM) as integrators of radar rainfall information and numerical weather prediction model (MM5) for rainfall prediction.

Specific Objectives

- To evaluate the performance of different DDM as model integrators for rainfall prediction using data sampling.

- To evaluate the integration of radar rainfall information and MM5 model outputs using DDM without sampling.

- Explore the use of rule based forecast on the integration of radar rainfall information and MM5 model outputs.

Scope

The scope of this study is to obtain understanding from a hybrid model that integrates weather radar (data oriented) and a numerical weather prediction NWP (based on physics). For this purpose, four data driven models (DDM) have been built: model tress (M5), linear regression model (LR), artificial neural networks (ANN) and modular neural networks (MNN).

In this research five rainfall events were selected 27 May 2007, 28 May 2007, 24 June 2007, 05 September 2008 and 08 November 2008. With those data 1, 2, 3, and 4 hours forecasts were produced applying the radar tracking method and two different MM5 model configurations: continuous MM5 simulation (24 hours) and MM5 online update every 6 hours.

The DDM were compared with radar tracking forecast and MM5 model forecast through their performance measures. It was considered the root mean squared error (RMSE), the correlation coefficient (R) and the persistence index (PERS). Probabilistic measures also were applied such as probability of detection (POD), false alarm ratio (FAR) and Critical success Index (CSI).

1.3 Outline of the thesis

Chapter 1 gives the introduction to the problem, the description of the work with research criteria and objective definition.

Chapter 2 describes briefly some aspects related to rainfall estimation. A general introduction to weather radars and numerical weather prediction are presented followed by the description of mesoscale weather models and MM5 modelling systems as well as the practical indications on the MM5 model set up are described respectively.

Chapter 3 Describes the main data driven models used in this thesis such as model trees (M5), linear regression (LR), artificial neural networks (ANN) and modular neural networks (MNN).

Chapter 4 presents the methodology that has been followed in this research. It starts with a flowchart summary followed by a data processing including a radar data validation and transformation (NIMROD). The chapter continues with a description of the mesoscale weather forecast model setup (MM5) and its respectively configuration. Afterwards, a hybrid model is built by using data driven modelling techniques combining a forecast based on correlation “tracking” and output precipitation data from the mesoscale weather model (MM5). Finally, three case studies are presented with five different rainfall events occurred during 2007 and 2008 in UK obtaining rainfall forecast 1, 2, 3 and 4 hours ahead.

Chapter 5 discusses de results of this research. The effectiveness of the decision trees, linear regression, artificial neural networks (ANN) and modular neural networks (MNN) are compared.

Finally, the conclusions based upon the study results and the recommendations are given in chapter 6.

2 RAINFALL FORECASTING

2.1 Introduction

The study of rainfall is a major factor especially for flash flood disaster management. The increase in damage caused by flood is due to the area vulnerable to higher risk of flood disaster. The use of conventional rain gauge measurements sometimes is limited because their distribution is sparse and not available in mountainous and remote areas. The concept of rainfall estimation, using weather radars or numerical weather models, provides guidance to hydrologist issuing flash flood watch, forecast and warning. As it was stated in chapter 1, this thesis focuses on the integration of two sources of information for rainfall prediction, before starting to describe de models to combine these rainfall data, this chapter briefly presents a general description and basic knowledge of weather radars and numerical weather prediction models.

2.1.1 Background

RADAR (Radio Detection And Ranging) is a system that uses electromagnetic waves to identify the range, altitude and direction of both moving and fixed objects. The radar used in meteorology (see Figure 2-1) works to locate precipitation and calculate its type and motion also to forecast its future position and intensity. After the II world war the American scientist David Atlas developed the first operational weather. In Canada J.S Marshall and Walter Palmer worked on the drop size distribution in mid-latitude rain that led to understanding of the Z-R relation. This relation correlates a given radar reflectivity with the rate at which water is falling on the ground.

Figure 2-1 *A weather radar for rainfall detection.*
(http://en.wikipedia.org/wiki/Weather_radar)

The radar rainfall information for this research were obtained from the UK met office Nimrod system (Golding, 1998).

Nimrod is a general purpose nowcasting system for the analysis and very short range prediction of hazardous weather variables, including precipitation amount, type, run-off and river flow; visibility, fog and low cloud; surface wind and lightning. Here we are mainly concerned with precipitation amount. For each variable or group of variables three main tasks are developed: the processing of remotely sensed imagery, data fusion and extrapolation forecast merging with Numerical Weather Prediction (NWP).

The radar processing includes image quality control, spurious echo removal, vertical profile correction, gain adjustment and composing. Meteosat satellite imagery is then calibrated against the radar composite and combined with lightning data, a short forecast and in situ observations to extend the precipitation distribution outside radar range. The match is optimised by defining range according to the ability of each radar to observe any rainfall event.

To extrapolate the rainfall distribution, the Nimrod technique selects the optimum vector for each precipitation object, from either cross-correlation or the NWP model winds. The current state is extrapolated forward using the computed motion vectors and then combined with latest output from the NWP model using weighted average. The Nimrod system covers the British Isles and surrounding areas on a 5km grid using radar data from France, Belgium, Ireland and the UK and other data sources over the more distant sea areas. This study uses 5-km grid resolution data using the radar data of the stations Chenies near Amersham, and Wardon Hill in Dorset in the southeast of England.

Routine quality monitoring is carried out within the UK. Statistics are computed from comparisons with rain gauges within range of each UK radar. Several developments are currently in progress, including use of Meteosat second generation data.

The Nimrod system has been used for flood forecast applications and to assess water resources. It provides radar measurements of rainfall enhanced the availability of data for input to hydrological models. Alternatively, distributed models have been used successfully with the use of radar data that originally were developed with rain-gauge network data.(Zhijia et al., 2004).

There are three basic types of numerical models to forecast rainfall they are stochastic, dynamical and statistical-dynamical models. In this sense, the radar data have been used to assess the validity of those approaches, to verify dynamical models and also to assimilate into them.

2.1.2 Rainfall Nowcasting Using Weather Radars

The weather radar transmits microwaves energy bounces off of objects and returns to the radar to be measured. In other words the radar sends directional pulses of microwave radiation, on the order of a microsecond long, using a cavity magnetron connected by a waveguide to a parabolic antenna. Depending on the radar's ability to scan up, down and in a circle, modern radars can measure the three-dimensional distribution of precipitation within 100 + miles of the radar. Weather radars systems scan at low elevation angles to obtain estimations of precipitation closer to the ground for meteorological and hydrological purposes (Rico-Ramirez et al., 2008). The wavelengths of 1 to 10 cm are approximately ten times the diameter of the droplets or ice particles of interest.

Figure 2-2 shows how weather radar works. Basically, it sends out a microwave pulse, 4-10cm, and listens for a return echo (A). This microwave radiation is scattered by precipitation particles (B). Both frozen and non-frozen, energy is sent in all directions (C). The radar has a listening period where it detects the radiation scattered in its direction, this radiation is called the echo (D).

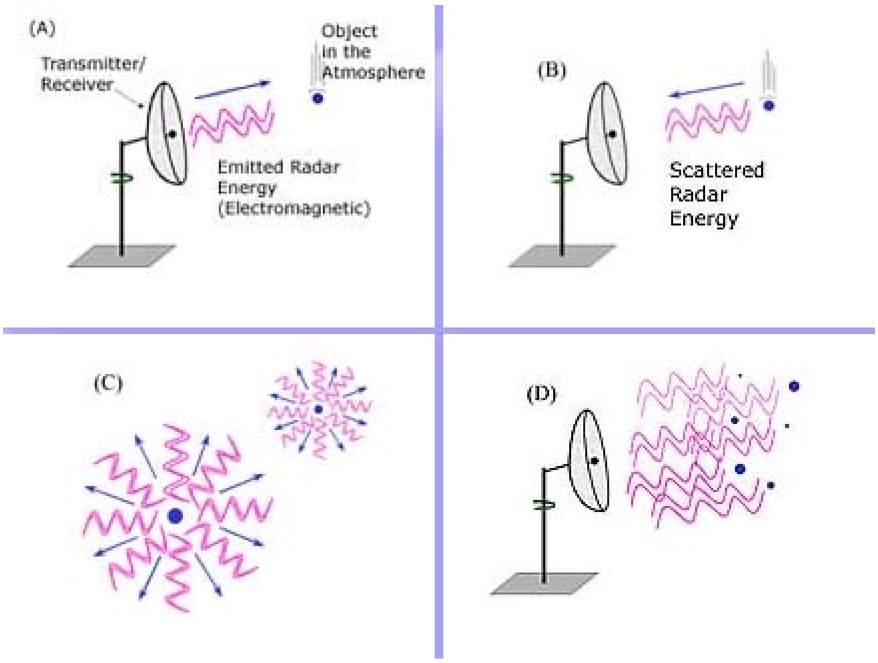


Figure 2-2 *The mechanism of weather radar.*
(http://www.aos.wisc.edu/~aalopez/aos101/wk6.html)

In hydrology, weather radars provide rainfall estimations over a region that can be processed to produce short-term quantitative precipitation forecast used for real-time flood forecasting. In meteorology, weather radars are used in conjunction with atmospheric observations to initialize weather models and also for cloud research (Rico-Ramirez et al., 2007a). Radars send a high-power pulse of electromagnetic energy that can be reflected back to the radar when precipitation particles lie along the path of the beam. This power reflected back to the radar gives a measure of the reflectivity (Z), which can be transformed to an estimate of precipitation (R) by using the empirical equation given by Marshall and Palmer (1955). A Z-R relationship relates radar reflectivity to rainfall rates based on a specific drop-size distribution.

$$Z = 200R^{1.6} \text{ (R in mm/hr)} \quad (2.1)$$

2.1.2.1 Related research

Several attempts to know the uncertainty in the rainfall estimation have been carried out (Lee, 2003; Rico-Ramirez and Cluckie, 2006; Rico-Ramirez et al., 2007b) evaluating the performance of different polarimetric rainfall estimation algorithms at C-band frequencies. C band scanning is a name given to certain portions of the electromagnetic spectrum, as well as a range of wavelengths of light used for communications. A dense network of rain gauges has been used to validate the polarimetric rainfall estimates and the overall results also have been compared with the conventional Marshall and Palmer (MP) relationship (Marshall et al., 1955).

Weather radars have practical limitations and the estimation of precipitation using these devices has several sources of errors. One important error is the variation of the vertical reflectivity profile of precipitation (VRP). Various efforts have been undertaken to reduce the effect of the VRP and to improve the estimation of precipitation using scanning weather radar measurements. The studies include correcting for the VRP measured from scanning radars (Joss and Lee, 1993) or from High-resolution X-band vertically pointing radars. Polarimetric radar measurements potentially provide important advantages over the conventional reflectivity factor. It offers the possibility to classify hydrometeors, which provides the possibility of applying different rainfall estimators and attenuation corrections within the rain region (Cluckie and Rico-Ramirez, 2004).

With respect to the radar calibration it is provided by the radar manufacturers such as the estimation of system parameters, matched filter losses, correct distance calibration and the careful adjustment of the radar antenna. Corrections of the velocity measurements based on the difference of two subsequent phase measurements and remote calibration of radar targets by removing clutter contamination are also provided. Subsystems such as antenna gain and beam, radar transmitter and the receiver are calibrated separately.

Calibration with punctual ground data may be in error because of different nature of the radar measurement (Bras, 1989). Some of the discrepancies in measurements are due to low altitudes may go undetected by radar. Detected precipitation at a given height may never reach the ground owing to evaporation or winds carrying it to a different location. Vertical variability of precipitation type and intensity may lead to incorrect integration of echoes and incorrect estimation precipitation. Ultimately, surface evaporation may lead to false radar echoes.

Radar errors have been divided into large and small scale errors. Inaccuracies in measuring large area and time scale events such as storms totals over a large catchment area are called large scale errors. This type of error is presented by the ratio of the radar measurement of storm total and the true storm total. The small scale errors are due to local variations in the relation between rainfall intensity and radar reflectivity. This error is measured in terms of the ratio of the radar measurement and rain-gage measurement (Bras, 1989).

The use of weather radars have been implemented as a component in water resources strategies in countries such as England, United States and Finland (Kuusisto, 2006) among others so that activities of water resources authorities are efficient, impartial and reliable. Forecasts are used for regulation, flood damage prevention and general information.

2.1.3 Precipitation forecast based on echo extrapolation

As it was stated before, extrapolation techniques are used to forecast rainfall radar data, the use of Lagrangian analysis techniques done by Collier and Hardaker (1996) obtained realistic estimates of probable maximum precipitation, Bayesian technique and kalman filtering (Todini, 2001) in which increase the credibility of the radar based rainfall estimates have demonstrated also the improvements in rainfall spatial distribution obtainable by means of weather radars.

The use of feature identification has been used in radar data and satellite images. Some of the simpler techniques which have been applied are level slicing and contouring, Fourier analysis, bivariate normal distribution and clustering. Fourier analysis is used within pattern recognition procedures which require knowledge of the shape of the feature for the purpose on assessing pattern motion. Bivariate normal distribution is useful in dealing with large intense thunderstorm but not practical for frontal rainfall. The difficult of using clustering arise in defining an initial grouping in which to begin allocating areas of radar echoes, and in determining the true number of clusters.

The extrapolation procedures, tracks individual features, or clusters of features, from one image to the next involve some kind of matching or correlation procedure. Once the movement is defined then it may be extrapolated to derive the forecast. Procedures which have been employed within operational systems include the use of wind velocities derived from independent observational or numerical systems, the use of centroid positions and the use of cross correlation. This last procedure has the advantage of taking into account the detailed shape of the feature being tracked, and decreases the changes of mismatching features.

It is difficult to assess objectively the relative performance of the several different methods of extrapolating feature movement as most of the techniques claim to be successful in a limited sense and have been used in different weather situations (Collier, 1994). The cross correlation technique for extrapolating features within images has been widely used in operational systems although it is likely that the reliability of objective extrapolation will continue to improve.

2.2 Numerical weather prediction

2.2.1 Background

Nowadays, weather forecasts are mainly conduced using computer models also known as the Numerical Weather Prediction (NWP). The models relies on sophisticated numerical solutions to a set of governing equations that describe atmospheric dynamics, with initial and boundary conditions settled using data obtained form global observations network. Mathematically, NWP models follow similar approaches applied to many other computer models, especially those in computational fluid dynamics (CFD). The complexity, however are normally several orders higher than those applied at local scale. Certainly, precipitation forecast is one of the most important products of NWP, with coverage ranging from counties to the entire globe and time starting from hours to days and even weeks. Post-processing and visualisation are also important in order to facilitate various downstream applications. An example of NWP precipitation forecast is given on Figure 2.3.

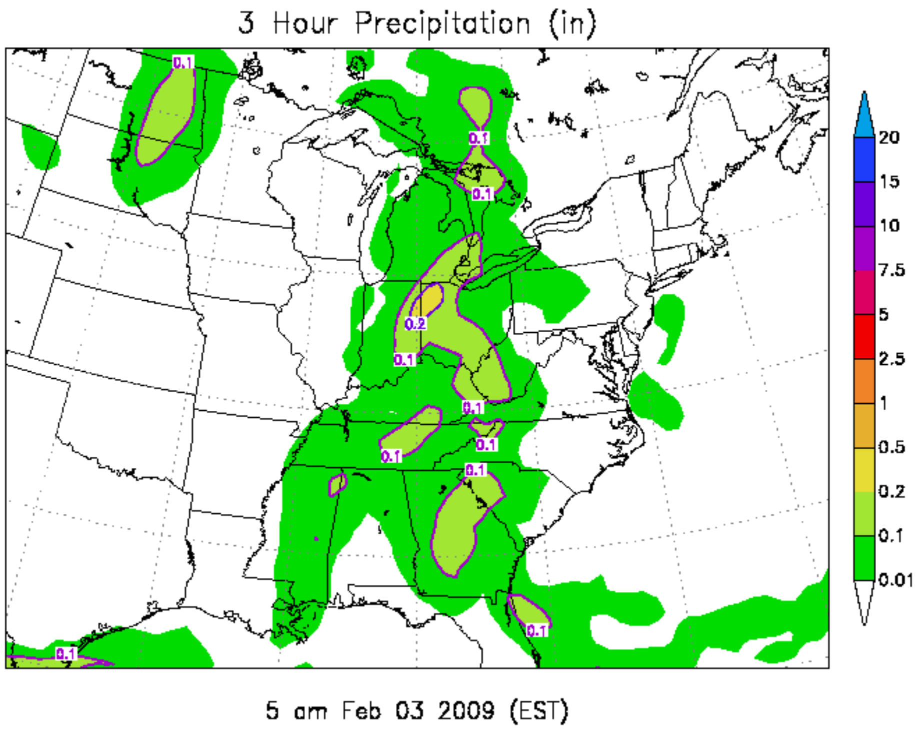


Figure 2-3 *Numerical weather prediction in real time*
(Source: Department of Atmospheric and Oceanic Science, University of Maryland, **http://www.atmos.umd.edu/)**

NWP started in 1920's but until the invention of the computer it was feasible to do in real-time. Most recently the use of a model ensemble forecast helps to define the forecast uncertainty and extend weather forecasting farther into the future. A number of forecast models, both global and regional in scale, are run to help to create forecasts for nations worldwide.

2.2.2 Meso-scale weather models

Generally speaking, numerical weather models have to be set globally as there is no physical (lateral) boundaries that separate/isolate air motions. Indeed, the models run at national weather services most likely belong to this category. These global weather models are able to provide forecast over the whole globe. However, limited by the computational resources, many important processes at smaller scales are not possible to be represented efficiently in global models. The immediate result is that many products are provided with coarser resolution compared with those needed by applications. For example, local flood warning centres might have difficulties in explaining/utilizing the rainfall forecast with 100 km resolution when considering local flood related to an area around hundred of square kilometres in size.

Apart from global weather models, the mesoscale (middle-scale) models are also developed to address the processes at smaller scales (10 – 100 km). A mesoscale model is a numerical weather prediction (NWP) with sufficiently high horizontal and vertical resolution to forecast mesoscale weather phenomena. These phenomena are often forced by topography, coastlines, or are related to convection. Essentially mesoscale models are very similar to those global models in terms of using physical laws of motion and conservation of energy govern the evolution of the atmosphere expressed by a series of mathematical equations. However, mesoscale model normally work at a limited area which means that the (lateral) boundary also need to be settled apart from the initial conditions. Normally, the boundary conditions of mesoscale models are specified using forecast results from global models. The higher resolution is achieved by using a higher (smaller grid size) one for the working area compared with the resolution of the global model that specifies the boundary conditions. Such a process can be repeated by embedding another smaller area inside with an even higher resolution. Therefore, the resolution can finally reach to the level suitable for other use, e.g., flood forecasting/warning. The mesoscale model is conceptually depicted in Figure 2.4.

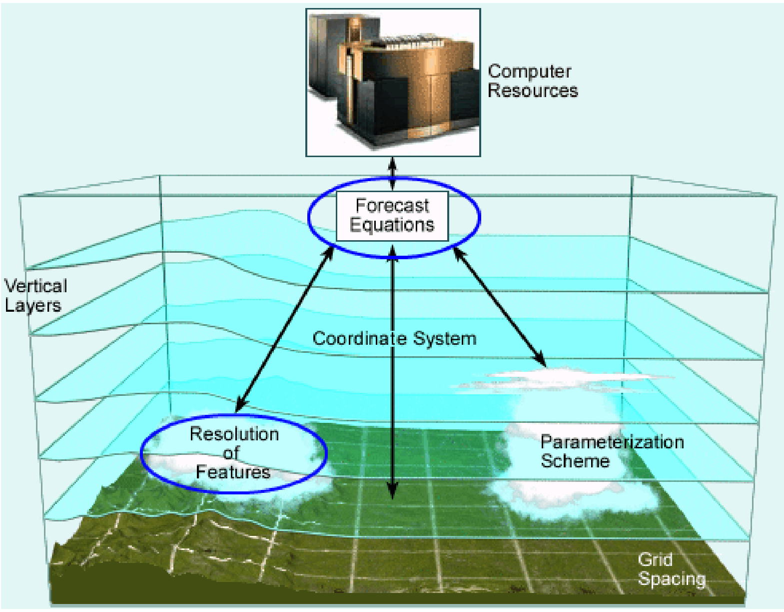


Figure 2-4 *Mesoscale model concept*
Source: UCAR, **http://www.meted.ucar.edu/**

Mesoscale models are run at varying horizontal resolutions with grid spacing typically less than 30 km. They have the capacity to provide great detail and often accurately represent the intensity of smaller-scale weather phenomena, often produce superior forecasts in coastal and mountainous regions when compared to traditional larger-scale models.

The horizontal resolution of an NWP model is related to the size of the weather feature. Grid-point models solve forecast equations at equally spaced grids and the variables forecasted are specified on a set of these grid points. NWP models also forecast the vertical structure and effect of a variety of meteorological events by its vertical resolution but cannot resolve processes that occur within the confines of a single grid box. As a result they must account for the total effect of these processes with a single number within the grid box. This method of accounting is called parameterization and one of the main reasons to do this is because of the processes are often not understood well enough to be represented by an equation.

According to University Corporation for Atmospheric Research (UCAR, 2008) precipitation schemes have much strength such as more realistic prediction of relative humidity fields, direct prediction of cooling from evaporating or melting precipitation particles and direct prediction of various ice crystal habits and their effects. Furthermore, there is a more realistic prediction of snow blowing downwind from its generation region.

2.2.3 MM5 Modelling system

The fifth-Generation NCAR / Penn State Mesoscale Model (MM5) (Dudhia et al., 2003) was chosen as the main weather model in this research. The availability of the source and the performance helped the MM5 model gain its popularity in both operational and research applications. It is a limited-area, nonhydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale atmospheric circulation. The model is supported by several auxiliary programs called the MM5 modelling system and it facilitates the development of various model inputs and outputs analysis. The algorithm uses full physics equations of momentum, thermodynamics and moisture. The model solves the full non-hydrostatic equations for the three-dimensional wind, temperature, water and pressure fields and the scheme allows for the vertical acceleration. It can be run with multiple nested grids in order to solve a range of atmospheric processes and circulations on spatial scales ranging from one to several thousands of kilometres.

The dynamic equations are resolved numerically by finite difference methods, the MM5 model uses the Arakawa B grid scheme for the horizontal grid (Arakawa and Lamb, 1981) and the sigma coordinate to define the vertical layers. In the modelling system several components use implicit parameterisations that give applicable range for a particular option. In terms of data assimilation it provides the facilities to make 3D and 4D variational assimilation so that the initial conditions can be improved by assimilating observed radiosonde data.

The use of mesoscale models demand for computing power, a PC-based cluster has been the popular choice accompanied by Linux operational system. MM5 modelling system programs are mostly FORTRAN programs that require compilation on a local computer. Some programs, Fortran 77 for example need recompilation each time the model configuration is changed; others like Fortran 90 need only once.

Many researches have successfully modelled atmospheric phenomena at high horizontal and vertical resolution using MM5 model (Cluckie and Rico-Ramirez, 2004; Lu et al., 2000; Xuan and Cluckie, 2006). They concluded that the mesoscale weather model MM5 is able to produce the forecast of the extreme precipitation events, even with a basic configuration, which implies its positive value in the flash-flood warning area. However, there are errors in precipitation prediction in terms of the spatial displacement and the disagreement in rainfall magnitude that need to be addressed.

2.2.4 Practical Indications on the MM5 Set Up

The MM5 modelling system has been well developed and maintained by a core team and features as a "community model" meaning that users can have sufficient and convenient support for setting up and running the model. Detailed information can also be referred to the user manual of MM5 as well as its homepage on the Internet. The structure of the MM5 modelling system is shown in Figure 2-5

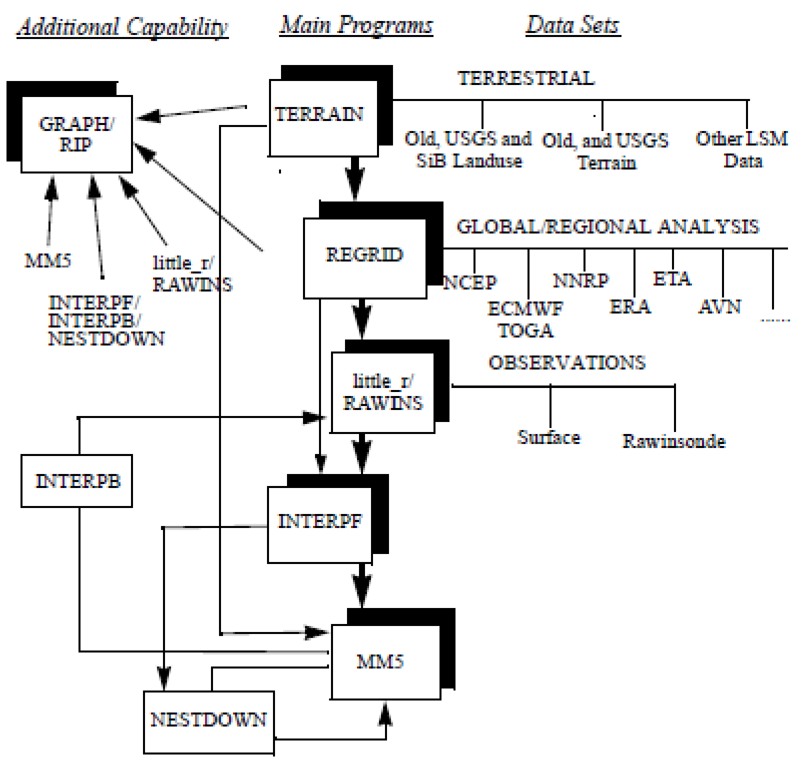


Figure 2-5 *The MM5 modelling system components (Dudhia et al., 2003)*

In this study, the MM5 model version 3.7 was set up on a Linux PC of which the hardware is sufficiently fast to support the research work. As there was no need for data assimilation in this case, only the following four components were used:

TERRAIN: The model domains that cover the study area were created and statistically initialised using this program.

REGRID: This program was used to convert datasets from ECMWF (global models) for setting initial and boundary conditions.

INTERPF: The converted data from REGRID was further interpolated onto the model domains and therefore the model was dynamically initialised with IC and LBC at this stage.

MM5 This is the real weather model. The MM5 program was compiled with support for parallel computing and therefore it was able to take full advantage of multi-core design of PC hardware. The results were generated by then running the M5 program RIP: This is a post-processing and visualising package for MM5 (and other weather models as well). Results from MM5 were processed and visualised. The intermediate files produced by RIP were further processes by other programmes of MATLAB as seen in following chapters.

The model dynamics and physics options is discussed in the user manual of MM5 (Dudhia et al., 2003) and a brief summary is provided in Appendix A.

3 DATA DRIVEN MODELLING TECHNIQUES

3.1 Introduction

In this study, the data-driven modelling techniques were employed in an effort to combine the prediction forecast from radar extrapolation and numerical weather predictions. This chapter gives an overview of the most common data driven modelling techniques used in hydrological modelling; model trees, linear regression, artificial neural networks (ANN) and modular model networks (MNN) used for integration of models.

3.2 Model trees and Linear Regression models

Classification and regression types algorithms are decision support tools that use a tree-like graph or model of decisions and their possible consequences, those were popularized by Breiman et al. (1984). Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal.

According to the nature of the model, it is possible to find two different approaches of trees for numerical prediction. If the model is based on average values of the instances (samples) for this range the approach is a regression tree but if the model is a linear regression function of the output variables the approach is a model tree. Like regression trees, a model tree can be learned efficiently from large datasets. However, model trees have advantages over regression trees in the prediction accuracy and the ability to use well the local data linearity (Quinlan, 1992).

A model tree is a classification method in which the data is split into classes or into a number of simple tasks. In other words, it is an algorithm whose inputs are a collection of instances and their correct classification. The outputs of this classification are class labels that can be used to classify each instance.

In a model tree each branch node represents a choice between a number of alternatives, and each leaf node represents a classification or decision. The algorithm operates over a set of training instances C. If all instances in C are in class P, create a node P and stop, otherwise select a feature or attribute F and create a decision node (Figure 3-1). The training instances are divided in C into subsets according to the values of V and apply the algorithm to each of the subsets C. (Wilson, 2008).

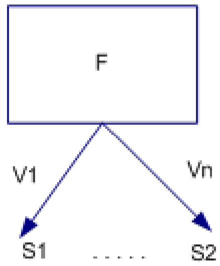


Figure 3-1 *Tree Induction Algorithm*

An M5 model tree is a system for learning models that predict values. One of the main characteristics of M5 is that it learns efficiently and can deal with very high dimensional tasks and hundreds of attributes. Moreover, the model trees in M5 are generally much smaller than regression trees and have showed more accurate in the tasks investigated (Quinlan, 1992).

Model trees are constructed by having a set T and associate it with a leaf. Some tests are chosen to splits T into subsets corresponding to the test outcomes. The same procedure is applied to other subsets. This division often produces over elaborate structures that must be pruned back. To build a model tree it is necessary to compute the standard deviation of the target values of cases in T. Whether T contains very few cases or their values vary only slightly, T can be split on the outcomes of a test. Each possible test is evaluated by determining the subset of cases associated with each outcome. If it is treated the standard deviation *sd* (T_i) of the target values of cases in T_i as a measure of error, the expected reduction in error could be obtained as follow:

$$\Delta error = sd(T) - \sum_i \frac{|T_i|}{|T|} * sd(T_i) \quad (3.1)$$

M5 algorithm analyse all possible tests, it chooses one that maximises this expected error reduction.

In respect of error estimates, M5 often needs to estimate the accuracy of a model on unseen cases. To estimate the error of a model derived from a set of training cases, M5 determines the average residual of the model that is the absolute difference between the actual target value and the value predicted by the model. M5 multiplies the value by (n+v)/ (n-v) where n is the number of training cases and v is the number of parameters in the model. The result is to increase the estimated error of model with many parameters constructed from small numbers of cases.

In order to avoid over fitting problems each non-leaf node of the model tree is examined, starting near the bottom. M5 selects as the final model for this node, depending on which has the lowest estimated error either the simplified linear model above or the model sub tree. The sub tree replacement is done in order to select some sub trees and replace them by single leaves.

M5 uses a smoothing process to compensate for the sharp discontinuities that will occur between adjacent linear models at the leaves of the pruned trees. Smoothing has most effect on a case when the models along the path predict very different values and when some models were constructed from few training cases. Experiments have shown that smoothing increases the accuracy of predictions. (Solomatine, 2007).

With respect to the linear regression models in which a target function is approximated, are constructed for the cases at each node of the model tree using standard regression techniques. However, instead of using all attributes, this model is restricted to the attributes that are referenced by tests or linear models somewhere in the sub tree at this node. The linear regression model is based on an assumption of linear dependencies between input and output.

The first objective of regression analysis is to best-fit the data by estimating the parameters of the model; it is built with the least squares criterion as follow:

$$y' = a_0 + \sum_{k=1}^{k} a_k x_k \qquad (3.2)$$

Where a_0 and a_k are the regression coefficients to be computed from a training dataset. As most natural processes are non-linear, there are non-linear regression methods in which the non-linearity is incorporated by a fixed non-linear functional form. This non-linearity is used in ANN by a set of basis function or in model trees by a combination of linear models.

Model trees have been applied successfully in water management. Haifa (2008) used the model tree to predict flow and water quality constituents combined with genetic algorithm for calibrating the model tree parameters. The method produced close fits in most cases, but limited peak flows and water quality loads. Solomatine and Xue (2004) and Kompare et al. (1997) showed improved predicting flows in a flood forecasting problems. Bhattacharya et al .(2007) applied them to solve sediment transport inaccuracies providing better accuracy than the existing models.

3.3 Artificial Neural Networks

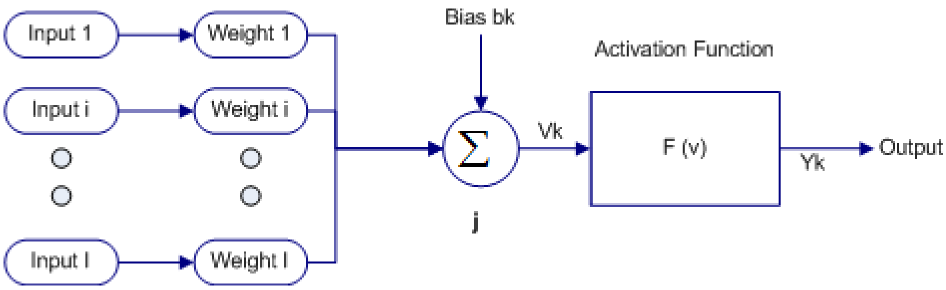
An Artificial Neural Network (ANN) is a mathematical methodology which describes relations between cause (input data) and effects (output data) without considering the process behind. The applications are widespread and vary from optimization of measuring networks, operational water management, prediction of drinking water consumption, wastewater treatment plants, sewage systems and to fill a gap between a primary model (physically based model) and a complementary model (data driven model) to obtain a better model for water level prediction (Abebe, 2004).

Neural networks have the ability to derive meaning from complicated or imprecise data, they can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions. Artificial Neural Networks (ANN) also have been known as the leading method in machine learning. ANNs are known as universal function approximations due to their ability to emulate complex processes and are considered to be one of the most important tools for knowledge encapsulation (Bhattacharya, 2005).

It is found that ANNs are robust tools for modelling many of non-linear hydrologic processes such as rainfall-runoff, stream flow, ground-water management, water quality simulation and precipitation. After appropriate training, they are able to generate satisfactory results for many prediction problems. However, it has been taken into account that the physical interpretation of ANN arquitecture, the optimal training data set and the adaptive learning tend to be very intensive and there appears to be no established methodology for design and successful implementation.

In general, the literature involving the application of ANNs to estimate rainfall includes data from satellite, radar observations and rain-gauge measurements (Da Silveira and Holt, 2001; Zhang et al., 1997). It has been used a standard back propagation neural network to estimate the rainfall rate using the satellite brightness temperatures as the inputs demonstrating that neural networks outperformed the regression model. Research involved the use of a neural network to determine the presence or absence of rain on the ground showing an improvement over the statistical techniques used as a comparison. More importantly the neural network has provided a robust spatial interpolation of the rainfall in areas without rain gauges (Bellerby et al., 2000).

A neural network is composed of processing elements (neurons or nodes) and connections between them (synapses). A single neuron is modelled with algorithms designed to alter values or weights used to adjust the network with a function, often non-linear that takes the weighted sum of neuron inputs and generates a desired output. An example of this structure is depicted in Figure 3-2. The bias changes the input range of the activation function.



$$Net = \sum_{i=1}^{I} I_i W_{ij}$$

Figure 3-2 *Example of a single neuron with I input signals*

There are three main categories of activation function (Haykin, 1999), Threshold function and Linear function where the amplification factor of the linear part is definitely large, and Non linear or sigmoid function described by:

$$F_v = \frac{1}{1+\exp^{-\alpha v}} \qquad (3.3)$$

Or by a hyperbolic tangent:

$$F_v = \tanh(\alpha v) \qquad (3.4)$$

One example of using the activation function is in the feed-forward Multi-Layer Perceptron (MLP) configuration. It uses several interconnected nodes, three types of layers: input, hidden and output layer with weighted connections between all nodes. Learning takes place in the MLP by changing connection weights after each piece of data processed based on the amount of error in the output compared to the expected output (back propagation). The error can be defined in the output node j in the nth data point by:

$$\begin{aligned} e_j(n) &= (measured_j(n) - predicted_j(n) \\ E(n) &= \frac{1}{2}\sum_j e_j^2(n) \end{aligned} \qquad (3.5)$$

There are some parameters used in MLP and in many of the learning algorithms. For instance, the momentum coefficient (μ) is used to prevent the system from converting to a local minimum. It also can help to increase the speed of convergence of the system if it has a high value. However, setting the momentum parameter too high can create a risk of overshooting the minimum and the system can become unstable. On the other hand, if the momentum coefficient is too low cannot reliably avoid local minima, and also can slow the training of the system.

The learning rate (η_s) affects the speed at which the ANN arrives at the minimum solution. In back propagation, the learning rate is analogous to the step-size parameter from the gradient-descent algorithm. If the step-size is too high, the system will either oscillate about the true solution, or it will diverge. Conversely, if the step-size is too low, the system will take a long time to converge on the final solution.

3.4 Modular Neural Networks (MNN)

Modular models (MM) are series of independent neural networks that serves as a module to solve subtasks. Each module of the network produces a specific output moderated by a combining unit. The intermediary takes the outputs of each module and processes them to produce the final output of the network (Petridis and Kehagias, 1998). One o f the major benefits of a modular neural network is the ability to reduce a large neural network to smaller, the subtasks will execute more efficiently than trying to deal with the whole tasks at once. A large neural network can suffer from interference as new data can alter existing connections so the training data used for each sub-network can be unique and implemented much more quickly.

There are a number of possible motivations for adopting a modular approach (Sharkey, 1999). A better performance can be achieved when the network is split into a number of specialist modules and when there is a switch to the most appropriate module, or blend of modules, depending on the current circumstances. Combinations of modules also produce greater noise robustness and a shorter training cycle due to the divide and conquer approach to problem solving. Each sub-problem could then be solved with a different neural net arquitecture or algorithms, making it possible to exploit specialist capabilities. A modular model scheme is showed in figure 3-3.

According to figure 3-3 the problem of the task components would be in finding the groups or clusters. A cluster analysis is based on a mathematical formulation of measure of similarity. There are a number of characteristics that distinguish different approaches to cluster analysis such as numerical, statistical and conceptual clustering. The clustering methods well known are: K-means, Jarvis-Patrick, agglomerative hierarchical and self organizing maps. In order to know the distance between data points in clustering the following options are used: Euclidean, Euclidean squared, Manhattan, Pearson Correlation, Pearson squared, Chebychev and Spearman (Abonyi and Feil, 2007).

The Levenberg-Marquardt algorithm (LMA) used in this research for MM, provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function. LMA is a robust algorithm; it means it finds a solution even if it starts very far of the final minimum. Detailed description of the algorithm can be found in (Press et al., 2007) .

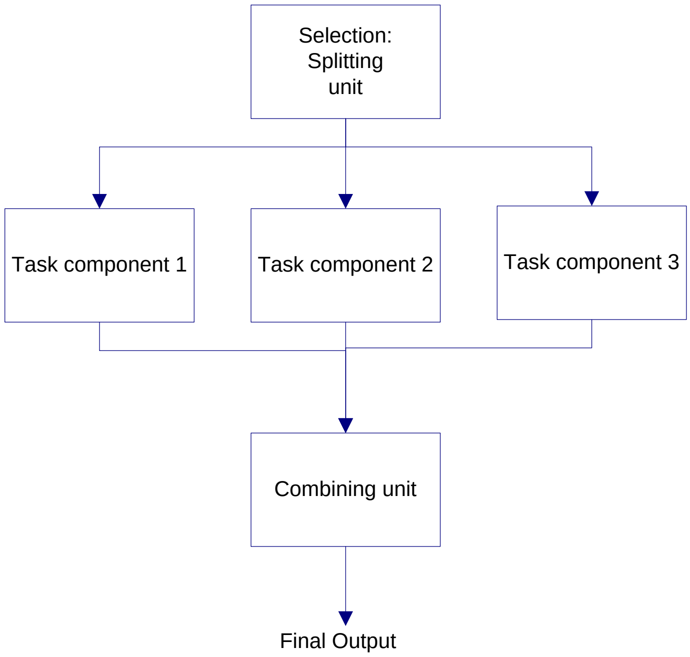


Figure 3-3 *Modular Model Scheme*

The literature regarding to the modular neural networks have shown good results. McCullagh (2005) explored the application of MNN to the rainfall estimation problem, demonstrating that an improvement in performance can be achieved by the application of MNN to this problem domain. The rainfall was classified as a No rain, Low, Medium and High, the four rainfall classes demonstrated the potential of the model developed. However, a less satisfactory result was found in the performance on the high rain examples. Corzo and Solomatine (2007a) built local data-driven models for different components of the water flow. A base flow separation technique was used as a basis for partitioning the data with local models for each flow component. The parameters of the separation were optimized. The results showed that the use of local models in a modular model structure is more accurate than the model trained using the whole data set.

4 METHODOLOGY

4.1 Flowchart

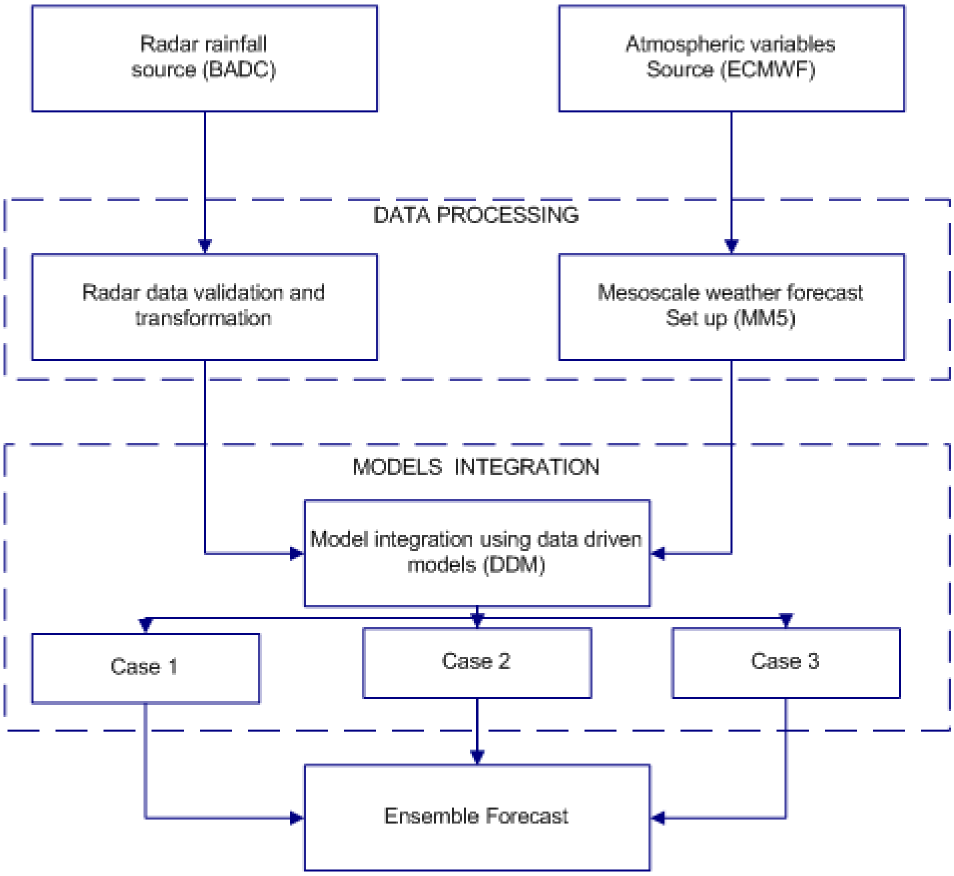


Figure 4-1 *Study Flowchart*

Figure 4-1 depicts the general schematization of the several tasks this study encompasses. The data were obtained from the British Atmospheric Data Centre (BADC) and from the European Centre for Medium-Range Weather Forecasts (ECMWF). This research was divided into two parts. The first part includes the radar data validation and transformation (NIMROD) and the mesoscale weather forecast model set up and configuration (MM5) explained in detail in section 4.2. In the second part the hybrid model is presented, spatial rainfall forecasts using data driven modelling techniques were analysed with three different cases.

4.2 Data Processing

4.2.1 Radar data validation and transformation (NIMROD)

The information used in this study comes from the UK Met Office Nimrod system (Golding, 1998). This system is oriented for nowcasting system purposes that analyse weather variables including precipitation amount in a very short-range prediction. The data is already processed. Nimrod is an automated system for weather analysis and nowcasting based around a network of C-band rainfall radars. The dataset has the fine-resolution analyses of rain rate for the UK and Europe (UK, 2003). The processing of the Nimrod radar data used in this study includes image quality control, echo removal and vertical profile correction (Golding, 1998). The imagery is calibrated against the radar composite and combined with lightning data.

In this study, the Nimrod composite images were used to build and verify extrapolation-based rainfall nowcasting. The field of presented by these radar images are precipitation intensity at every scan time. This study uses 5-km grid resolution and a temporal resolution of 15 minutes using the radar data of the stations Chenies near Amersham, and Wardon Hill in Dorset located in the Southeast of England (100-250 km N, 500-650 km E in the UK). The horizontal coordinate used in Nimrod is the National Grid Reference (NGR) which specifies locations by easting and northing pairs in metres relative to the origin on the map of the UK.

The data available was formatted on an hourly basis and the area was gridded into 31 by 31 cells of 5 km each. Figure 4-2 shows the study area for this research.

Figure 4-2 *Study Area*
(Source: Google Earth Platform)

4.2.1.1 Rainfall Events

Radar information (data sets) were downloaded from the British Atmospheric Data Centre (UK, 2003). Events 27 May 2007, 28 May 2007, 24 June 2007, 05 September 2008 and 08 November 2008 were selected and used for training and verification of the models. May and June 2007 were both exceptionally wet months. The main feature of the summer was the high rainfall experienced in many regions especially during June and July. It was the wettest summer for the whole of the UK since the rainfall series began in 1914. The areal precipitation value from 22 May to July 2007 in the Southeast of England was 364.7mm. 5 and 6 September 2008 brought prolonged heavy rainfall whilst 8 November were reported showers and thunderstorms moved east across central and southern England (Source:UK, 2007).

The Nimrod system data comes into binary files followed by the Nimrod file format. The data had to be decoded before its use. A code written in MATLAB was used for this purpose. The rainfall information given by Nimrod usually comes as rainfall intensity which is the simultaneous measurement from the radar scan. Figure 4-3 shows the schematization of the tasks for the radar data processing.

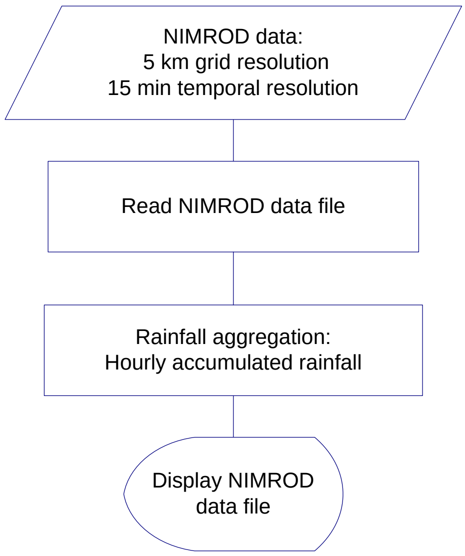


Figure 4-3 *Chart of the radar data processing*

4.2.2 Forecasts based on correlation "Tracking"

The radar echo tracking was used as a model reference in this study; this methodology was based on the persistence of the rainfall pattern (Lagrangian persistence). It permits to predict the new development, displacement and dissipation of the rainfall in existing cells by means of a cross-correlation between consecutive radar scans. Its maximum correlation shows the displacement of the rainfall event. The radar tracking echo is successful when it has been predicted one or two steps ahead but the error increases with longer lead times. Previous studies using this method have shown that an improved QPF with a three hour lead time can be reached (Cluckie et al., 2006).

The tracking of the radar echo starts from two successive scans A and B, the velocity vector V is obtained from the displacement from A to B taking into account the constant time interval ∂t between two successive scans. The radar echo extrapolation is shown graphically in Figure 4-4 in Cartesian coordinates. The rains A, B, C are radar echoes, the current and the extrapolated one stop ahead. B is different from A due to growth/decay effects. The velocity vector is derived from the displacement from A to B and is expressed as $V = \partial x \vec{i} + \partial y \vec{j}$. One step ahead forecast can be obtained by advecting the rain B with the same amount of displacement to location C. B needs to be found in such a way that it directly corresponds to the identity A at the previous time step.

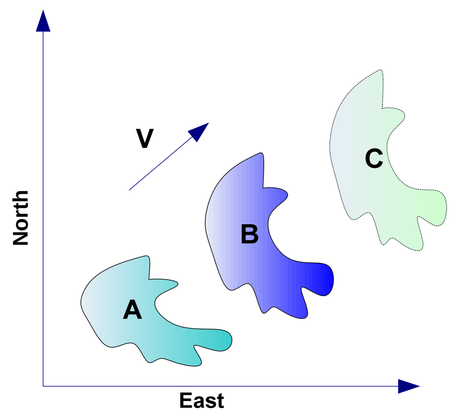


Figure 4-4 *Radar echo extrapolation in Cartesian coordinates*

The weakness of the method comes from the errors for longer lead-times (longer than 3 hours), also from the tracking error with respect to the tracking technique and the error of the unaccounted growth and decay of precipitation intensities. The precipitation might still be so fast that tracking only scheme may completely lose its skill. As the displacement correction is still based on the tracking method, similar difficulties are expected to be encountered when the precipitation system is a multi-cell one rather than a clearly defined frontal system, or the system is located near the image boundary so that the best-process match would fail. The success of the algorithm depends upon the quality of the forecast from the numerical weather model and/or the reliability of the radar observation as the adjustment would be difficult to find for patterns that do not resemble each other. An approach to reduce the uncertainty of the rainfall forecast in terms of location and pattern matching can be found in Xuan (2007).

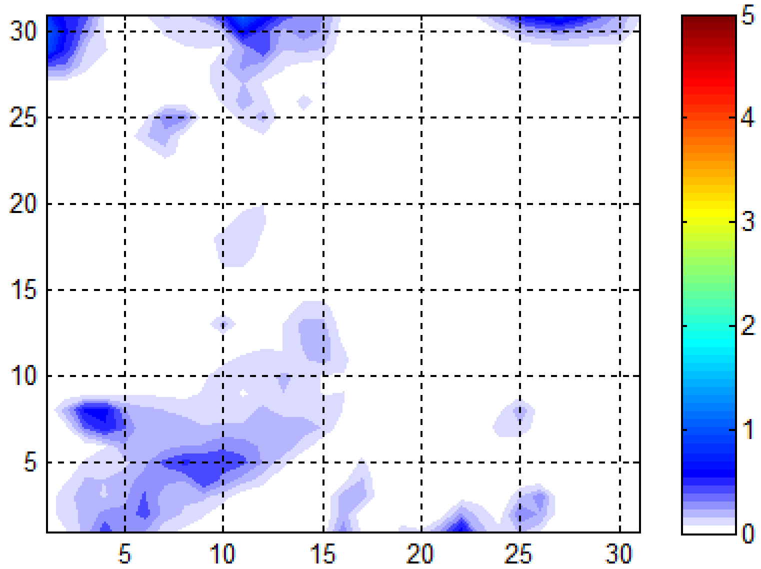
For the cross-correlation scheme, the search of area B is made by finding the area having the highest value of cross-correlation with the previous area A. For this research The MATLAB function called xcorr2 was applied to return the cross-correlation of matrices A and B. Matrix A has dimensions ($m1$, $n1$) and matrix B has dimensions ($m2$, $n2$). Equation (4.1) shows the two-dimensional discrete cross-correlation:

$$C(i,j) = \sum_{m=0}^{(m1-1)} \sum_{n=0}^{(n1-1} A(m,n) * conj(B(m+i, n+j)) \qquad (4.1)$$

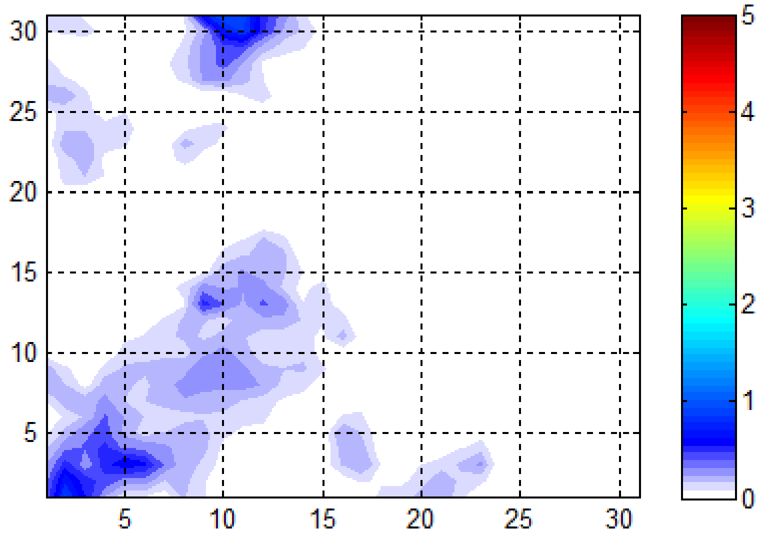
Where $0 \le i < m1 + m2 - 1$ and $0 \le j < n1 + n2 - 1$.

Extrapolation of hourly accumulated radar scans (24 hours) was applied as a first step to obtain spatial distributed rainfall forecast. The radar data obtained from NIMROD system and processed as it is described in 4.2.1 was used as rainfall measured. The displacement (or the advection vector) was determined by the difference between the image at t, and the corresponding part of the image at t+1 that matched with the initial one with the maximum cross correlation. In the hybrid modelling (section 4.3), for case 1 the event 27 May 2007 was used to forecast rainfall 1, 2, 3 and 4 hours ahead. Similarly, for case 2 the radar tracking method was applied for extended data set (5 events) to forecast rainfall as it was done in case 1.

Figure 4-5 shows an example of cross correlation nowcasting compared with the radar rainfall measured 1 hour ahead for the event 24 June 2007. For this example, although the system is located near the image boundary so that the best-process match fails, the scheme was able to produce good nowcast with respect to the radar measured.



(a)



(b)

Figure 4-5 *Example of cross correlation nowcasting based on the echo extrapolation*
(a)Rainfall measured (mm) 24/06/2007 03:00hrs
(b) Prediction based on correlation tracking (mm)
1 hour ahead 24/06/2007 04:00hrs

4.2.3 Mesoscale weather forecast set up MM5

The MM5 model was initialised with the initial condition (IC) and lateral boundary condition (LBC) taken from the outputs of global models at the European Centre for Medium-Range Weather Forecasts (Persson, 2003). The essential variables included in the data sets are:

1. SST, the sea surface temperature.
2. MSLP, the mean sea level pressure.
3. GHT, geopotential height or its equivalent geopotential.
4. T, the temperature.
5. RH, relative humidity.
6. V, the wind velocity field.

The MM5 model was configured to have four nested domains with a grid size of 54 km, 18km, 6km and 2km from the outermost to the innermost domain respectively, as seen in Figure 4-6. The parameters needed in order to define mesoscale domains are: Location of grid point (1,1) in its mother domain, Mother domain ID, domain size (number of grid points in each direction) and grid distance in Km. For this case Domain 4 nearly matches the radar coverage.

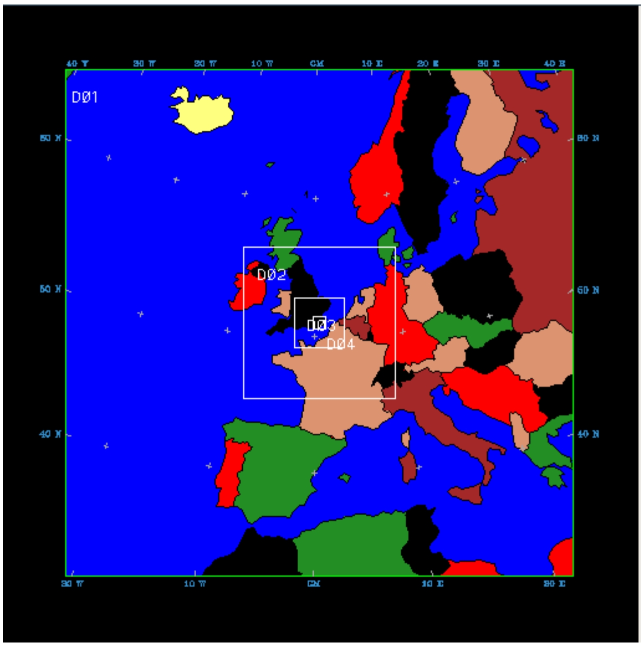


Figure 4-6 *MM5 Nesting Configuration*

Because of the NGR is not one of the three projections supported by MM5, the data had to be processed through a projection transformation from MM5 to NGR. The MATLAB function called griddata was used to interpolate the MM5 surface specified by the radar grids using nearest neighbour interpolation. The MM5 data were re-gridded onto the target grids (Radar data), process called regridding.

Hourly rainfall accumulation was also produced in order to compare the hourly available radar and tracking data.

The RIP4 program from the weather research and forecasting model(WRF, 2008) was used to visualise the mesoscale model outputs. This RIP4 was compiled using LINUX operative system. When it was executed, rainfall data was obtained. Figure 4-7 shows the schematization of the tasks for the numerical weather model (MM5) data processing for this study.

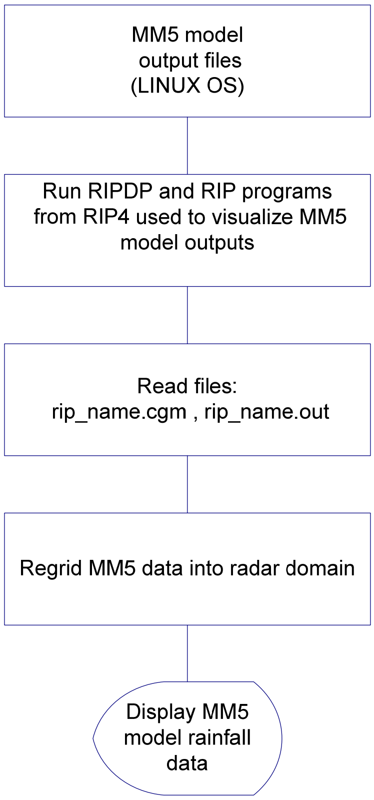


Figure 4-7 *Chart of the MM5 model data processing*

4.2.4 Configuration of Mesoscale weather model (MM5)

The MM5 model was initialised at 00 UTC for 24-hour forecast of the selected events. For case 1: event 27 May 2007 was run continuously to produce 24 hours forecast. For cases 2 and 3: Events 27 May 2007, 28 May 2007, 24 June 2007, 05 September 2008 and 08 November 2008 MM5 model was run every six hours to produce 6 hours forecast.

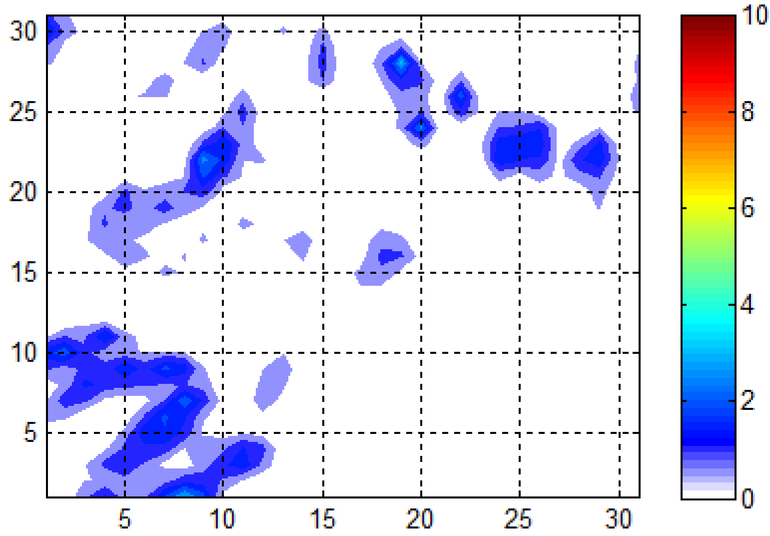
As it was stated in chapter 2 section 2.2.4, the model domain was obtained using the TERRAIN program, the data available as input of the program include terrain elevation, land use/vegetation, land-water mask, soil types, vegetation fraction and deep soil temperature available at six resolutions: 1 degree, 30, 10, 5 and 2 minutes and 30 seconds.

The pregrid program, the first component of REGRID was run using the gridded meteorological analyses obtained from the TERRAIN program. It created files for every time between starting and ending dates. The regridder program, a second component of REGRID was run using temperature, horizontal wind components, relative humidity and height of pressure levels. At the end, it was generated a file called “REGRID_DOMAIN #”. This file contains the data at every time period for a single domain.

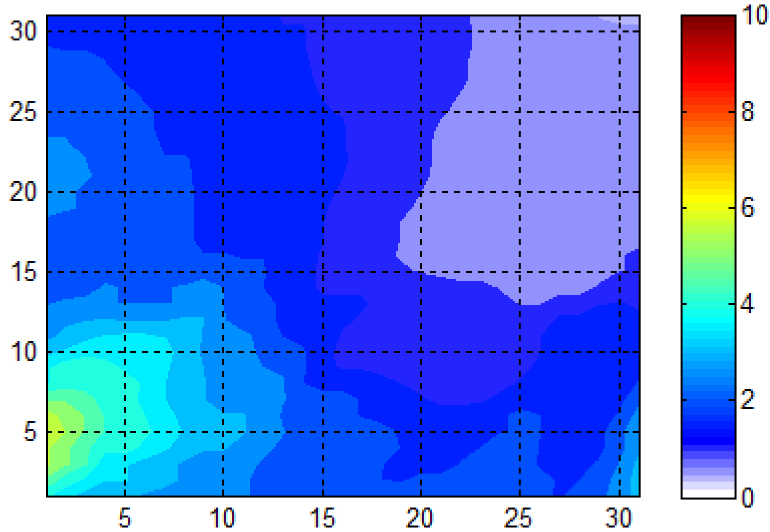
With the meteorological data on pressure levels INTERPF was run and the output files obtained were the initial condition for MM5, the lateral boundary condition for MM5 and the lower boundary condition. They were fed once and were run over the entire prediction interval. Afterwards, those files were used to configure the MM5 model. The MM5 model run produced a number of raw output files naming MMOUT_DOMAIN4.

Finally, a post-processing was needed in order to evaluate and utilize the model results. A code written in MATLAB was used to read the rainfall output files as the “raw” outputs are written in a predefined binary format. The geo-referenced process included the interpolation of the MM5 into the radar domain using the nearest-neighbour method because MM5 does not use the local projection of the radar part. The interpolation method assigns a sample value to a point in space. The value assigned is equal to the value of the spatially nearest sample data point.

Figure 4-8 in the next page depicts an example of the MM5 model configuration process and it is compared with the radar rainfall measured. For the event 24/06/2007, it can be seen that the MM5 model (b) created more “smooth” or flat rain bands due to the relative-fast-varied weather system. Radar rainfall measured (a) advection cannot simulate the process easily.



(a)



(b)

Figure 4-8 *(a) Radar rainfall measured (mm) 24/06/2007 12:00hrs (b) Rainfall prediction (mm) based on –NWP model MM5 24/06/2007 12:00hrs*

4.3 Hybrid Modelling

4.3.1 Introduction

In chapter 2 rainfall nowcasting obtained from either weather radars or numerical weather prediction models has been presented. Weather radars do not use important physical information, like wind, temperature, pressure and others, to calculate the rainfall information. This information obtained is commonly based on tracking spatial patterns of rain. Therefore, on extended forecast horizon the high variability of physical variables (e.g. wind, humidity and others) are contemplated by the NWP and not by the tracking of radar rainfall information. On the other hand, NWP models do not generally capture well the initial rainfall distribution. Nevertheless, over longer time scales, NWP models would perform better than weather radars as they resolve dynamically the large scale flow.

A hybrid modelling approach was the method adopted in this research in order to combine the best features of both radar rainfall information and numerical weather prediction model (MM5) through the integration of their outputs. To achieve this objective, four data-driven models were used in three different cases. The motivation of working with three cases is to evaluate the DDM performance of one rainfall event using data sampling (case 1). Also, to evaluate the DDM performance with an extending data sets (5 events) using expert knowledge (case 2). Finally, to explore the use of a rule based forecast on the integration of weather radar and MM5 model outputs (case 3). For the three cases rainfall prediction 1, 2, 3, and 4 hours ahead were produced. Table 4-1 shows detailed information of the three cases used in this research.

Table 4-1 *General information of the hybrid modelling*

Case	Data partition	Amount data set	MM5 mode Used	Forecasts	DDMs	Storm Events
1	Data sampling	1 day	Fed once in 24h	1 - 4 hours	M5,LR,ANN,MM	27/05/2007
2	Expert knowledge	5 days	Fed every 6 hours	1 - 4 hours	M5,LR,ANN	27/05 28/05 24/06/2007 // 05/09 08/11 2008
3	Expert knowledge / Rule based forecast	5 days	Fed every 6 hours	1 - 4 hours	ANN	27/05 28/05 24/06/2007 // 05/09 08/11 2008

4.3.2 Performance measures

There are a wide variety of statistical forecasting measures for the performance analysis; here it was considered the root mean squared error (RMSE) and a detailed analysis of the variability of the results in each cell of the forecast grid. In this study the radar information was used as reference value of the measured precipitation. The RMSE was used to compare the differences between the radar measured and the rainfall forecast obtained from tracking forecast and MM5 model. The correlation coefficient (R) was also used to measure the strength between variables and the persistence index (PERS) to know the relationship of the model performance and the performance of the naïve model. Naïve model in this case represents the forecast at each time step equal to the current value by the equation 4.2:

$$PERS = 1 - \frac{SSE}{SSE_{naive}} \qquad SSE = \sum_{t=1}^{n} (R_{forecast,t} - R_{measured,t})^2 \qquad (4.2)$$

$$SSE_{naive} = \sum_{t=1}^{n} (R_{measured,t+L} - R_{measured,t})^2$$

t = time, L = lead time, n= number of steps for which the model error is computed.

Conditional probabilities for categorical forecasts of a binary event were applied. Event is defined as a value of rainfall above threshold. Performance measures such as Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Sussex Index (CSI) were considered from the schematic contingency table (Jolliffe and B., 2003). A two-way contingency table is a two-dimensional table that gives the discrete joint sample distribution of forecasts and observations in terms of cell counts. Table 4-2 gives de cell counts for each of the four possible combinations of forecast and observed event represented by a, b, c and d.

Table 4-2 *Schematic contingency: table for categorical forecasts of a binary event.*

Forecast	Observed		
	Yes	No	Total observed
Yes	a (hits)	b (false alarms)	a+b
No	c (misses)	d (correct rejections)	c+d
Total	a+c	b+d	a+b+c+d=n

Cell count *a* is the number of event forecasts that correspond to event observations, or the number of hits; cell count *b* is the number of event forecasts that do not correspond to observed events, or the number of false alarms; cell count *c* is the number of no event forecasts corresponding to observed events, or the number of misses; and cell count *d* is the number of no-event forecasts corresponding to no events observed, or the number of correct rejections.

The hit rate also known as probability of detection (POD) is the proportion of occurrences that were correctly forecast. This quantity is defined by the total number of correct event forecasts (hits) divided by the total number of events observed:

$$POD = \frac{a}{a+c} \qquad (4.3)$$

The false alarm ratio (FAR), an event that is associated with no event observed, it is equal to the number of false alarms divided by the total number of event forecast:

$$FAR = \frac{b}{a+b} \qquad (4.4)$$

Finally, the Critical success index (CSI) is equal to the total number of correct event forecasts (hits) divided by the total number of event forecasts plus the number of misses (Hits+false alarms+ misses):

$$CSI = \frac{a}{a+b+c} \quad (4.5)$$

4.3.3 Case 1: Rainfall Forecasts using Data Driven Models (Continuous MM5 Simulation).

Data

As a first analysis 27 May 2007 event was chosen, the rainfall prediction in the case of the weather radar was obtained by implementing an extrapolation of hourly-accumulated radar images as mentioned in 4.2.2. The maximum cross-correlation of two radar scans at time *t* and *t* – 1 indicated the displacement of the rainfall event. The objective of the forecast was to obtain 1, 2, 3, and 4 hours ahead. Later on, The MM5 model was initialised at 00 UTC for 24 hours continuous simulation, the information was fed once and it was run over prediction interval. MM5 model forecast data were interpolated into the radar domain in order to get the local projection of the radar.

Figure 4-9 show the images for the radar measured, radar tracking forecast and MM5 model data respectively obtained for 2 hours forecast. There is a displacement among the actual (a) and forecasted rainfall (b). Although, the MM5 model (c) captured the precipitation system observed by the radar, similarly to Figure 4-8, it created flat rain bands showing growth and decay of precipitation compared with the nowcasting system.

Pre-processing

Afterwards, the data set was split using random sampling into a training (9130 instances) and verification (9130 instances). The approach used for this split was based on the research conduced by (Corzo, 2009). The criteria adopted for partitioning of data into training and validation should result in statistically similar sets. The sampling was done in two steps; firstly, it was generated 100 different groups of randomly generated training and validation sets. Secondly, the group in which the two mentioned sets are statistically similar were chosen by comparing the absolute difference in terms of a combination of statistical measures. In other words, the relative absolute difference between the standard deviation of the training and validation data set ($RADS_{tv}$) and the relative absolute difference between the mean of the training and validation data set ($RADM_{tv}$) were evaluated as follows:

$$RADS_{tv} = \left| \frac{\sigma_{tr} - \sigma_{va}}{\sigma_{tr}} \right| \quad (4.2)$$

$$RADM_{tv} = \left| \frac{\mu_{tr} - \mu_{va}}{\mu_{tr}} \right| \quad (4.3)$$

Models Set Up

After the splitting was done, four different models based on data driven modelling techniques explained in chapter 3 were constructed. The models are: model trees (M5), linear regression model (LR), artificial neural networks (ANN), and modular neural networks or modular models (MM). The models use rainfall prediction from radar tracking and the MM5 model outputs. The rainfall prediction from the models can be described as:

$$R_{t+\tau} = f\left(MM5_{t+\tau}, RT_{t+\tau}\right) \quad (4.4)$$

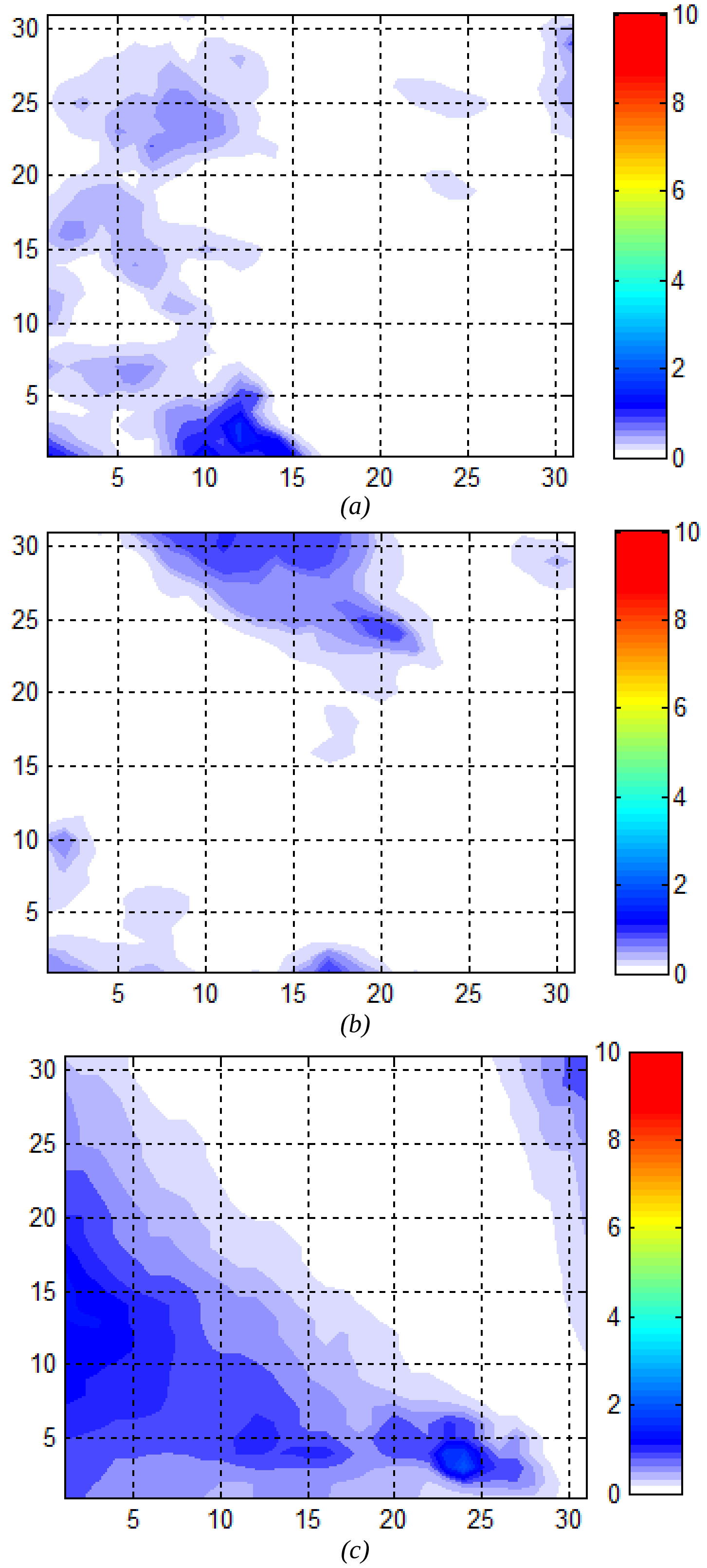


Figure 4-9 *Rainfall prediction (mm) (a) Radar measured (b) Radar tracking forecast (c) MM5model forecast 2 hours ahead 27/05/2007 09:00hrs*

Where $R_{t+\tau}$ is the rainfall prediction at time t+τ (lead time) in function of MM5 model and radar tracking (RT). For the model integration the radar and MM5 models where indexed in such a way that the results of each forecasted image can be graphically evaluated independently. Figure 4-10 shows the indexed mapping used.

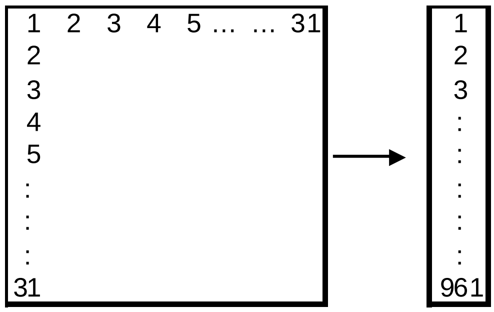


Figure 4-10 *Radar and MM5 image split*

Two software's were used in this study to run the data driven models: The Waikato Environment for Knowledge Analysis (WEKA) which uses Java class libraries and implements many state of the art machine learning and data mining algorithms. It has been developed by Witten and Fran in 1999. MATLAB (Matrix laboratory) a users interface to call some FORTRAN matrix routines developed by Math works.

Two inputs variables of rainfall were used: MM5 model forecast and radar tracking forecast, they were compared with the radar measured data. The models were run every hour to obtain from 1 to 4 hours forecasts. ANN models were multi-layer perceptron (MLP), the correct number of nodes in the hidden layer for each model was obtained by iteration procedure and it uses along with the output layers a tangent transfer function. The values of the learning parameters were Momentum coefficient (μ) =0.2, adaptive learning rate (η_s) =0.3 with a maximum epochs of 500.

With respect to the Modular Models a standard k-means algorithm was used to find groups (clusters) of rainfall (Corzo and Solomatine, 2007b) . K-means clustering generated a specific number of non-hierarchical clusters. The data sets were partitioned into K clusters and the data points were randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. For each data point, it was calculated the distance from the data point to each cluster. If the data point was closest to its own cluster, it was left where it was but if the data point was not closest to its own cluster, it was moved into the closest cluster. The process was repeated until a complete pass through all the data points resulted in no data point moving from one cluster to another. At this point the clusters were stable and the clustering process ended. The number of clusters in this study was optimized such that the best number of clusters to use was found. The distance metrics used to build the clusters was sq Euclidean. It can be described as follows:

Distance between a point X(X_1, X_2, etc) and point Y (Y_1, Y_2, etc) is

$$d = \sum_{i=1}^{n} \left(x_i - y_i \right)^2 \qquad (4.5)$$

4.3.4 Case 2: Rainfall Forecast using Data Driven Models (MM5 online update, extended data set)

Data

The rainfall data used for this case include five storm events 27 May 2007, 28 May 2007, 24 June 2007, 05 September 2008 and 08 November 2008. 6 hours forecasts were produced applying the radar tracking method. The MM5 model was initialised at 00 UTC for 6 hours forecast, every hour. The information was fed at hours 6, 12, 18 and run over prediction interval. MM5 model forecast data were interpolated into the radar domain in order to get the local projection of the radar similar to case 1.

Figure 4-11 shows the images for the radar measured, radar forecast and MM5 model forecast respectively for 3 hours ahead. The storms can be classified as a stationary convective system as the overall cloud and rainfall patterns are mostly round in shape. Those figures are described in more detail in chapter 5.

Models Set Up

Three data driven models were run: model trees (M5), linear regression model (LR) and Artificial Neural Networks (ANN) using WEKA and MATLAB every hour (from 1 to 4 hours forecasts). ANN models were multi-layer perceptron (MLP), the correct number of nodes in the hidden layer for each model was obtained by iteration procedure and it uses along with the output layers a tangent transfer function. The values of the learning parameters were Momentum coefficient (μ) =0.2, adaptive learning rate (η_s) =0.3 with a maximum epochs of 500.

Two input variables of rainfall were used: MM5 model forecast and radar tracking forecast. Similarly to case 1, the radar and MM5 models where indexed in such a way that the results of each forecasted image can be graphically evaluated independently.

This case was analysed as follow: Firstly the models were trained using the whole data from the five events. Secondly the data was split using four events 27 May 2007, 28 May 2007, 24 June 2007, and 08 November 2008 for training and one event 05 September 2008 for validation. The motivation of working with the whole data for training was to analyse the improvement of the models performance and also due to the amount of data. For each case, the performance of the models was analysed using RMSE, conditional probabilities and visual inspection of the images.

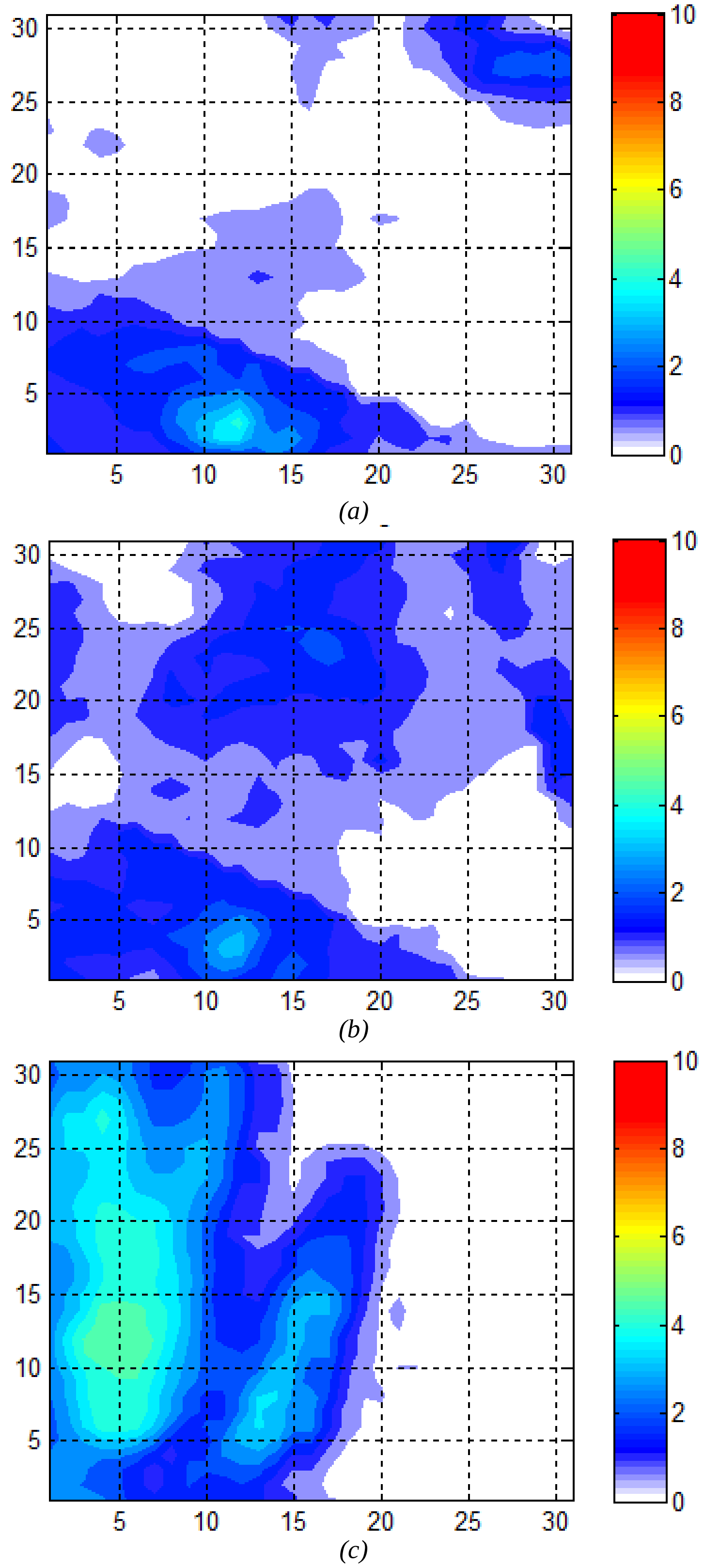


Figure 4-11 *Rainfall prediction (mm) (a) Radar Measured (b) Radar tracking forecast (c) MM5 model forecast 3 hours ahead 05/09/2008 22:00hrs*

4.3.5 Case 3: Using rule based Forecast (MM5 online update, extended data set).

This experiment consists in predicting the class of rainfall from 1 to 4 hours ahead based on five storms events used in case 2, four events 27 May 2007, 28 May 2007, 24 June 2007, 08 November 2008 for training and one event 05 September 2008 for validation. 6 hours forecasts were obtained with the radar tracking method and 6 hours forecast, every hour initialised at 00 UTC using the MM5 model. ANN model was built using MLP with sigmoid function and three hidden layers. This model works to classify rainfall into three classes (No, Low or High). The input variables used were MM5 model forecast and radar tracking forecast. The model uses values of the learning parameters (μ) =0.2, adaptive learning rate (η_s) =0.3 and maximum epochs of 500. 15376 samples were used for training and 3844 for verification. The model was run for each hour forecast.

This case was based on the research conducted by McCullagh (2005). It consists as it was stated above of classifying rainfall into one of the three classes (no rain, low rain, high rain). This process called rainfall classifier consist of three ensembles of neural networks developed to make one of the three binary rainfall decisions. In this research, it was used a simplified approach due to the characteristics an amount of data using one ANN model. The following rules were done to build the model:

If MM5 and radar tracking (RT) forecast = 0 **Then**
Forecast = 0

Else If MM5 forecast>0 and RT = 0 **Then**
ANN model - Rain
Integration (trained with only non-zero values)

Else If MM5 Forecast=0 and RT>0 **Then**
ANN model - Rain
Integration (trained with only non-zero values)

Else If MM5 forecast >0 and RT>0 **Then**
ANN model - Rain
Integration (trained with only non-zero values)

The range for each of the classes is outlined as follows:

Class 1	No Rain	>=0 mm and <0.3 mm
Class 2	Low Rain	>=0.3 mm and < 3 mm
Class 3	High Rain	>=3 mm

These rules use ANN model with two input variables MM5 model forecast and radar tracking forecast. The predicted rainfall depends on MM5 rainfall forecast and radar tracking forecast. MM5 and radar tracking samples were initially passed through the Rain/No Rain. The samples classified as No Rain maintained with this classification. Samples classified as Rain were passed through to the Low/High. Samples classified as Low Rain maintained this classification. Finally no Low Rain samples were classified as a High Rain to complete the classification process. Figure 4-12 depicts the rainfall classifier chart.

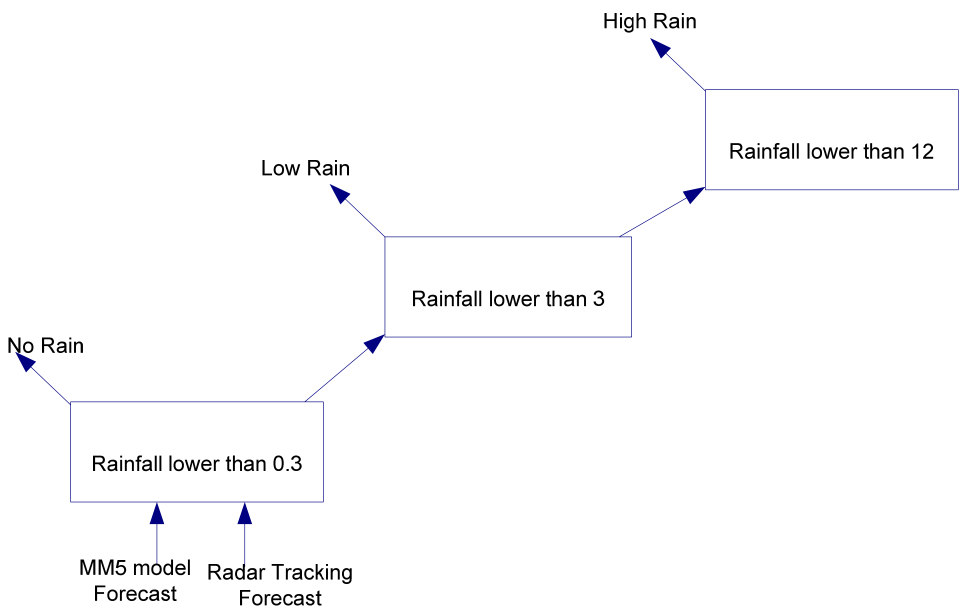


Figure 4-12 *Rainfall Classifier chart (Hierarchical ensemble of ANN)*

5 RESULTS AND DISCUSSION

5.1 Case 1: Rainfall Forecasts using data driven models (DDM) (Continuous MM5 simulation)

A detailed analysis was made to analyse the cell results for the period considered. Each image was analysed independently and the cell's results for each hour forecast were plotted. Since the results have been indexed and added together only a fragment of the forecast is shown in Figure 5-1. Here we can see the first 100 cells results (e.g. first image forecast), for a 3-hours forecast. The figure shows the DDM: linear regression model, model trees, and ANN forecasts. The prediction results using DDM underestimate the rainfall due to the combination of the radar forecast with the MM5 model.

Figure 5-2 shows the scatter plots comparing the observed and predicted rainfall. The red line means the best linear fit. ANN and linear regression models show larger under prediction of rainfall amounts compared with those obtained from radar tracking and model trees model. This was also anticipated from the comparison with figure 5-1. MM5 presents a negative correlation due to the difference of values between observed and forecasted.

The RMSE of the verification results is used to asses the performance of the integration. The DDM integration are better able to reduce the RMSE values for all the forecast steps (Figs. 5-3, 5-4, 5-5). The results show values of around 30% reduction in all the forecasts made.

In terms of persistence index (PERS) the figures 5-3, 5-4 and 5-5 also show the DDM are better than the naïve model where the closer to 1 the better. Values of PERS less than 0 means the forecast is less worthy than the naïve model. According to this the results show an improvement of PERS values, for instance using an ANN model it was obtained a value of PERS equal to 0.48 compared with -0.38 with a radar tracking forecast.

Regarding the correlation coefficient (R) of the verification results, although the values obtained are still low they have been improved with all DDM used and compared with those obtained with radar tracking forecast. The reason of obtaining a low correlation coefficient is due to the use of zero rainfall inputs (no rain) present in the cells of the images. For three hours forecast with radar tracking forecast the R obtained was 0.13 and using model trees the result was 0.39 giving an improvement of 26% over the radar tracking.

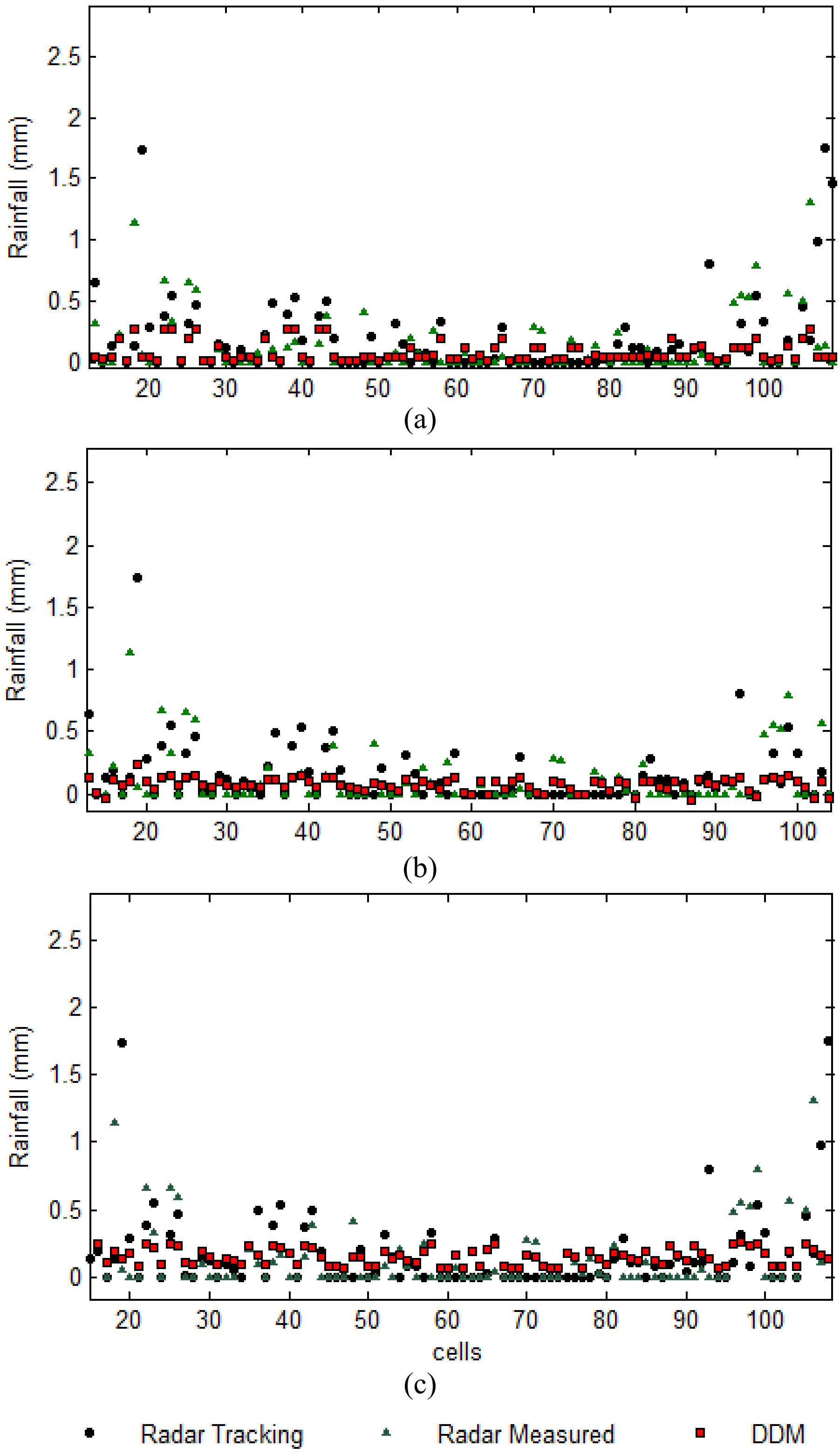


Figure 5-1 *Rainfall prediction (mm) (a) M5P (b) LR (c) ANN*
3 hours forecast event: 27/05/2007 Indexed image in 961 cells

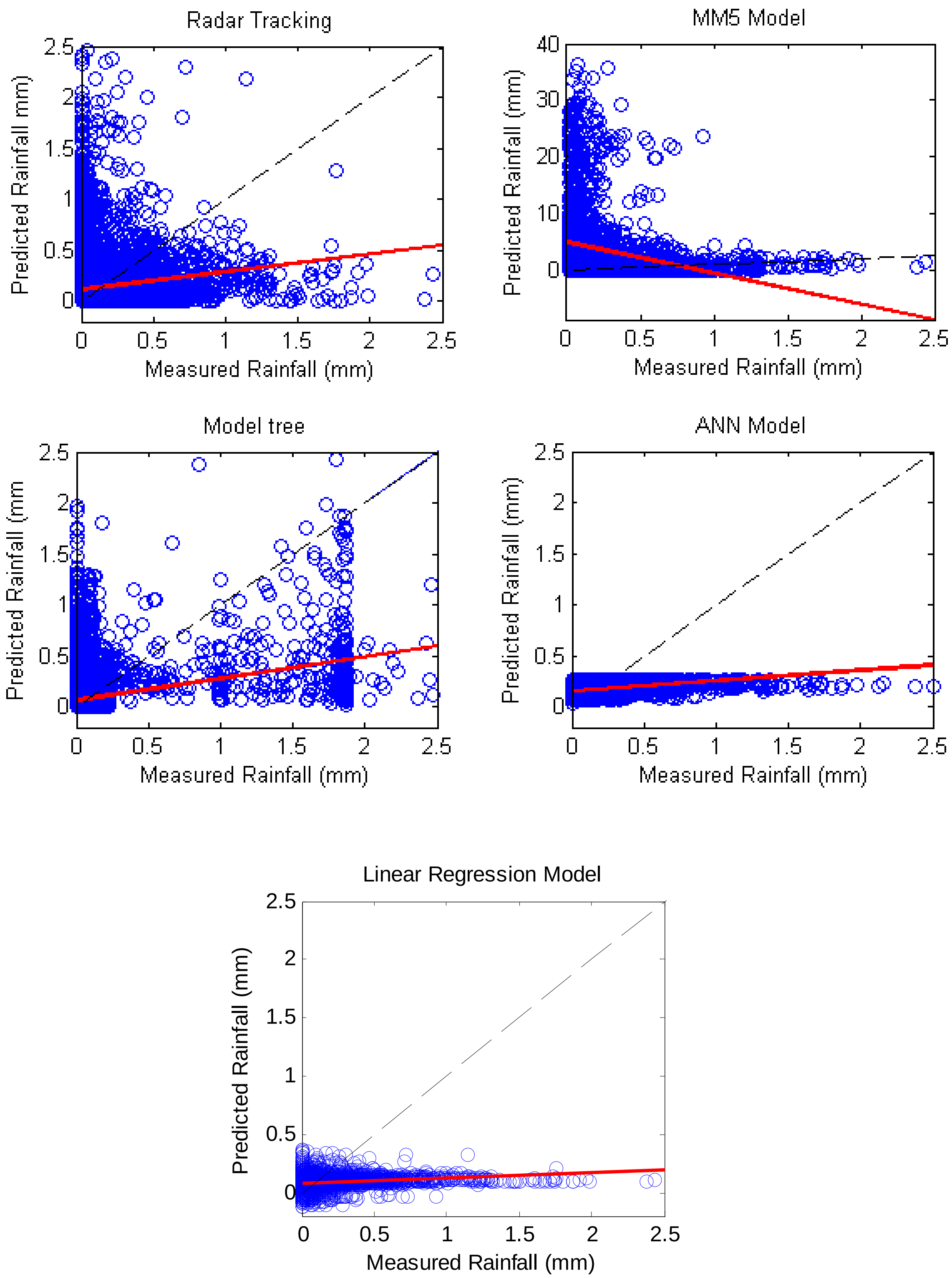


Figure 5-2 *Scatter plot comparison between predicted and measured rainfall in (mm)*
3 hours forecast event: 27/05/2007

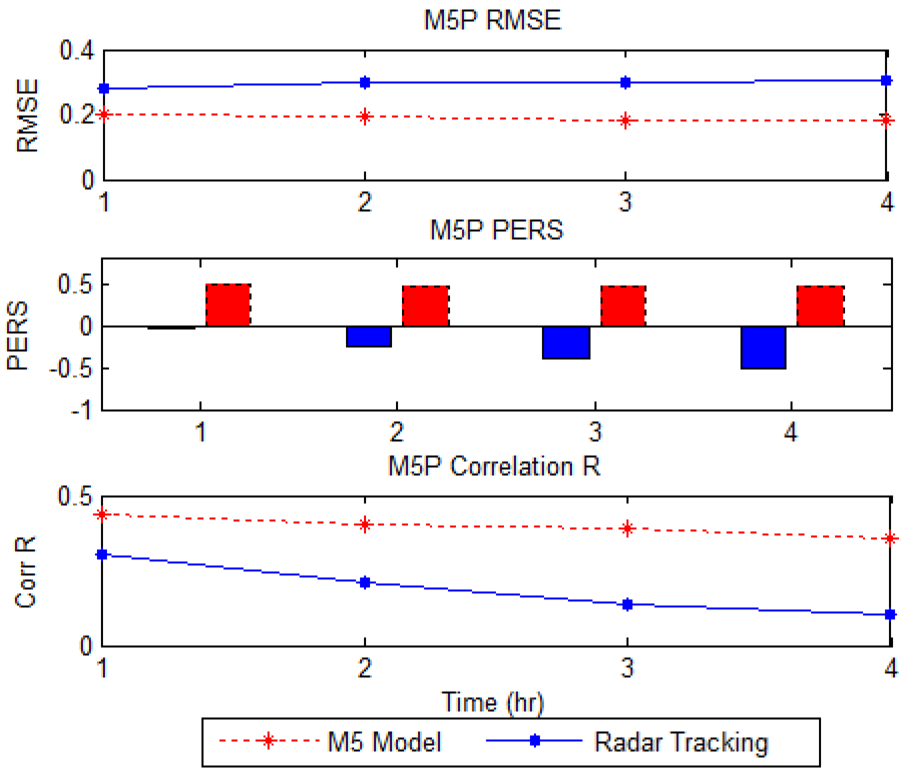


Figure 5-3 *Performance Comparison between Radar Tracking and M5 Model tree*

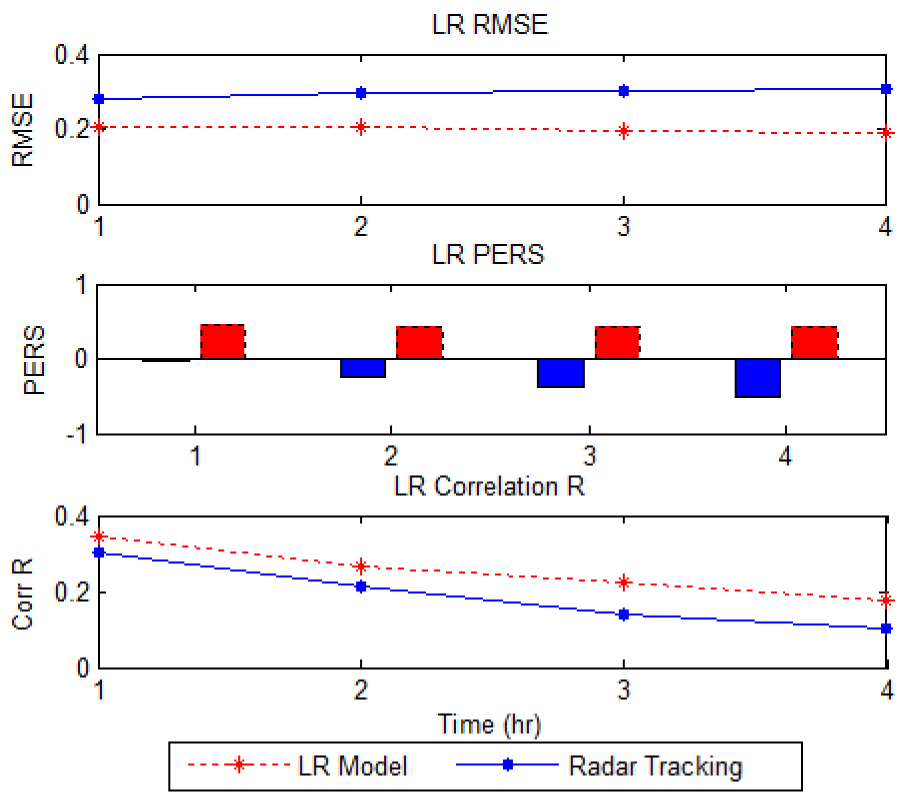


Figure 5-4 *Performance Comparison between Radar Tracking and linear regression model*

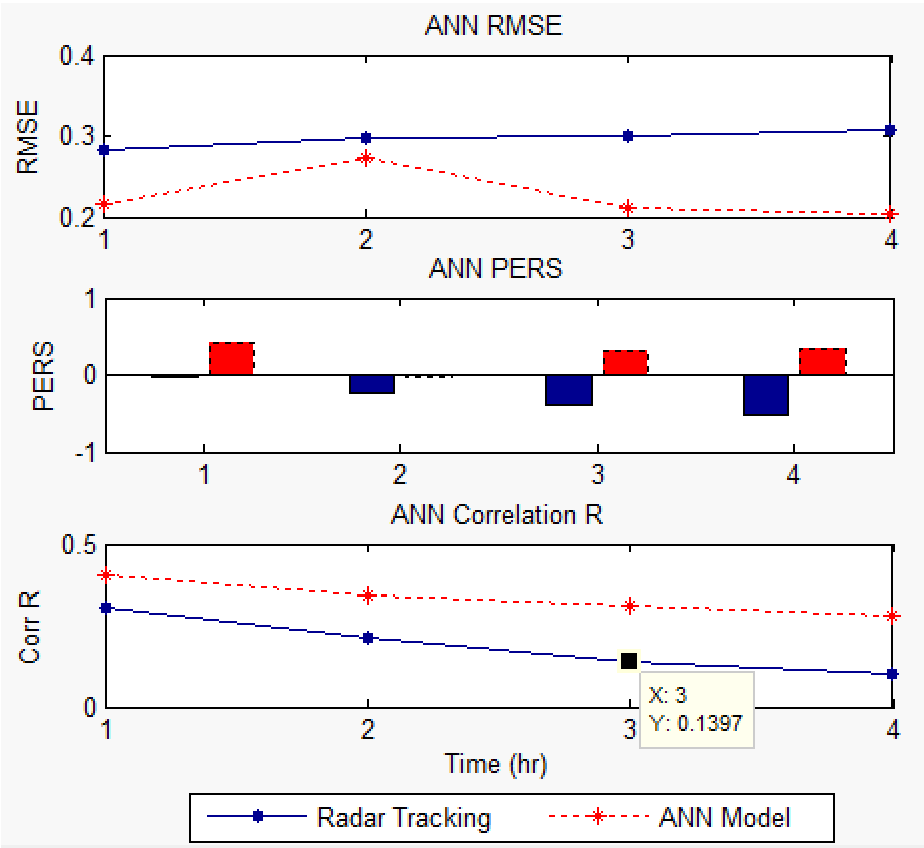


Figure 5-5 *Performance Comparison between Radar Tracking and ANN model*

Table 5.1 summarizes the skill score statistics, a comparison between rainfall prediction from DDM forecasts, radar tracking and MM5 model forecasts using a threshold of 0.1mm. The reason of using this threshold value was due to the low values of rainfall found through this event. The ideal forecasting system has a POD and a CSI of 1 with FAR values of 0.

Table 5-1 *Skill scores comparison between radar forecast, MM5 model and DDM Case 1*

hour	POD					FAR					CSI				
	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR
1	0.63	0.88	0.96	0.85	0.96	0.45	0.67	0.65	0.47	0.61	0.41	0.31	0.35	0.48	0.38
2	0.60	0.91	0.97	0.85	0.95	0.54	0.70	0.69	0.53	0.66	0.35	0.29	0.31	0.43	0.33
3	**0.59**	**0.91**	**0.98**	0.84	0.93	**0.59**	**0.73**	**0.73**	0.58	0.69	**0.32**	**0.26**	**0.27**	0.39	0.30
4	0.53	0.92	0.98	0.68	0.88	0.66	0.75	0.74	0.59	0.73	0.26	0.24	0.26	0.34	0.26

*Radar Tracking

The results show a better skill in detecting rainfall (POD) from 1 to 4 hours forecasts using three DDM and comparing with radar tracking and MM5 model. For instance, for 3 hours forecast the POD of the radar tracking and MM5 model are 59% and 91% respectively whereas with ANN model the value obtained was 98%.

The FAR values of the DDM in general are similar to those obtained for radar tracking and MM5 model as well. Nevertheless, it was found some high values and the reason is that the MLP in the normalization procedure assigned a small amount of rain value to those cells with zero rainfall value so the probability of obtaining a false alarm radio increased. The values of the CSI score which considers hits, misses and false alarms are better using model trees and represents the 17% of the improvement for the 1st hour forecast.

Figure 5-6 shows the RMSE and correlation coefficient (R) for a Modular Model Networks (MMN). The results show that the optimum number of clusters found where 3. The model trained with four clusters shows a reduction in terms of RMSE (7%) compared with the radar tracking forecast. Using 3 clusters there was an improvement in the correlation coefficient of 26%.

Figure 5-7 depicts a clusters plot using K-means algorithm and sq Euclidean distance metric. Cluster 3 includes a high number of data points which have high values of rainfall. Cluster 2 has low values of rainfall whereas cluster 1 has no uniform values of rainfall.

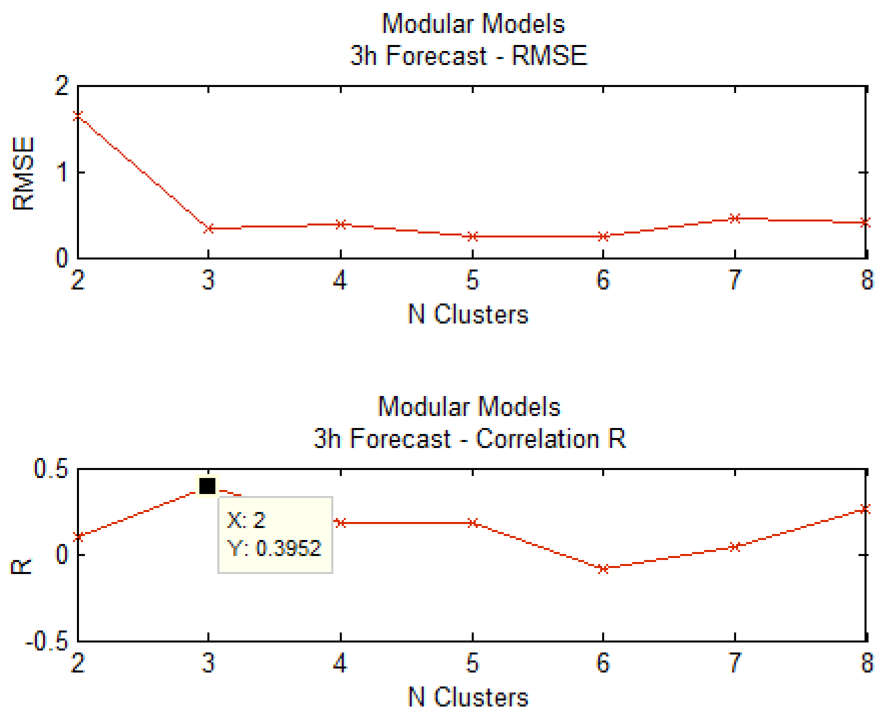


Figure 5-6 *RMSE and correlation Coeff. (R) for a Modular Model Network (MMN) 3hours forecast and 8 clusters*

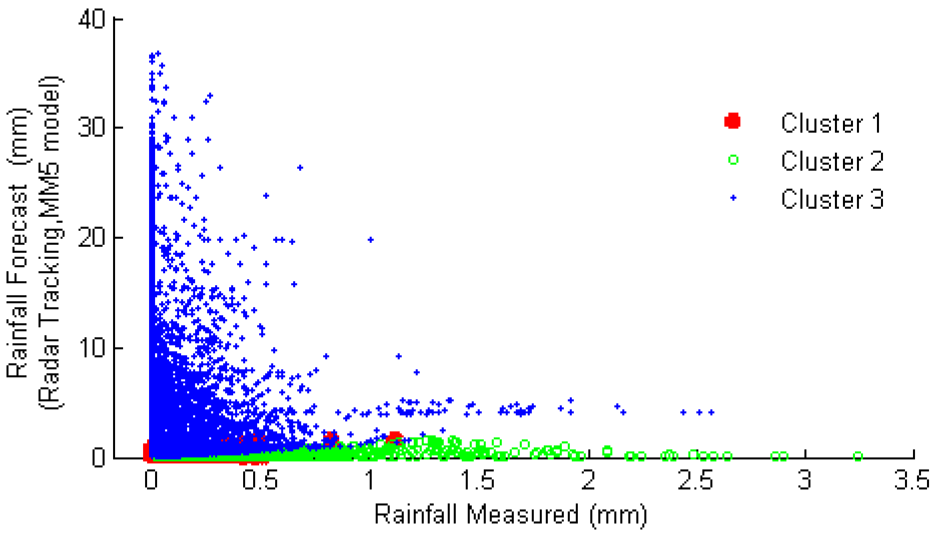


Figure 5-7 *Clusters plot using K-means algorithm and sq Euclidean distance metric*

5.2 Case 2: Rainfall forecast using Data driven models (MM5 online update, extended data set)

As was described previously on chapter 4 this case was built with two different data sets: the models were run firstly training the whole data of the five events and after that using the storm events 27 May, 28 May, 24 June 2007 and 08 September 2008 for training and the event of 05 September for validation.

Tables 5-2 and 5-3 show the performance comparison between radar tracking forecast, MM5 model and DDM with the whole data training and validation data respectively.

From table 5-2 it is possible to deduce the improvement in terms of RMSE over the radar tracking and MM5 model in around 11% of reduction for 4 hours forecast using model trees. The MM5 model shows the highest values of error with a maximum of 4.41 four hours ahead.

On the other hand, using validation data table 5-3 shows that the RMSE values of DDM are similar to those obtained by radar tracking forecast. The MM5 model presents lower values compared with those using whole data training.

The correlation coefficient (R) for this case was improved using the whole data training and it presents similar values using validation data. The correlation coefficient 4 hours ahead using whole data training was improved in 17% using model trees.

Table 5-2 *Performance Comparison between Radar forecast, MM5 model and DDM (whole data training)*

hour	RMSE					R				
	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR
1	0.43	2.66	0.40	0.39	0.41	0.45	-0.09	0.48	0.52	0.46
2	0.46	3.19	0.40	0.37	0.41	0.36	-0.02	0.44	0.54	0.36
3	0.46	3.73	0.50	0.38	0.42	0.35	0.00	0.41	0.55	0.35
4	**0.46**	**4.41**	0.42	**0.35**	0.40	**0.30**	**0.00**	0.39	**0.57**	0.30

Table 5-3 *Performance Comparison between Radar forecast, MM5 model and DDM (validation)*

hour	RMSE					R				
	R.T*	MM5	ANN	M5	LR	R.T*	MM5	ANN	M5	LR
1	0.30	0.96	0.32	0.36	0.30	0.80	0.25	0.80	0.61	0.78
2	**0.52**	1.48	**0.51**	0.56	0.52	**0.53**	0.55	**0.50**	0.30	0.47
3	0.61	1.91	0.67	0.64	0.65	0.47	0.39	0.47	0.35	0.45
4	0.70	2.45	0.77	0.73	0.75	0.37	0.38	0.31	0.23	0.30

*Radar Tracking

Figures 5-8 to 5-13 show the images of the actual radar rainfall and the forecasted rainfall obtained from radar tracking, MM5 model, Model trees, ANN and Linear regression for 3 hours forecast. As shown, in general the storms can be classified as a stationary convective system as the overall cloud and precipitation patterns are mostly round in shape. The figures also show their structures in such away that storms repeatedly move over the same area and the system exhibits little or no movement for the four hours forecasted.

In figures 5-8 and 5-9 with validation data and whole data training respectively is shown the forecast made using ANN. The model captured well the precipitation system observed by the radar in special the stationary small storm (supercell) persisted in the south part of the image which eventually was overtaken by a convective line which moved from the east.

Figures 5-10 and 5-11 show the forecast made with Model trees, the model fairly reproduces the features of both the stationary small storm described above and the rain band or the precipitation structure associated with the rainfall area and observed by the radar. There is a displacement of the precipitation formation from the north part of the model trees results compared with the radar observation.

Figures 5-12 and 5-13 depict the forecast made with a linear regression model. Unlike the other two models presented (ANN and M5) it could be considered as a frontal-type flash flood event (Chappell, 1986). There existed a quasi-stationary frontal boundary which remained nearly stationary for the four hours period forecasted.

With respect to conditional probabilities Tables 5-4 and 5-5 contain the skill scores statistics for radar forecast and DDM for whole training data and with validation data using a threshold of 0.2 mm.

Table 5-4 *Skill score statistics for Radar tracking, MM5 model and DDM Case 2 (whole data training)*

hour	POD					FAR					CSI				
	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR
1	**0.72**	**0.44**	**0.87**	0.63	0.73	**0.04**	**0.40**	**0.04**	0.05	0.04	**0.71**	**0.34**	**0.84**	0.61	0.71
2	0.63	0.67	0.95	0.50	0.56	0.12	0.39	0.51	0.34	0.13	0.58	0.48	0.49	0.40	0.52
3	0.58	0.84	0.32	0.45	0.41	0.20	0.40	0.24	0.36	0.24	0.51	0.53	0.29	0.35	0.36
4	0.50	0.95	0.13	0.30	0.21	0.26	0.37	0.21	0.20	0.22	0.53	0.61	0.12	0.28	0.20

Table 5-5 *Skill score statistics for Radar tracking, MM5 model and DDM Case 2 (Validation)*

hour	POD					FAR					CSI				
	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR	R.T*	MM5	ANN	M5P	LR
1	0.59	0.27	0.71	0.73	0.64	0.28	0.79	0.35	0.36	0.31	0.48	0.14	0.51	0.52	0.50
2	**0.46**	**0.49**	**0.60**	0.73	0.57	**0.40**	**0.70**	**0.45**	0.51	0.43	**0.36**	**0.23**	**0.40**	0.42	0.39
3	0.39	0.67	0.17	0.59	0.46	0.55	0.70	0.37	0.55	0.56	0.26	0.26	0.15	0.34	0.29
4	0.33	0.79	0.22	0.41	0.36	0.62	0.72	0.46	0.48	0.64	0.22	0.26	0.19	0.30	0.22

*Radar Tracking

The three data driven models perform significantly better than radar tracking and MM5 model in both cases. The probability of detection (POD) is improved over radar tracking and MM5 model in around 15% and 43% respectively; using ANN model for 1 hour forecast with whole data training and 27% 2 hours forecast using validation data, which demonstrate the ability of the DDM to predict rain versus no-rain areas. The values of false alarm ratio (FAR) are lower than those from MM5 model and radar tracking using validation data. The critical success index (CSI) has been improved in general terms in around 10%.

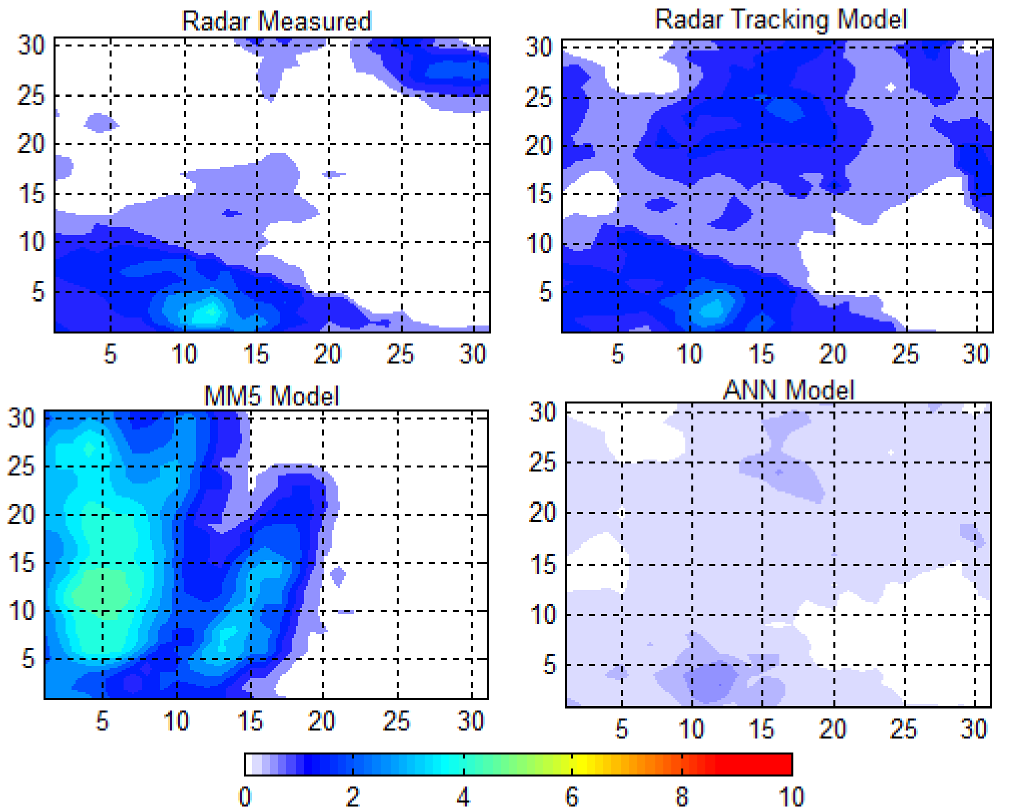


Figure 5-8 *Rainfall prediction (mm) validation data ANN 3 hours ahead 05/09/2008 22:00hrs*

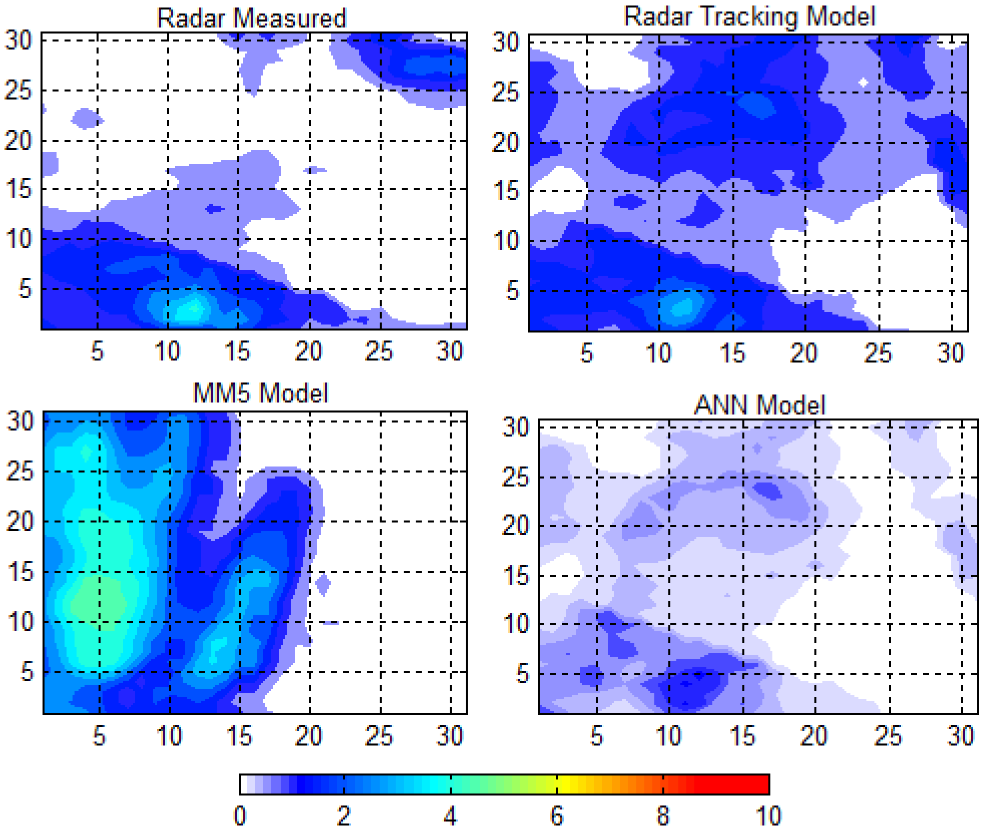


Figure 5-9 *Rainfall prediction (mm) whole data training ANN 3 hours ahead 05/09/2008 22:00hrs*

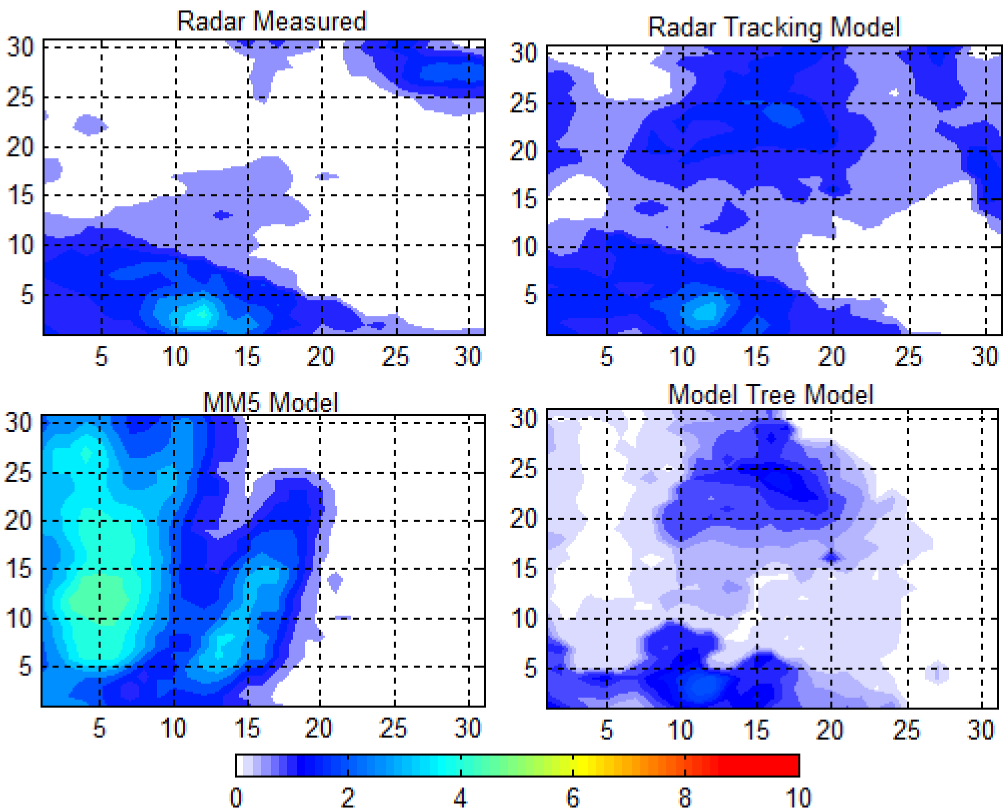


Figure 5-10 *Rainfall prediction (mm) M5 validation data 3 hours ahead 05/09/2008 22:00hrs*

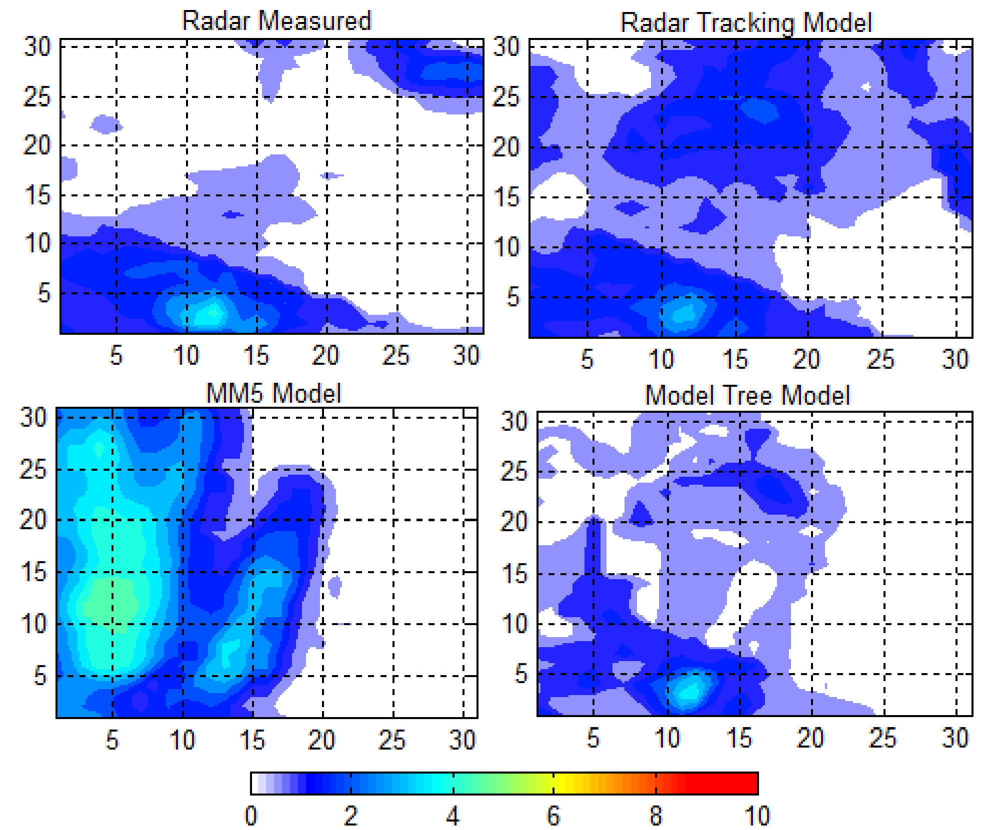


Figure 5-11 *Rainfall prediction (mm) M5 whole data training 3 hours ahead 05/09/2008 22:00hrs*

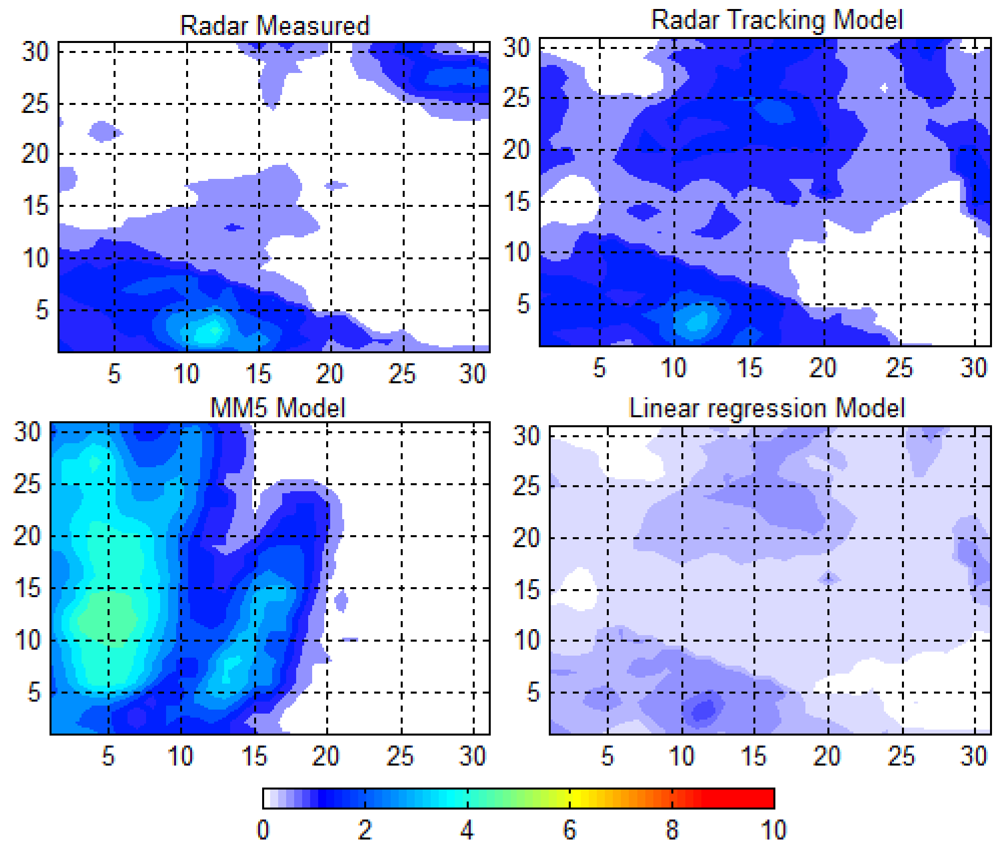


Figure 5-12 *Rainfall prediction (mm) LR validation data 3 hours ahead 05/09/2008 22:00hrs*

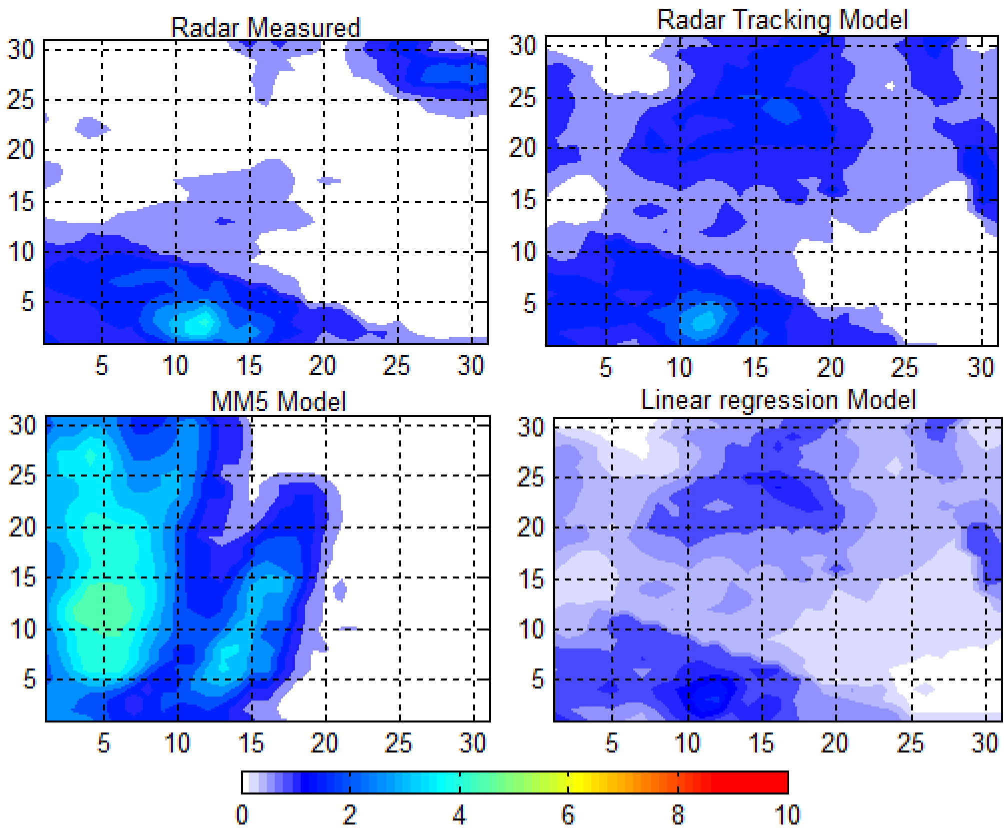


Figure 5-13 *Rainfall prediction (mm) LR whole data training 3 hours ahead 05/09/2008 22:00hrs*

5.3 Case 3: Using rule based forecast (MM5 online update, extended data set)

Tables 5-6 and 5-7 show the results achieved by the rainfall classifier.

Table 5-6 *Confusion matrix for training data*

Forecast 1h	Observed 1h: No	Low	High	Total
No	12884	0	0	12884
Low	0	2424	68	2492
High	0	0	0	0
Total	12884	2424	68	15376
% Correct	83.79	15.76	0.44	100

Forecast 2h	Observed 2h: No	Low	High	Total
No	12733	0	0	12733
Low	0	2595	48	2643
High	0	0	0	0
Total	12733	2595	48	15376
% Correct	82.81	16.88	0.31	100

Forecast 3h	Observed 3h: No	Low	High	Total
No	13132	0	0	13132
Low	0	2215	29	2244
High	0	0	0	0
Total	13132	2215	29	15376
% Correct	85.41	14.41	0.19	100

Forecast 4h	Observed 4h: No	Low	High	Total
No	13411	0	0	13411
Low	0	1953	12	1965
High	0	0	0	0
Total	13411	1953	12	15376
% Correct	87.22	12.7	0.08	100

Table 5-7 *Confusion matrix for validation data*

Forecast 1h	Observed 1h: No	Low	High	Total
No	2285	0	0	2285
Low	0	1557	2	1559
High	0	0	0	0
Total	2285	1557	2	3844
% Correct	59.44	40.50	0.05	100

Forecast 2h	Observed 2h: No	Low	High	Total
No	2161	0	0	2161
Low	0	1681	2	1683
High	0	0	0	0
Total	2161	1681	2	3844
% Correct	56.22	43.73	0.05	100

Forecast 3h	Observed 3h: No	Low	High	Total
No	2206	0	0	2206
Low	0	1626	12	1638
High	0	0	0	0
Total	2206	1626	12	3844
% Correct	57.39	42.3	0.31	100

Forecast 4h	Observed 4h: No	Low	High	Total
No	2095	0	0	2095
Low	0	1719	30	1749
High	0	0	0	0
Total	2095	1719	30	3844
% Correct	54.5	44.72	0.78	100

The highlighted numbers in tables 5-6 and 5-7 indicate the number of each class which has been estimated correctly. For instance, for three hours forecast in table 5-7 it is possible to see that 2206 out from 3884 samples were correctly classified as No Rain with 57.39% whereas 1626 out from 3884 were correctly classified as Low Rain with 42.3% but the 12 samples left were incorrectly classified as a Low Rain as they should have been classified as a High Rain. There was a poor performance of the model to forecast High Rain due to the short amount of high rain both forecasted and observed through the five events.

An analysis of Low Rain and High Rain was carried out in order to know the performance of the model. Real values of MM5 model and radar tracking were used as inputs and they were compared with the observed classes finding similar results than the MM5 and radar tracking inputs using classes previously described. In general, a rule based forecast outperforms the MM5 model and radar tracking method; this can also be illustrated graphically as it has been done in Figure 5-14 with the validation of the model. There is a probability of almost 80% to predict rainfall 3 hours ahead compared with 30% of probability from MM5 model and 20% using a radar tracking forecast.

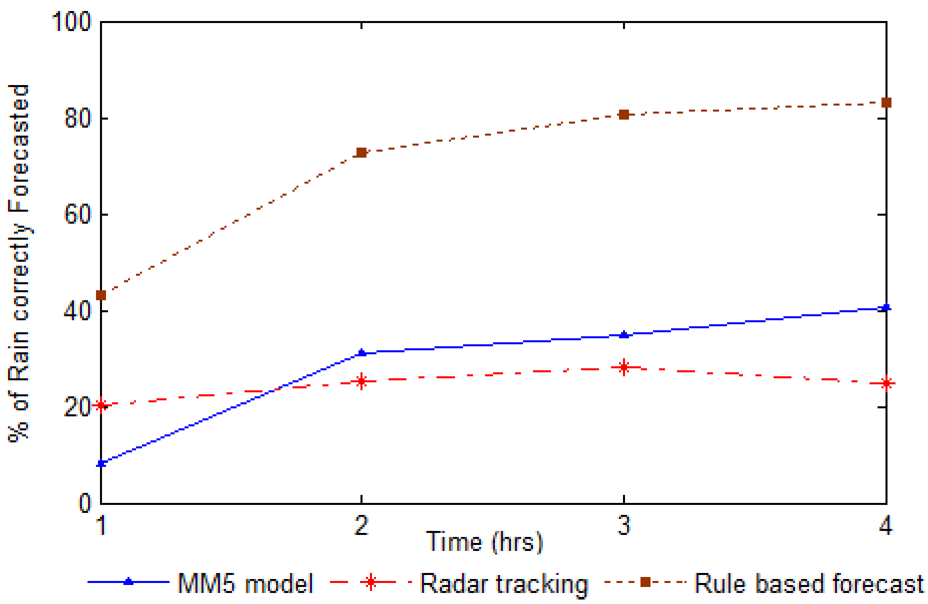


Figure 5-14 *Percentage of Rain correctly forecasted using a Rule based forecast*

6 CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary and conclusions

This thesis had as a purpose to explore the combination of rainfall information from weather radars and numerical weather model outputs (MM5 model) using data from South-east of England. The models were combined using data driven modelling techniques (DDM). The motivation of using a hybrid modelling approach was to combine the best features of both radar rainfall information and numerical weather prediction model (MM5) through the integration of their outputs.

Five different data driven models were used to predict rainfall 1, 2, 3, and 4 hours ahead namely model trees (M5), linear regression model (LR), artificial neural networks (ANN) and modular neural networks (MNN). The results of those models have been compared with radar tracking forecast and MM5 model forecast.

The study was carried out firstly analysing one single event 27 May 2007. The information contained in the MM5 model configuration 24 hours continuous simulation has been valuable for complementing a radar tracking technique as it helped in reducing the magnitude of the error forecasts. It was found that the integration using ANN, M5, LR were better able to reduce the RMSE values for all the forecast steps in around 30% if compared with radar tracking forecast technique. In addition, using MNN it was found an improvement in the correlation coefficient of 26% and a reduction in terms of RMSE 7%.

Nevertheless, one single event was not enough to conclude this research. Therefore, a new MM5 model configuration was done every 6 hours to produce 6 hours forecast using extended data sets. It was found the improvement in terms of RMSE over the radar tracking and MM5 model in around 11% of reduction using the whole data in the training process. Afterwards, it was obtained similar values of RMSE compared with radar tracking training the models leaving one event for validation. Consequently, it is possible to improve the models performance by adding more data (events) for training and validation.

Images of the actual radar rainfall and the forecasted rainfall obtained from radar tracking, MM5 model, M5, ANN and LR for 3 hours forecast were produced using five events. As a conclusion the storms were classified as a stationary convective system as the overall cloud and precipitation patterns were mostly round in shape. They also showed their structures in such away that storms repeatedly move over the same area and the system exhibited little or no movement for the four hours forecasted. However, an accurate analysis of those images requires a better understanding of the storm motion and its evolution.

Performance measures were carried out focusing on the relationship between forecasts and observations such as hit rates (POD), false alarm ratio (FAR) and critical success index (CSI). It was found that data driven models perform significantly better than radar tracking and MM5 model. The probability of detection (POD) was improved over radar tracking and MM5 model in around 20%. FAR in the case of ANN increased due to the normalization process as it assigned a small amount of rain value to those cells with zero rainfall values.

As a final stage in this research, a rule based forecast was carried out based on classifying rainfall in three classes (No Rain, Low Rain and High Rain). As a result there is a probability of almost 80% to predict rainfall per cells 3 hours ahead compared with 30% of probability from MM5 model and 20% using a radar tracking forecast for the five events studied. It was not possible to show any improvement predicting High Rain due to the low amount of high rainfall values found through the events.

The use of data-driven models as integrators has shown the capacity to improve and to complement the information content in weather radar and NWP outputs in rainfall prediction.

In this research the use of data sampling was an important step in order to build data-driven models. However, it might be vulnerable to sampling error because the randomness of selection may result in over represent or under represent low or high rainfall values. This selection can be improved through optimization methods.

The integration of weather radar information and MM5 model outputs using DDM without sampling shows an improvement of the models in terms of performance. It is possible to improve the process by adding seasonal data to the DDM.

The results of predicting rainfall using a rule based forecast were satisfactory in a simplified model approach. Nevertheless, the rainfall classifier presented problems to predict High rain events due to the large amount of low rainfall values.

In general, it can be concluded that a hybrid approach to combine a data oriented radar information with any numerical weather prediction based on physics using DDM can be an important addition to improve rainfall forecast.

6.2 Recommendations

It is important for future studies on this topic, to include as an input variable to the DDM wind vector components as this variable influence rainfall. Furthermore, a more detailed study of the configuration of the MM5 model is needed and to do a better analysis of its outputs and compare them with those obtained from weather radars.

Similarly, a more extensive study of the data partition or amount of data for training and validation is still required, for some events as it was stated above were not possible to get a good performance of the models in a specific region. Hence, it is reasonable to consider alternative number of events for different seasons. In term of errors they should be analyzed at different scales.

Another useful suggestion is to include an identification of rainfall regimes, duration of rainfall per event and extend the amount of data using at least one month period of events. In this sense, it is possible to get a better performance of any data driven model having a remarkable different between multiple rainfall classes.

In this study, Model trees, ANN, Linear regression model and MNN were chosen. However, there are other methods in machine learning and used in data driven modelling that might be used as a complement of this study such as naïve Bayes classifier, support vector machine and fuzzy logic and to compare its performance with the previous mentioned.

It is suggested to explore the implication of the spatial and temporal rainfall forecast of those results by means of an ensemble in a simple distributed hydrological model and to study the distribution effect of rainfall forecast at the catchment scale.

Ultimately, this hybrid approach can help as the basis for future researches aiming to improve in a better way, rainfall prediction. Therefore, this tool is useful especially to assess flash floods. High predictability of rainfall will make hydrological models more reliable in terms of decision making.

7 REFERENCES

Abebe, A., J., 2004. Information theory and artificial intelligence to manage uncertainty in hydrodinamic and hydrological models. PhD Thesis, UNESCO-IHE Institute for water education.

Abonyi, J. and Feil, B., 2007. Cluster analysis for data mining and system identification.

Arakawa, A. and Lamb, V.R., 1981. A potential entropy and energy conservating scheme for the shallow water equation. Monthly Weather review(109): 18-36.

Bellerby, T., Todd, M., Kniveton, D. and Kidd, C., 2000. Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network. Journal of Applied Meteorology, 39(12): 2115-2128.

Bhattacharya, B., 2005. Learning from data for aquatic and geotechnical environments. PhD Thesis, UNESCO-IHE Institute for water education.

Bhattacharya, B., Price, R.K. and Solomatine, D.P., 2007. Machine learning approach to modelling sediment transport. Journal of Hydraulic Engineering, 133(4): 440-450.

Bowler, N., Pierce, C. and Seed, A., 2006. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. Quarterly Journal of the Royal Meteorological Society 132: 2127-2155.

Bras, R.L., 1989. Hydrology: An introduction to hydrologic science.

Breiman, L., Friedman, J., Olshen, R. and Stone, C., 1984. Classification and Rregression Trees 1, Chapman and Hall/CRC.

Chappell, C.F., 1986. Quasi-stationary convective events, Mesoscale Meteorology and Forecasting. American Meteorological Society.

Cluckie, I.D. and Rico-Ramirez, M.A., 2004. Weather radar technology and future developments. GIS and remote Sensing in Hydrology. Water resources and Environment. IAHS, 289.

Cluckie, I.D., Rico-Ramirez, M.A. and Xuan, Y., 2006. An experiment of rainfall prediction over the Odra catchment by combining weather radar and numerical weather model. 7th International conference on Hydroinformatics HIC 2006, Nice, France.

Collier, C.G., 1994. Objective methods of forecasting precipitation using weather radar data Advances in radar hydrology. Proceedings of the international workshop held in Lisbon, Portugal. Nov. 1991.

Collier, C.G. and Hardaker, P.J., 1996. Estimating probable maximum precipitation using a storm model approach Journal of Hydrology 183: 277-306.

Corzo, G., 2009. Hybrid models for hydrological forecasting: Integration of data-driven and conceptual modelling techniques. PhD Thesis, UNESCO-IHE Institute for Water Education, Delft The Netherlands.

Corzo, G. and Solomatine, D., 2007a. Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. Hydrological Sciences Journal, 52(3).

Corzo, G. and Solomatine, D., 2007b. knowledge-based modularization and global optimization of artificial neural network models in hydrological forecasting. Journal Neural Networks, 20: 528-536.

Da Silveira, R.B. and Holt, A.R., 2001. An Automatic identification of clutter and anomalous propagation in polarization-Diversity weather radar using neural networks. IEEE transactions on geoScience and remote sensing 39(8): 1777-1788.

Dudhia, J., 1993. A Nonhydrostatic version of the Penn State-NCAR Mesoscale model: Validation Tests and simulation of an Atlantic Cyclone and Cold front. American Meteorological Society, 121(5 May 1993).

Dudhia, J., Gill, D., Manning, K. and Bruyene, C., 2003. PSNU/NCAR Mesoscale modelling system tutorial class notes and User's Guide: MM5 modelling system version 3. . Mesoscale and Micro scale Meteorology Divison, National Centre for Atmospheric research

Ganguly, A.R., 2002. Distributed quantitative precipitation forecasts combining information from radar and numerical weather prediction model outputs. Massachusetts Institute of Technology.

Golding, B., 1998. Nimrod: A system for generating automated very short range forecast. Meteorological Applications, 5(01): 1-16.

Haifa, I.T., 2008. A coupled model tree - genetic algorithm scheme for flow and water quality predictions in watersheds. Journal of Hydrology, 349(3-4): 364-375.

Haykin, S., 1999. Neural Networks - A comprehensive Foundation. Second Edition, Prentice Hall, New Jersey.

Jolliffe, I.T. and B., S.D., 2003. Forecast Verification: A practitioner's Guide in Atmospheric Science. john Wiley & sons, Ltd.

Joss, J. and Lee, G.W., 1993. Scan strategy, clutter suppression,calibration and vertical profile corrections.Preprints of the 26th radar meteorology conference. Amer.Meteor.Soc.: 390-392.

kompare, B., Steinman, F., Cerar, U. and Dzeroski, S., 1997. Prediction of rainfall runoff from catchment by intelligent data analysis. Acta hydrotechnica.

Kuusisto, E., 2006. Water management and water quality in Finland. Finnish Environment Institute.Rotterdam.

Lee, G.W., 2003. Errors in rain measurement by radar: effect of variability of drop size distributions. PhD Thesis, McGill University,Montreal, Canada.

Lu, D., Croft, P., Fitzpatrick, P. and Reddy, S., 2000. Data assimilation and model evaluation for MM5 and COAMPS. Jackson State University.

Marshall, J.S., Hitschfeld, W. and Gunn, K.L.S., 1955 Advances in weather radar. Advances in Geophysics(2): 1-56.

Maskey, S., Guinot, V. and Price, R., 2004. Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach. Advances in water resources, 27: 889-898.

McCullagh, J., 2005. A modular neural network arquitecture for rainfall estimation. 23rd IASTED International Multi- Conference. Artificial Intelligence and Applications.

Persson, A., 2003. User Guide to ECMWF forecast products, Meteorological Bulletin M3.2, ECMWF, (2003).

Petridis, V. and Kehagias, A., 1998. Predictive Modular Neural Networks. Applications to time series.

Pierce, C. et al., 2005. Use of stochasic precipitation nowcast scheme for fluvial flood forecasting and warning. Atmospheric science letters(6): 78-83.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 2007. Numerical Recipes: The art of scientific computing. Third edition.

Quinlan, J.R., 1992. Learning with Continuous classes. 5th Australian joint conference on Artificial intelligence, Singapore: 343-348.

Rico-Ramirez, M., Cluckie, I.D. and Gonzalez-ramirez, E., 2007a. An overview on Dual-polarization weather radars Encuentro de Investigacion en Ingenieria Electrica. Zacatecas Mex.

Rico-Ramirez, M.A. and Cluckie, I.D., 2006. Assessment of polarimetric rain rate algorithms at C-band frequencies.ERAD.

Rico-Ramirez, M.A., Cluckie, I.D., Shepherd, G. and Pallot, A., 2007b. A high-resolution radar experiment on the island of Jersey. Royal Meteorological Society.

Rico-Ramirez, M.A., Gonzalez-Ramirez, E. and Cluckie, I.D., 2008. On-line monitoring of weather radar antenna pointing using high-resoulution DTM and AI techniques.

See, L. and Stan, O., 2000. A hybrid multi-model approach to river level forecasting. Hydrological Sciences Journal, 45: 523-536.

Sharkey, A., 1999. Combining Artificial Neural Nets. Ensemble and Modular Multi-Net Systems.

Solomatine, D.P., 2007. Data-Driven Modelling: Machine learning, Data mining and knowledge discovery. Lecture notes. UNESCO-IHE Institute for Water Education. Delft, The Netherlands.

Solomatine, D.P., See, L.M. and Abrahart, J.R., 2008. Data-driven modelling: Concepts, approaches and experiences. Practical Hydroinformatics: Computational intelligence and technological developments in water applications.

Solomatine, D.P. and Xue, Y., 2004. M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai river in China. Journal of Hydrologic Engrg., 9(6): 491-501.

Todini, E., 2001. A bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. Hydrology and Earth systems Sciences., 5(2): 187-199.

UCAR, 2008. Understanding NWP models and their processes. Distance learning course. University Corporation for Atmospheric research. From http//www.meted.ucar.edu/.

UK, M.O., 2003. Rain radar products (NIMROD),[Internet]. British Atmospheric Data Centre, 2009. Available From http://badc.nerc.ac.uk/data/nimrod.

UK, N.W.S., 2007. Weather news. Available from www.metoffice.gov.uk.

Wilson, B., 2008. Induction of Decision Trees. School of Computer Science and Engineering. University of New South Wales.

Wilson, J.W., 2003. Precipitation Nowcasting: past, present and future. . Sixth International symposium on Hydrological applications of weather radar. National center for atmospheric research, Boulder Colorado.

WRF, 2008. The Weather Research and Forecasting Model, WRF model graphic tools RIP4 (version4.2). Available from http://www.mmm.ucar.edu/wrf/users/graphics/RIP4/RIP4.htm.

Xuan, Y., 2007. Uncertainty propagation in complex coupled flood risk models using numerical weather prediction and weather radars. Phd Thesis, University of Bristol, Uk.

Xuan, Y. and Cluckie, I.D., 2006. Uncertainty propagation in ensemble rainfall prediction systems used for operational real-time flood forecasting. Hydroinformatics in practice: Computational Intelligence and technological developments in water applications.

Zhang, M., Fulcher, J. and Schofield, R.A., 1997. Rainfall estimation using Artificial neural networks group. . Neurocomputing(16): 97-115.

Zhijia, L., Wenzhong, G., Jintao, L. and Zhao, K., 2004. Coupling between weather radar rainfall data and a distributed hydrological model for real-time flood forecasting. Hydrological Sciences Journal, 49(6): 945-958.

Appendix A

This appendix is an extension of chapter 2 and includes a brief summary of the MM5 model dynamics and their equations.

MM5 program is the numerical weather prediction part of the modelling system. Its equations are for the nonhydrostatic model's basic variables excluding moisture. The basic model equations can be written in this (x, y, σ) coordinate system after subtracting large cancelling terms from the vertical momentum equation. (Adapted from (Dudhia, 1993)):

Pressure:

$$\frac{\partial p'}{\partial t} - \rho_0 g w + \gamma p \nabla \cdot V = -V \cdot \nabla p' + \frac{\gamma p}{T}\left(\frac{\dot{Q}}{c_p} + \frac{T_0}{\theta_0} D_0\right) \quad (2.2)$$

Momentum (x-component):

$$\frac{\partial u}{\partial t} + \frac{1}{\rho}\left(\frac{\partial p'}{\partial x} - \frac{\sigma}{p^*}\frac{\partial p^*}{\partial x}\frac{\partial p'}{\partial \sigma}\right) = -V \cdot \nabla u + fv + D_u \quad (2.3)$$

Momentum (y-component):

$$\frac{\partial v}{\partial t} + \frac{1}{\rho}\left(\frac{\partial p'}{\partial y} - \frac{\sigma}{p^*}\frac{\partial p^*}{\partial y}\frac{\partial p'}{\partial \sigma}\right) = -V \cdot \nabla v + fu + D_v \quad (2.4)$$

Momentum (z-component):

$$\frac{\partial w}{\partial t} - \frac{\rho_0}{\rho}\frac{g}{p^*}\frac{\partial p'}{\partial \sigma} + \frac{g}{\gamma}\frac{p'}{p} = -V \cdot \nabla w + g\frac{p_0}{p}\frac{T'}{T_0} - \frac{gR_d}{c_p}\frac{p'}{p} + D_w \quad (2.5)$$

Thermodynamics:

$$\frac{\partial T}{\partial t} = -V \cdot \nabla T + \frac{1}{\rho c_p}\left(\frac{\partial p'}{\partial t} + V \cdot \nabla p' - \rho_0 g w\right) + \frac{\dot{Q}}{c_p} + \frac{T_0}{\theta_0} D\theta \quad (2.6)$$

From the equations above 0 denotes reference state and primes denote deviations from the reference state. Density and potential temperature are described by ρ and θ respectively, the heating due to diabatic processes is $\dot{Q}$ and sub grid-scale eddy terms are represented by D_ϕ. The constants g, f, R_d, c_p, and γ represent acceleration due to gravity, Coriolis term, gas constant for dry air, heat capacity at constant pressure of air and the radio of heat capacity for air at constant pressure.

The advection terms may be expanded for any variable A as follows:

$$V \cdot \nabla A = u\frac{\partial A}{\partial x} + v\frac{\partial A}{\partial y} + \dot{\sigma}\frac{\partial A}{\partial \sigma} \qquad (2.7)$$

Where the coordinate velocity used for advection $\dot{\sigma}$, is related to the model-predicted velocity components by:

$$\dot{\sigma} = -\frac{\rho_0 g}{p^*}w - \frac{\sigma}{p^*}\frac{\partial p^*}{\partial x}u - \frac{\sigma}{p^*}\frac{\partial p^*}{\partial y}v \quad (2.8)$$

The upper and lower boundaries are taken to be rigid surfaces so $\dot{\sigma} = 0$ is applied as the boundary condition at $\sigma = 0$ and $\sigma = 1$. The divergence term in 2.2 can be expanded in (x, y, σ) coordinates as:

$$\nabla \cdot V = \frac{\partial u}{\partial x} - \frac{\sigma}{p^*}\frac{\partial p^*}{\partial x}\frac{\partial u}{\partial \sigma} + \frac{\partial v}{\partial y} - \frac{\sigma}{p^*}\frac{\partial p^*}{\partial y}\frac{\partial v}{\partial \sigma} - \frac{\rho_0 g}{p^*}\frac{\partial w}{\partial \sigma} \qquad (2.9)$$

p^* is constant, the second and fourth terms on the right disappear.

GRAPH program generates simple diagnostics and plots for some standard meteorological variables. Its code provides simple vertical interpolation capability, cross-section figures and skew-T plots. RIP is a FORTRAN program that uses NCAR Graphics routines in order to visualize output from gridded meteorological data sets. It uses a control file to manage the plot it produces. These control files are also called "in" files as they need to have an extension ".in" in their file names. The control file specifies the variables to be plot, the plot type and background maps. The detail of how to prepare the control file can be found in the RIP package at: http://www.mmm.ucar.edu/mm5/documents/ripug_V4.html

List of Figures

List of Tables

Acknowledgements

First of all, I would like to acknowledge the Netherlands Fellowship Programme NUFFIC-NFP for giving me the opportunity to come here for my MSc. and providing me with the necessary financing.

I would also like to express my appreciation and respect for Dr. Yunqing Xuan from whom I received excellent guidance, support, nice suggestions and for having faith in me in every stage of this study. I am thankful also to Prof. Dimitri Solomatine our head of core for his continuous effort in maintaining the standard of our Hydro informatics programme and for his accurate comments and suggestions to this study.

I want to express my gratitude to Dr. Gerald A. Corzo P. who has been my day-to-day supervisor, as well as friend. He conceived the initial ideas and motivation for this thesis. He not only specially helped me in this research with his knowledge in data driven modelling but also he gave me his time at weekends with extra MATLAB classes and further discussions related to this study. I learnt a lot from you Gerald God bless you!

It is also my pleasure to express my gratitude to: Dr. Biswa Bhattacharya, Dr. Zoran Vojinovic and the entire Hydro informatics and Knowledge Management Department staff, not only because of their excellent educational labour, but also for their human support. Thanks to Dr Miguel A. Rico-Ramirez at the University of Bristol for the data sets provided to initialise the MM5 mesoscale model.

Extensive thanks also to Dr. Nelson Obregon Neira at National University of Colombia who provided me with all the basic knowledge in hydraulics and hydrology and opened my mind to the Hydro informatics world.

Also I would like to thanks Oscar Hernandez my classmate and friend for spending his time with me discussing and sharing his experience and interests in this field through our MSc. studies. Thanks to: my friend Alexandra Arevalo, my Colombians and Latin-American friends, my classmates in hydro informatics and to my colleagues at the computer help desk.

My gratitude is immense with UNESCO-IHE for all the financial support that has been offered to me via the ICT group. Thanks to Ineke, Claudia, Maria, Silvia and Marielle for their enormous help during my time at the Institute.

Finally and very important, I would like to thanks to my family for their prayers, love and spiritual support.

Carlos A. Martinez C.
Delft, The Netherlands
September 2009