On the derivation of vector fields from series of radar images and forecast rain fields

Chrysostomos-Spyridon Milas

Meteovista

2015

Date: 10/03/2015

Internship Report

Author: Chrysostomos-Spyridon Milas

Education: Master Climate Studies, Meteorology and Air Quality (MAQ), Wageningen University (WUR)

Wageningen University Supervisor: dr. LJM (Leo) Kroon

Company Supervisor (Meteovista): Joralf Quist

Code: MAQ-70824

Abstract

Perform a short-term weather forecast is a challenging task and many different attributes should be taken into consideration. The use of radar images can be used for such a research but this implies the application of many techniques to be combined in order to obtain the desirable result. One should comprehend well in advance those techniques and theoretically build a model that later on could be developed under a programming language. The model built for this project uses a combination of techniques and calculations and is developed in such a way that it can be put on a live platform to run in a time interval the developer will choose. This will serve as an informative tool for people who want to be informed about the upcoming weather situation over a specific area in a time interval of 5 or 15 minutes. The data used here to test the model come from different radars installed all around Europe for the specific date of 30th September 2014 from 10:15 am (GMT) until 11:00 am (GMT) with a time step of 15 minutes. A great advantage of this model is the fact that it can give the user the option not only to plot the data of his desire but also to zoom in and create a weather forecast on a local scale. Meteorologically speaking the results show an average agreement with the observations but the model is open for improvements on every step of it.

List of Figures

Summary

A model has been built in order to perform short-term weather forecast using a series of radar images. The model has the intention to identify rain clusters over a specified area using radar data under the time interval of 15 minutes and afterwards derive the displacement vectors for every rain cluster. Such a task requires the model to firstly correlate the clusters between themselves and afterwards create the displacement vectors at the position where each cluster is correlated with its own at a previous time step. The correlation is performed using the minimum Mean Absolute Error value between all possible positions and the location of this position defines the offset position of the rain clusters at the next time step. This is the way to perform a rain field forecast and test it at the same time. Between these two steps comes the interpolation of the vector field. When acquiring a vector field out of some data we need to interpolate the rest of the field in order to obtain vectors also at points where data are not available. Barnes scheme is used for this purpose and the interpolated field at the end shows towards where the rain field will propagate at the next time step of the analysis. The project performed here points out the striking difficulties that someone needs to overcome when dealing with such an analysis. Great importance should be given on the clustering method that will be chosen. In addition the interpolation of the vector field should be performed under a scheme that is not dependent on every data point and a choice of the more important ones should be done carefully. The research showed great dependency on the input data and the model is highly affected by the situation is trying to analyze. Weather systems that are fully developing have greater chance of being interpreted in a more accurate way than a calm situation especially when there is not much change on the rain intensity or on the enhancement/decay rate of rain clusters.

Table of Contents

1. Introduction

Performing a short-term forecast might generally be even more demanding and essential than performing a long-term one. This has to do with the fact that even though a long-term forecast might be useful and provide everyone with the needed information for the coming weather, a short-term one can be better in terms of practicality as it can inform people and businesses what the weather will look like in a shorter period of time and in greater detail (De Lannoy et al., 2005). The major difference between these two types of forecast can be seen in the fact that long-term forecasts can inform someone about the orientation towards where weather systems propagate but with less detail upon the exact location whether the short-term ones can indicate the exact locations where a weather system may have an impact and the exact timing (Schmid W., 2000). The most crucial drawbacks of a short-term forecast analysis are the big computational time and the complexity of the analysis itself. In order to produce a forecast with an interval of 15 minutes or either 5 minutes we need to have a model that runs under these time intervals so it can produce the forecast fast enough. Such forecasts can be done by using images taken from radars which are installed all around the world and are able to produce these images in a very short time range. Depending on the capabilities of each radar system, the images can be taken on a 5 minutes basis and thus it is very important for their analysis to be performed fast in order to have an accurate weather forecast. By using radar data though any analysis will experience difficulties because radars are dependent on the topography of the area for example (Nakakita et al., 1996).

The way to perform such a weather forecast is to analyze the radar images and ultimately derive the wind field from the propagation of the rain clusters. In that way and out of a series of radar images we can predict towards where the rain clusters will move identifying the areas which will be affected. Such a train of thought seems simple but the interpretation of the model which needs to be built is very demanding. There are many parameters that need to be taken into account and many thresholds to be applied if we want our results to be scientifically correct and accurate. The complexity of this research demands fast computational techniques and systems to be used. Programming is the only way to handle the amount of data needed for such a project and therefore the analysis was performed in PHP and by using Object Oriented Programming techniques (OOP).

Tackling the difficulties that would arise and were known from the beginning in addition with the high importance of such research were the reasons that motivated me the most in order to concern myself with this subject. It was well understood from the beginning that such a project would be half meteorological half programming and the combination of both fields made it even more intriguing. Therefore the following research question id formed and this report tries to addresses it and give an insight on further improvements and additions to be made in order to improve the result.

Can a series of radar images produce an accurate rain field forecast?

The structure of the report will be in such way that in the first chapter the theory and information about the concepts included in the analysis will be given. Afterwards, in the second chapter the methodology followed in the analysis will be introduced, explained and justified. In the third chapter the results will be presented mentioning in great details what can be done in the future and for further improvements. Finally the conclusions of the research will be mentioned followed by the last section which will be the general discussion on the topic.

2. Theoretical Analysis

The basic concept of this project is to analyze a series of radar images but before doing that we need to explain what they are and what they contain. This chapter will deal with the theory of radar images and rain clusters as long as with the reasoning behind using several techniques in the experimental analysis. These techniques are a vital step towards the successful completion of the analysis and their absence would cost the analysis its scientific credibility.

2.1 Radar Images

The use of radar images in order to perform a better weather forecast is widely used in today's meteorological society and it has multiple applications depending on their size, their covered area and their resolution. For this project specifically we are going to use radar images of size 620x700 pixels with each pixel representing a 5x5 km area. The area covered is mainly the northwest part of Europe and more in details the radar images are covering France, Belgium, Netherlands, Luxembourg, Germany, United Kingdom, Denmark, Switzerland and the northern parts of Spain and Italy as it can be seen in Figure 2.1 together with the western parts of Poland, Czech Republic and Austria. The radar images can be taken every 15 minutes and they show the evolution of the rain field above these areas. This is achieved through the depiction of the rain clusters that are formulated at each time step. Their propagation can be identified by observing the same rain clusters at another time step for example in the next radar image after 15 minutes. In that way the use of radar images is a useful diagnostic tool on how rain fields evolve through time and according to this evolution and their intensity we can identify which areas will be more affected and by how much.

Figure 2.1: Area covered by the radar image in use.

These radar images can be scaled down to dimensions of 413x413 in order to exactly fit the map shown in Figure 2.1 of northwest Europe leaving out the parts outside these borders. There is also the possibility to focus on specific regions or countries by defining the boundaries of the analysis domain. Nevertheless, the use of a radar image for this kind of research and purposes is convenient for the reason that we can choose on our own the exact size of the image we want to analyze and further study.

2.2 Rain Clusters

The areas in a radar image that experience rainfall can be identified by the existence of rain clusters above them or not. These clusters contain the rain intensity in mm/hour which can be also seen in their representation in different color intensity (darker parts of the clusters represent higher rain intensity and vice versa in Figure 2.2 and Figure 2.3). Due to the fact that some rain clusters are positioned really close to each other we are forced to include them in the same cluster area under study and this is achieved through the use of the recursive search for clusters method which is explained in the next section. The result of this procedure will be a rectangular area which will contain all these smaller rain clusters and some of their attributes will be also calculated such as the mean weight of the cluster, the dimensions, the absolute center and the weighted center of the cluster. With this thought in mind we can define the cluster area as the rectangular area which contains a cluster, even though this can be a group of smaller clusters with similar attributes, and exactly fits each cluster. Each of these rectangular areas extends from the western point to the eastern point of each cluster and the same applies in the north-south direction. By definition we are going to use the top left corner as the starting point of any calculation and it's going to be the reference point of every cluster, a decision which is going to be justified later. Since the desirable outcome of the analysis is a better rain forecast we need to consider the fact that even though each radar image will provide us with many rain clusters we can never be sure which cluster is correlated to which in a previous radar image in order to be sure how the rain field propagates. Generally speaking by observing a radar image we can identify over which areas there is an active rain field and what its intensity is. By getting two successive radar images though and correlating them we can investigate the evolution and propagation of a rain field and due to the fact that the radar images are taken every 15 minutes we can provide everyone with a short term rain forecast.

Figure 2.2: Representation of rain clusters derived from a radar image.

Figure 2.2 is a depiction of how these cluster look like for the area described before and the date of this situation is the 30th of September 2014 at 10:30 am (GMT) whilst on Figure 2.3 we can see the exact position of these rain clusters and which areas are influenced by their existence. This is a typical representation of a radar image and consecutive images like the one presented in Figure 2.3 will lead us to the desirable outcome.

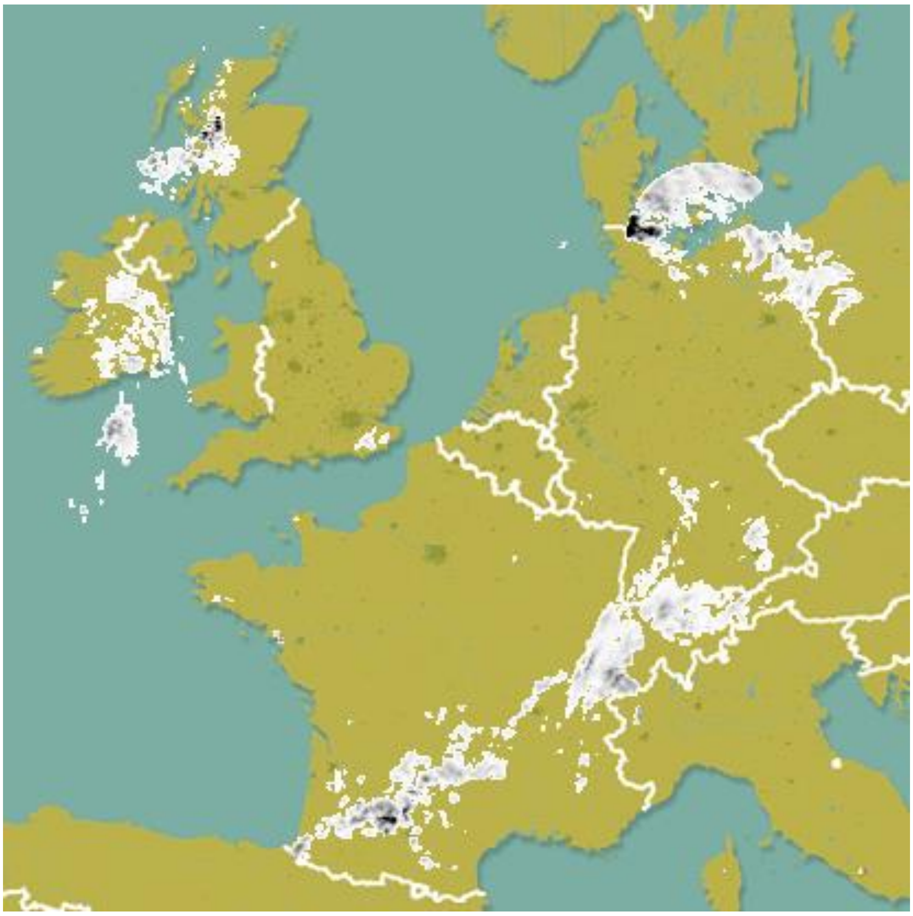

Figure 2.3: Rain clusters over the area under study.

2.3 Clustering Methods

Clustering is essential when dealing with identified areas experiencing rainfall with the use of radar images. When a radar scans its area of influence, it will identify all the points where rainfall is present. At the end all these points need to be clustered and basically create the cloud formations which can be seen for example in Figure 2.3. There are several methods when it comes to clustering, a procedure which connects and groups all rain clouds with similar properties and consider them as one. The translation model is one of them with a relatively easy mathematical framework to be implemented and being flexible in identifying movements of horizontal rainfall distribution. It has a pretty high evaluating accuracy in predicting rainfall patterns as its mathematical framework is rather simple and easy to implement (Takasao et al., 1994). On the other hand using such a model does not take into account any possible growth or decay of rain cells (KIM et al., 2006) which are usually the case and in addition with the fact that not many topographic information are included (KIM et al., 2006) it makes it a rather unreliable model to use.

The use of the Autocorrelation/Cross correlation model might be computationally heavy and with a rather sophisticated mathematical framework but it offers a real advantage when it comes to credibility as it is a very straightforward model and theoretically motivated (Peura and Hohti, 2004). To apply this kind of model someone needs to possess also knowledge of optical flows procedures which is not essential to go into details for this project. That makes it an even more complicated model and it should be taken into consideration before deciding whether it's a rational decision to implement it or not.

One of the most used methods of clustering, in fields not only of meteorology, is the K-means clustering one. The method can be basically used for identification of clusters at different scales (Lakshmanan et al., 2003). Its main attribute is that it groups together points that have similar properties under the nonstop iterating over the total amount of points until convergence is reached. The difficulty about implementing this method lies on the fact that the more data point are observed the larger the computational time will be. All observation points must be separately studied in order to find in which cluster they belong.

Finally a method that can be used, which is more of a method depended on the user's willingness to define the searches' attributes, is the recursive search for clusters which has better application to arbitrary distribution of data points on a grid and especially when the inter spacing distance is large. The method will continue searching for data points to be included in the same cluster under certain conditions defined by the user until convergence is reached once again.

One of the most important features the model we are going to use must have is to be suitable for short term weather forecast. Thus it becomes important to know in advance which one best suits our needs in regard to computational time and which one has the best performance for short term analysis. The translation model is not performing very good at time scales greater than an hour (Takasao et al., 1994) whereas the autocorrelation/cross correlation model should be preferred when forecast of time scales of 2 to 3 hours want to be performed (Peura and Hohti, 2004). For now it seems that K-means clustering method is suitable for different time scales forecast (Lakshmanan et al., 2003) but the application of a recursive search for clusters has no limitations when it comes to time scales as it is more of a method which is time independent.

2.4 Correlation of Clusters

One of the most important steps in this kind of analysis is the correlation of clusters between different time steps. When different radar images are observed then it is an easy task for the user to identify which cluster correlates to which at a different time step. But when using a model for any kind of analysis this task must be performed also from the model. This procedure will be performed after the determination of rain clusters and the identified clusters in the radar image need to be compared only with the same clusters at the previous time step. In order for the model to realize and identify which one is the same in the previous radar image a correlation technique must be applied. The correlation can be achieved with the calculation of Mean Absolute Error (hereafter MAE) between the clusters at the different time steps. The technique will be explained in a following paragraph but it is worth mentioning that it is an empirical method which is widely used in every statistical or not problem and it is expected to connect the clusters between them.

2.5 Interpolation

A derived vector field in general should be depicted in all grid points so it is visible towards where and by how much a weather system propagates. Sometimes though the data used in a research may not be equally distributed on a grid. The necessity to interpolate the vector field in such cases becomes important when using radar images because then we can only get one displacement vector for each cluster. The fact that some radar images exclude some points in the grid without representation makes it essential to interpolate. By interpolating we can obtain an approximation of values which will serve as the missing data points and only in that way the vector field can be complete. If no interpolation scheme is used then the vector field will only depict the movement from each cluster and such a depiction probably will be chaotic because each cluster has the capability to propagate towards every possible direction especially when smaller intervals are taken into account. The final result we want to reach includes the forecast of a rain field and because of the fact that we are getting a wind vector for every cluster in a radar image consequently we can only get a vector field which is not complete. For this reason and also in order to smooth the wind field by ignoring small chaotic movements, we need to interpolate the whole field and calculate the vectors in positions where there are no clusters present. Strictly speaking this is going to be an approximation and not a representation of the real propagation of the rain clusters but it will give the necessary information towards our goal i.e. forecast the whole rain field. Depending on the type of data, grid structure, magnitude of the variables generally there are many schemes (David L. et al., 2002) and techniques that can be used and one of them is Barnes scheme which will be explained later on.

3. Methodology

This chapter includes the methodology that is going to be followed in this project step by step and an explanation of each one of these steps. Several techniques should be applied regarding different problems that need to be tackled according to the theory. The general flowchart of the model is presented in the next section where more details on the correlation of clusters and interpolation of the vector field can be found.

3.1 Flowchart of the Model

The model we are going to use needs some specific requirements and in order for it to be neat and easily adaptable it will be designed under the flowchart presented in Figure 3.1. The whole model will be divided in different classes in PHP and each of them will be responsible for a different task. Therefore the classes which are going to be used will be responsible for reading the data, the mathematical framework and the interpretation of the outcome, beside other tasks. The logic we will follow requires that firstly the data (rain intensity) should be read and from them the radar images to be created. The radar images will contain the rain clusters with the rain intensity included. For that reason a class that determines the position of each cluster will be used. Additionally by using the method of the recursive search of the clusters and by applying the MAE calculation we will be able to correlate the clusters with themselves in a previous time step. This step will require the implementation of a method which will perform the calculation of the MAE. After the completion of this step the vector field will be derived from the radar images and the visualization will be made by using Barnes interpolation scheme. The final step will be the forecast of the rain field using the calculated distance which the correlated clusters have covered. The performance of the model at the end will be tested by comparing the actual situation with the one we obtained from the analysis.

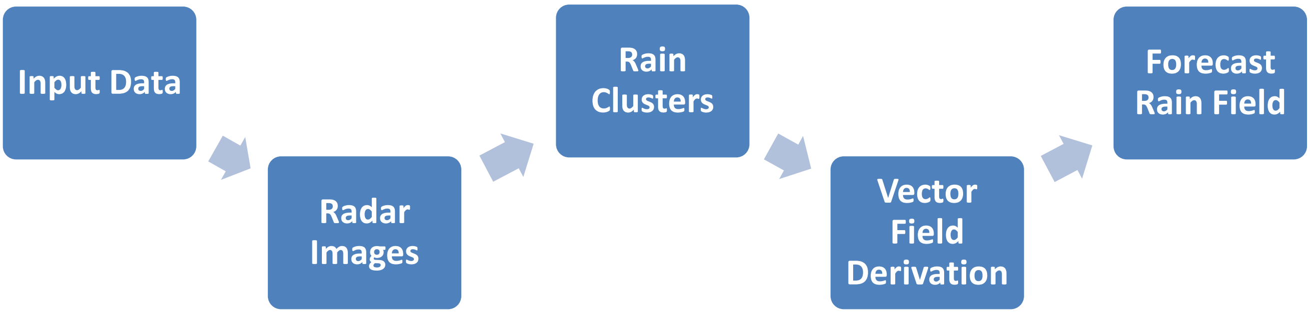

Figure 3.1: General flowchart of the model.

The class which will be used for the rain clusters will perform the calculations of each cluster's stats such as the detection of the cluster center, the amount of pixels it covers in the radar image and the implementation of weights in order to take into account the rain intensity at different parts of the cluster. The code will be able also to print and save each radar image accompanied with the individual clusters in it in a png format for each time step.

For the whole model to correctly perform there are some other essential parts that need implementation, such as those shown in Figure 3.2. Those requirements are essential in order for the code to be executed and produce the desirable outcome. For example the Matrix class will contain all the initial requirements for the matrices to be created and the implementation of the calculations. This class will provide the code with the appropriate tools in order to perform certain tasks on the matrices which will hold the data. Even though the matrices will hold the data and they will be available for calculations, another part of the code will be needed in order to visualize the results of our research. This part, the templates, will contain all the display options and the means to visualize the radar images in the appropriate form.

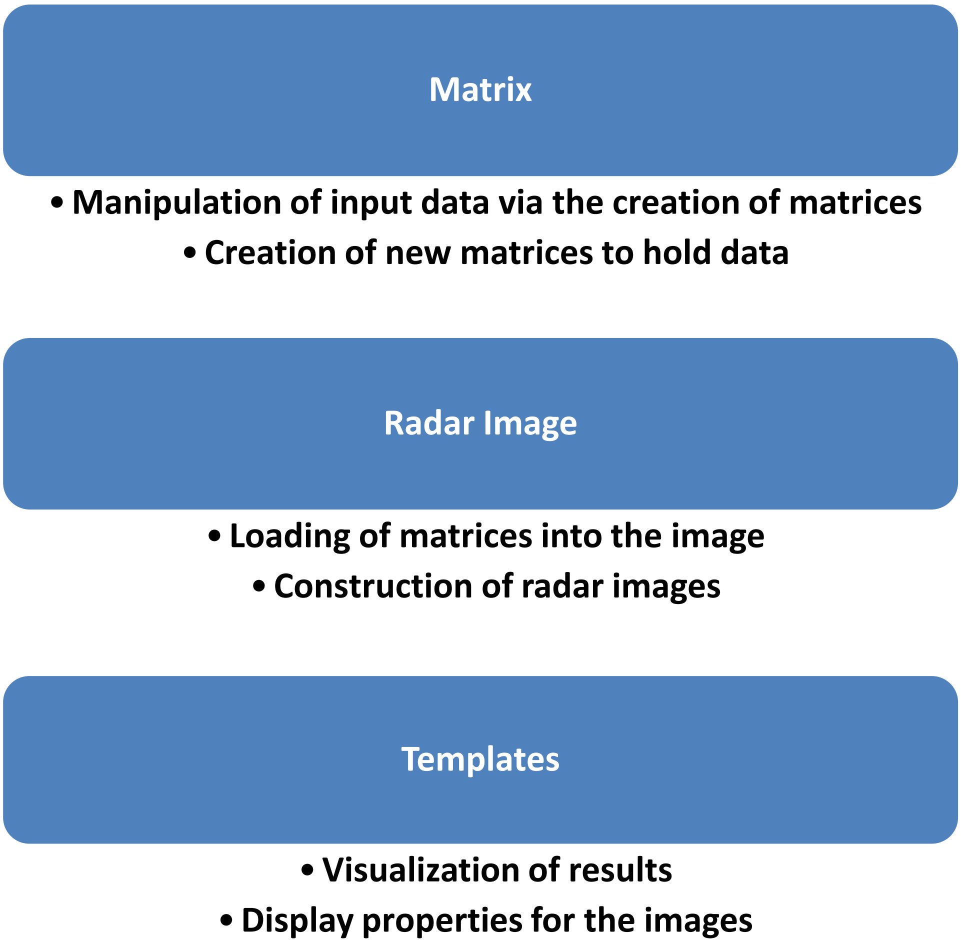

Figure 3.2: Necessary tools for the model.

In addition to the previous general steps the analysis will be performed by using small scripts and each one of them will be connected to the others. In that way the code will be clearer and easier for someone to further develop it. More in details, the creation of the radar images along with the correlation of the clusters will be performed in such a way where the position of where the MAE is found to be saved. These values will serve as the input data for the interpolation of the vector field as long as for the performance testing at the end. Therefore the interpolation of the vector field is performed in a different script and the same applies for the performance check.

3.2 Mean Absolute Error

In order to correlate each identified rain cluster at a specific moment t with the same rain cluster at t-1 the MAE method will be applied. With the obtained radar images we have the capability to derive all the rain clusters included in the images in a time interval of 15 minutes. Our goal here is to calculate the MAE for every rain cluster at different time steps so that the minimum value of it will correlate the clusters by identifying them as the same clusters at two consecutive time steps. In order to achieve that we will take one radar image and for each rain cluster in it we will apply the following technique.

At the height of around 700 hPa the maximum wind speed that can be met is around 160 km/h which is translated to 40 km/15 min. Taking into account the fact that each pixel in our radar images has dimension of 5x5 km we can fairly say that the maximum distance each cluster can cover in 15 minutes should be approximately 8 pixels in the radar image. This is a rough estimation and in that way the clusters that will correlate to each other will be located fairly close to each other. By placing a radar image on top of the radar image of the previous time step in the model and iterating for all possible positions we will calculate the MAE (equation 1).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \qquad (3.1)$$

In equation 3.1 f_i is the forecasted and y_i is the observed value of a variable while *n* is the number of iterations. If we adapt this equation according to our needs we can measure the distance of the top left corner of each area that contains a rain cluster from its original position while we iterate around the outer area as can be seen in Figure 3.3.

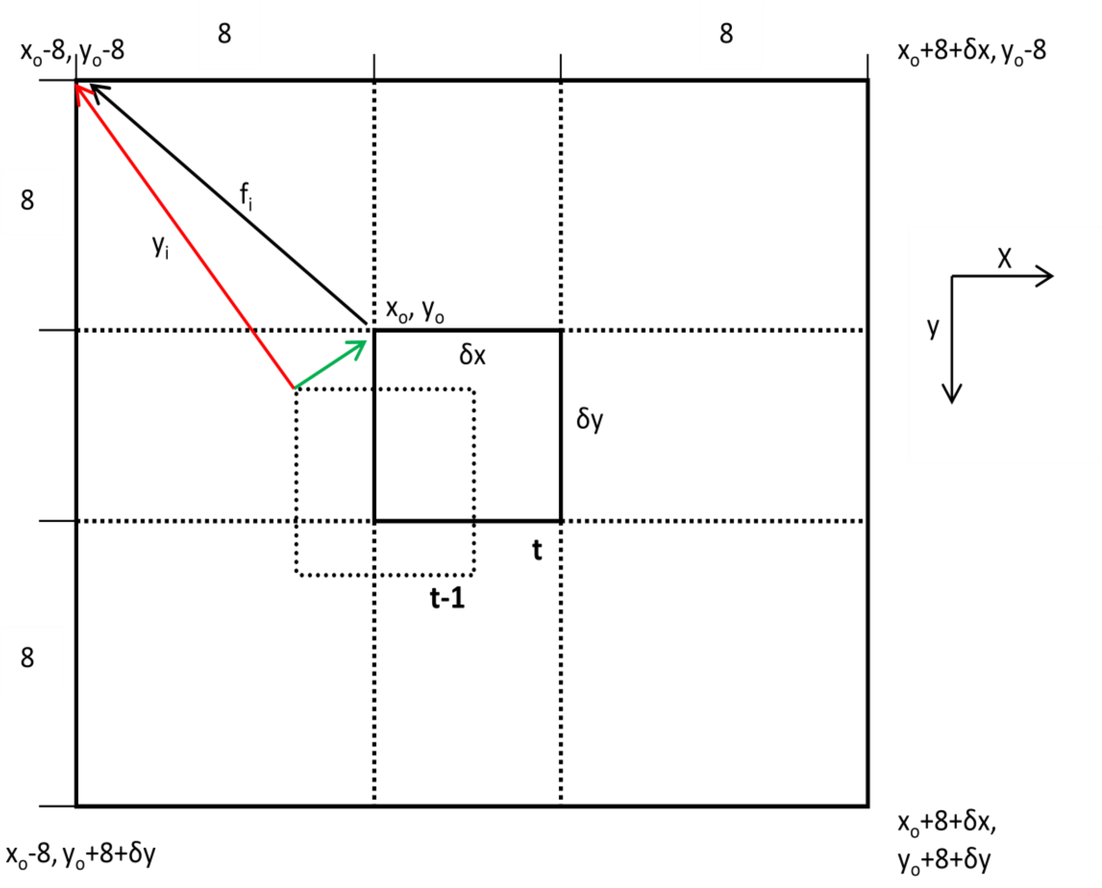

Figure 3.3: Graphical interpretation of the cluster area under study (inner box-solid line) at t, the cluster area at t-1 (dotted box) and the iteration area (outer box). δx and δy denote the dimension of the rain cluster.

In Figure 3.3 the outer box denotes the area in which we will iterate for the rain clusters with the inner boxes denoting the area which contains the rain cluster under study at t and at t-1 (solid and dotted line respectively). If this is the radar image at t then we are searching for the location of the same rain cluster 15 minutes ago i.e. at t-1. The calculation of the minimum value of the MAE will lead us to find the correlated rain cluster. This minimum value has to be found theoretically where the rain cluster at t-1 is located. Even though it is pretty obvious to correlate the rain clusters by just observing the radar images, this procedure must be followed in order to acquire better correlation in case for example a rain cluster breaks into two smaller clusters in between those 15 minutes. The iteration area will be from top left corner of the outer box with coordinates x_o,y_o until the bottom right corner with coordinates $x_o+16+\delta x$,$y_o+16+\delta y$.

In equation 3.1 the "forecasted" value will be the distance of the top left corner of the rain cluster area at t to the top left corner of the iteration area and the "observed" value will be the distance of the top left corner of the cluster area at t-1 to the top left corner of the iteration area.

After the calculation of the MAE and its minimum value is found, the distance between the correlated rain clusters can be easily measured in order to obtain the true vector displacement. This distance is shown in Figure 3.3 by the green arrow.

The choice of the top left corner for the calculation of the MAE and all distances can be justified by the fact that clusters may change in dimension and not be covered by a square area at the next time step of the study. Moreover, it would be inconvenient to use any other point of the cluster area such as the centre of it as in that case we will also have to calculate the distance between the centre and the edges of the cluster area.

3.3 Recursive Search of Clusters

The recursive search for clusters method which is applied in this project basically checks every single point in the given radar image whether it contains data of precipitation or not. In that way all adjoin points that contain data are grouped together to form a rain cluster. There is a constant checking for an already existing cluster and should this be false or not then a new rain cluster is created or the point is added to the already existing cluster depending on its position. This is the reason why the method is called *recursive* due to the fact that after the first checking, then the procedure is performed all over again from the beginning, adding up points or creating new clusters. Because each point is analyzed separately it is easy to calculate its attributes as mentioned in a previous paragraph. In the end through these calculations we can calculate the mean values of every attribute needed. These stats are for example the total points included in each cluster, the absolute center of the cluster, the weight of the cluster and consequently we can calculate the mean weight and the mean center of the cluster.

3.4 Derivation of Vector Field

The derivation of the vector field will be performed using the displacement of the vectors between the radar images at different time steps. Once the rain clusters are defined then by using the position where the MAE for every cluster is found we can create a vector which is practically a representation of the distance each cluster covered in the specific time interval. The direction of this vector will be the identification towards where the wind vector points to. The magnitude of each vector though will not be the actual value of the wind. This derivation will be performed using distances between the points where the MAE is found and the actual offset position of each cluster. The larger this distance will be, the larger the magnitude of the displacement vector. At the end we are going to have one displacement vector for every cluster in the radar image showing the direction towards where the cluster moved in time. The starting point of these vectors will be the offset position of the clusters at t-1 and the ending point will be the position where the minimum MAE is found on the radar image at t.

3.5 Barnes Scheme

The interpolation will be performed by using a very popular scheme when it comes to terms of interpolating between data points and grid points and was first introduced by Barnes. This scheme is based on the application of weights on every data point according to their distance from every possible grid point at which we want to find the interpolated values (Barnes, 1964). The applied weights and theirs dependency on the distance from the grid points is given in equation 3.2 with k_o being a parameter determined from the inter-point spacing between the data points (equation 3.3 with Δn being the distance between two data points). This distance is basically dependent on the grid we define across the data points and is subject to our definition due to the fact that usually data points can never be found on an even spaced grid. The reader should keep in mind the fact that the calculated values here are not the actual magnitude of the wind in m/s but the interpolated values of the vector displacement which represents the wind vector at any point.

$$w_m = e^{-\frac{{r_m}^2}{k_0}} \tag{3.2}$$

$$k_0 = 5.052\left(\frac{2\Delta n}{\pi}\right)^2 \tag{3.3}$$

The next step will be to calculate the interpolated values of the wind field. Equation 3.4 shows how this calculation will be performed. Here g_o is the interpolated field as it holds one datum for every grid point.

$$g_o = \frac{\sum_{i=1}^{M} w_m f}{\sum_{i=1}^{M} w_m} \tag{3.4}$$

The analysis of Barnes suggests a second pass interpolation in order for corrections on the interpolated field to be applied and for faster convergence to be reached. A procedure like this which requires only one correction pass can diminish the computational time compared to many passes. This is one great advantage of Barnes interpolation scheme due to the fact that calculated values converge with those that are interpolated much faster. Summing up the procedure, after the first interpolated field is calculated we need to interpolate one more time over the grid area and the product of this calculation will be the final interpolated wind field. Equation 3.5 holds this calculation with g_1 giving the values for the final interpolated field.

$$g_1 = g_o + \frac{\sum_{i=1}^{M}(w'_m f - w'_m g_o)}{\sum_{i=1}^{M} w'_m} \tag{3.5}$$

In equation 3.5 specifically we have to use the corrected weights for every data point according to equation 3.6. The difference lies on the constant k_1 (equation 3.7) which is subject to our definition of the constant γ varying from 0.2 to 1. Using lower values of this constant will result to bigger differentiations from the original field values. It is up to the user to choose the most appropriate value for γ and by how much should the interpolated values differ from the calculated ones. The choice of the 0.2 value is suggested and justified by Barnes as higher convergence is reached, together with greater detail on the result, when $\gamma<0.5$ (Koch et al., 1983).

$$w'_m = e^{-\frac{{r_m}^2}{k_1}} \tag{3.6}$$

$$k_1 = \gamma k_o \tag{3.7}$$

4. Experimental Analysis

The theory that was presented in the previous sections will be applied at the end with real time data and for this project the analysis will be tested with past data and for the region of Europe. The expectation of such testing is to evaluate the performance of the model and put the theory into practice. This kind of testing will be a useful tool in order for us to determine where the implementation of the theory lacks in performance and suggesting corrections.

4.1 Data Description

The data used in the project come from MetOffice and they consist of radar observations of rain intensity in mm/hour divided by a factor of 10 meaning that data of rain intensity of 5 mm/hour will actually be 0.5 mm/hour in the model. Across north-west Europe there are numerous radars installed which cover a circular area around them. The combination of the data obtained by all these radars is our input towards the generation of the radar images and the visualization of the rain clusters. The radar images are generated with an interval of 15 minutes so the propagation of the rain clusters can be observed. More specifically the series of radar images that will be created will depict the situation on 30^{th} of September 2014 at 10:15 am, 10:30 am, 10:45 am and 11:00 am (GMT). The data chosen to be displayed have no specific significance and they are a diagnostic tool regarding the performance of the model.

4.2 Thresholds

When designing a computational method which has to deal with a large amount of data it is crucial to understand that we have to apply some thresholds on our calculations. These restrictions aim towards a more accurate result in terms of scientific credibility and for better visualization of the results. For this project those thresholds determine the format of the outcome and are applied throughout the whole process.

These thresholds must be applied at different stages of the model and through them certain requirements must be fulfilled. For example the rain intensity in a rain cluster which is usually not the same all over the cluster must be taken into consideration. For that reason we have to apply weights on the clusters and take into account the fact that some parts of the cluster may be of greater significance than others. Another restrain may be the appearance of small clusters due to radar noise. In that case a cluster may not correlate to any cluster at t-1 because of the inexistence of this cluster at t-1. So a threshold can be implemented to the model so that some clusters i.e. which are smaller than 3 pixels, will be ignored. That can be done by applying a filter which will filter out all clusters smaller than the threshold or it can be done after the calculation of the MAE where all clusters that will give the maximum potential MAE will be ignored. The latter is applicable because when the MAE calculation will result in maximum values in the whole iteration area it means that all subtractions between the distances are zero leading to the conclusion that there is no cluster at t-1.

The recursive search for clusters method continuously searches for points in a radar image which can be connected to an already existing cluster. Some areas though may have not many points grouped together leading to the creation of really small clusters. The existence of clusters with area less than 9 pixels in total can be ignored mainly for two reasons. Firstly, this has to do with the fact that they can be considered as radar noise and secondly because they are not sufficiently large enough to affect the interpolated vector field. Taking into account the fact that one pixel represents an area of 5x5 km it is reasonable to justify our choice and ignore these clusters. This kind of threshold means to neglect clusters that are small compared to the overall situation. If we need to study a much smaller area and not the whole radar image of north-west Europe this threshold must be even smaller if not absent.

In addition, we have to take into account the fact that not all clusters move a significant distance in between the time interval of 15 minutes. The creation of some weather systems which include rain clusters may have the characteristic that they don't propagate much in order to have a visible effect on the vector field. In the model only clusters with a vector displacement larger than 1 pixel on any direction were taken into consideration. Following this thought we can exclude from the interpolation of the field those clusters that are almost stationary or their movement is not significant compared to the whole system yet they would have an impact on the interpolated field.

Beside the thresholds that are necessary and applied in the creation of the vector field, new thresholds should be applied on the interpolation scheme in use. The vector field will never be uniform due to the fact that every cluster can move towards every possible direction. Thus the interpolation of the vector field will assist on the creation of a more uniform field. The method which is used relies on the definition of several constants which are discussed in a previous paragraph and in the paragraph we are interested on the thresholds and applications of several values for the constants. Parameter gamma (γ) defines the level of convergence at which the analysis is done. Here it is chosen to get the lowest possible value of 0.2 according to Barnes suggestion. On the other hand the other two important thresholds are the inter-spacing distance of the data and the spacing of the grid area in which we want to interpolate. The inter-spacing of the data is chosen to 30 pixels whilst the interpolated field will be displayed every 15 pixels. These values have been chosen for better visualization results compared to others but the general rule of the inter-spacing distance of data to be doubled than the grid area spacing has been suggested by Barnes once again.

4.3 Limitations

An analysis like this has two major limitations beside others which influence the outcome in a negative way. First of all, the fact that the radars in use can cover a specific area around them leaving parts unmapped. A typical example of such activity can be seen in Figure 2.3 on the top right corner of the radar image. The circular covered area of this radar contains rain data but because of the position of this radar no more data are observed and mapped. Such behavior can be expected by radars which are close to the borders of the analysis and each radar image because there aren't any more radars left to map the rest of the area. Thus the vector field is limited close to the borders and of course the interpolated field will not be the most accurate representation of the rain cluster propagation.

The other major limitation of the project is the dependency on the input data. With different sets of data tested in the same model and values of the variables the conclusion that can be drawn is that the model is highly dependent on the quality of the input data. Situations with severe storms or larger weather systems would be better represented and forecasted. When trying to apply such a model in a situation with not much of rainfall or in a situation with not too many clusters involved then the result wouldn't have a great significance.

Another limiting factor against the performance of the analysis can be considered the fact that we are using many methods in order to make calculations at different stages of the model. Every method used though introduces errors on the calculations. These errors are transmitted further on the next calculations making the overall error bigger and more significant. Apart from the statistical errors though, the real limitation relies on the credibility of them methods themselves. More in details the MAE calculation is a very useful tool in order to correlate the clusters as described before but it lacks in performance when one cluster that is observed in one radar image is missing from the previous one. When it comes to interpolation there are many schemes that can be used with each one having its own advantages and disadvantages. According to the desirable outcome it relies on us to decide which scheme to implement. Moreover, the grid area in which the interpolation will take place needs to be defined with caution. Choosing a larger grid to display the interpolated field means that we will need more grid points to be interpolated and therefore these larger distances from every actual data point will affect the interpolated field. Something like that though will just influence the orientation of the field by a very small factor and also the magnitude of the vectors. Finally the most crucial step on the model which introduces the biggest limitation on the performance is the method by which we group together the clusters found in a radar image. From the models and methods previously presented K-means seems to be the most appropriate one but it is one of the most difficult ones to implement. This method goes really deep in details and needs a lot of computational time in order to be applied. The recursive search for clusters on the other hand it gathers up all points that can be considered as a cluster in an efficient way but it is more empirical.

5. Results

This section will deal with the results of the analysis. The interpolated wind field will be presented for the case study of this research and the forecast of the rain field will be evaluated as mentioned before. The results being presented here are just an example who a model like this has to operate. The forecast of the wind field will be an ongoing procedure for as much as someone wants and it can be used as an informative tool for a weather forecast.

5.1 Wind Field

By following the methodology described in Chapter 3 we will acquire the vector field out of the series of radar images. These radar images are shown in Figure 5.1. At first sight they might look to be the same but this is not the case. Beside the fact that the clusters are moving through time, we can observe changes in the rain intensity over some areas and the appearance or dissipation of clusters in other areas. The change in the rain intensity can be seen through the decay or growth of clusters for example over the southern part of France (changes in darkness in Figure 5.1). On the other hand we can observe clusters being created through time west of the United Kingdom or above Ireland.

Figure 5.1: Radar images over Europe compiled every 15 minutes on 30th of September 2014 at 10:15 am, 10:30 am, 10:45 am and 11:00 am (left to right and top to bottom respectively). All times are in GMT.

All kind of changes in the appeared clusters would be much greater if the time interval of the observations would be larger. In a calm day there is no sufficient time for clusters to be developed or be destroyed. A situation with much more rain activity such as thunderstorms would have a clearer result and the propagation of the rain clusters would be more easily observed.

Another fact that is worth mentioning at this stage is the dependency of the research on the boundaries of the radar images and the dependency on the influence area of the radars. This is clearly depicted over Denmark where a cluster with a perfect round boundary is observed. This is a typical situation over areas where there are no more radars to collect data or there are radars left out from the research. For these radar images we can safely conclude that the radar used for this area is just the last one to provide data for the case study we are conducting thus leading to the round shape of the cluster.

The comparison between these radar images will give us the desirable vector field which is going to be a representation of how much the clusters moved in the last 15 minutes. The procedure has been described in a previous section and here are the results presented in Figure 5.2 and Figure 5.3. Each figure is derived from the comparison of two successive radar images and that means that Figure 5.2 is produced using the radar images at 10:15 am and 10:30 am (GMT). The relative position of clusters chosen to be displayed in Figure 5.2 is from the radar image at 10:30 am (GMT). In this figure the vector is denoted by the black line which shows the vector's magnitude by the size of it. The red dot shows the direction towards where the vector points at. Practically the red dot denotes the position where the minimum value of the MAE was found when comparing the two radar images which resulted in this vector field.

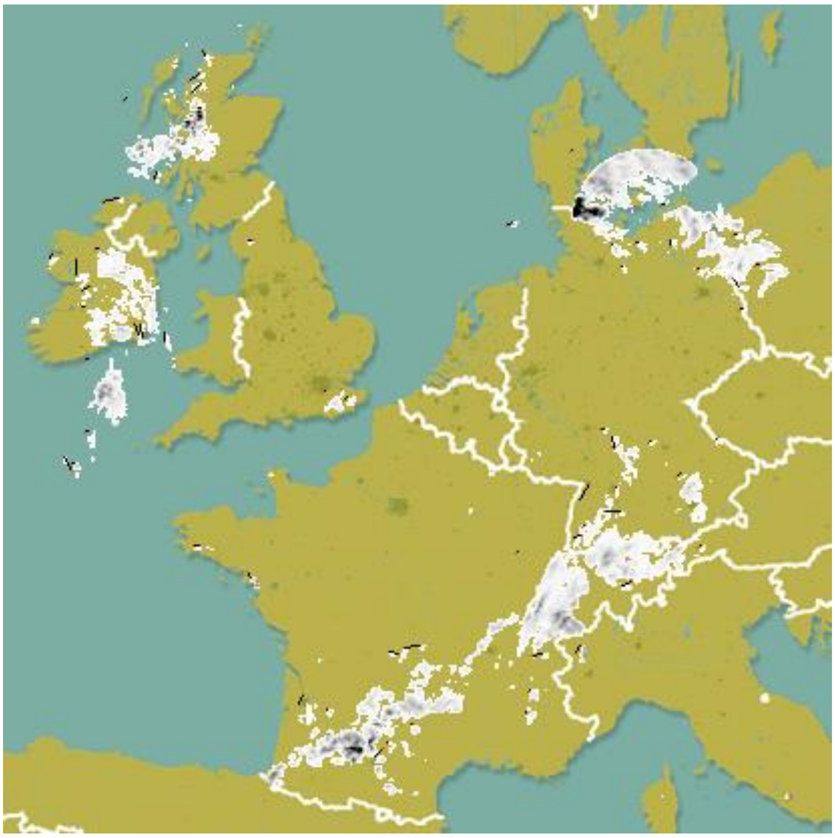

Figure 5.2: Derived vector field between radar images at 10:15 am and 10:30 am (GMT). Black line denotes the vector and the red dot shows the direction of the vector.

The original wind field we got from this procedure is chaotic as expected. Each rain cluster is predicted to propagate towards different direction and the magnitude of each vector seems to have big differentiations from cluster to cluster. This is a normal situation especially at these small time intervals of study. Similar results are found also when comparing the radar images at 10:30 am and 10:45 am and the radar image at 10:45 am with the radar image at 11:00 am (GMT). The results are shown in Figure 5.3.

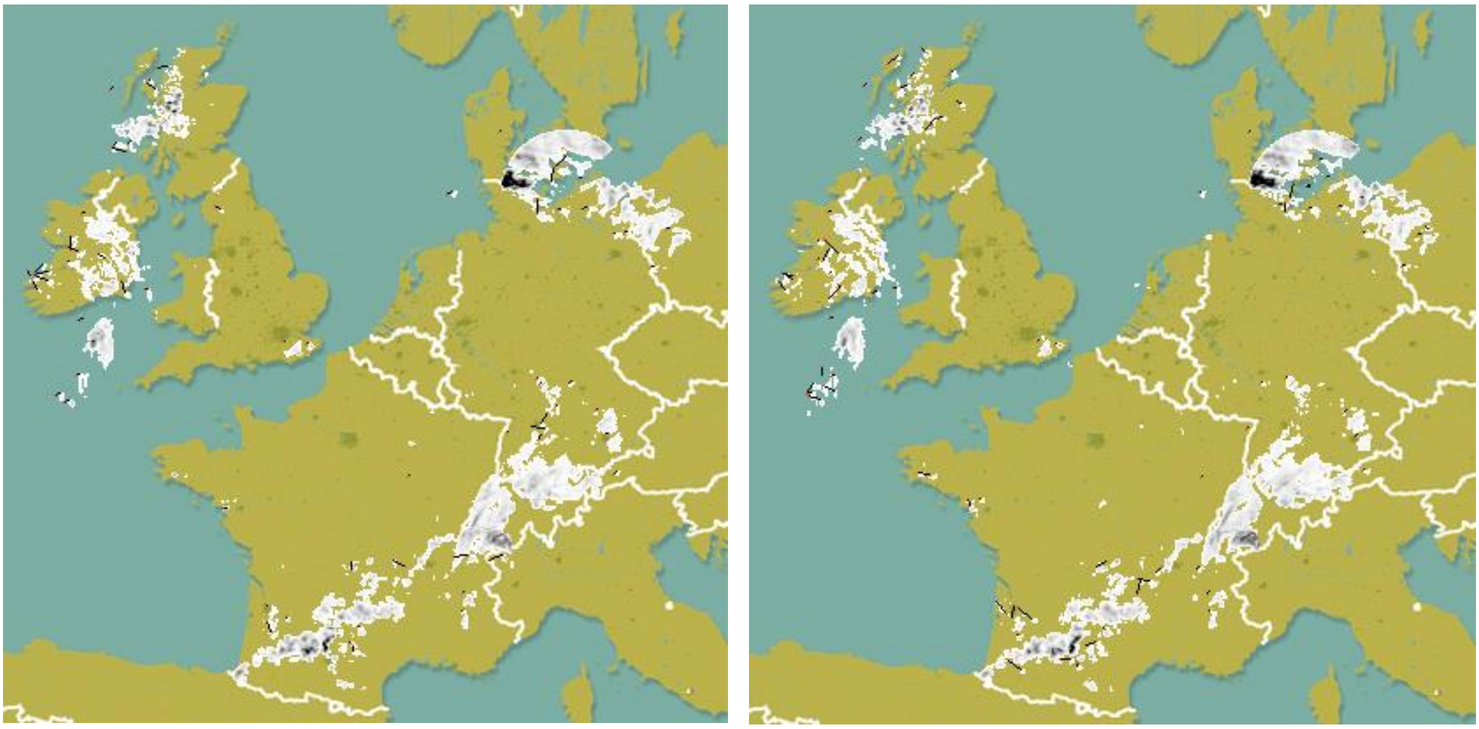

Figure 5.3: Derived vector field between radar images at 10:30 am and 10:45 am and at 10:45 am and 11:00 am (GMT) (left and right respectively). Black line denotes the vector and the red dot shows the direction of the vector.

Even though the whole wind field is no interpolated yet we can still observe patters in every figure which can be much more easily interpreted with the interpolated field. But the most important outcome out of this step of the analysis is the finding of the position where the MAE value for each cluster is found. This position will serve as the identification of the forecasted field shown later.

5.2 Interpolated Wind Field

The next vital step towards the forecast of a rain field would be the interpolation of the wind field which is derived already. When we want to forecast we need to know towards which direction a weather system is propagating and by how much. As we saw already the original wind field is rather abstract and chaotic giving just a first estimation on the areas which will be affected by the rain clusters. The interpolated field though is expected to clarify these uncertainties by giving us a more homogeneous field. Using such a method such as the one described in the theory under Barnes scheme will not be as accurate as plotting the vectors in every grid point but it will give the average displacement vectors for every position we are interested in.

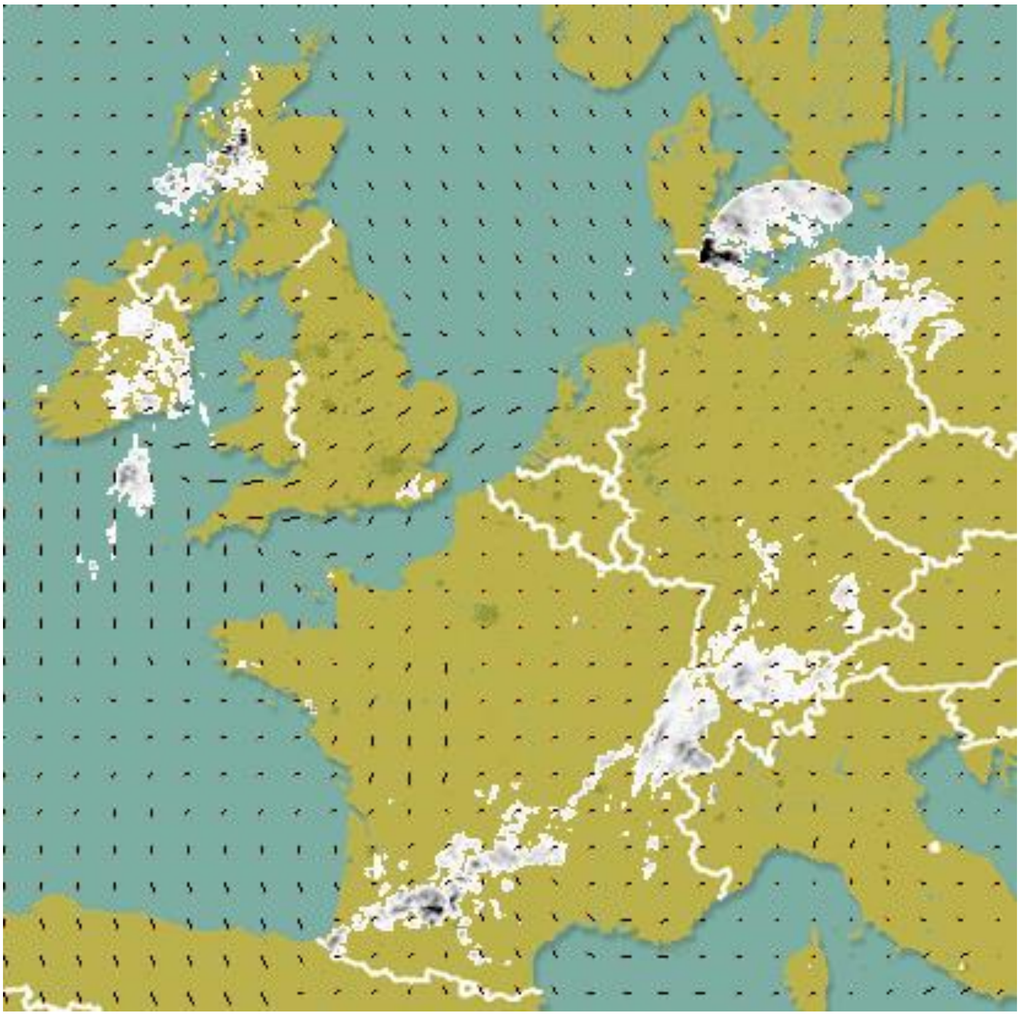

Figure 5.4: Interpolated vector field between radar images at 10:15 am -10:30 am (GMT). Black line denotes the vector and the red dot shows the direction of the vector.

The interpolated wind field between 10:15 am and 10:30 (GMT) can be seen in Figure 5.4. If we exclude the parts of the area where the clusters are located we can safely conclude that there is an average field showing an anti-clockwise movement over Europe on the 30th of September 2014. The field over the cluster areas is disrupted compared to the rest and this has to do with the high influence of the actual data points in the nearby grid points. It is highly important not to neglect the fact that the interpolation is performed by using the applied weight of every data point which is relative to their distance from the grid points with missing data. Thus, data points which are located far away from grid points will have a small influence on them but still they could cause a disruption of the field.

Similar situation is met when studying the next two radar images as we did in the rest of the analysis before. A small difference observed here is that the wind field is showing some strange patterns in both cases seen in Figure 5.5. An example of such strange situation is the area east of the United Kingdom and north of the Netherlands where the wind field is interpolated pointing towards the West without any tilting. Another example can be considered the area west of France in the next radar image where the wind field seems to be breaking up into two parts, one pointing towards the north-west direction and the other one towards the south-west. Observing the fields in general though and neglecting these minor changes in the overall orientation we can see that the wind field follows the same pattern as in Figure 5.4.

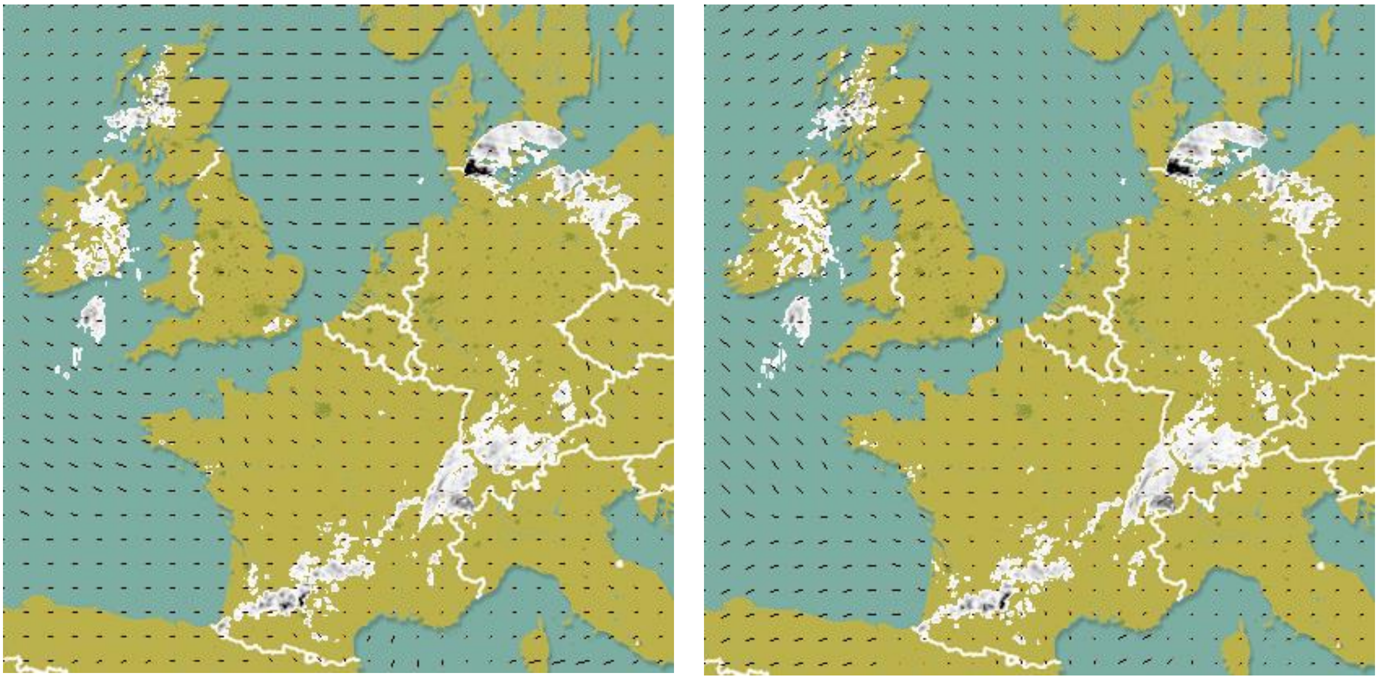

Figure 5.5: Interpolated vector field between radar images 10:30 am -10:45 am and 10:45 am -11:00 am (GMT) (left to right respectively). Black line denotes the vector and the red dot shows the direction of the vector.

Concluding, the interpolated wind field can be perceived as an estimation of the direction towards where the system will move to. Therefore we can determine our expectation on the possible positions where the cluster will have to appear in the next time step and determine th quality of the analysis by comparing them with the observed situation. From Figure 5.4 and Figure 5.5 we can expect the rain field to move in an anti-clockwise direction by ignoring as mentioned before small differentiations from the average field. More in details the rain clusters over France are expected to propagate towards the north-east direction while those over the United Kingdom towards the south-west direction.

5.3 Forecasted Rain Field

When conducting a research on this kind of topic we are not only interested in the final result of the derivation of a wind field but we are also interested to move the research one step forward and create something that is going to serve as an informative tool on the upcoming weather but also as an evaluating tool of the model. These two points that seem unconnected to each other can be met by the creation of the forecasted rain field.

From the derived vector field between 10:15 am and 10:30 am (GMT) and the vector field between 10:30 am and 10:45 am (GMT) (Figure 5.2 and Figure 5.3) we can forecast the rain field position at 10:45 am and 11:00 am (GMT) respectively. This can be done if we compare the observed situation over the area with the derived vector field. In specific, in Figure 5.2 the red dot points towards where each cluster is supposed to move to. It is the position where the minimum value of the MAE is found and therefore in the next time step each cluster ideally has to move exactly on this position. Keeping in mind that when we mention the cluster area or we refer to a cluster we actually mean the rectangular area that perfectly fits each identified cluster, we can clarify that when each cluster is expected to move to the red point of the displacement vector it is actually expected to have this position as a new offset at the next time step.

The forecasted situation at 10:45 am (GMT) can be seen in Figure 5.6 where at the same time we chose to display the clusters of 10:45 am but the displacement vector of the previous two radar images. In that way we can compare the actual observed situation with the expectation created from the displacement vectors. The displacement vectors derived from the comparison of the radar images at 10:15 am and 10:30 am (GMT) are supposed to ideally fit the position of the clusters as presented in Figure 5.6. This is exactly the case for some clusters whereas for some others the observed field is way off the forecasted one. For example for the big rain cluster over Denmark the forecasted rain field is in agreement with the observed field. The displacement vector is positioned where the cluster is observed and the dame applies for the small clusters over south-west France. On the other hand over south Germany the clusters are not forecasted in a good way.

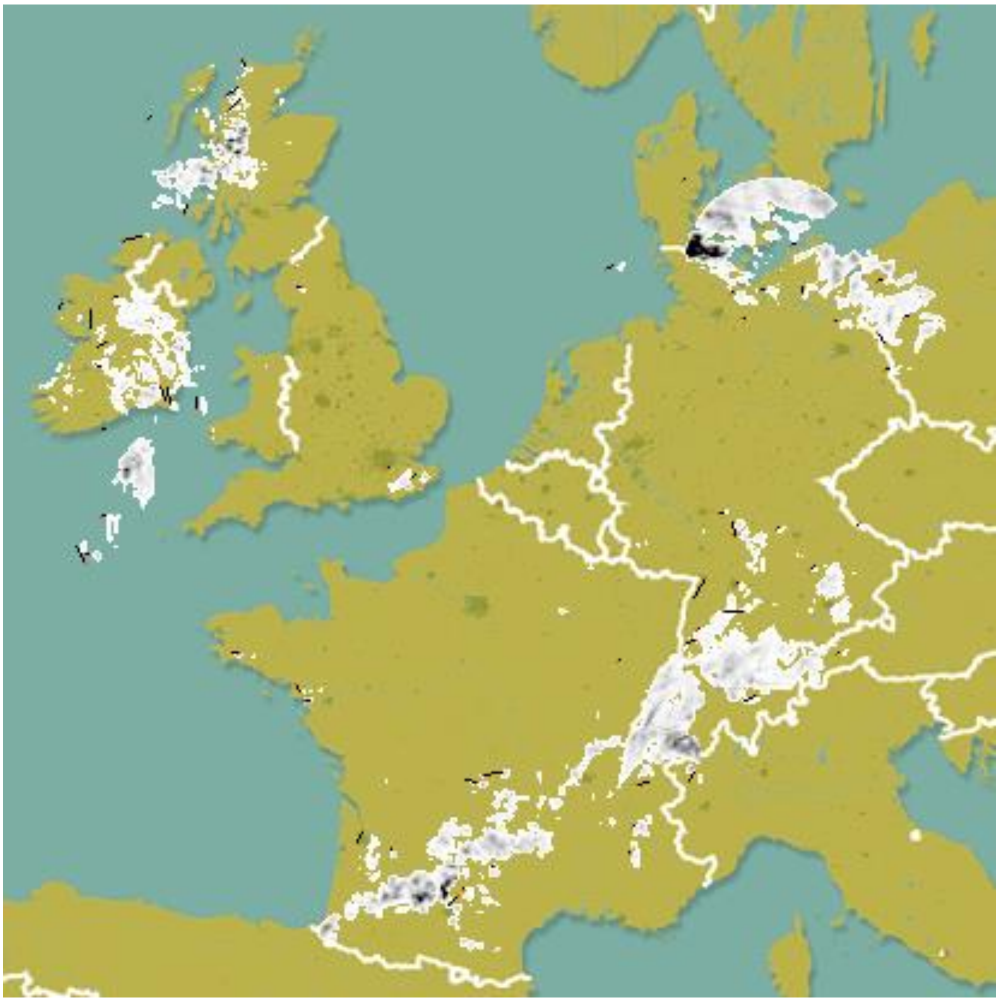

Figure 5.6: Forecasted wind field for 10:45 am (GMT). Black line denotes the vector and the red dot shows the direction of the vector.

In Figure 5.7 we have the same result for 11:00 am (GMT). The conclusions derived from this radar image can be considered as of the same quality as in Figure 5.6. There are clusters which are forecasted to be present in the right position according to the model bust there are clusters that are way off from the expected position.

There are many reasons for something like that to happen and they are mentioned in the discussion on the limitations of the model. The creation of clusters that weren't previously observed create problems like these and from the results we can also realize that the more significant a cluster is the more accurate the forecast is. We have to keep in mind that the data chosen here to test the model are not a depiction of an extreme situation but they show a rather calm situation without many changes in the position or the significance of each cluster.

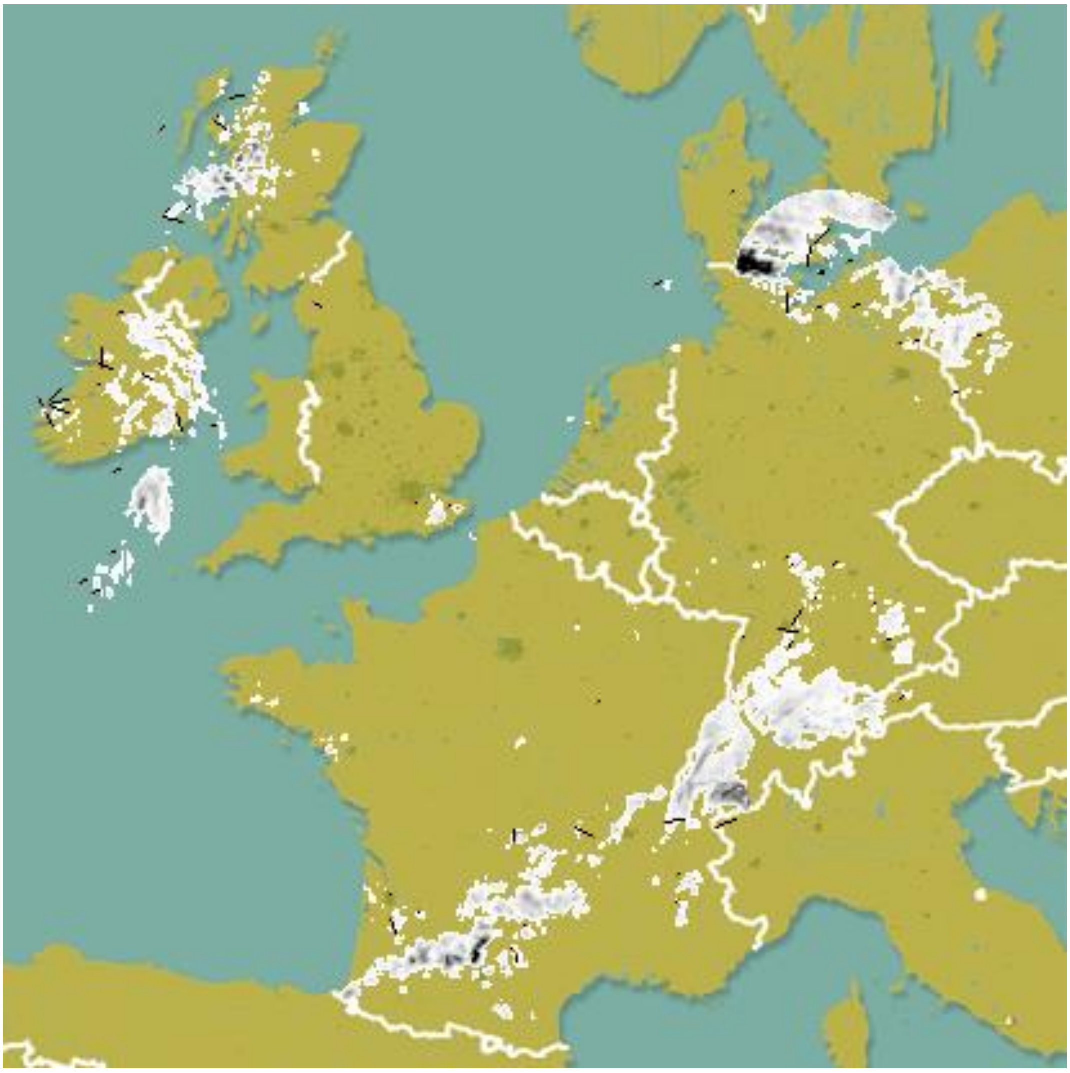

Figure 5.7: Forecasted wind field for 11:00 am (GMT). Black line denotes the vector and the red dot shows the direction of the vector.

Some examples of good agreement between observations and forecast in Figure 5.7 are the clusters over south Germany and those over Denmark and East Poland. The situation over the United Kingdom is not accurately forecasted but the biggest problem lies on the fact that we find displacement vectors where there are no clusters present. When a cluster breaks up into two smaller clusters or the opposite happens then the model doesn't perform as expected. This is something normal though as the comparison is done between different radar images and the recursive search for clusters is performed on each radar image separately. Typical example of such behavior is the appearance of a displacement vector close to the borders between England and Scotland. The cluster that was observed there in the previous time steps is not longer observed by the radars at 11:00 am (GMT).

Discussion

The derivation of a vector field out of a series of radar images is not a straightforward procedure and this can be seen on the results we got on this project. The combination of different techniques that have to be applied on different tasks throughout the model makes it hard to diminish the errors and minimize the possibility to discover large deviations between the forecast and the observations. A successful short-term forecast implies the adequate combination of the cluster identification procedure with the correlation of them. Furthermore, the interpolation scheme used on this project had some limitations which cannot be neglected and highly influence the form of the derived interpolated vector field. Beside the analytical part, the use of radar data itself has some limitations as well. Their cover area is very strict and using them as input data for any research makes it a very demanding task to correctly implement any theoretical part especially close to the borders of the research area. That is to say that the focus should be targeted on the improvements that can be made on the model itself and not on the actual performance of it with the chosen techniques. The application of K-means clustering method will provide the model with much more credibility even though it will need much more computational time and it would require a faster programming language. The amount of calculations is huge and considering the fact that this model will be launched on a live environment it becomes clear that the performance of it should be not only accurate but also fast. After all it's also a matter of precision of these calculations which differs from one programming language to another. Further improvements can be also done on the forecast of the rain field where a better result can be obtained if someone choose to compare the interpolated vector field with the observed interpolated vector field. The cluster by cluster comparison performed on this project gives some credible results but the interpolated field is the one that has the potential to better depict the expected situation. In any case the model is depended on the input data we have. This is why the user needs to think really deeply and decide on the thresholds he will apply on every step of the model. Just an example of such behavior is the choice someone will make on the grid size to be interpolated on the vector field. If we choose to plot the displacement vectors on a more detailed and less wide grid the results will differ compared to what was presented here. This is one of the many adjustments that can be made on the model and the most appropriate combination of them is still part of the research that has to be done. Furthermore, we have to mention that the quality and performance control of this model would be better if performed using data acquired from forecast models and not the actual observations. No model can reach a perfect agreement with the observations thus it would be more accurate and precise to test the outcome of two different models for example. Even though, at first, completing a project like this might look a simple task to do, we have to realize the amount of factors that have to be taken into account and afterwards decide on the programming techniques that have to be applied in order to make the model run faster and produce more credible results. The biggest drawback of this model is that even though we may eventually reach the point where the result depicts perfectly an observed field, we can never be sure that this will be reproduced on another situation as well. Weather systems behave in a very chaotic way and forecasting is a very demanding task. The ultimate goal will be to approach as much as possible the observations in advance even if this will happen on a 5 or 15 minutes based analysis.

Conclusions

Considering the fact that this project is due towards the creation of a more accurate short-term forecast the major conclusions out of it are discussed here. The general conclusion is that a series of radar images can produce an accurate vector field which at the end can forecast the propagation of the rain field in an accepted way. Such an analysis though has many limitations due to the fact that we have to use many different schemes and methods. Every method used in the analysis is subject to improvements and accuracy enhancement. It is obvious from the analysis that the outcome is dependent on a high degree to the input data. A situation with not much or intense rain will not produce a really accurate forecast.

Firstly, the search for clusters method that is used in this project even though it produces the rain clusters in the radar images, it could be more precise and provide us with a better representation of them. This could be done by either improving the actual technique itself either by choosing to apply a different one such as the K-means clustering.

Secondly, the correlation of the clusters and the way it was performed could not take into account the possible break of clusters into smaller ones or the assembly of two clusters into one after the time interval has passed. Because of this limitation sometimes the displacement vectors could not be calculated and the model wouldn't perform in an accurate way over that region.

Thirdly, after the experimental analysis and the results we got out of the model it is pretty obvious that the interpolated field is highly dependent on many variables that need to be determined from the user. Therefore anyone conducting such a research may get different results when it comes to interpolate on the wind field just by choosing different values for the constants or even better account only for the nearby data points for any grid point that wants to be interpolated.

Fourthly, a recommendation should be made on the display options someone might have. The way chosen here in this report to display the displacement vectors was a fast and easy one in terms of programming skills and designing. One should consider the fact of actually printing the vectors in a form of an arrow, a procedure which hides a lot of difficulties but it will provide every image with much more clarity.

Last but not least, after the whole model was complete the conclusion that was derived for the programming language chosen to be used, i.e. PHP, is that it is not the most suitable or wise option for such a project. PHP performs slowly and as the amount of calculations were increasing it became even slower. Needless to say that with increasing complexity on the different scripts that were combined to create the general model, the performance of PHP was slowly declined.

Acknowledgments

For the preparation and the completion of this project I would like to show my appreciation to everyone in the IT department of Meteovista who gave me the strength and insight to prepare and finish the analysis. They were always there for me to give me any kind of help I needed in the problems I encountered on my time here in Meteovista. My biggest appreciation though goes to Joralf Quist and Gareth Williams who assisted me the most on the programming part of the project and all the meteorological aspects of it. Without their contribution this project wouldn't be finished in the time limit of my internship and their contribution to it was more than essential.

References

Barnes, S. L. (1964). A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, *3*(4), 396-409.

David, L., Esnault, A., & Callaud, T. D. (2002, July). Comparison of interpolation techniques for 2D and 3D velocimetry. In *Eleventh International Symposium on Application of Laser Techniques to Fluid Mechanics, Lisbon*.

De Lannoy, G. J., Verhoest, N. E., & De Troch, F. P. (2005). Characteristics of rainstorms over a temperate region derived from multiple time series of weather radar images. *Journal of Hydrology*, *307*(1), 126-144.

KIM, S., TACHIKAWA, Y., SAYAMA, T. and TAKARA, K., 2006. Ensemble rainfall-runoff prediction with radar image extrapolation and its error structure. 水工学論文集, 50: 43-48.

Koch, Steven E., Mary Desjardins, and Paul J. Kocin. "An interactive Barnes objective map analysis scheme for use with satellite and conventional data."*Journal of Climate and Applied Meteorology* 22.9 (1983): 1487-1503.

Lakshmanan, V., Rabin, R. and DeBrunner, V., 2003. Multiscale storm identification and forecast. Atmospheric research, 67: 367-380.

Nakakita, E. et al., 1996. Short-term rainfall prediction method using a volume scanning radar and grid point value data from numerical weather prediction. Journal of Geophysical Research: Atmospheres (1984–2012), 101(D21): 26181-26197.

Peura, M. and Hohti, H., 2004. Motion vectors in weather radar images, Proceedings of the 7th International Winds Workshop, Helsinki, Finland.

Schmid, W. (2000). Nowcasting winter precipitation with radar. In *Proceedings of the 10th SIRWEC conference, Davos* (pp. 17-24).

Takasao, T., Shiiba, M. and Nakakita, E., 1994. A real-time estimation of the accuracy of short-term rainfall prediction using radar, Stochastic and Statistical Methods in Hydrology and Environmental Engineering. Springer, pp. 339-351.