UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Ingegneria Civile, Edile e Ambientale

Laurea Magistrale in Ingegneria Civile

The small scale spatial variability in rainfall data collected in a small urban catchment area

Relatore: Prof. Ing. Andrea Marion

Correlatore: Prof. Simon Tait

Correlatrice: Prof. Alma Schellart

Laureando: Antonio Patania

Matricola: 1019978

Anno Accademico 2013 – 2014

Alla mia famiglia,
per aver sempre creduto in me,
A Mery,
per avermi sempre incitato.
Ai miei nonni,
che vorrei avere qui con me.

Abstract

The urban drainage system could be studied using a hydrological model to simulate the flow inside the sewers and the surface runoff. It is not so easy to find appropriate methods of representing and predicting rainfall; rainfall is usually obtained from rain gauges set up in a larger area, or derived using weather radar data. The currently is no method to estimate the real trend of the data in a small (sub-kilometre) catchment, so it is difficult to have an accurate analysis with hydrological models about what happens in terms of surface runoff in a small urban catchment.

This work explores the spatial variability of rainfall on a sub kilometre scale, in an urban catchment of 1 km^2, in terms of correlation between the rain gauges according to the dependence on time scale (ranging from 5 minutes to 3 hours) and dependence on rainfall intensity.

It is also showed that it is possible to use a good interpolating method like Kriging to have a good trend of the rainfall, starting from some ground truth (rainfall data measured using rain gauges) and how it is operative according to the different rainfall event.

In the end it is possible to argue that there are different behaviours of the spatial variability according to the different events and time scales, which can be described with the correlation.

CONTENTS:

6

List of figures:

8

List of Tables:

Introduction

There is a growing interest to determine the performance of urban drainage systems both in terms of water quality and urban flood risk.

To have a good pattern to study these performances the input data used are taken by radar and rain gauges.

However it is not so easy to have good cumulative temporal scales of data, because the two instruments indicated above have some problems respectively.

The data from the radar can give back only the quantitative of 5 minutes of rainfall and even if it covers a big area, it is able to show the same value of rainfall in all the area covered by a pixel, which is usually a square kilometre.

The rain gauge on the other hand is able to give back a more specific value of rainfall data for every minute, but it is only a punctual data and we will need a lot of rain gauges to know the trend of the event.

So the radar data gives the spatialization of the event and the rain gauge data gives the punctual precision of the rainfall, the best condition would be to have a fusion of the two items.

At the moment, researchers are working on finding a solution to have a good value from radar data in an area lower than one kilometre square, and also to find a link between the real data of rainfall collected by rain gauge and the radar data, because the latter are usually bigger or smaller than the real data value that is measurable by rain gauge.

The main aim of this study is to analyse the correlation between the rain gauges for different events, to understand how many rain gauges we need to have a good study in a small urban catchment and to know which are the best conditions of correlation to search for the lower cumulative temporal scales to use on hydrological models.

This study is a follow up on previous studies in this field and in particular on the study "Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale", where the researchers have analysed the behaviour of some hydrological models using as data input the radar data and the rain gauge data.

They have compared the results from their hydrological model, using both data input from radar and rain gauge, with the data measured inside the sewer.

They have noticed that the flow simulation results provide a generally good fit, but the results are changed if they use lower temporal resolution rainfall data as input to the model.

In fact, in the graphs that they show it is possible to notice that when the rainfall resolution is reduced, peak intensities are smoothed, and it is noticed for 15 minutes resolution rainfall data as an input; when the resolution is further reduced to 60 minutes, the peak is almost entirely missing, and the simulated flow mimics few of the features of the measured flow.

In the end, although the hydrodynamic model showed a tendency to underestimate the flow peaks for the larger events, comparison of the flow peaks simulated using radar or rain gauge data with measured flow peaks in the sewer system provided a valuable insight into the effect of rainfall measurement technique and the effect of spatial variability of rainfall. On the small spatial and temporal scales of the hydrodynamic model, the spatial representativeness error in the rainfall estimates translated to considerable differences in simulated flow peaks. The results highlight that there were very significant differences in the prediction of the peak flow rates within the sewer system. The events studied in this paper illustrated the difficulty in accurately simulating the occurrence of sewer flow peaks.

It is concluded that, to ensure optimum accuracy of peak flow predictions, it is necessary to use rainfall data that has both high spatial and temporal resolution.

This is the reason why we are looking for the best trend of a correlation for every event that we would like to categorize into some fixed groups. So if we have good results, we can use a number of rain gauges in every location, which will be enough to estimate the spatialization of all the data on the study area. In these conditions we know which kind of resolution of cumulative temporal scale we can use in every case.

This study will provide a good estimation of the data to use inside hydrological models to study the trend of the drainage system.

Chapter 1: Measurement instruments

The rain gauge and the radar are two instruments that researchers and meteorologists use to provide the rainfall data. In this work it will be used only the rain gauge that is able to give back a punctual value and is more useful in an analysis of small urban area than the radar.

1.1. Rain gauges

A rain gauge, also known as a pluviometer, is a type of instrument used by meteorologists and hydrologists to gather and measure the amount of liquid precipitation over a set period of time.

Most rain gauges generally measure the precipitation in millimetres. The level of rainfall is sometimes reported as inches or centimetres.

The rain gauge amounts are read either manually or by automatic weather station. The frequency of readings will depend on the requirements of the collection agency. Some countries will supplement the paid weather observer with a network of volunteers to obtain precipitation data for sparsely populated areas.

Rain gauges have their limitations; attempting to collect rain data in a hurricane can be nearly impossible and unreliable (even if the equipment survives) due to the wind power. Also, rain gauges only indicate rainfall in a localized area. For virtually any gauge, drops will stick to the sides or funnel of the collecting device, such that amounts are very slightly underestimated, and those of .01 inches or .25 mm may be recorded as a trace.

Another problem encountered is when the temperature is close to or below freezing. Rain may fall on the funnel and ice or snow may collect in the gauge and not permit any subsequent rain to pass through.

Rain gauges should be placed in an open area where there are no obstacles, such as buildings or trees, to block the rain. This is also to prevent the water collected on the roofs of buildings or the leaves of trees from dripping into the rain gauge after a rain, resulting in inaccurate readings.

1.2. Types

In the following section there is a general description of the different rain gauges. Types of rain gauges include graduated cylinders, weighing gauges, tipping bucket gauges, and simple buried pit collectors. Each type has its advantages and disadvantages for collecting rain data.

1.2.1. Standard rain gauge

The standard rain gauge, developed at the start of the 20th century, consists of a funnel emptying into a graduated cylinder, 2 cm in diameter, which fits inside a larger container which is 20 cm in diameter and 50 cm tall. If the rainwater overflows the graduated inner cylinder, the larger outer container will catch it. When measurements are taken, the height of the water in the small graduated cylinder is measured, and the excess over flow in the large container is carefully poured into another graduated cylinder and measured to give the total rainfall. In locations using the metric system, the cylinder is usually marked in mm and will measure up to 250 millimetres (9.8 in) of rainfall. Each horizontal line on the cylinder is 0.5 millimetres (0.02 in). In areas using Imperial units each horizontal line represents 0.01 inch.

Figure 1: Standard rain gauge.

1.2.2. Weighing precipitation gauge

A weighing type precipitation gauge consists of a storage bin, which is weighed to record the mass. Certain models measure the mass using a pen on a rotating drum, or by using a vibrating wire attached to a data logger. The advantages of this type of gauge over tipping buckets are that it does not underestimate intense rain, and it can measure other forms of precipitation, including rain, hail and snow. These gauges are, however, more expensive and require more maintenance than tipping bucket gauges.

The weighing type recording gauge may also contain a device to measure the quantity of chemicals contained in the location's atmosphere. This is extremely helpful for scientists studying the effects of greenhouse gases released into the atmosphere and their effects on the levels of the acid rain.

Figure 2: Weighing precipitation gauge.

1.2.3. Optical rain gauge

These have a row of collection funnels. In an enclosed space below each is a laser diode and a photo transistor detector. When enough water is collected to make a single drop, it drops from the bottom, falling into the laser beam path. The sensor is set at right angles to the laser so that enough light is scattered to be detected as a

sudden flash of light. The flashes from these photo detectors are then read and transmitted or recorded.

Figure 3: Optical rain gauge.

1.2.4. Tipping bucket rain gauge

The tipping bucket rain gauge consists of a funnel that collects and channels the precipitation into a small seesaw-like container. After a pre set amount of precipitation falls, the lever tips, dumping the collected water and sending an electrical signal. An old-style recording device may consist of a pen mounted on an arm attached to a geared wheel that moves once with each signal sent from the collector. In this design, the wheel turns the pen arm moves either up or down leaving a trace on the graph and at the same time making a loud click. Each jump of the arm is sometimes referred to as a 'click' in reference to the noise. The chart is measured in 10 minutes periods (vertical lines) and 0.4 mm (0.015 in) (horizontal lines) and rotates once every 24 hours and is powered by a clockwork motor that must be manually wound.

The tipping bucket rain gauge is not as accurate as the standard rain gauge because the rainfall may stop before the lever has tipped. When the next period of rain begins it may take no more than one or two drops to tip the lever. This would then indicate that pre-set amount has fallen when in fact only a fraction of that amount has actually fallen. Tipping buckets also tend to underestimate the amount of rainfall, particularly in snowfall and heavy rainfall events. The advantage of the tipping bucket rain gauge is that the character of the rain (light, medium, or heavy)

16

may be easily obtained. Rainfall character is decided by the total amount of rain that has fallen in a set period (usually 1 hour) and by counting the number of 'clicks' in a 10 minutes period the observer can decide the character of the rain. Correction algorithms can be applied to the data as an accepted method of correcting the data for high level rainfall intensity amounts.

Figure 4: Tipping-bucket rain gauge, internal view.

Figure 5: Tipping-bucket rain gauge, external view.

1.3. Tipping bucket rain gauge errors

There are some studies to characterize the errors associated with tipping-bucket rain gauges when used to provide rainfall intensity estimates at small temporal scales. The main source of the tipping bucket gauge sampling error is its sampling mechanism and its inability to capture the small temporal features of the rainfall time series. The study has found significant error levels in the one minute estimates especially at low rain rates, but it also has found that as the time scale of the computed rain rates increased, the error decreased substantially. With time scales longer than 15 minutes, the error becomes negligible. The gauge's performance and its associated errors are sensitive to the applied sampling interval and the bucket volume. Therefore, we recommend sampling intervals on the order of 5–10 s, along with a bucket size no larger than 0.254 mm. finally, there are also approximate formulas of the tipping bucket sampling error obtained under various operational schemes. The analysis was based on rainfall observations that were mostly dominated by convective storms with few strati form events. Similar analysis may be needed to investigate the scope of applicability of the developed formulas under different climatological regimes. However, the formulas serve as a first-order approximation that can be used in practical applications.

1.4. Tipping-bucket rain gauge for our research

1.4.1. General description

Research has shown that a conventionally shaped rain gauge interferes with the airflow and that the flow accelerates and turbulence increases over the top of the funnel than otherwise would have fallen on the ground. In most cases this is ignored, but it may be corrected arithmetically or overcome physically by placing the gauge in a pit so that the rim of the funnel is level with the ground, the pit is covered by a grating to simulate the aerodynamic roughness of the ground surface while preventing any splash into the funnel. There are obvious advantages with this method, but it is not always practical. The body of the rain gauge has a profile that has been designed to reduce the drag and turbulence and therefore be sited conventionally on exposed sites with some confidence.
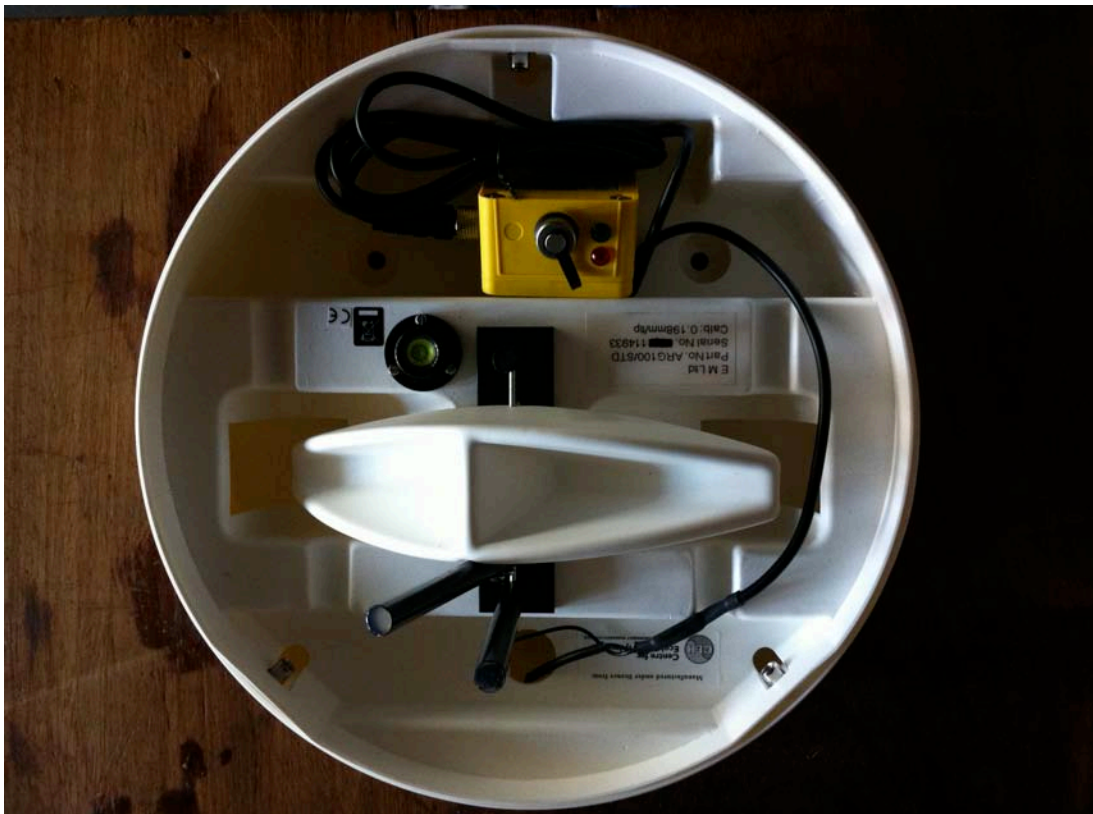
Figure 6: Tipping-bucket used for our research.

Figure 7: Tipping-bucket used for our research.

Rain gauges which operate on the tipping bucket principle provide a switch closure output, which may be connection to a digital data logger. The pulses returned during rainfall may be counted over any time interval desired allowing accurate determination of the rainfall rate.

1.4.2. Description of the ARG100 rain gauge

The ARG100 aerodynamic rain gauge has been designed to minimise the above effect by presenting a reduced area to the wind. Rainfall is measured by the well-proven tipping bucket method. Precipitation is collected by a funnel and is passed to one of the two buckets situated at either end of a short balance arm. The balance arm tips when the first bucket is full, emptying this bucket and positioning the second bucket under the funnel. The tipping process repeats indefinitely as long as the rain continues to fall, with each tip corresponding to a fixed quantity of rainfall; at each tip the moving arm forces a magnet past a reed switch, causing contact to be made for a few milliseconds.

1.4.3. Choosing a site

The site chosen to install the rain gauge will depend in part on the application to which it is being put and in part on the particular circumstances at the site. But if possible site the rain gauge so the distance between the rain gauge and obstruction, such as trees or buildings is at least as great as twice the height of obstruction.

1.5. Calibration of tipping-bucket rain gauge

The sensitivity of the rain gauge is accurately calibrated at manufacture to a nominal 0,2 mm/tip. Each rain gauge being supplied with its own calibration figure.

The figure below shows the calibration number of the rain gauge that we have recalibrated. To do this there are two different methods, the static calibration and the dynamic calibration.
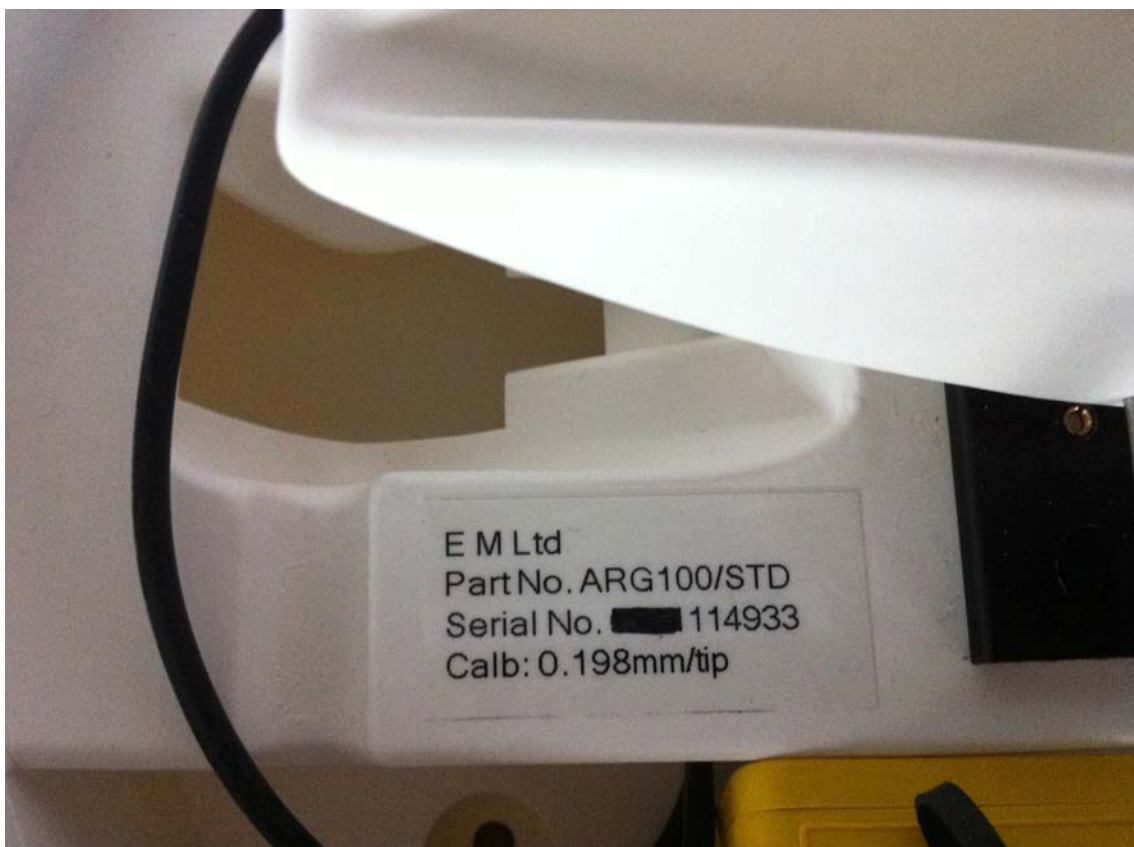
Figure 8: Calibration number

1.5.1. Static calibration

Install the gauge over a sink as illustrated in figure, ensuring that it is correctly levelled; using a burette or pipette, slowly drip in 10,13 cm3 of water for 0,2 mm/tip. The bucket should tip on the last drip of water. Adjust the relevant thumb screw, located under each bucket, until the above condition is met. We have to repeat for the other side of bucket. It is not possible to set screws very precisely using this method, but it should be done with as much care as possible. It is obviously very important that both buckets tip in response to the same amount of water. However, a dynamic test is required to check this calibration precisely after reach readjustment and the process becomes very time-consuming. In any case, it is virtually impossible to get the adjustments absolutely correct, and it is generally preferable to adjust the settings as closely as is reasonably practical, and then derive a calibration factor for each rain gauge individually after a dynamic calibration. And this is what we did for several rain gauges. It is explained better in the following under section.
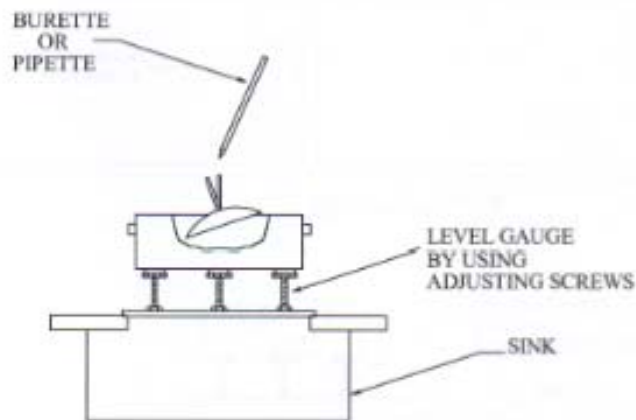
Figure 9: static calibration

1.5.2. Dynamic calibration

Configure the gauge as in figure, ensure gauge is levelled and connected to a data logger or counter. Fill a container with 1000 cm3 of water for 0,2 mm/tip calibration, but we have seen that is not enough, so we prefer to do it with 3000 cm3, this is usually achieved most precisely and consistently by weighing the water on a balance capable of measuring to 0,1 g that correspond to 0,1 cm3. Allow the water to drip slowly into the gauge, taking at least 60 minutes to empty, approximately 40 seconds for each tip, but for us that we have used a bigger value

of three times we need to wait at least 3 hours, to do it we have used a pump that had different intensities to simulate different situation. At the end of this period approximately 98 tips will have occurred (we have found approximately 294 tips), but the exact number is obtained from the data logger or counter. To this add on an accurate estimation of what fraction of a tip is left in the bucket when the water stops flowing, a graduated syringe is ideal for this.

Figure 10: Dynamic calibration in laboratory.

1.5.3. Calculating the calibration factor

To calculate the value of calibration factor, we know the nominal amount of the tips for a 0,2 mm bucket is 98,7167 tips, let N as the number of tips, the calibration factor (C.F.) is then calculated as follows:

C.F. = Gauge size (0,2 mm) X Nominal amount (98,7167) / N

Numbers of tips	Calibration factor
96,7 to 96,9	0,204 mm/tip
97,0 to 97,4	0,203 mm/tip
97,5 to 97,9	0,202 mm/tip
98,0 to 98,4	0,201 mm/tip
98,5 to 98,9	0,200 mm/tip
99,0 to 99,4	0,199 mm/tip
99,5 to 99,9	0,198 mm/tip
100 to 100,4	0,197 mm/tip

Table 1: range of calibration factors.

in other words, each tip correspond to a value of rainfall. Provided the calibration factor falls between 0,197 mm and 0,204 mm it is acceptable for most purposes. If the calibration factor lies outside these limits, repeat the static and dynamic calibration procedure.

Chapter 2: Location of the study

Bradford lies at the heart of the City of Bradford, a metropolitan borough of West Yorkshire, in Northern England. It is situated in the foothills of the Pennines, 8,6 miles (14 km) west of Leeds, and 16 miles (26 km) northwest of Wakefield.

Figure 11: Location of the city of our studies.

2.1 General description of location

2.1.1. Geography

Bradford is located at 53° 45′ 00″ N 01° 50′ 00″ W (53.7500, -1.8333)1. Topographically, it is located in the eastern moorland region of the South Pennines.

Bradford is not built on any substantial body of water but is situated at the junction of three valleys, one of them, that of the Bradford Beck which rises in moorland to the west, and is swelled by its tributaries, the Horton Beck, Westbrook, Bowling Beck and Eastbrook. At the site of the original ford, the beck turns north, and flows towards the River Aire at Shipley. Bradfordale is a name given to this valley. It can be regarded as one of the Yorkshire Dales, though as it passes though the city, it is often not recognised as such. The beck's course through the city centre is interred and has been since the mid 19th century. On the 1852 Ordnance Survey it is visible as far as Sun Bridge, at the end of Tyrrell Street, and then from beside Bradford Forster Square railway station on Kirkgate. On the 1906 Ordnance Survey, it disappears at Tumbling Hill Street, off Thornton Road, and appears north of Cape Street, off Valley Road, though there are culverts as far as Queens Road.

The Bradford Canal, built in 1774, linking the city to the Leeds and Liverpool Canal took its water from Bradford Beck and its tributaries. The supply was often inadequate to feed the locks, and the polluted state of the canal led to its temporary closure in 1866: the canal was closed in the early 20th century as uneconomic.

2.1.2. Geology

The underlying geology of the city is primarily carboniferous sandstones. These vary in quality from rough rock to fine, honey-coloured stone of building quality. Access to this material has had a pronounced effect on the architecture of the city.

2.1.3. Climate

As with the rest of the UK, Bradford benefits from a maritime climate, with limited seasonal temperature ranges, and generally moderate rainfall throughout the year. Records have been collected since 1908 from the Met Office's weather station at Lister Park, a short distance north of the city centre. This constitutes one of the nations longest unbroken records of daily data.

The absolute maximum temperature recorded was 32.2 °C in August 1990. In an average year, the warmest day should attain a temperature of 27.5 °C, with a total of 6 days rising to a maximum of 25.1 °C or above.

The absolute minimum temperature recorded was −13.9 °C during January 1940. The weather station's elevated suburban location means exceptionally low temperatures are unknown. Typically, 41.4 nights of the year will record an air frost. Rainfall averages around 870 mm per year with over 1mm falling on 139 days.

2.2 Locations of rain gauge stations

The rain gauge stations are placed on Bradford University city campus and they are sixteen.

All rain gauges are ARG100 tipping bucket rain gauges, they have been setted to read tips every 1 minute; we manually read out the data approximately every 3 weeks and we also check and clean the rain gauges.

The period of study is long from 6th of April to 2nd of November, the rest of the year we are unable to have a good value of data for the problem of the tipping bucket rain gauge to catch the data of snowy and haily days.

The picture below could give a general idea of the location of all the rain gauges, but it will be explained better in the following section.
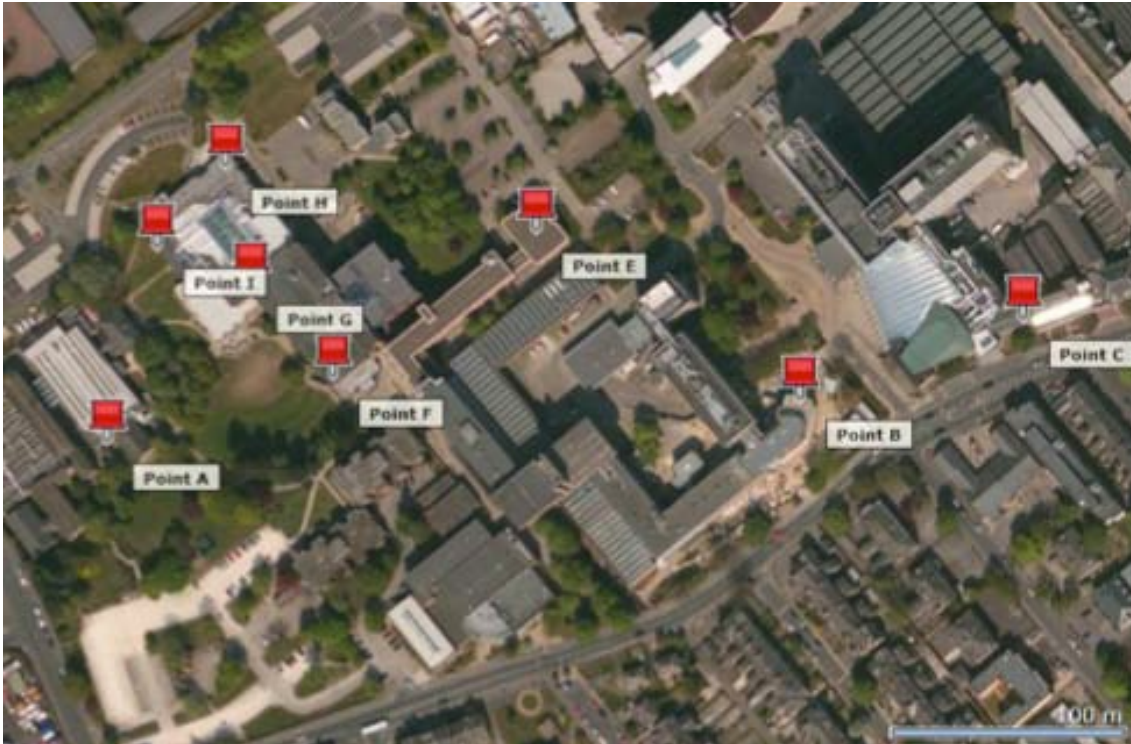
Figure 12: Location of measurement places.

2.2.1. Location of every point

Point A – two rain gauges clamped on the railings of Phoenix building, distance of more than twice the height of the surrounding building, but they do not satisfy the 5 metres away from roof edge.

Figure 13: Rain gauges of point A, in Phoenix building.

Point B – School of Health. Two rain gauges clamped to the railings of a balcony, they can't be placed in middle, as there is an approximately 4 m high wall. Rain gauges are not at 5 m away from roof edge.

Figure 14: Rain gauges of point B, in School of Health.

Point C – Richmond building, behind the atrium wing. Two rain gauges clamped to the white railings, because there is a high building behind us, the blue cladded Richmond building on right of the photo above, we are not sure about exact height of the building, at the edge of this roof we're probably not quite twice the height of the building away. The rain gauges are also not 5 m away from roof edge.

Figure 15: Rain gauges of point C, in Richmond building.

Point E – Pemberton building. Two rain gauges in middle of the flat roof, so it is satisfying all the criteria.

Figure 16: Rain gauges of point E, in Pemberton building.

Points F and G, Library roof, 2 pairs of rain gauges, one on either side of flat roof, located several meters from the edge. The rain gauges have more than 5 metres away from roof edge.

Figure 17: Rain gauges of point F and G, both in Library roof.

Point I and H – There are four rain gauges, one pair either side, on roof of Student Central, several metres from edge, distance of more than twice the height of surrounding buildings.

Figure 18: Rain gauges of point I and H, on roof of Student Central.

In the end to give clarity of the all locations, there are some pictures taken from aerial photography.

Location of points: E - F - G - H - C
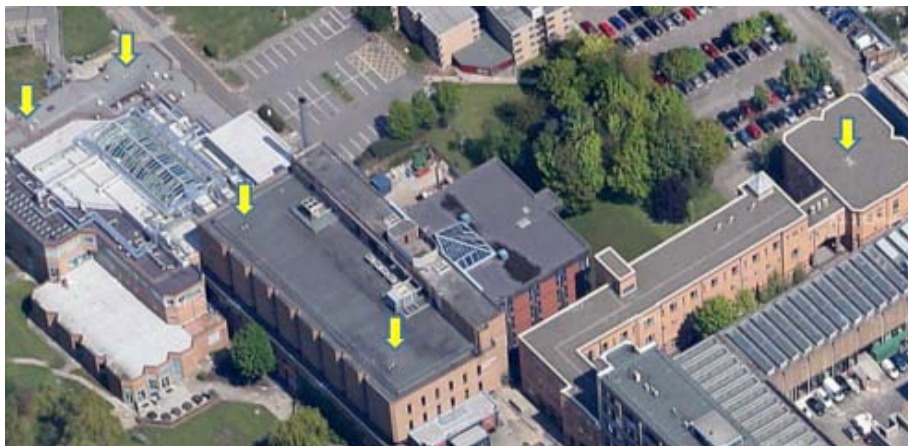
Figure 19: Location of the rain gauges in Pemberton Buildings, Student Central and Library.

Location of points: A - F - G - H - C
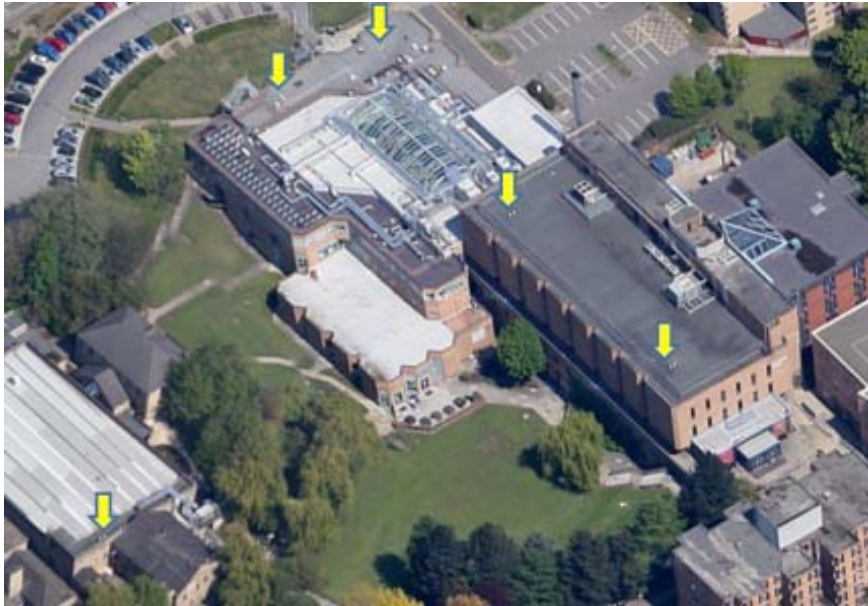
Figure 20: Location of the rain gauges in Phoenix Buildings, Student Central and Library

Location of point B and C:

Figure 21: Location of the rain gauges in Richmond Buildings and School of Health.

2.2.2. Distances and coordinates

Here below there are the coordinates of every point, just analysed.

- Phoenix building, point A, rain gauges 7 and 6: x = 415325, y = 432740;
- Student central, point H&I:

    Rain gauges 8 and 9: x = 415356, y = 432847;

    Rain gauges 10 and 12: x = 415374, y = 432858;

- Library, point F&G:

rain gauges 15 and 16: 415395, y = 432811;

rain gauges 13 and 14: 415426, y = 432774;

- Pemberton, point E, rain gauges 4 and 5: x = 415513, y = 432834;

- School of Health, point B, rain gauges 17 and 18, x = 415626, y = 432762;

- Richmond building, point C, rain gauges 1 and 2, x = 415720, y = 432795;

The following table shows the distances between all the points.

Stations	Distance [m]
AB	304
AC	404
AI	88
AE	209
AF	105
AG	95
AH	131
BC	104
BI	290
BE	139
BF	204
BG	246
BH	270
CI	380
CE	216
CF	304
CG	338
CH	355
IE	166
IF	96
IG	44
IH	46
EF	111
EG	127
EH	139
FG	56
FH	102
GH	52

Table 2: distances between all the stations.

Chapter 3: Models for the analysis

3.1. Correlation

In statistics, dependence is any statistical relationship between two random variables or two sets of data. Correlation refers to any of a broad class of statistical relationships involving dependence.

Familiar examples of dependent phenomena include the correlation between the physical statures of parents and their offspring, and the correlation between the demand for a product and its price. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice.

Formally, dependence refers to any situation in which random variables do not satisfy a mathematical condition of probabilistic independence. In loose usage, correlation can refer to any departure of two or more random variables from independence, but technically it refers to any of several more specialized types of relationship between mean values. There are several correlation coefficients, often denoted ρ or r, measuring the degree of correlation. The most common of these is the Pearson correlation coefficient, which is sensitive only to a linear relationship between two variables (which may exist even if one is a nonlinear function of the other). Other correlation coefficients have been developed to be more robust than the Pearson correlation – that is, more sensitive to nonlinear relationships. Mutual information can also be applied to measure dependence between two variables.

3.1.1. Pearson's product-moment coefficient

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient, or "Pearson's correlation coefficient", commonly called simply "the correlation coefficient". It is obtained by dividing the covariance of the two variables by the product of their standard deviations. Karl Pearson developed the coefficient from a similar but slightly different idea by Francis Galton.

3.1.2. Population and sample methods

Pearson's correlation coefficient when applied to a population is commonly represented by the Greek letter ρ and may be referred to as the population correlation coefficient or the population Pearson correlation coefficient. The formula for ρ is:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where, *cov* is the covariance, which is a measure of how much two random variables change together. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the smaller values, i.e., the variables tend to show similar behaviour, the covariance is positive. In the opposite case, when the greater values of one variable mainly correspond to the smaller values of the other, i.e., the variables tend to show opposite behaviour, the covariance is negative. The sign of the covariance therefore shows the tendency in the linear relationship between the variables; σ_X is the standard deviation of X; μ_X is the mean of X; and E is the expectation, which refers, intuitively, to the value of a random variable one would "expect" to find if one could repeat the random variable process an infinite number of times and take the average of the values obtained. More formally, the expected value is a weighted average of all possible values.

Pearson's correlation coefficient when applied to a sample is commonly represented by the letter r and may be referred to as the sample correlation coefficient or the sample Pearson correlation coefficient. We can obtain a formula for r by substituting estimates of the covariances and variances based on a sample into the formula above. That formula for r is:

$$R = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

An equivalent expression gives the correlation coefficient as the mean of the products of the standard scores. Based on a sample of paired data (X_i, Y_i), the sample Pearson correlation coefficient is:

$$R = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(X_i - \bar{X})}{S_X} \frac{(Y_i - \bar{Y})}{S_Y}$$

where:
The standard score is:

$$\frac{(X_i - \bar{X})}{S_X}$$

the mean is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and the sample standard deviation is:

$$S_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

The absolute values of both the sample and population Pearson correlation coefficients are less than or equal to 1. Correlations equal to 1 or -1 correspond to data points lying exactly on a line, or to a bivariate distribution entirely supported on a line (in the case of the population correlation). The Pearson correlation coefficient is symmetric: corr(X,Y) = corr(Y,X).

3.1.3. Interpretation
The correlation coefficient ranges from −1 to 1. A value of 1 implies that a linear equation describes the relationship between X and Y perfectly, with all data points lying on a line for which Y increases as X increases. A value of −1 implies that all

data points lie on a line for which Y decreases as X increases. A value of 0 implies that there is no linear correlation between the variables.

More generally, note that $(X_i - X)(Y_i - Y)$ is positive if and only if X_i and Y_i lie on the same side of their respective means. Thus the correlation coefficient is positive if X_i and Y_i tend to be simultaneously greater than, or simultaneously less than, their respective means. The correlation coefficient is negative if X_i and Y_i tend to lie on opposite sides of their respective means.

3.1.4. Interpretation of the size of a correlation

Several authors have offered guidelines for the interpretation of a correlation coefficient. However, all such criteria are in some ways arbitrary and should not be observed too strictly. The interpretation of a correlation coefficient depends on the context and purposes. A correlation of 0.8 may be very low if one is verifying a physical law using high-quality instruments, but may be regarded as very high in the social sciences where there may be a greater contribution from complicating factors.

3.2. Kriging

Optimal interpolation based on regression against observed z values of surrounding data points, weighted according to spatial covariance values.

The interpolation is an estimation of a variable at an unmeasured location from observed values at surrounding locations. For example, estimating ϕ_u based on values at nearest six data points in our zone data:
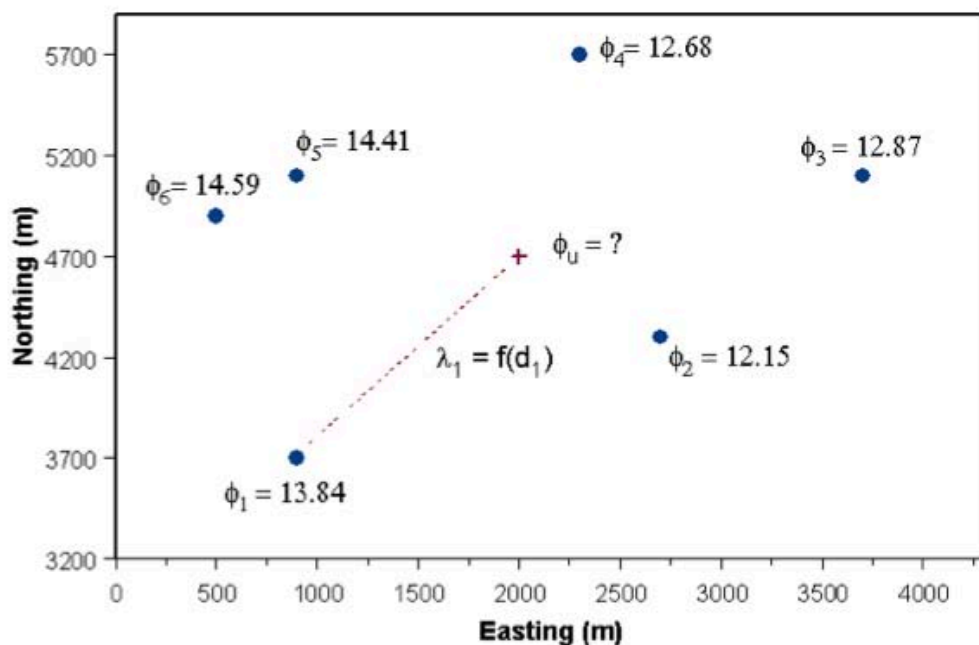
Figure 22: Estimation of the value ϕ_u, using data values of the surrounding locations.

It would seem reasonable to estimate ϕ_u by a weighted average $\Sigma \lambda_\alpha \phi_u$, with weights λ_α given by some decreasing function of the distance, d_α, from u to data point α.

All interpolation algorithms (inverse distance squared, splines, radial basis functions, triangulation, etc.) estimate the value at a given location as a weighted sum of data values at surrounding locations. Almost all assign weights according to functions that give a decreasing weight with increasing separation distance. Kriging assigns weights according to a (moderately) data-driven weighting function, rather than an arbitrary function, but it is still just an interpolation algorithm and will give very similar results to others in many cases. In particular:

- If the data locations are fairly dense and uniformly distributed throughout the study area, you will get fairly good estimates regardless of interpolation algorithm.

- If the data locations fall in a few clusters with large gaps in between, you will get unreliable estimates regardless of interpolation algorithm.

- Almost all interpolation algorithms will underestimate the highs and overestimate the lows; this is inherent to averaging and if an interpolation algorithm didn't average we wouldn't consider it reasonable.

Some advantages of kriging:

- Helps to compensate for the effects of data clustering, assigning individual points within a cluster less weight than isolated data points (or, treating clusters more like single points)

- Gives estimate of estimation error (kriging variance), along with estimate of the variable, Z, itself (but error map is basically a scaled version of a map of distance to nearest data point, so not that unique)

- Availability of estimation error provides basis for stochastic simulation of possible realizations of Z(u)

3.2.1. Kriging approach and terminology

All kriging estimators are but variants of the basic linear regression estimator $Z^*(u)$ defined as

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)]$$

with:

$\mathbf{u}$ and $\mathbf{u}_\alpha$: location vectors for estimation point and one of the neighbouring data points, indexed by α;

$n(\mathbf{u})$: number of data points in local neighbourhood used for estimation of $Z^*(u)$

$m(\mathbf{u})$, $m(\mathbf{u}_\alpha)$: expected values (means) of $Z(\mathbf{u})$ and $Z(\mathbf{u}_\alpha)$

$\lambda_\alpha(\mathbf{u})$: kriging weight assigned to datum $z(\mathbf{u}_\alpha)$ for estimation location $\mathbf{u}$; same datum will receive different weight for different estimation location.

$Z(u)$ is treated as a random field with a trend component, m(u), and a residual component, $R(\mathbf{u}) = Z(\mathbf{u}) - m(\mathbf{u})$. Kriging estimates residual at $\mathbf{u}$ as weighted sum of residuals at surrounding data points. Kriging weights, λ_α, are derived from

covariance function or semivariogram, which should characterize residual component. Distinction between trend and residual somewhat arbitrary; varies with scale.

3.2.2. Basics of kriging

Again, the basic form of the kriging estimator is

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)]$$

The goal is to determine weights, λ_u that minimize the variance of the estimator

$$\sigma_E^2(\mathbf{u}) = Var\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$$

under the unbiasedness constraint $E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0$

The random field Z(**u**) is decomposed into residual and trend components, $Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u})$, with the residual component treated as an random field with a stationary mean of 0 and a stationary covariance (a function of lag, **h**, but not of position, **u**):

$$E\{R(\mathbf{u})\} = 0$$
$$Cov\{R(\mathbf{u}), R(\mathbf{u} + \mathbf{h})\} = E\{R(\mathbf{u}) \cdot R(\mathbf{u} + \mathbf{h})\} = C_R(\mathbf{h})$$

The residual covariance function is generally derived from the input semivariogram model, $C_R(\mathbf{h}) = C_R(\mathbf{0}) - \gamma(\mathbf{h}) = Sill - (\mathbf{h})$.

Thus the semivariogram we feed to a kriging program should represent the residual component of the variable.

The three main kriging variants, simple, ordinary, and kriging with a trend, differ in their treatments of the trend component, $m(\mathbf{u})$.

3.2.3. Simple kriging

For simple kriging, we assume that the trend component is a constant and known mean, $m(\mathbf{u}) = m$, so that

$$Z_{SK}^*(u) = m + \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)]$$

This estimate is automatically unbiased, since $E[Z(\mathbf{u}_\alpha) - m] = 0$, so that $E[Z_{SK}^*(\mathbf{u})] = m = E[Z(\mathbf{u})]$. The estimation error $Z_{SK}^*(\mathbf{u}) - Z(\mathbf{u})$ is a linear combination of random variables representing residuals at the data points, $\mathbf{u}_\alpha$, and the estimation point, $\mathbf{u}$:

$$Z_{SK}^*(\mathbf{u}) - Z(\mathbf{u}) = [Z_{SK}^*(\mathbf{u}) - m] - [Z(\mathbf{u}) - m]$$

$$= \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(\mathbf{u})R(\mathbf{u}_\alpha) - R(\mathbf{u}) = R_{SK}^*(\mathbf{u}) - R(\mathbf{u})$$

Using rules for the variance of a linear combination of random variables, the error variance is then given by

$$\sigma_E^2(\mathbf{u}) = Var\{R_{SK}^*(\mathbf{u})\} + Var\{R_{SK}(\mathbf{u})\} - 2Cov\{R_{SK}^*(\mathbf{u}), R_{SK}(\mathbf{u})\}$$

$$= \sum_{\alpha=1}^{n(u)}\sum_{\beta=1}^{n(u)} \lambda_\alpha^{SK}(\mathbf{u})\lambda_\beta^{SK}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + C_R(0) - 2\sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u})$$

To minimize the error variance, we take the derivative of the above expression with respect to each of the kriging weights and set each derivative to zero. This leads to the following system of equations:

$$\sum_{\beta=1}^{n(u)} \lambda_\beta^{SK}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1, \dots, n(\mathbf{u})$$

Because of the constant mean, the covariance function for $Z(\mathbf{u})$ is the same as that for the residual component, $C(\mathbf{h}) = C_R(\mathbf{h})$, so that we can write the simple kriging system directly in terms of $C(\mathbf{h})$:

$$\sum_{\beta=1}^{n(u)} \lambda_\beta^{SK}(\mathbf{u})C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1, \dots, n(\mathbf{u})$$

This can be written in matrix form as:

$$\mathbf{K}\lambda_{SK}(\mathbf{u}) = \mathbf{k}$$

where $\mathbf{K}_{SK}$ is the matrix of covariances between data points, with elements $\mathbf{K}_{i,j} = C(\mathbf{u}_i - \mathbf{u}_j)$, $\mathbf{k}$ is the vector of covariances between the data points and the estimation point, with elements given by $\mathbf{k}_i = C(\mathbf{u}_i - \mathbf{u})$, and $\lambda_{SK}(\mathbf{u})$ is the vector of simple kriging weights for the surrounding data points. If the covariance model is licit (meaning the underlying semivariogram model is licit) and no two data points are collocated, then the data covariance matrix is positive definite and we can solve for the kriging weights using:

$$\lambda_{SK} = \mathbf{K}^{-1}\mathbf{k}$$

Once we have the kriging weights, we can compute both the kriging estimate and the kriging variance, which is given by:

$$\sigma_{SK}^2(\mathbf{u}) = C(0) - \lambda_{SK}^T(\mathbf{u})\mathbf{k} = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(\mathbf{u})C(\mathbf{u}_\alpha - \mathbf{u})$$

after substituting the kriging weights into the error variance expression above.

So what does all this math do? It finds a set of weights for estimating the variable value at the location u from values at a set of neighbouring data points. The weight on each data point generally decreases with increasing distance to that point, in accordance with the decreasing data-to-estimation covariances specified in the right-hand vector, $\mathbf{k}$. However, the set of weights is also designed to account for redundancy among the data points, represented in the data point-to-data point covariances in the matrix $\mathbf{K}$. Multiplying $\mathbf{k}$ by $\mathbf{K}^{-1}$ (on the left) will down weight points falling in clusters relative to isolated points at the same distance.

Some characteristics to note:

Smoothness: Kriged surface will basically be as smooth as possible given the constraints of the data; in many cases, probably smoother than the "true" surface.

Bullseyes: Because kriging averages between data points, local extremes will usually be at well locations; bullseyes are inevitable. This is true of almost all interpolation algorithms. Extreme form of this is artefact discontinuities at well locations when semivariogram model includes significant nugget.

Error map reflects data locations, not data values: Map of kriging standard deviation depends entirely on data configuration and covariance function; essentially a map of distance to nearest well location scaled by covariance function.

3.2.4. Ordinary kriging

For ordinary kriging, rather than assuming that the mean is constant over the entire domain, we assume that it is constant in the local neighbourhood of each estimation point, that is that $m(\mathbf{u}_\alpha) = m(\mathbf{u})$ for each nearby data value, $Z(\mathbf{u}_\alpha)$, that we are using to estimate $Z(\mathbf{u})$. In this case, the kriging estimator can be written:

$$Z^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)]$$

$$= \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{u})Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})}(\mathbf{u})\right] m(\mathbf{u})$$

and we filter the unknown local mean by requiring that the kriging weights sum to 1, leading to an ordinary kriging estimator of

$$Z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(\mathbf{u})Z(\mathbf{u}_\alpha) \quad \text{with} \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{OK}(\mathbf{u}) = 1.$$

In order to minimize the error variance subject to the unit-sum constraint on the weights, we actually set up the system minimize the error variance plus an additional term involving a Lagrange parameter, $\mu_{OK}(\mathbf{u})$:

$$L = \sigma_E^2(\mathbf{u}) + 2\mu_{OK}(\mathbf{u})\left[1 - \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{u})\right]$$

so that minimization with respect to the Lagrange parameter forces the constraint to be obeyed:

$$\frac{1}{2}\frac{\partial L}{\partial \mu} = 1 - \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{u}) = 0$$

In this case, the system of equations for the kriging weights turns out to be

$$\begin{cases} \displaystyle\sum_{\beta=1}^{n(u)} \lambda_\beta^{OK}(\mathbf{u})C_R\big(\mathbf{u}_\alpha - \mathbf{u}_\beta\big) + \mu_{OK}(\mathbf{u}) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1, \dots, n(\mathbf{u}) \\[2ex] \displaystyle\sum_{\beta=1}^{n(u)} \lambda_\beta^{OK}(\mathbf{u}) = 1. \end{cases}$$

where C_R (**h**) is once again the covariance function for the residual component of the variable. In simple kriging, we could equate C_R (**h**) and C (**h**), the covariance function for the variable itself, due to the assumption of a constant mean. That equality does not hold here, but in practice the substitution is often made anyway, on the assumption that the semivariogram, from which C(h) is derived, effectively filters the influence of large-scale trends in the mean.

In fact, the unit-sum constraint on the weights allows the ordinary kriging system to be stated directly in terms of the semivariogram (in place of the CR (h) values above). In a sense, ordinary kriging is the interpolation approach that follows naturally from a semivariogram analysis, since both tools tend to filter trends in the mean.

Once the kriging weights (and Lagrange parameter) are obtained, the ordinary kriging error variance is given by

$$\sigma_{OK}^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{OK}(\mathbf{u})C(\mathbf{u}_\alpha - \mathbf{u}) - \mu_{OK}(\mathbf{u})$$

3.2.5. Kriging with a Trend

Kriging with a trend (the method formerly known as universal kriging) is much like ordinary kriging, except that instead of fitting just a local mean in the neighbourhood of the estimation point, we fit a linear or higher-order trend in the (x,y) coordinates of the data points. A local linear (a.k.a., first-order) trend model would be given by

$$m(\mathbf{u}) = m(x,y) = a_0 + a_1 x + a_2 y$$

Including such a model in the kriging system involves the same kind of extension as we used for ordinary kriging, with the addition of two more Lagrange parameters and two extra columns and rows in the K matrix whose (non-zero) elements are the x and y coordinates of the data points. Higher-order trends (quadratic, cubic) could be handled in the same way, but in practice it is rare to use anything higher than a first-order trend. Ordinary kriging is kriging with a zeroth-order trend model. If the variable of interest does exhibit a significant trend, a typical approach would be to attempt to estimate a "de-trended" semivariogram using one of the methods described in the semivariogram lecture and then feed this into kriging with a first- order trend.

3.3. Cross validation

Cross validation allows us to compare estimated and true values using only the information available in the sample data set
Cross validation may help us to choose between different weighting procedures, search strategies, variogram models, or estimation methods

3.3.1. Role of Cross Validation

The Cross Validation results are often used simply to compare the distribution of the estimation errors or residuals from different estimation procedures and choose the one that works better

A careful study of the spatial distribution of cross validated residuals (estimated minus true values) can provide insights into where an estimation procedure may run into trouble.

3.3.2. Cross Validation Method

Starting from all the data inside the data set, we need to know if the kriging method is an accurate method to estimate the values of the interest points.

So, the sample value at a particular location, which is located on the interest place, is temporarily removed from the sample data set. The graphic below explains that condition:
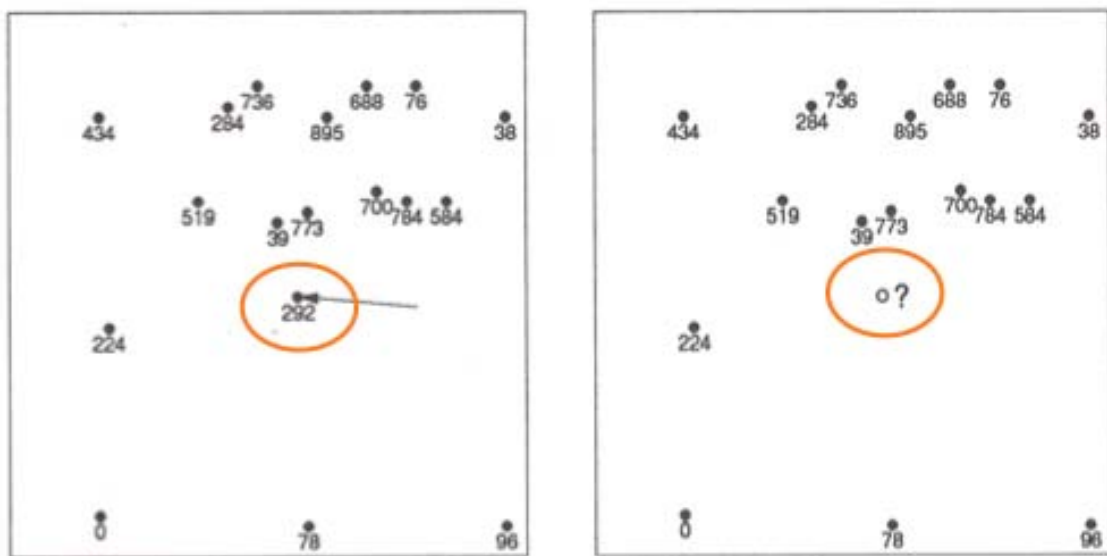
Figure 23: Principle of the cross validation, calculation of the value in a point and make the comparison between the estimate value and the real value.

The value at the same location is then estimated using the remaining samples, and in this case we have used the kriging interpolation method.

Once the estimation is calculated we can compare it to the true sample value that was initially removed from the sample data set

This procedure is repeated for all available samples, which are sixteen points that correspond to the sixteen rain gauges or to the eight average values of the rain gauges pairs.
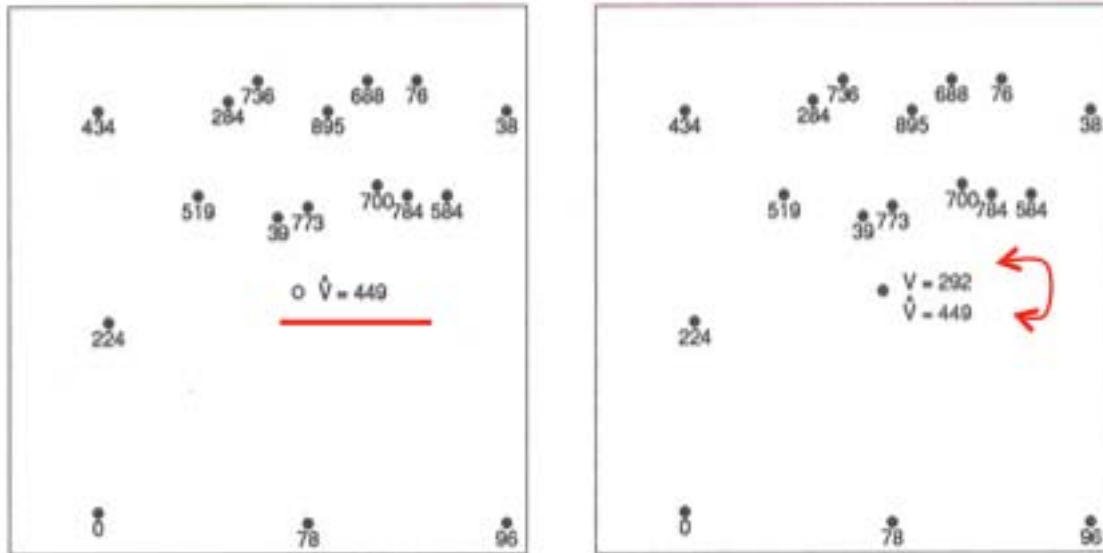
Figure 24: Compariosn between the real value and the estimate value.

How the figure 3 shows, in the second illustration where there was only the original datum we have calculated the datum using the kriging interpolation method, and we can have a comparison to understand how accurate is the method. One of the factors that limits the conclusions that can legitimately be drawn from a cross validation exercise is recurring problem of clustering

In fact, If our original sample data set is spatially too clustered, then so are our cross validated residuals. Therefore, some conclusions drawn from it may be applicable to the entire map area, others may not.

3.3.3. Limitations of Cross Validation Method

- CV can generate pairs of true and estimated values only at sample locations;
- Clustering problem in the sample data set;
- In practice, the residuals may be more representative of only certain regions or particular ranges of values.

Clustering problem can be overcome either by calculating declustered mean of residuals or by performing CV at a selected subset of locations that is representative of the entire study area. If very close nearby samples are not available in the actual estimation, it makes little sense to include them in CV

The problem areas identified by cross validation may warrant additional sampling, especially when there are major consequences.

Chapter 4: Atmospheric events in 2012

4.1. Introduction of the events

This work is going to analyse rainfall data collected in 2012.

Starting from the rainfall data of every rain gauge, we have in total 8 pairs of rain gauges, which take the value of the rainfall every minute, so it is possible to see the trend for the whole period of every rain gauge.

There have been some periods, to be more precise from the end of August 2012 until the end of 2012, in which not all the rain gauges were working because there have been some problems noticed while catching the data, which showed some problems of calibration. This occurred because there were building works on the library roof, which meant the rain gauges had to be moved out of the way of the building works, and some got damaged during the works. That is why for the last period it is possible to use the data of only 6 pairs of rain gauges.

It is possible to see the graphic in the following page that represents the trend of the daily rainfall for the whole period.

The location of the studies has not given the possibility to catch the data for all the year, but only for the period ranging from the beginning of April to the end of October. The reason of this condition is that the winter season is too cold and there is too much snow, and if it is not snowing, the rainfall is replaced by hail and also the rainfall inside the funnel and the bucket freezes.

This type of tipping-bucket rain gauge was unheated, and hence not able to catch these kind of data also because the hail usually bounces out of the bucket, and as regards the snow, the temperature is usually under 0 degree, so we have to wait for it to thaw; which means the evaporated part of the rainfall and information regarding real time of the event is lost.

Hence there are no accurate data for the winter period.

Watching the graph it is possible to notice that for the measurement period there are lot of rainy days, and there are only two weeks when it does not rain: the last week of May and the beginning of September. For the rest of the time, there are only one or two days without rain.

Dividing the analysing period in accordance with the seasons:

- Winter: 23rd of December to 20th of March;

- Spring: 21st of March to 21st of June;

- Summer: 22nd of June to 22nd of September;

- Autumn: 23rd of September to 22nd of December.

There are in average 830 mm of rainfall for the whole examined period, most of this during the spring and the summer, the following Table 1 shows the division for the period.

	Total	Spring	Summer	Autumn
[mm]	831.1	303.5	345.7	176.5

Table 3: Rainfall for the whole period and every season.

Most of the rainfall has fallen down during the summer and the first days of autumn and the last days of spring.

Every day in average there are 4 mm of rainfall, during the spring 1.4 mm, during the summer 1.7 mm and in the end, during the autumn only 0.8 mm.

The highest precipitation was on the 22nd of June when it rained 48 mm in 22 hours.
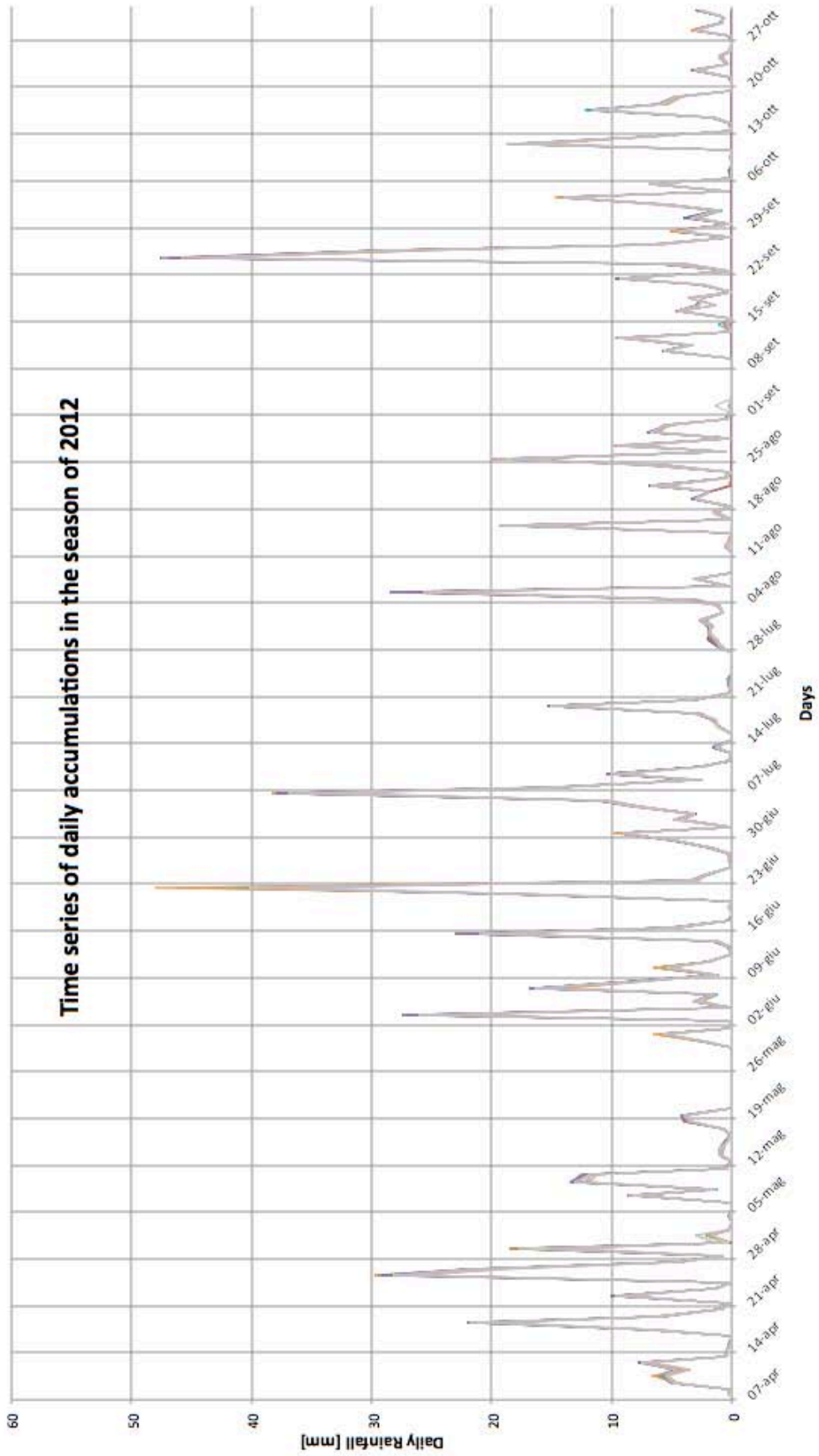
Figure 25: Trend of the rainfall for the whole period and every season.

4.2. Events analysed

To start our analysis we have chosen some events for every month according to the depth of the rainfall.

We have begun analysing the month of April and we have chosen three days: 25th, 26th and the 18th, which are characterized by 20 mm or more of rainfall.

Figure 26 shows with a histogram the trend of every rain gauge for every day, where every rain gauge is expressed by a different colour.
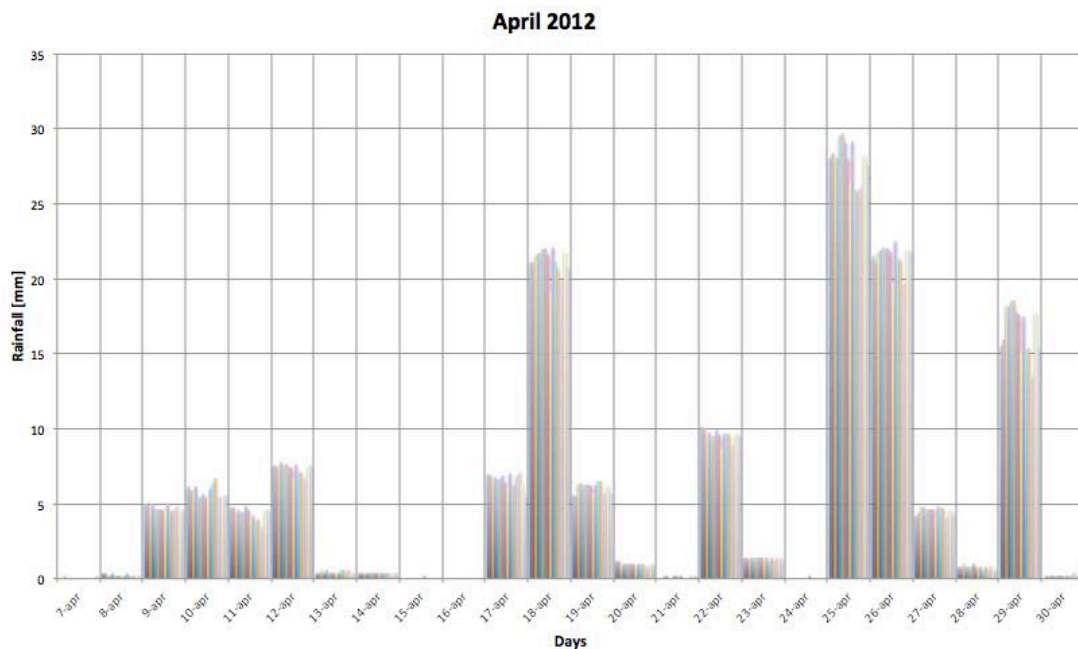
Figure 26: Trend of every rain gauge of every day of April

Using the same method, and watching the trend for all rain gauges for every month, we have chosen some events for every month.

The following graphs, figures 27 and 28, show the months of June and July. In the first, it is possible to notice better the event mentioned above: the 22nd of June, when there are more than 40 mm of rainfall.
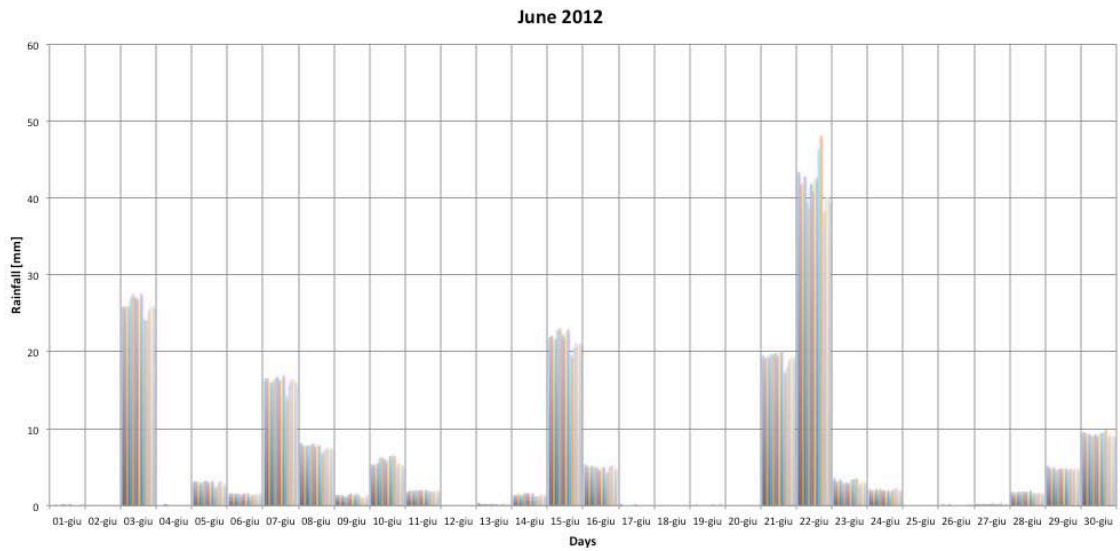
Figure 27: Trend of every rain gauge of every day of June

The second one, the graph of July, shows a high depth of rainfall for the 6th, but from the 2nd to the 10th we can see a group of small events.
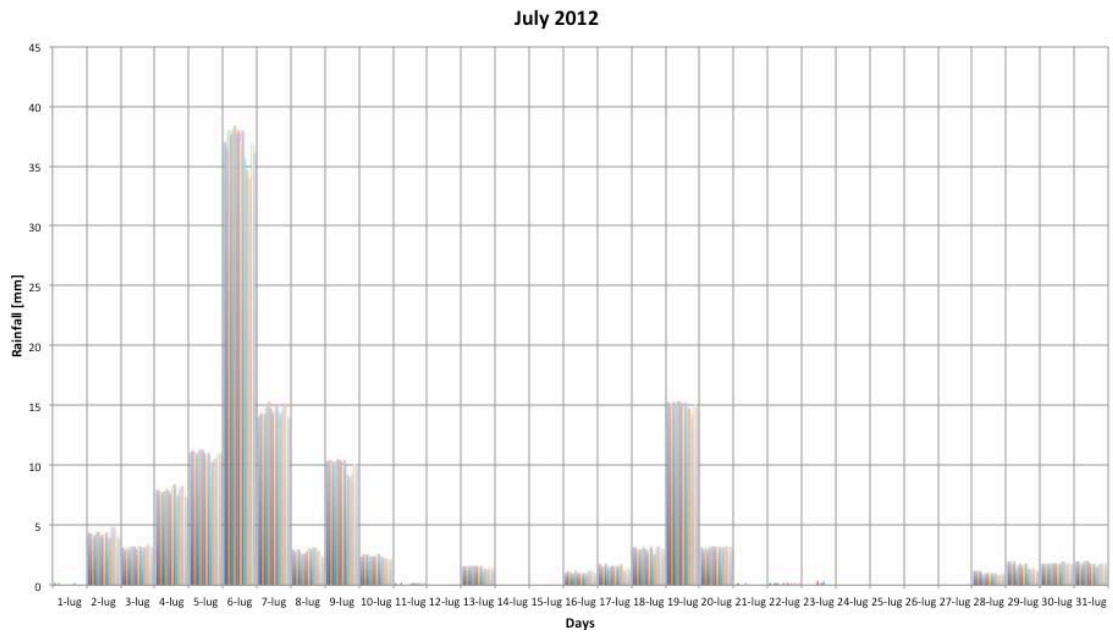
Figure 28: Trend of every rain gauge of every day of July

To conclude, the list of all the events is the following:

- Event 1: 25th of April for a duration of 13 hours;
- Event 2: 26th of April for a duration of 18 hours;
- Event 3: 9th – 10th of May for a duration of 13 hours;
- Event 4: 22nd of June for a duration of 22 hours;
- Event 5: 6th of July for a duration of 10 hours;
- Event 6: 5th of August for a duration of 4 hours;
- Event 7: 24th of September for a duration of 14 hours;
- Event 8: 3rd of October for a duration of 5 hours;
- Event 9: 11th of October for a duration of 8 hours;
- Event 10: 16th of October for a duration of 4 hours;
- Event 11: 18th of April for a duration of 12 hours;
- Event 12: 3rd of June for a duration of 20 hours;
- Event 13: 15th of June for a duration of 3 hours;
- Event 14: 15th of August for a duration of 4 hours;
- Event 15: 25th of August for a duration of 2 hours;
- Event 16: 25th of September for a duration of 22 hours.

4.3. Errors of rain gauges

Before starting the analysis, it is necessary to pay attention to downloaded data.

In fact, after the downloading, we have to check if there are some errors on the measurements of every rain gauge. This is the reason why we have 8 pairs of rain gauges and not only one rain gauge in every place, so it is easier to verify if there are some problems in one of those making a comparison between the elements of the pair.

As we have already observed in the chapter on the rain gauge, there is a range within which, during a period of use, the calibration factor could change and it could give a wrong data in output.

As it can be seen in table 4, the usual value of the calibration factor is ranging from 0,197 to 0,203 mm/Tip, over these ranges there are some problems, and in extreme cases we cannot use the data catched.

Range:	0.197 vs 0.203
Usual:	0.198 vs 0.202
Biggest:	0.196 vs 0.204
worst case:	0.195 vs 0.205

Table 4: range of an accurate calibration

So, if difference between rain gauge pairs is bigger, approximately 4.1% (it corresponds to the range 0,195 vs 0,205), then could be a small blockage or other random error, so we have to cut off these data from the analysis.

Figure 29 and 30 show the examination of the pair of rain gauges of the months of April and August, where there have been some problems.
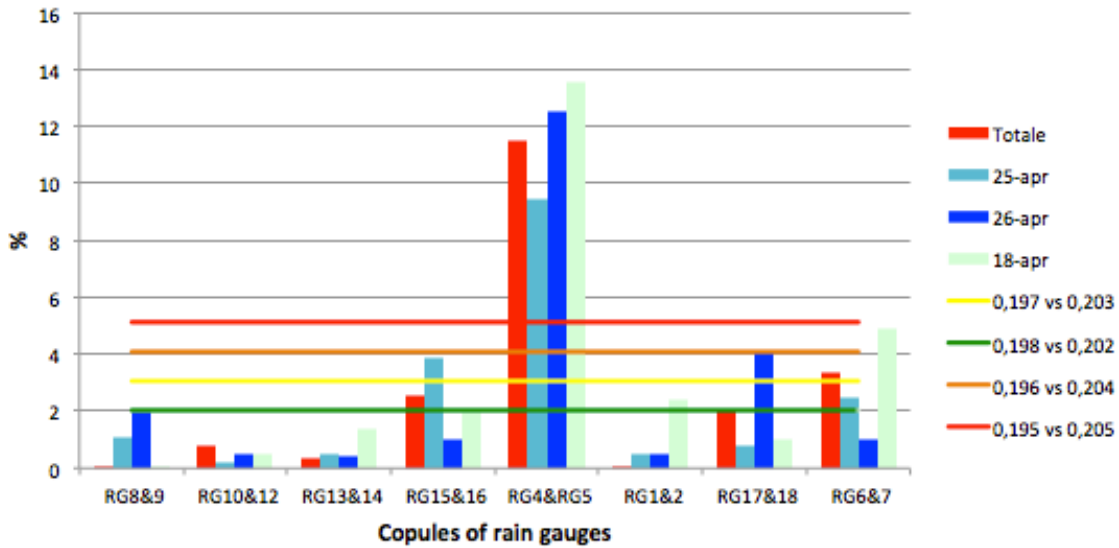
Figure 29: errors of the pair of rain gauges.

In this period all the rain gauges have a good behaviour and they are always near the first or second line (green line and yellow line), but for the couple of the rain gauges 4 and 5 there are always problems during the whole period. One of these is probably badly calibrated and it is not a blockage, because the problem persists for the whole period and it is not characteristic of only one event.
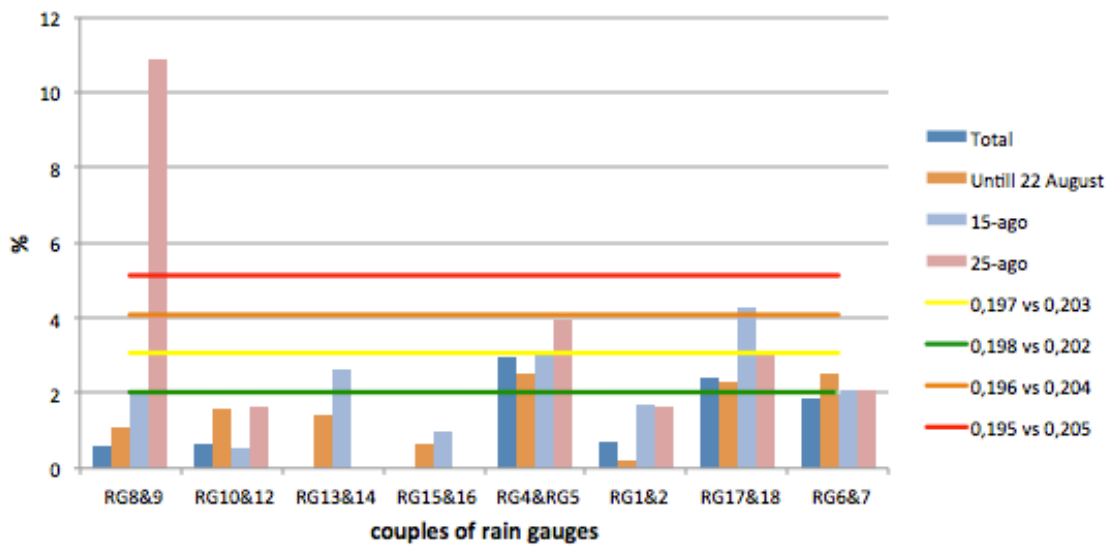
Figure 30: errors of the pair of rain gauges.

Figure 30 shows the period of August, where there is also a problem in one pair; it is not the same of Figure 29 but it is the pair of rain gauges 8 and 9.

This condition is different from the other one, in that it could be only a temporary blockage of one rain gauge, because the problem is for only one event as opposed to the situation described above.

During the analysis we cannot consider the data of those rain gauges, which have registered an error.

4.4. Intensity

After choosing the events, there is one more characteristic to analyse: the intensity of all the events. In this chapter we cannot show any reference value, but in the following ones, by studying the correlation of rain gauges, it will be possible to explain the behaviour of the events.
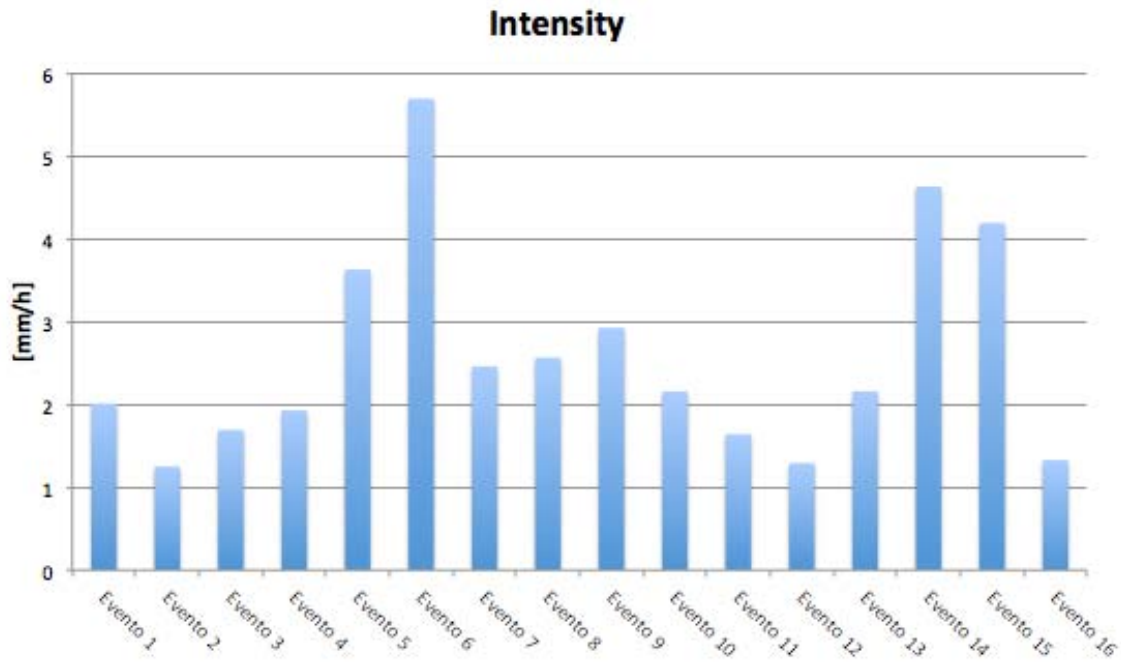
Figure 31: Intensity of all the events

Chapter 5: Analysis of correlation between rain gauges

This chapter analyses the correlation between the rain gauges, studying the different behaviour of each pair of rain gauges, and how the correlation between pairs of rain gauges changes with different typologies of rainfall events and with the distance between the rain gauge pairs.

The axes of the diagram, which will explain the trend, represent the correlation on the vertical axis, which is going from zero where there is not correlation, to one where there is perfect correlation between the values; on the horizontal axis there is the distance between rain gauges in kilometres.

Every event will be analysed for a range of time scales, so it is possible to see how the correlation changes with different cumulative time scales; starting with the cumulative time scale of 5 minutes followed by 6, 10, 12, 15, 20 and 30 minutes respectively, as well as 1, 2 and 3 hours.

Before starting it is necessary to notice that we are using the character "," instead of "." on the correlation graphs, to express decimal numbers.

5.1. Correlation analysis

The first step consists in studying the correlation between each single rain gauge and the others; at this stage we are not using the second rain gauge in the same position, because we have a pair of rain gauges in every location, and we do not consider the correlation between the couple, since it will always be near to 1.

We are using the data without the errors and the results that we obtain are showed in the following graphs.

The first graph, figure 32, shows the state of the event 2; the different colour points indicate the different cumulative time scales as the legend suggests.

It shows how for the lowest cumulative time scale, which is of 3 minutes, the correlation is low and it is near the 0.5 and is going down with the increasing of the distance; the second trend line of the cumulative time scale of 6 minutes shows that the correlation is better than before, and it is near 0.65 and it is always going down with the increasing of the distance.
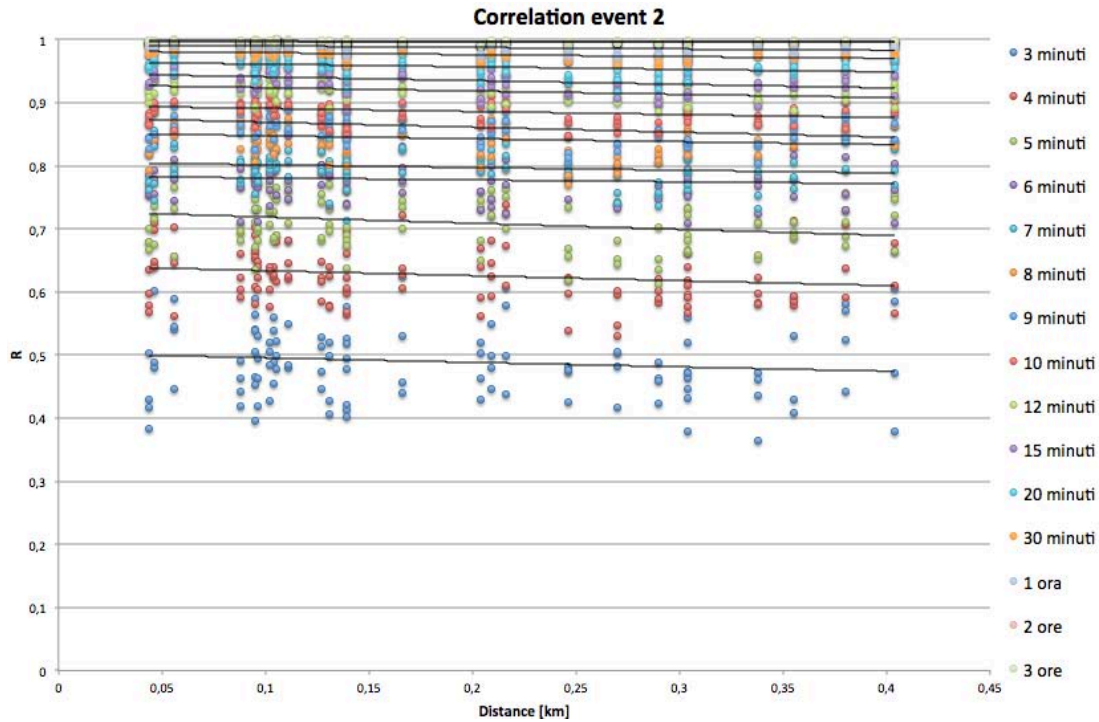
Figure 32: Correlation of event 2 – 26/04/2012, where there are cumulative temporal scales from 5 minutes to 3 hours.

To find a good state of correlation we have to observe the cumulative time scale of 7 minutes, which is on the line of 0.8 and it is a good line for the correlation.

The best one that we can use for the estimation data in this event is the cumulative time scale of 10 minutes, which is on the line of 0.9 and over this line there is a very high correlation.
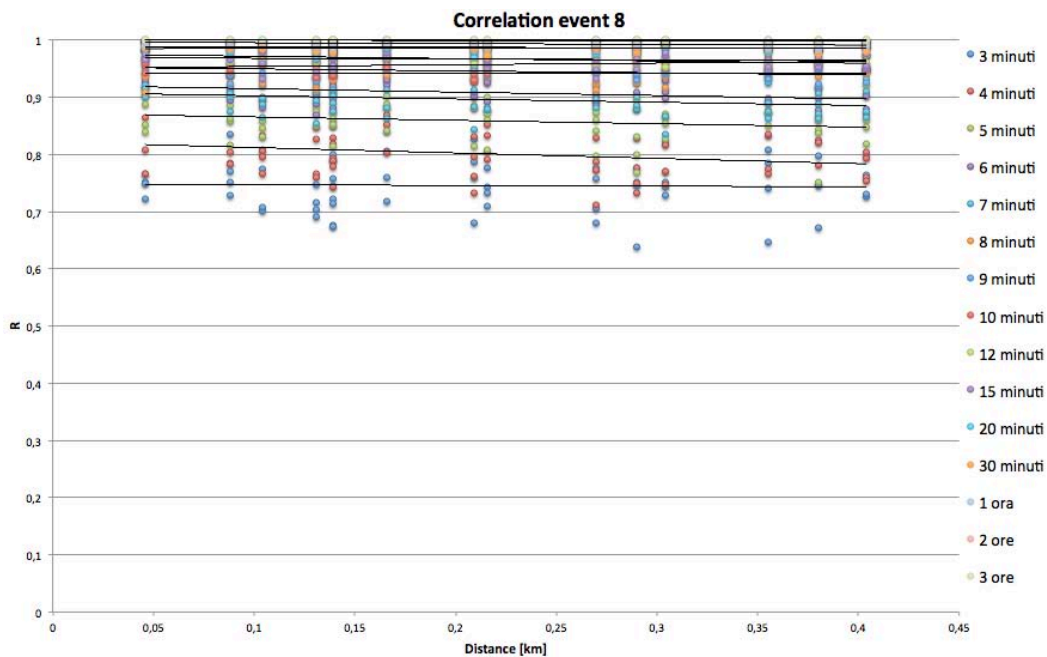
Figure 33: Correlation of event 8 – 3/10/2012.

The second graph, in figure 33, shows the state of the event 8, which represents a different behaviour in comparison with the event 2.

At the beginning we already have a rather high correlation. In fact, the cumulative time scale of 3 minutes is already between the values of 0.7 and 0.8 of correlation. We had to wait for the cumulative time scale of 10 minutes to have a good degree of correlation for the event 2, whereas in the event 8, the cumulative time scale of 6 minutes is enough to be on the 0.9 line.

The last graph of correlation in figure 34, which represents the event 14, shows a third different state of the correlation.

In fact, we have a new different behaviour in comparison with the other events.

It has the best behaviour because all the trend lines are above the 0.9 line of the correlation, so there is a good chance to have a very good estimation for the data also for the lowest cumulative time scale.
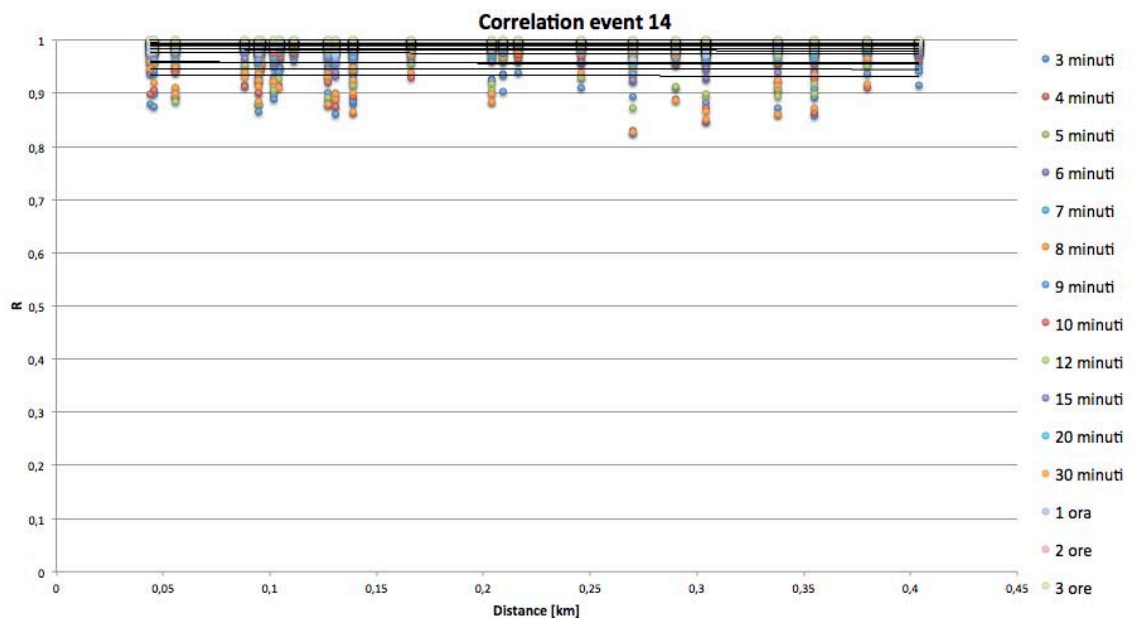
Figure 34: Correlation of event 14 – 15/08/2012.

5.2. Group division

In the section above (5.1.), we have described three different behaviours of the events, which represent the three groups in which we can insert all the events.

In fact, we have started the analysis studying sixteen events, but we cannot consider the event 7 and the event 16 because there are too many problems with the rain gauges in that period that is September, so it may have been due to the same problem. The rest of the events have been studied and the results showed that each one of them could be included in the three groups as showed above.

The studies indicate that the three groups can be divided in the following way:

- ➢ group A: event 1, 2 ,3, 4, 11 and event 12;
- ➢ group B: event 8, 9, 10 and event 13;
- ➢ group C: event 5, 6, 14 and event 15.

At the end of this chapter, it is possible to find all the graphs of all the events in order to make a comparison.

At this point, the aim is to find what these events have in common, in order to elaborate an objective method to insert a general event into one of these groups.

The first thing that we have considered is the period of occurrence, because there might be a link between the events. In fact, at the moment there is no accurate method of dividing them, but the seasons give a first idea of the group of they belong to.

As regards the method of division into groups according to the seasons, we can see that the events of group A happen during the spring or over the days immediately before or after the last days of spring, as the event 4 shows, which happens on the 22nd of June; the events of group B occur in autumn and those of group C can be observed in the summer.

A more objective method to divide the events into groups could be analysing the intensity of the rainfall, so now we are going to discuss more thoroughly the values of the intensities that we have already analysed in chapter 4.

Starting from the total cumulative of every event we make a ratio between it and the duration of the event.

The results give us three different groups appropriately selected. The tables below show the range values of the intensity according to which there is a division of the events into three groups.

Range	Group
0-2 [mm/h]	A
2-3 [mm/h]	B
over 3 [mm/h]	C

Table 5: group division based on the range of the intensity.

	Intensity [mm/h]	Groups
Event 1	2,010	A
Event 2	1,249	A
Event 3	1,690	A
Event 4	1,927	A
Event 5	3,626	C
Event 6	5,687	C
Event 8	2,563	B
Event 9	2,924	B
Event 10	2,157	B
Event 11	1,640	A
Event 12	1,290	A
Event 13	2,158	B
Event 14	4,627	C
Event 15	4,189	C

Table 6: values of intensity for each event and the groups they belong to.

The figure 35 shows the intensity values expressed in the table 6; the colours help to give a faster perception of the division of the events. The yellow one indicates the events which belong to group A, the pink one is for group B and the last one, the blue one, for the events belonging to group C.
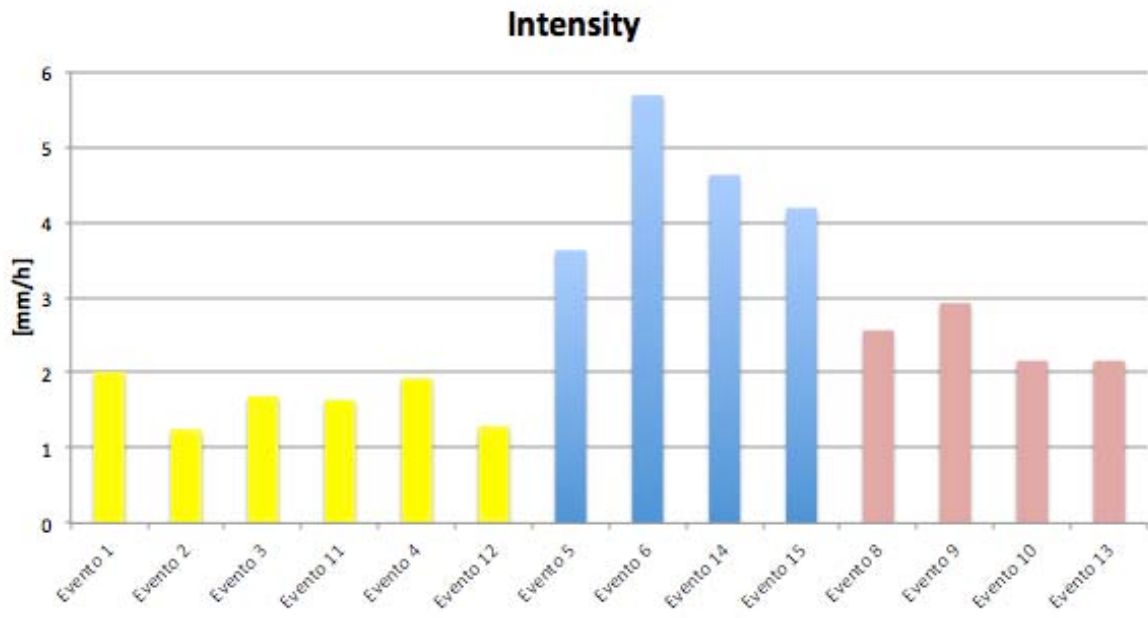
Figure 35: Intensity of every event and group division.

5.3. First considerations

After this first step, where we have studied the correlation for sixteen events starting from the data value of every single rain gauge correlated with the others, without using the data values of the other rain gauge of the pair in the same position, we have seen that all the correlations of all the events follow three different trends of behaviour. As concerns these three trends and the corresponding groups, the first one has the worst behaviour because the trend line that interpolates the lowest points starts from the 0.5 line of correlation and we have to use the cumulative time scale of 10 minutes to have a good degree of correlation; for the other two groups there is a better trend and the second one gave us a good correlation data starting from 6 minutes of cumulative temporal scale; the last one, which is the best one, is always over the 0.9 line of correlation also for the 3 minutes of cumulative time scale.

We noticed that these three groups could be divided according to the season of occurrence and the intensity, which offer a first indication about the behaviour of the events.

Event	Date	Group	Intensity range [mm/h]	Correlation at 5 min timescale		Correlation at 15 min timescale	
				100 m	400 m	100 m	400 m
1	25/04/12	A	0-2	0,72572	0,68010	0,96108	0,93261
2	26/04/12	A	0-2	0,71154	0,66392	0,93612	0,91131
3	09/05/12	A	0-2	0,61103	0,60003	0,91058	0,89209
4	22/06/12	A	0-2	0,93523	0,70552	0,93523	0,88581
5	06/07/12	C	over 3	0,94658	0,93643	0,98241	0,97964
6	05/08/12	C	over 3	0,97073	0,95129	0,99904	0,99440
8	03/10/12	B	2-3	0,89851	0,84753	0,97731	0,97590
9	11/10/12	B	2-3	0,92769	0,91276	0,98612	0,96609
10	16/10/12	B	2-3	0,71628	0,61743	0,96102	0,93851
11	18/04/12	A	0-2	0,72578	0,66711	0,94023	0,93787
12	03/06/12	A	0-2	0,45295	0,38179	0,84144	0,79988
13	15/06/12	B	2-3	0,91019	0,78009	0,92388	0,84761
14	15/08/12	C	over 3	0,94649	0,97689	0,98936	0,99822
15	25/08/12	C	over 3	0,96492	0,78691	0,97828	0,72877

Table 7: Summary of the events, with their group of belonging, date, intensity range and the values of correlation for 3 and 15 time scales at the distances of 100 and 400 metres.

5.4. Correlation analysis for average values

In this section we are going to study the behaviour of the correlation considering the average values of the rain gauge couple. So, if before we had sixteen rain gauges and we have considered each one independently, now we have eight rain gauges placed in the average coordinates of the original position of the two rain gauges.

We are going to examine the same three events that we have studied in section 5.1., where we used the correlation counting all the rain gauges.

The first graph in figure 36 regards the event 2, which belongs to group A of division. The trend of correlation is going better than before and all the trend lines have increased their degree of correlation. In fact, there is a good value of correlation for the cumulative time scale of 7 minutes, which is already on the line of 0.9.

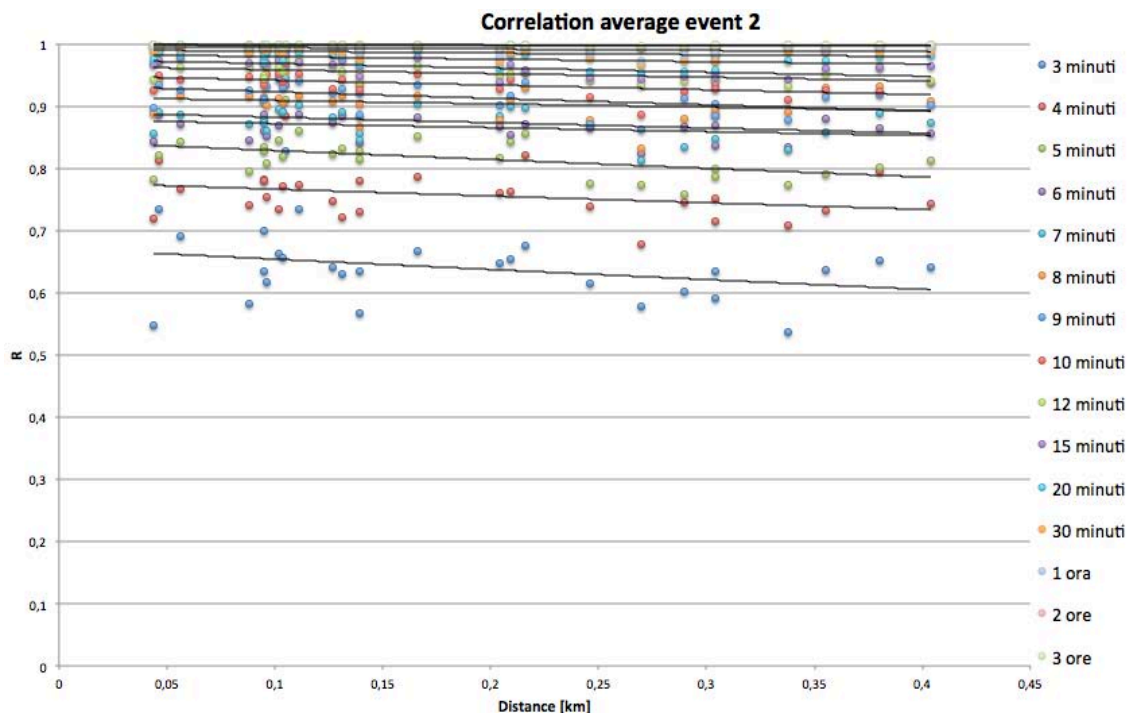In comparison with the trend of the results obtained before, there is an improvement of the method.

Figure 36: Correlation of the event 2 – 26/04/2012, using the average data value of the pairs of rain gauges.

The second graph in figure 37 regards the event 8 and also here we have a better trend. The first line related to the cumulative time scale of 3 minutes is already

66

between the 0.8 and 0.9 lines, and with the cumulative time scale of 4 minutes we are over the 0.9 line.
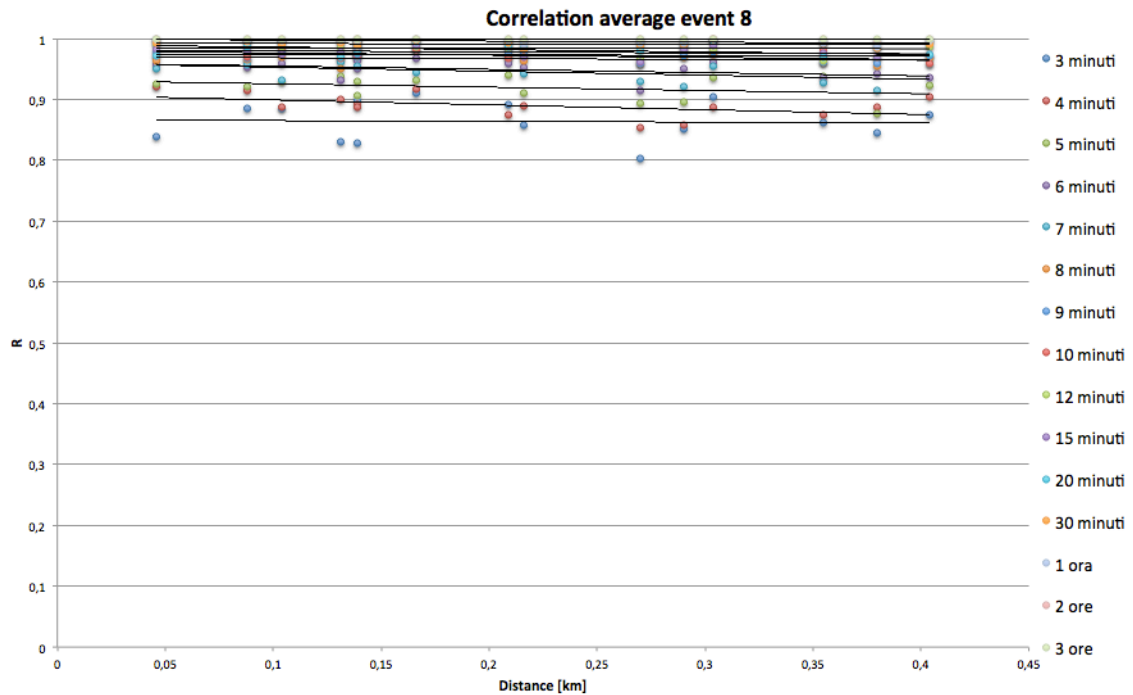
Figure 37: Correlation of the event 8 – 3/10/2012, using the average data value of the pairs of rain gauges.

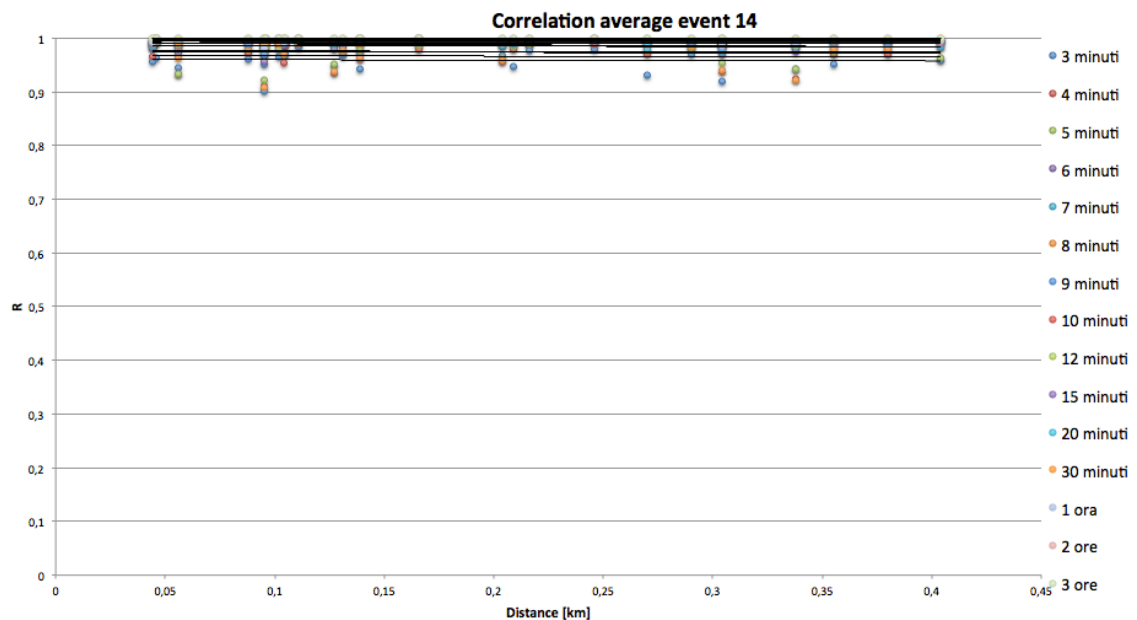*Figure 38: Correlation of the event 14 - 15/08/2012, using the average data value of the pairs of rain gauges.*

The last graph in figure 38 regards the event 14, which belongs to group C. It is not easy to establish if there is also an improvement of the position of the trend lines, but according to what we have seen in the other two graphs, figures 36 and 37, we can conclude that it should have the same improvement.

5.5. Comparison between the two analysis methods

5.5.1. Comparison between the two methods for the Event 8

To be more specific, we can make a comparison between the same events analysed with the different input values to see graphically how they change.

What we are going to explain will be referring to the graph below, which is a comparison between the two methods of analysis concerning the event 2 examined in the previous sections. The reason why we have chosen this event is because it had the best improvements.

There are only four cumulative time scales in order to have a clearer view of what is happening.

Starting with the cumulative time scale of 3 minutes it is possible to notice how the red line, which indicates the trend of the average data, is upper than the black line, which indicates the trend of single rain gauge, and there is an improvement of 0.15 of the correlation. There is also an improvement for the other trend lines, but the range is reduced, i.e. there is always a better trend with the red lines but the range of improvement decreases with an increase of the cumulative time scale.
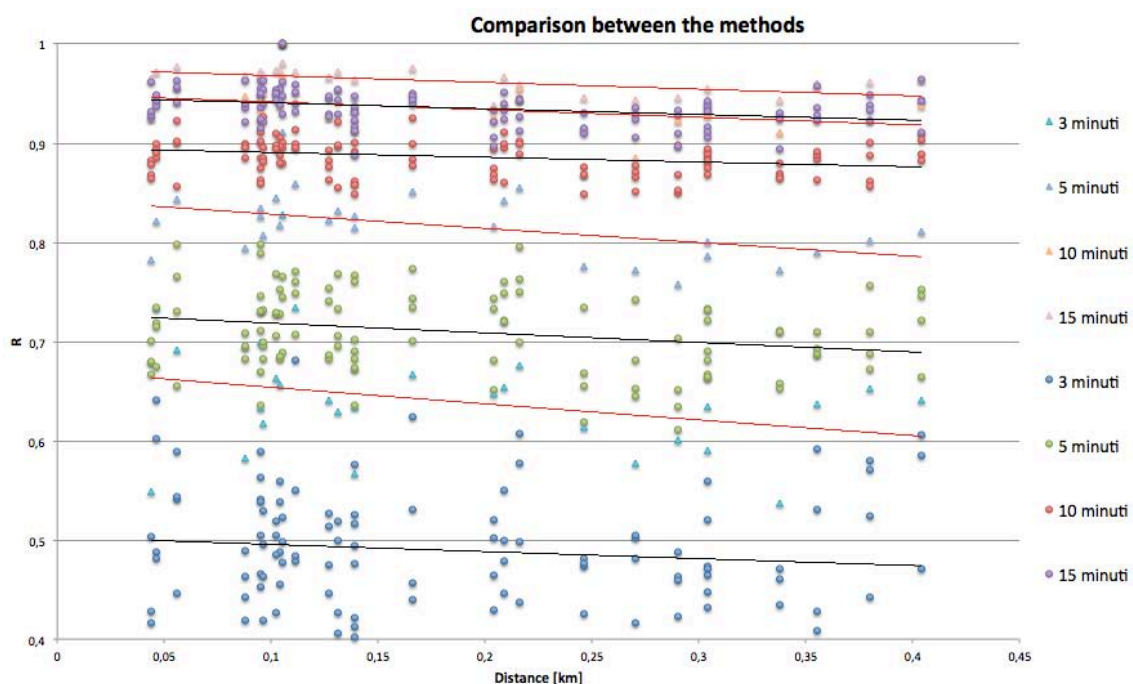

Figure 39: Comparison between the correlations of the event 2, using the two methods of analysing.

5.5.2. Comparison between the two methods for the Event 8

The following graph in figure 40 shows the behaviour of the correlation for the event 8 and for the all events, which belongs to group B.

As we have already examined in the section above for the event 2, we can notice an improvement of the correlation degree.

In fact, watching the trend lines, the black line indicate the trend of the points which are provided from a single rain gauge, the red line on the other hand indicate the trend of the points from the average of the rain gauge couples; the increase of the correlation degree is bigger for the smallest cumulative time scales and it smaller for the higher cumulative time scale.
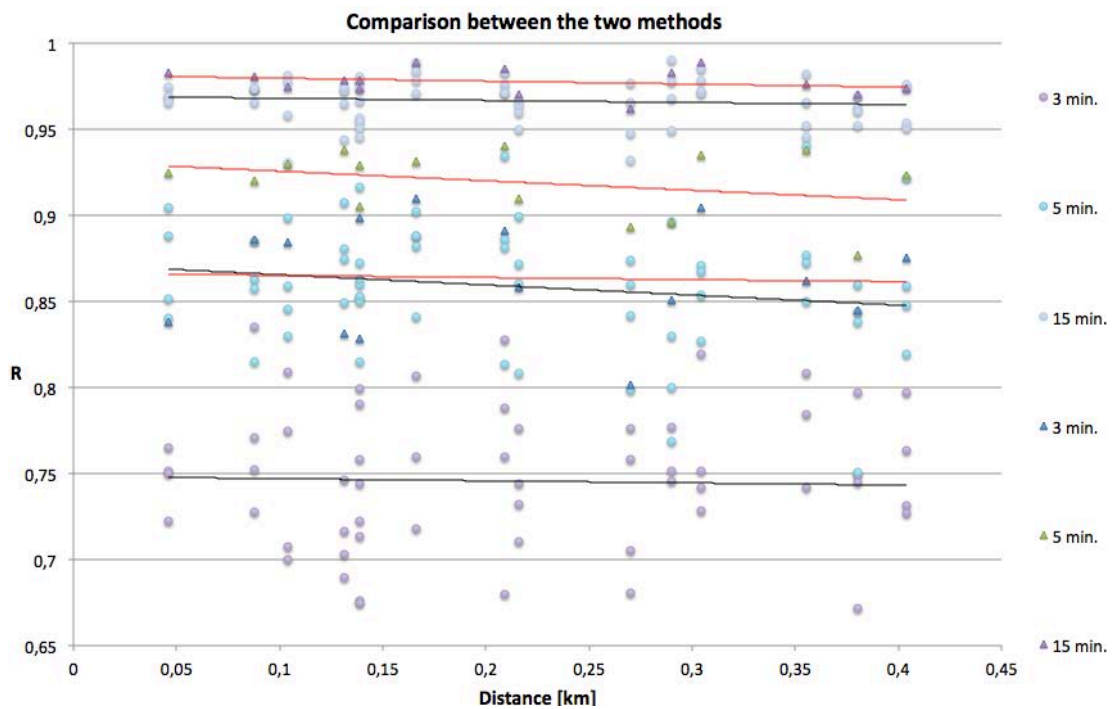

Figure 40: Comparison between the correlations of the event 8, using the two methods of analysing.

The values are:

- cumulative time scale of 3 minutes: R = 0.75 → R =0.87;
- cumulative time scale of 5 minutes: R = 0.85 → R =0.93;
- cumulative time scale of 15 minutes: R = 0.97 → R = 0.98.

5.5.3. Comparison between the two methods for the Event 14

These increases of the correlation degree become lower if we examine the events belonging to the group C, as the graph in figure 41 shows.
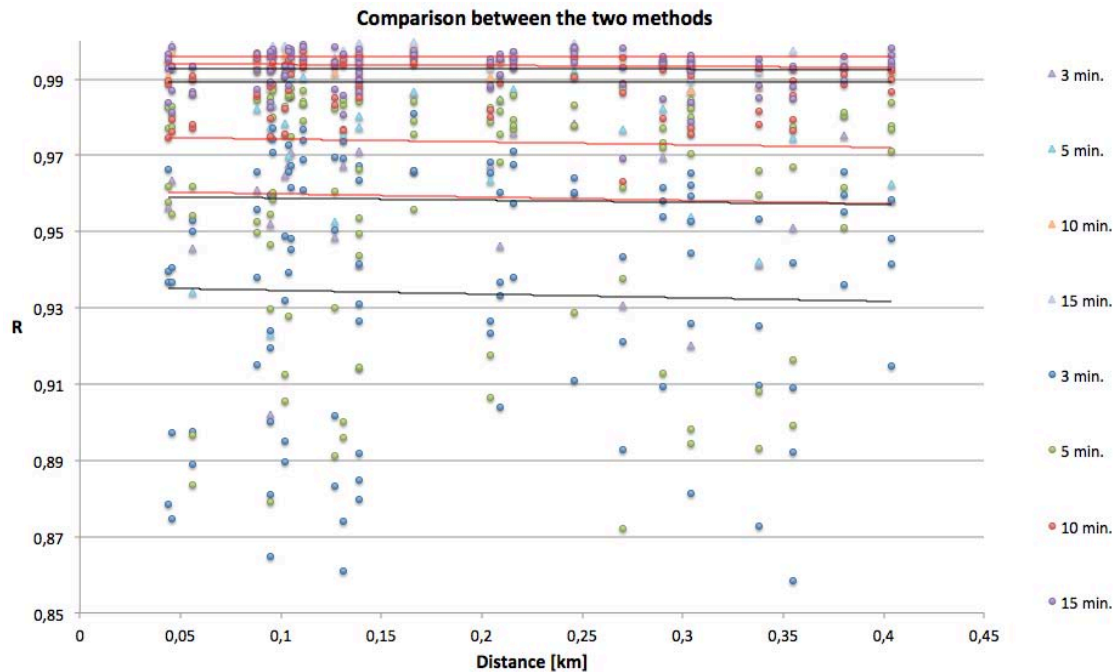
Figure 41: Comparison between the correlations of the event 14, using the two methods of analysing.

The vertical axis is starting from a higher value than the other two events examined, in fact, it starts from 0.93 that is already a good correlation degree.

So the increase of this method for these types of events it is not so useful.

The values of correlation degree are:

- cumulative time scale of 3 minutes: R = 0.93 → R = 0.96;
- cumulative time scale of 5 minutes: R = 0.96 → R = 0.975;
- cumulative time scale of 10 minutes: R = over 0.99;
- cumulative time scale of 15 minutes: R = over 0.99.

5.6. Considerations

In the end we can say that there is a great difference between the two models, though it is shown only for a group of events. In fact, the events belonging to the group C are not characterized by a great difference in their behaviour, because there is a high correlation in both models, but the advantage of the average data model is clear for the events of group B and even more so for the events of group C. The reason why there is this increase is due to the competence to reduce the errors of the single rain gauge. In fact, in our study area there are a lot of other effects, which exasperate the errors of the measurement, for example the wind that is really strong for the most of the year.

Using more than one rain gauge, there is more possibility to decrease the errors of measurement, so this method provides a higher degree of correlation.

5.7. Correlation analysis for average values of four rain gauges

In this section we are going to study the behaviour of the correlation considering the average values of the four rain gauges less distant. So, if before we had sixteen rain gauges and we have considered each one independently, and in the second step we have eight rain gauges placed in the average coordinates of the original position of the two rain gauges, now we have four clusters of rain gauges. We are going to examine the same three events that we have studied in section 5.1., where we used the correlation counting all the rain gauges. In comparison with the trend of the results obtained before, there is an improvement of the method, which is showed in the following section.

The first graph in figure 42 shows the event 2 of the 26/04/2014, and we expected the best improvement for these types of the events.
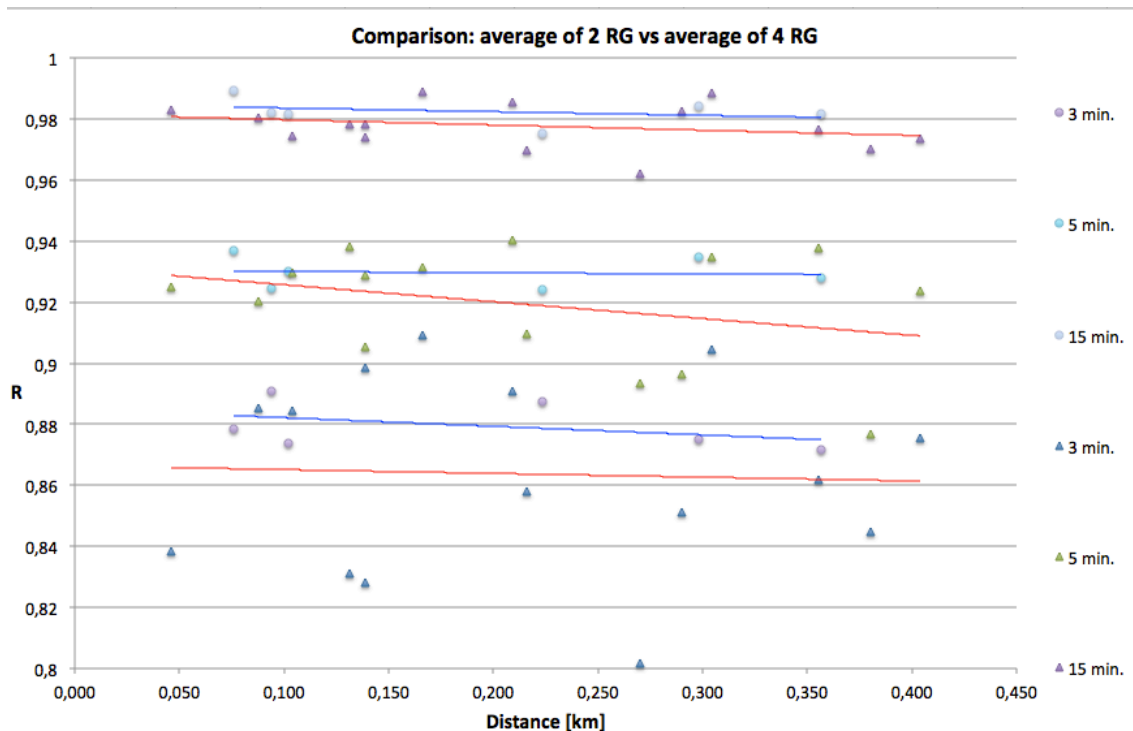
Figure 42: Comparison between the correlations of the event 2, using the second and the third method of analysing.

We can see another improvement of the correlation trend lines, which are represented by the blue line for the average of the four rain gauges and by the red line for the average of the couple of the rain gauges.

The values of correlation degree are:

- cumulative time scale of 3 minutes: R = 0.87 → R =0.88;
- cumulative time scale of 5 minutes: R = 0.93 → R =0.935;
- cumulative time scale of 15 minutes: approximately the same value.

We are closer to the 0.9 degree than before, even if the increase is not so big.

Using a group of 4 rain gauges we are able to defeat more than 0.4 of the errors for the correlation degree than using only one rain gauge.

It is not necessary to show how this method is able to change the correlation degree for the other two events (the event of group B and group C) because the best results are shown on the first graph, where there were the strongest problems.

As we have already examined in the section 5.4, the method of the average of the couples is enough to provide a good value of rainfall data for the events of the group B and C.

5.8. Graphs of correlation analysis for values of single rain gauge

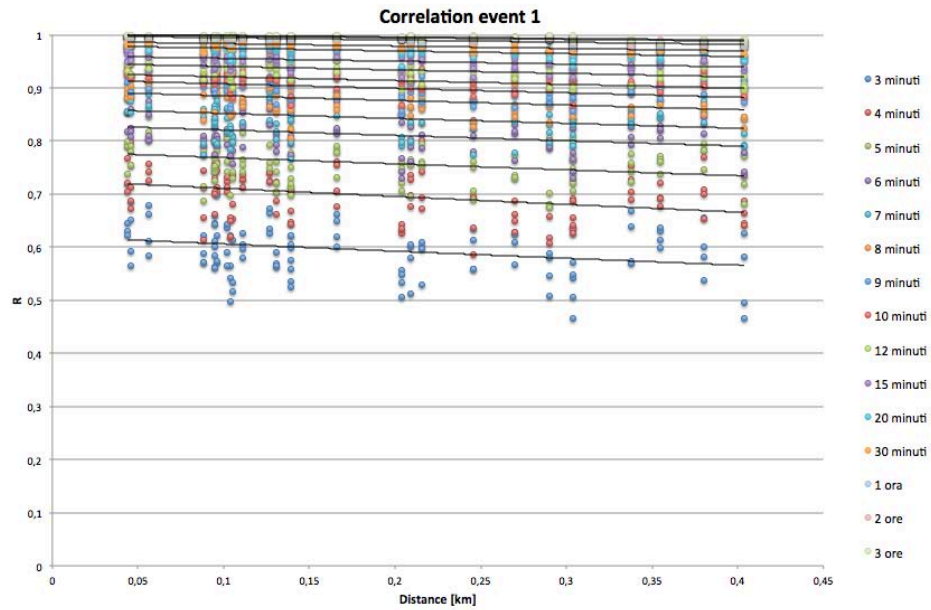### 5.8.1. Events belong Group A

Figure 43: Correlation of the event 1 – 25/04/2012.

Figure 44: Correlation of the event 3 – 10/05/2012.

Figure 45: Correlation of the event 4 – 22/06/2012.
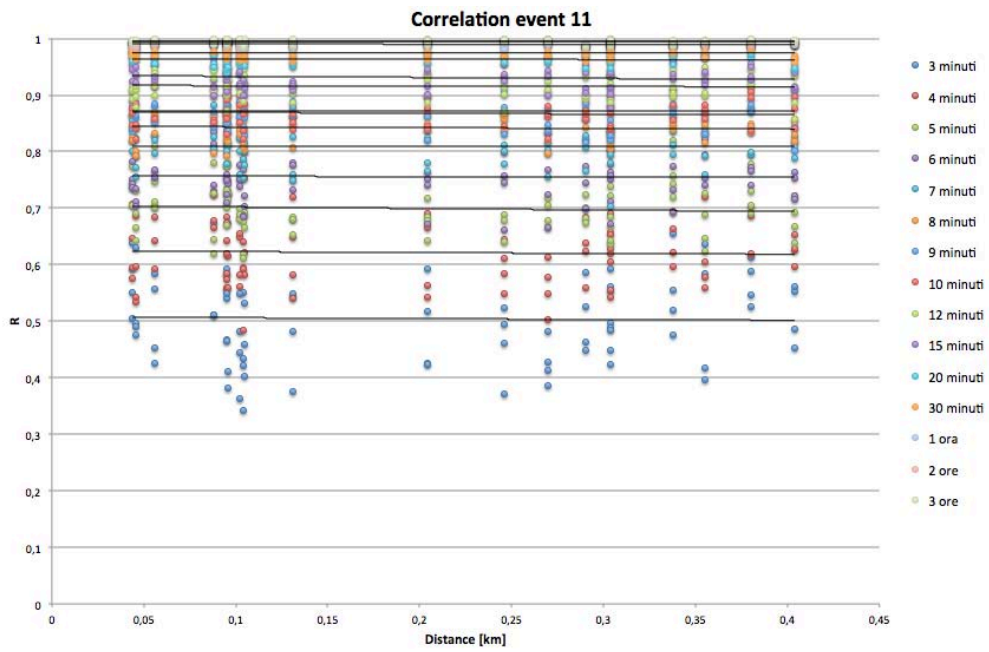
Figure 46: Correlation of the event 11 – 18/04/2012.
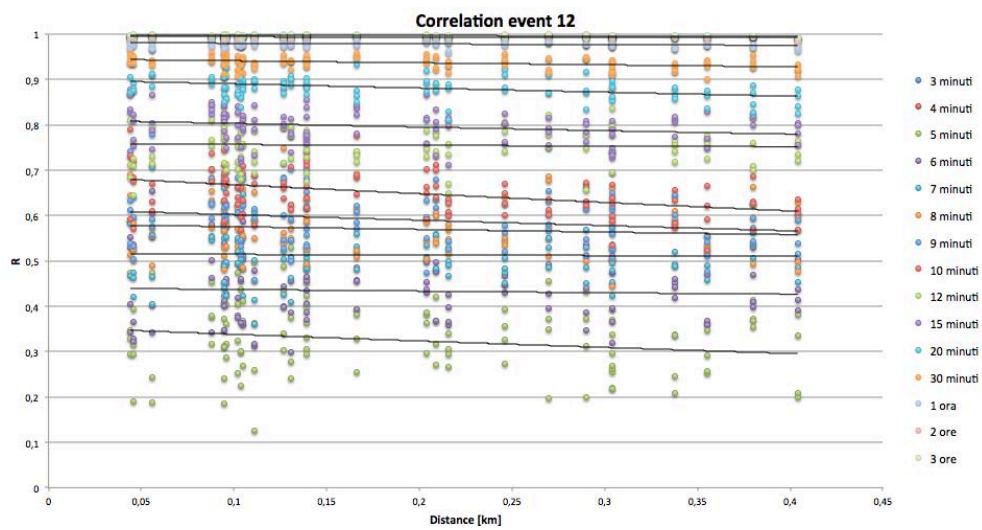
Figure 47: Correlation of the event 12 – 03/06/2012.

5.8.2. Events belong Group B
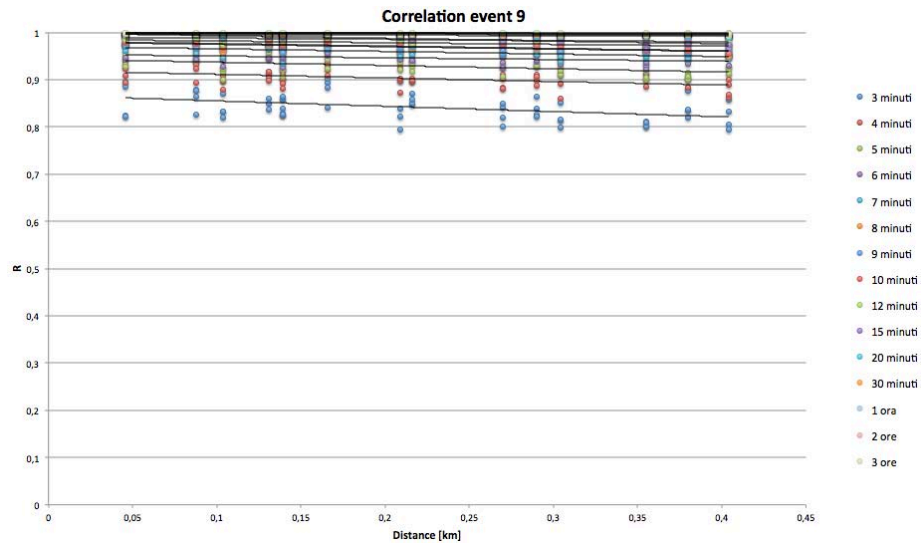
Figure 48: Correlation of the event 9 – 11/10/2012.
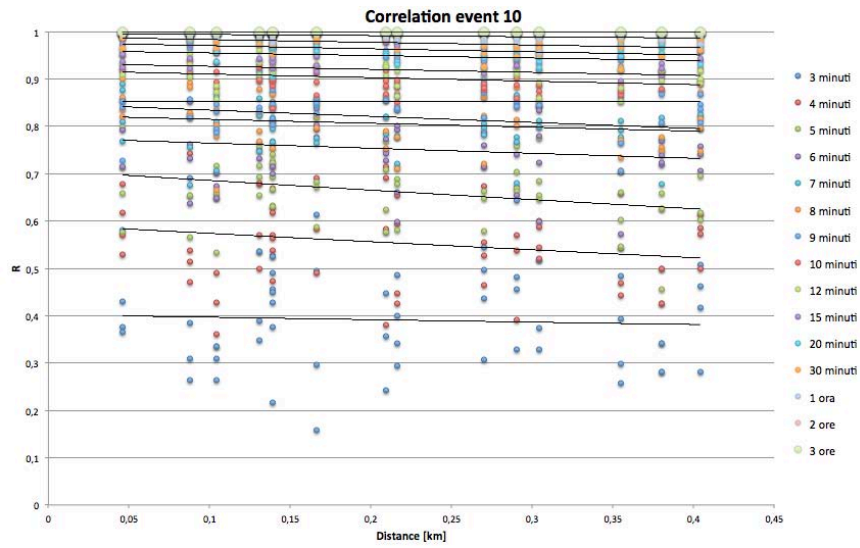
Figure 49: Correlation of the event 10 – 16/10/2012.
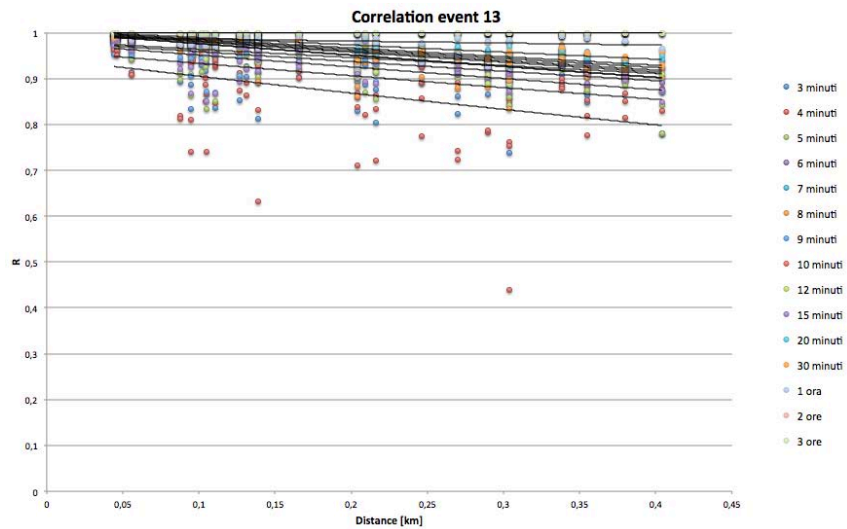
Figure 50: Correlation of the event 13 – 15/06/2012.

5.8.3. Events belong Group C

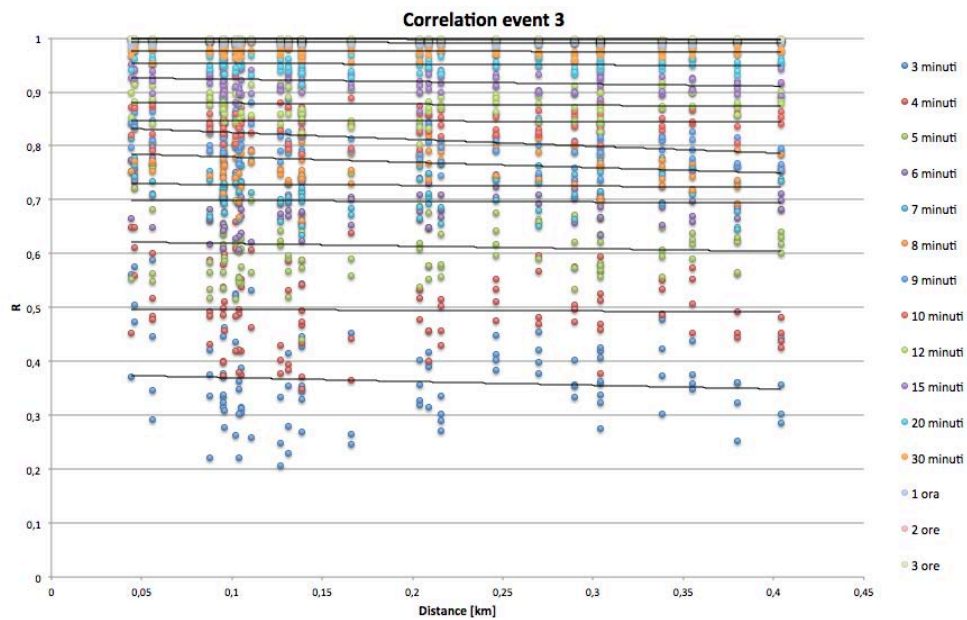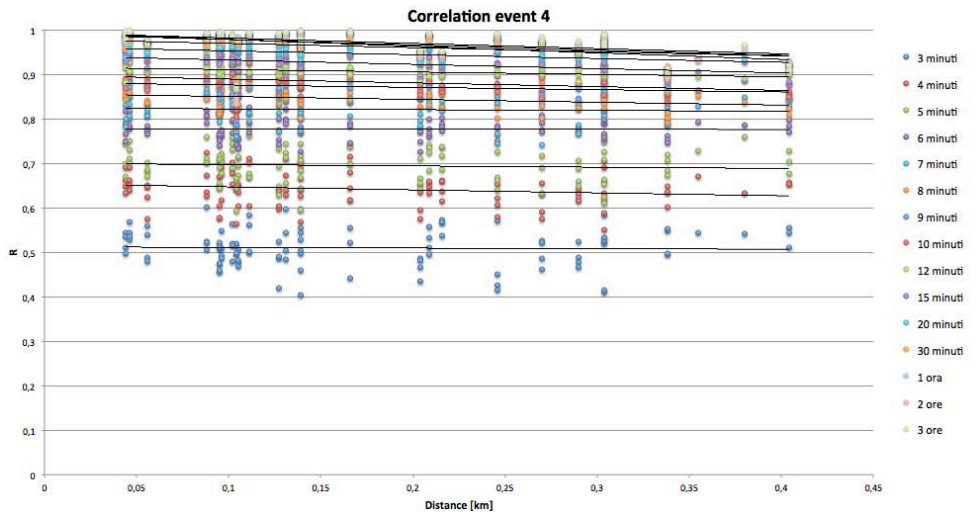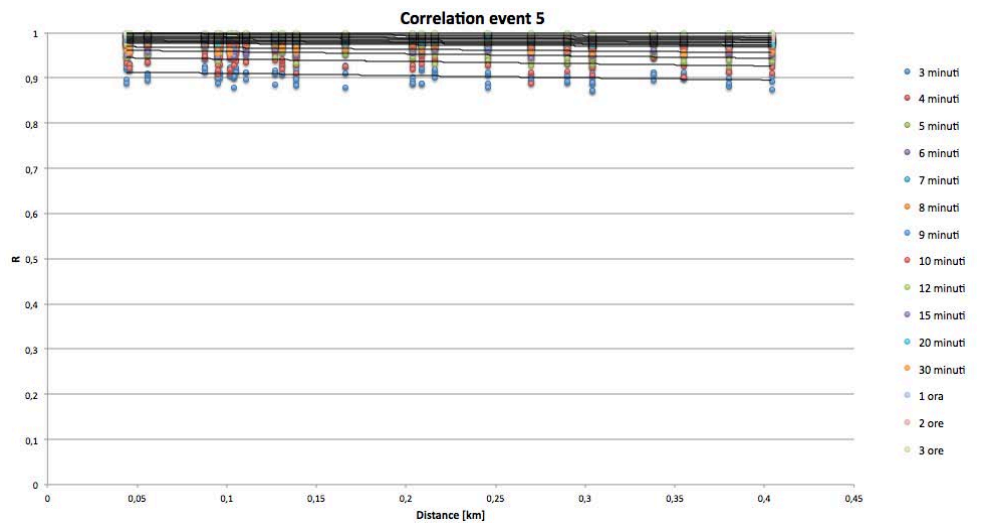*Figure 51: Correlation of the event 5 – 06/07/2012.*
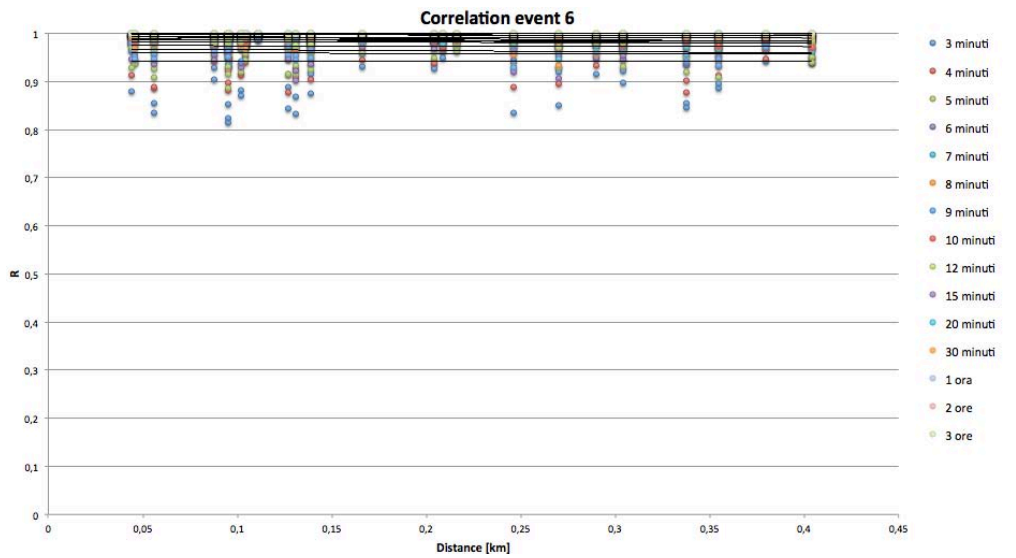
Figure 52: Correlation of the event 6 – 05/08/2012.
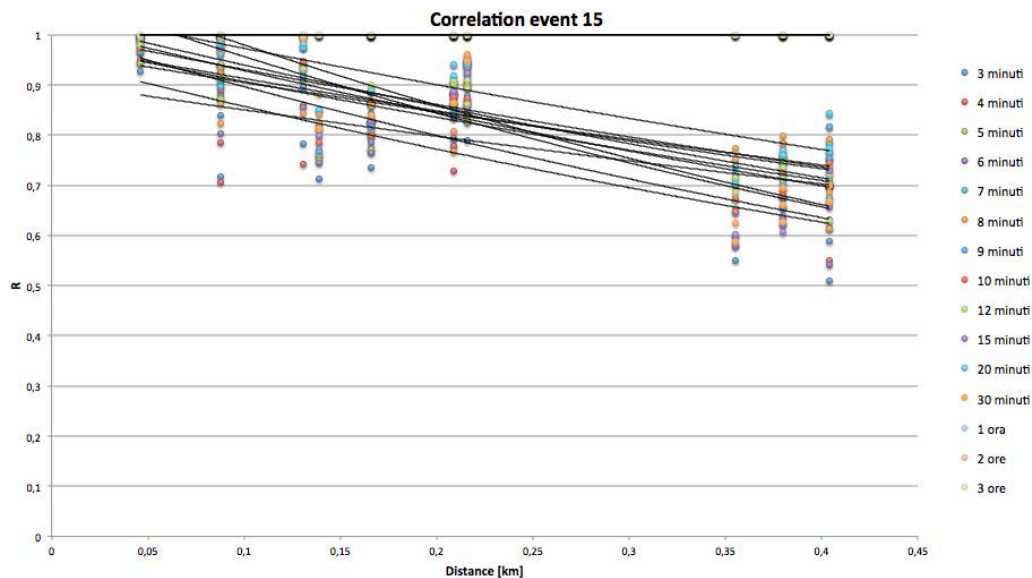
Figure 53: Correlation of the event 15 – 25/08/2012.

5.9. Graphs of correlation analysis for average values

5.9.1. Events belong Group A

Figure 54: Correlation using the average data for the event 1 – 25/04/2012.

Figure 55: Correlation using the average data for the event 3 – 10/05/2012.

Figure 56: Correlation using the average data for the event 4 – 22/06/2012.

Figure 57: Correlation using the average data for the event 11 – 18/04/2012.

Figure 58: Correlation using the average data for the event 12 – 03/06/2012.

5.9.2. Events belong Group B
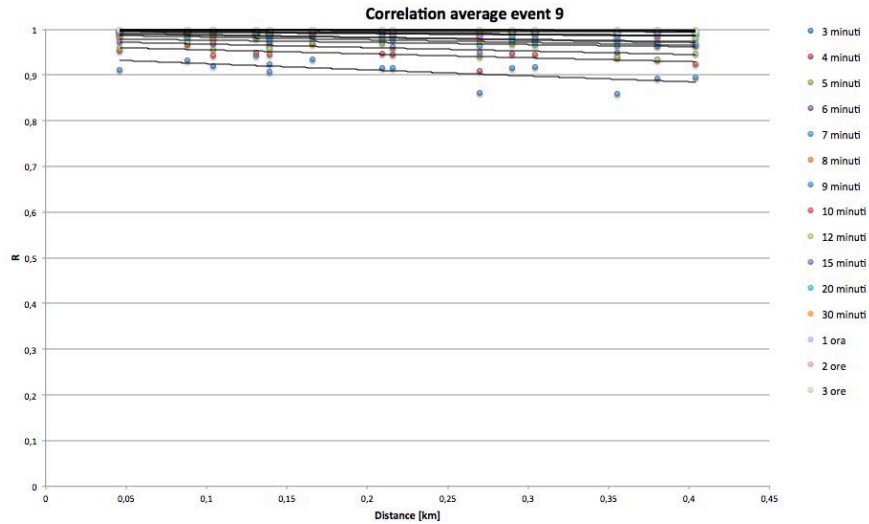
Figure 59: Correlation using the average data for the event 9 – 11/10/2012.
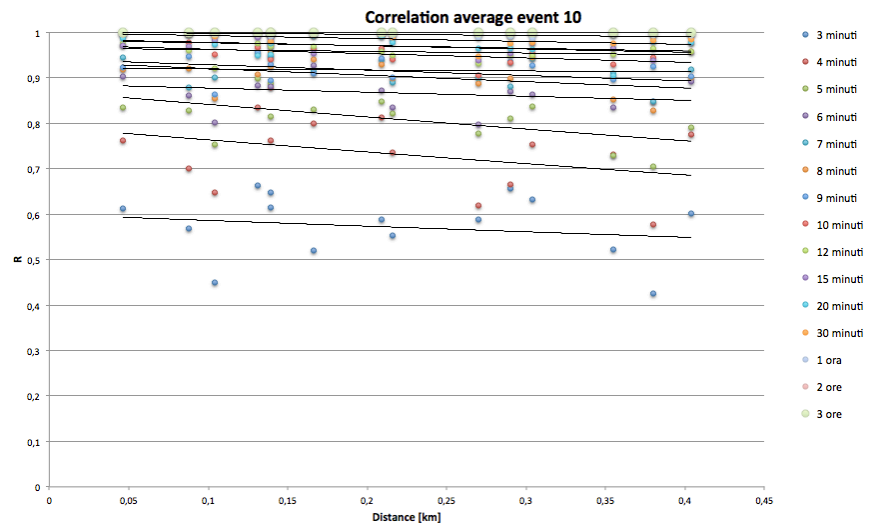
Figure 60: Correlation using the average data for the event 10 – 16/10/2012.
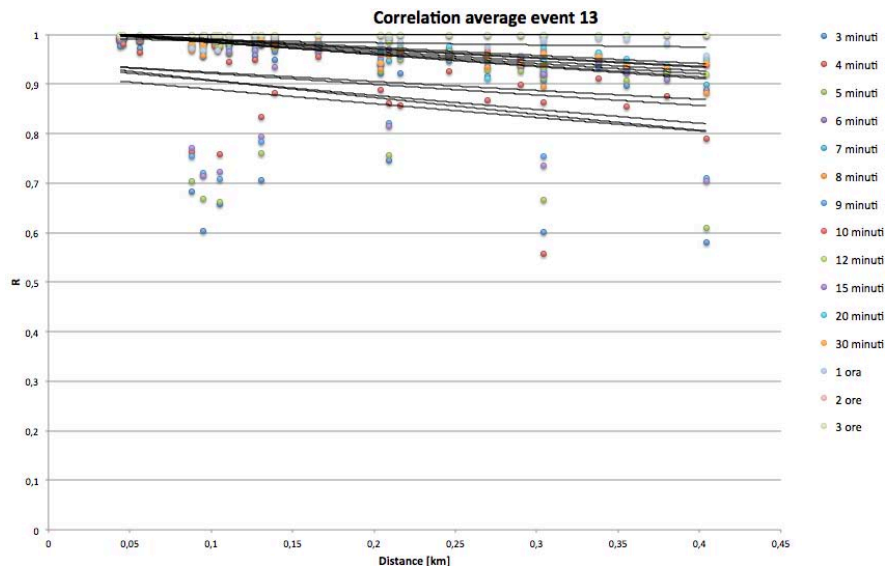
Figure 61: Correlation using the average data for the event 13 – 15/06/2012.

5.9.3. Events belong Group C
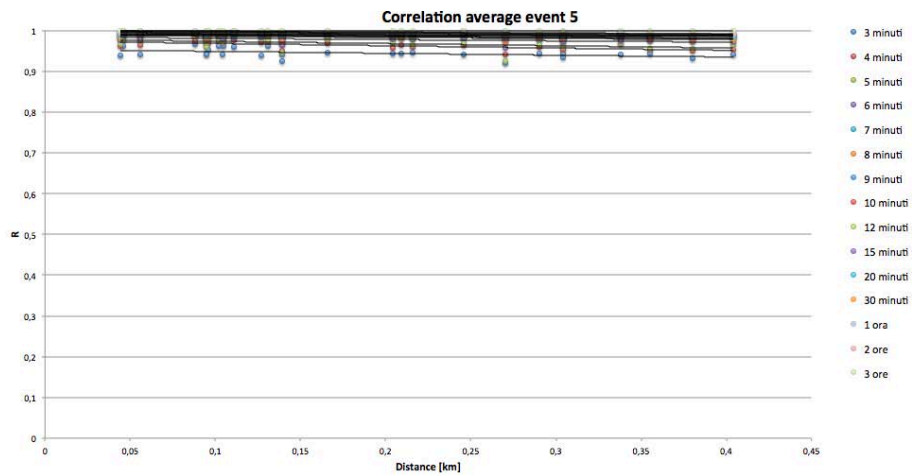
Figure 62: Correlation using the average data for the event 5 – 06/07/2012.
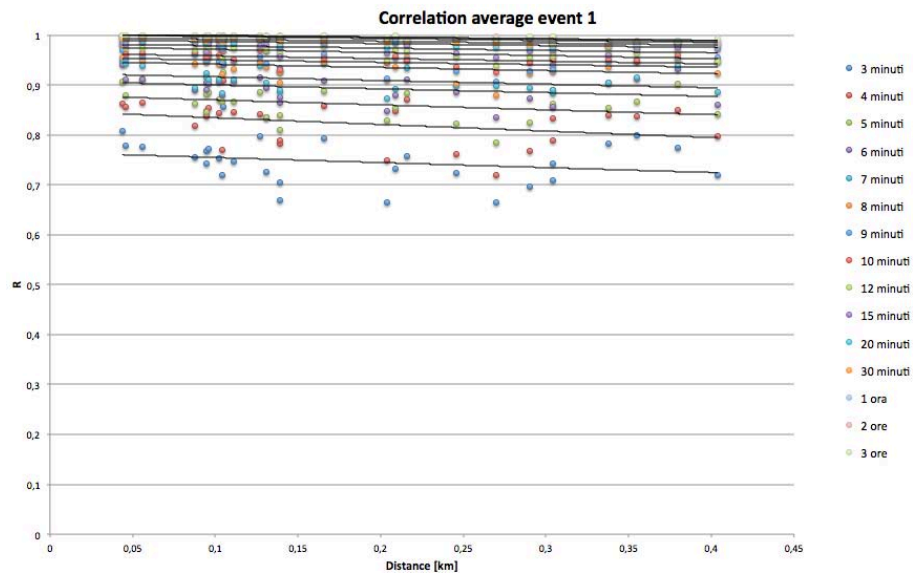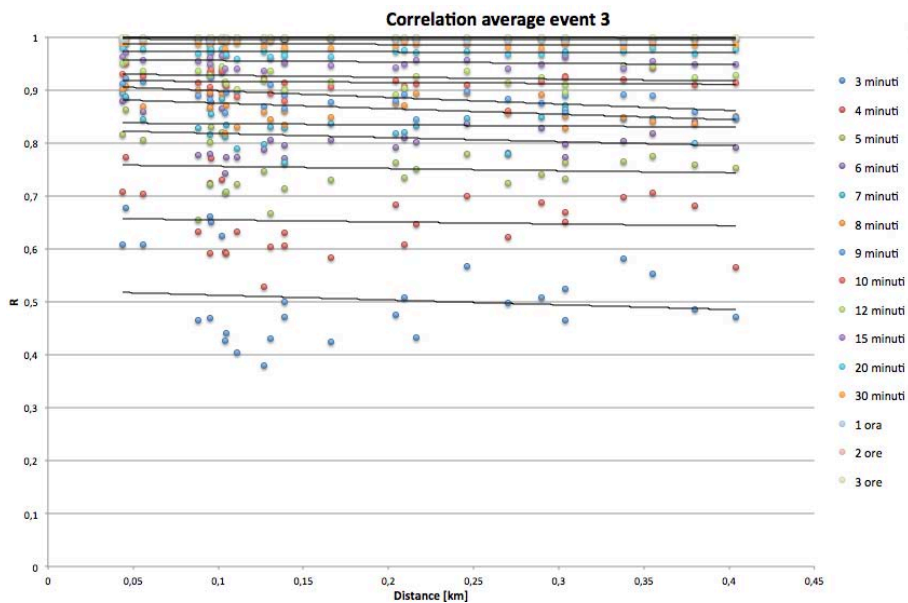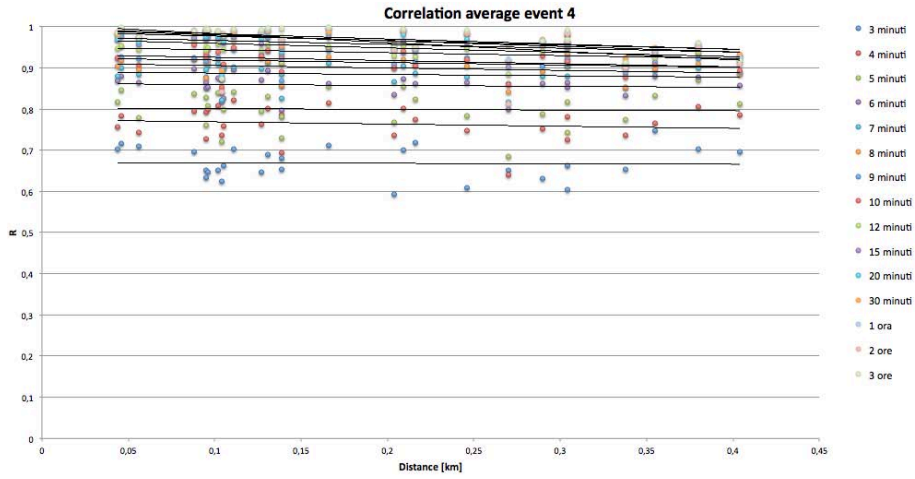

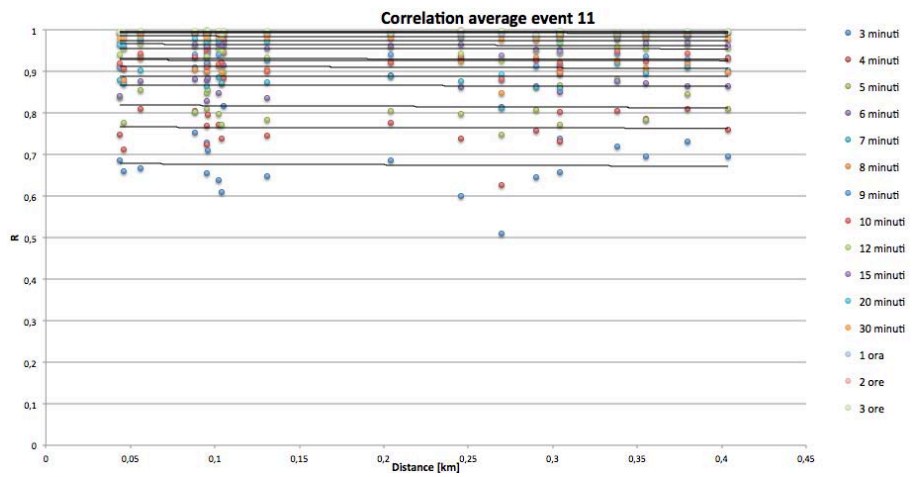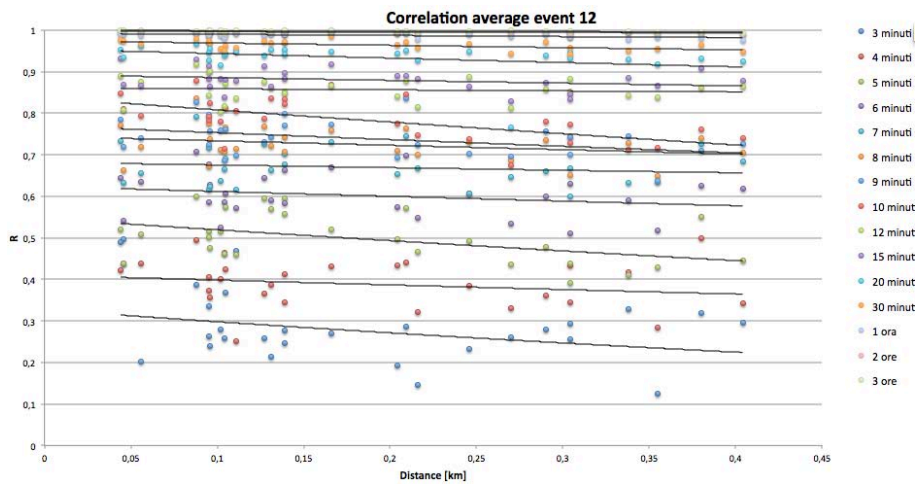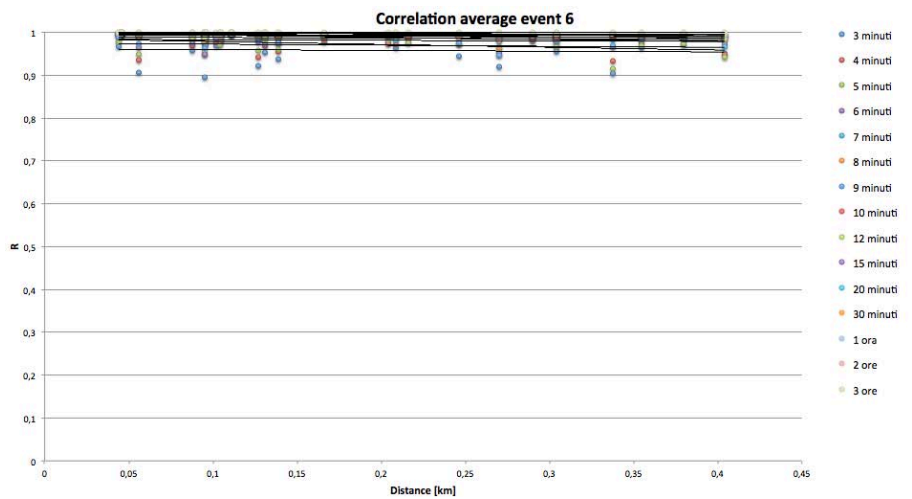*Figure 63: Correlation using the average data for the event 6 – 05/08/2012.*
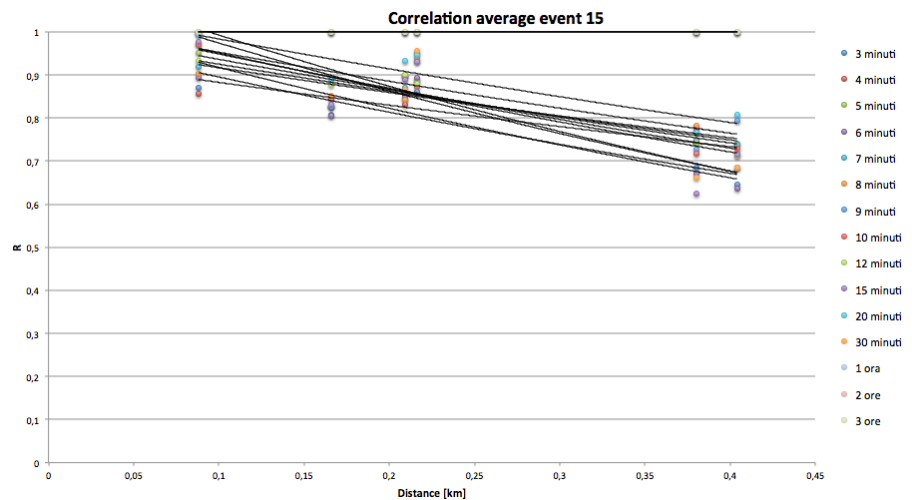
Figure 64: Correlation using the average data for the event 15 – 25/08/2012.

Chapter 6: Analysis for the spatial variability

6.1. Introduction

In this chapter we are going to analyse the method of cross validation, using the spatial interpolation model of Kriging. This analysis uses the data from the rain gauges, and we call these data points "ground truth" because in that place we are sure about that value.

Our latest analyses of correlation have shown the best way to use the average data of the pair of rain gauges, or of the four rain gauges, instead of the data of every single rain gauge, so for this study we are going to use only the average data of the pair of the rain gauges.

So we can show, comparing the ground truth data with the theoretical data, the trend of kriging interpolation for different events and cumulative time scales.

By "theoretical data" we intend the data of rainfall, which are estimated using the interpolation model of kriging in an area of 500 square metres; the ground truth data correspond to the empirical data, which are provided by the rain gauges.

The interpolation model of Kriging provides a trend of spatial variability of every event, starting from the ground truth data; in this way it is able to estimate the data everywhere inside the area of study.

In this chapter we analyse how the data values of the rainfall in the points where there are rain gauges change if they are calculated with the interpolation model of kriging. So, we can see the importance of a high correlation to estimate the data as closely as possible to the real data.

This method is very important because when the correlation is high and the kriging model is accurate, we are able to provide the data for every place inside our area with a high probability of obtaining correct values.

6.2. Comparison between the methods

In the following section, the behaviour of the cross validation test of the kriging spatialization method will be analysed for every cumulative time scale.

6.2.1. Cumulative time scale of 5 minutes

We are starting with the lower cumulative time scale, it is 5 minutes because if we had used 3 minutes there was the risk to have too zero data, and also because the cumulative of 5 minutes is really good to have a valid estimation.

To approach this analysis we are using the event 4 for group A, which represents the worst case of behaviour. The first graph, in figure 65, represents the typical trend of the group A using a single rain gauge as a measure of the rainfall, we can notice that the spread of the points is bigger than the spread of the graphs on the other two methods, which are represented in figure 66 and 67.

Figure 65: Comparison between the data estimate by kriging, the theoretical ones, and the real data of rainfall provide by rain gauges. It is referred to the event 2 of the group A, for 5 minutes timescale using as data input the single rain gauge.

The behaviour in the other two, figure 66 and 67, which represent the estimate of the data using the average of two and four rain gauges, it is not so different, but thinking at the results of the correlation they suggest that the third graph should have the best spread, i.e. that the points should be closer to the equality line than the other two groups.

Figure 66: Comparison between the data estimate by kriging, the theoretical ones, and the real data of rainfall provide by rain gauges. It is referred to the event 2 of the group A, for 5 minutes timescale using as data input the average of 2 rain gauges.
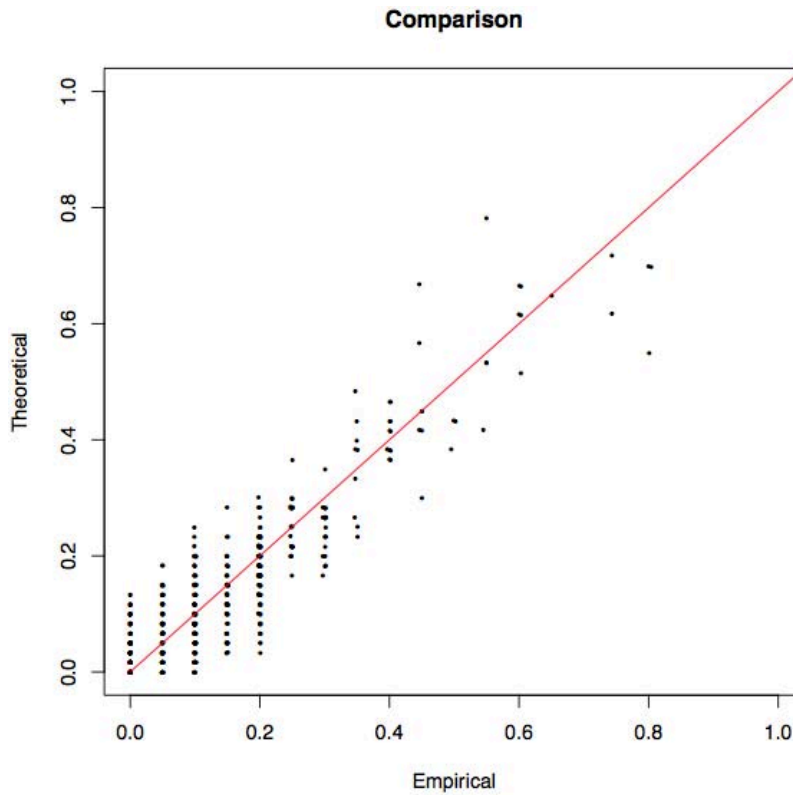
Figure 67: Comparison between the data estimate by kriging, the theoretical ones, and the real data of rainfall provide by rain gauges. It is referred to the event 2 of the group A, for 5 minutes timescale using as data input the average of 4 rain gauges.

6.2.2. Cumulative time scale of 15 minutes

About the comparison between the groups analysing the cumulative time scale of fifteen minutes, we can notice that the spread of the data for the first graph, figure 68, is always broader than the other two graphs, in figures 69 and 70.

Also here, it is clear that location of the data provided from the average of two and four rain gauges have a better trend and in particular the last one; in fact, it is possible to see that the last graph has a bigger number of points on the equality line than the second graph. So we have the best spatialization using the data from the average of 4 rain gauges.

According to what there is for the cumulative time scale of 5 minutes, the behaviour of kriging for 15 minutes is better than before, and this happen because for a bigger cumulative scale there are more data than for 5 minutes of cumulating and the correlation is always upper the 0.9 line.
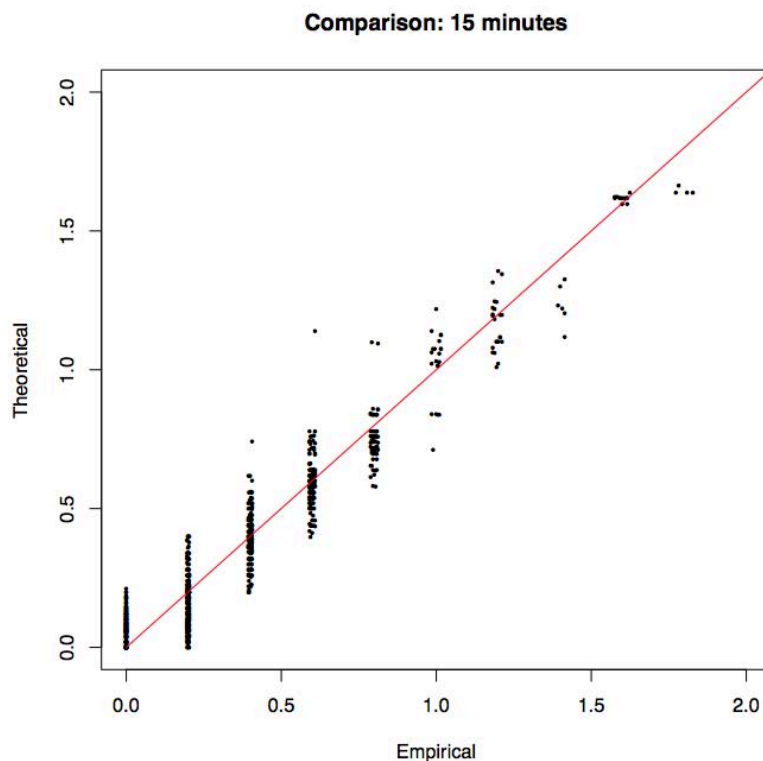
Figure 68: Comparison between the data by cross validation method, the theoretical ones and the real data of rainfall provide by rain gauges. It is of the event 2 of the group A, for 15 minutes timescale using as data input the single rain gauge.
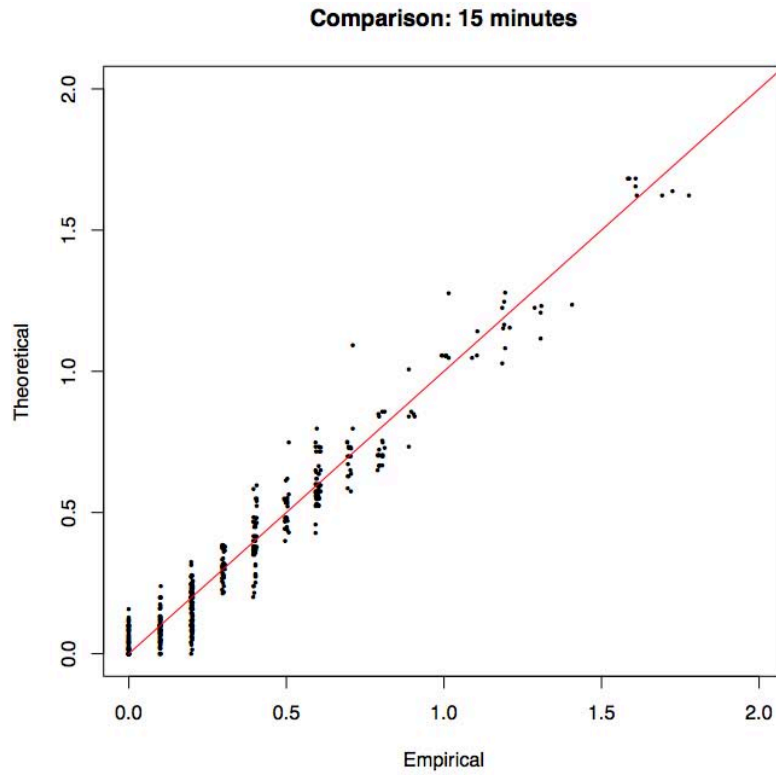
Figure 69: Comparison between the data by cross validation method, the theoretical ones and the real data of rainfall provide by rain gauges. It is of the event 2 of the group A, for 15 minutes timescale using as data input the average of 2 rain gauges.
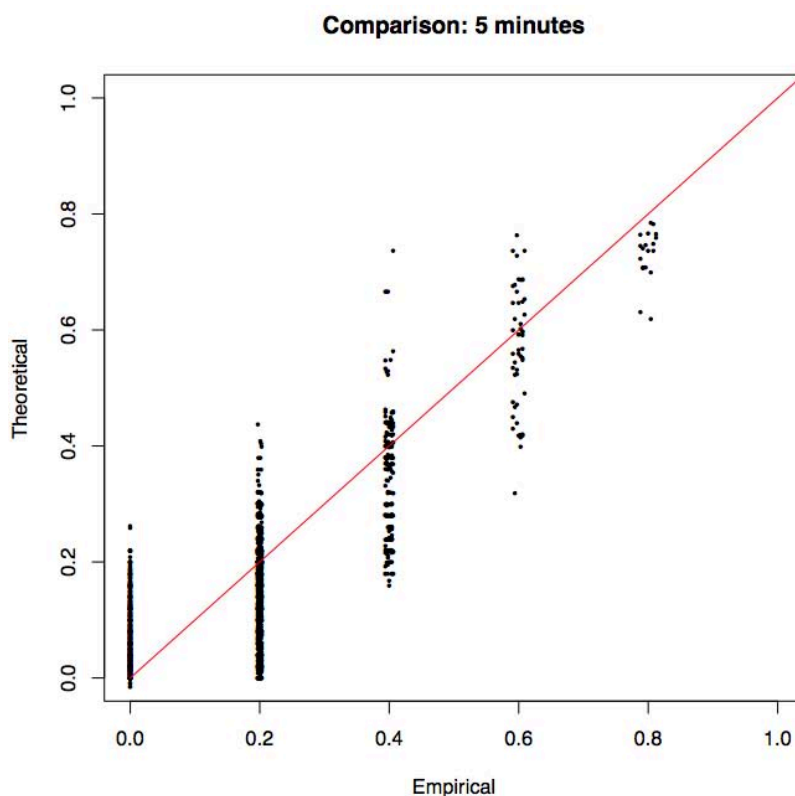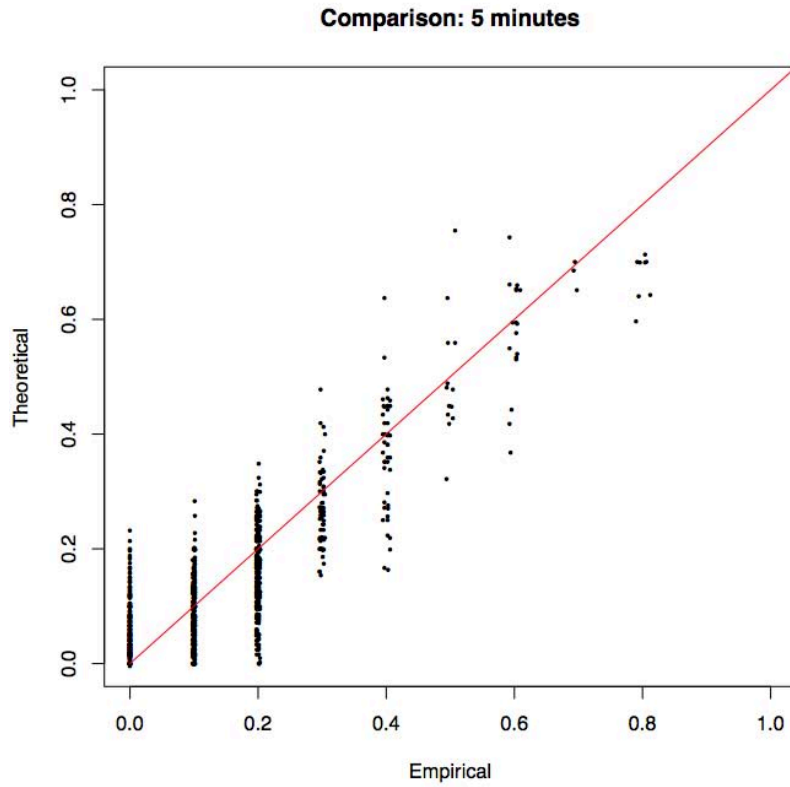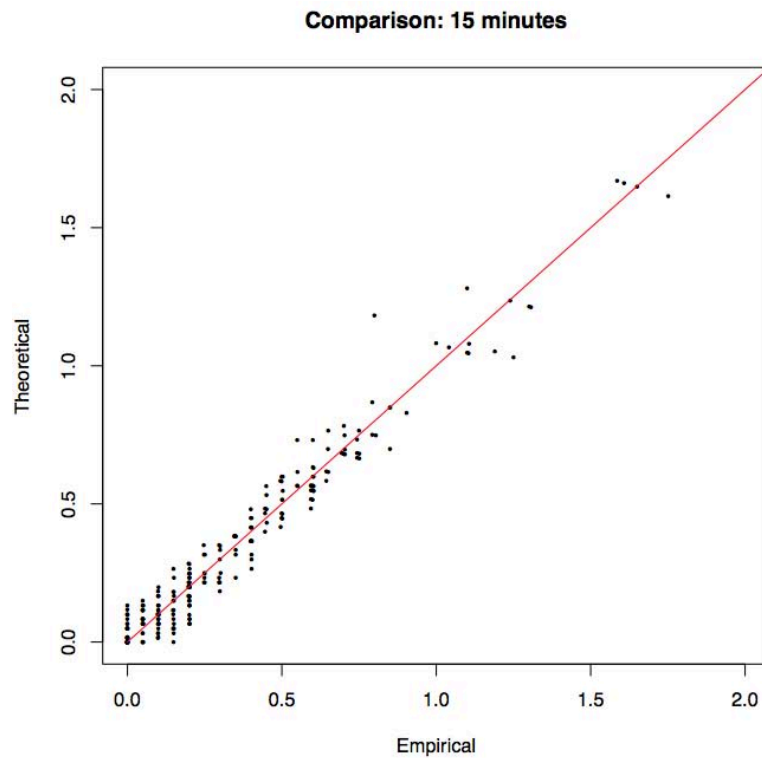

Figure 70: Comparison between the data by cross validation method, the theoretical ones and the real data of rainfall provide by rain gauges. It is of the event 2 of the group A, for 15 minutes timescale using as data input the average of 4 rain gauges.

6.3. Accuracy of the method

The next section speaks about the accuracy of the interpolation method of kriging to estimate rainfall depth data in all the area of the study.

We have used a different graphic, which is more qualitative than quantitative, to show how the interpolation method of kriging is able to provide an accurate spatial variability. In the graphic there are: the horizontal axis, which shows the difference among real data of rainfall, which are provided from the rain gauges, and predicted data, which are provided from the interpolation method of kriging; on the vertical axis there is the weighted cumulative distribution, which represent the percentage of the rainfall weighted in order to more intensive event of rainfall.

6.3.1. Using 5 cumulative time scale

The first under section analyses what happen for the 5 minutes cumulative time scale, the three green lines indicate the values of the mean and the first and third quartile, if they are approximately near zero there is a very good trend of the spatialization. The value of the mean shows us how the predicted data and the real data are as close as possible between them; the other two green lines show us the distance between the mean line, which can indicate if the kriging interpolation is a good method to estimate data inside our boundary lines.

The following graphs represent the accuracy of estimation using as data input the data from single rain gauge, the data from the average of the couples of the rain gauges and the last one represents the behaviour for data of the average of four rain gauges.

Figure 71: Trend of kriging estimation to provide the rainfall data for the area analysed. It is about the event 2, for a time scale of 5 minutes using a single rain gauge.

Figure 72: Trend of kriging estimation to provide the rainfall data for the area analysed. It is about the event 2, for a time scale of 5 minutes using as data input the average of 2 rain gauges.

All the three case studies indicate that the kriging interpolation method is a good method to estimate the value of data in a small urban catchment; there is a good trend in all of them, but the best one is the graphic in figure 73, which is the graph refers to the using as data input the average of 4 rain gauges.

Analysing the spread, which is the distance between the first and third quartile, we can see that for the figure 71 it is 0.137, for the figure 72 it is 0.1095, for the figure 73, it is 0.0974. The reason why the last one is the best one is that with the average data of 4 rain gauges the trend of the correlation is already high for 5 minutes of cumulative time scale.
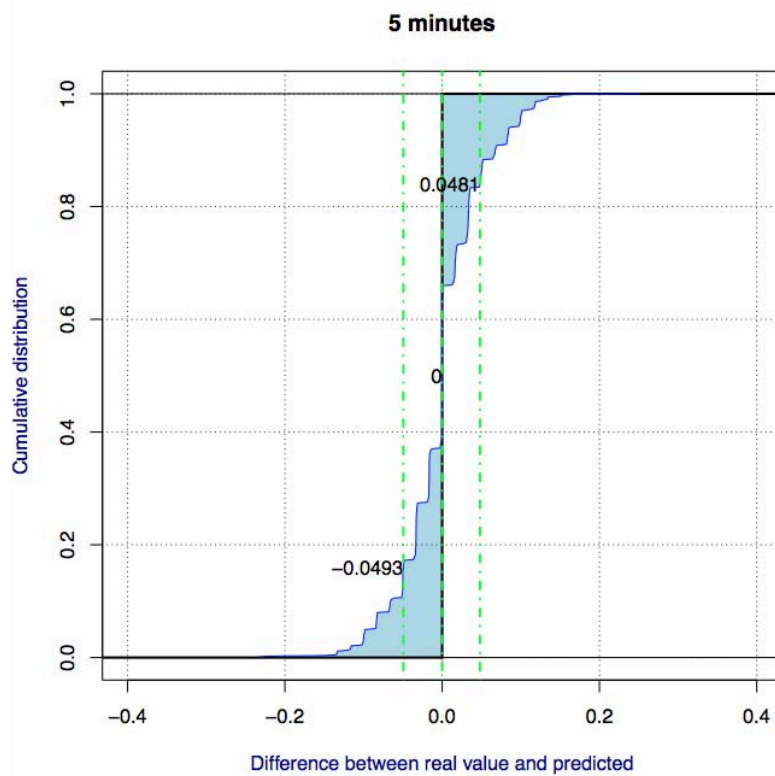
Figure 73: Trend of kriging estimation to provide the rainfall data for the area analysed. It is about the event 2, for a time scale of 5 minutes using as data input the average of 4 rain gauges.

6.3.2. Using 15 cumulative time scale

To finish, there are the graphs of the 15 minutes cumulative temporal scale; they show a homogenization results on the graphs in figures 74,75 and 76 that is an evolution than what we have seen above for the cumulative time scale of 5 minutes. There is a good trend for all the three groups, and this improvement is due to increasing values of rain on the cumulative data.

So there are not big differences between the estimation from the average of 2 and 4 rain gauges, in fact, analysing the spread again, which can provide the trend of the method, we can find:

- group A, spread = 0.1827;
- group B, spread = 0.1295;
- group C, spread = 0.1209;

Now the trend is approximately the same as expected, because for a cumulative temporal scale of 15 minutes, we have seen that using the correlation way of the average data we have an improvement of the correlation, and so with the 15 minutes cumulative time scale there is already a high correlation.
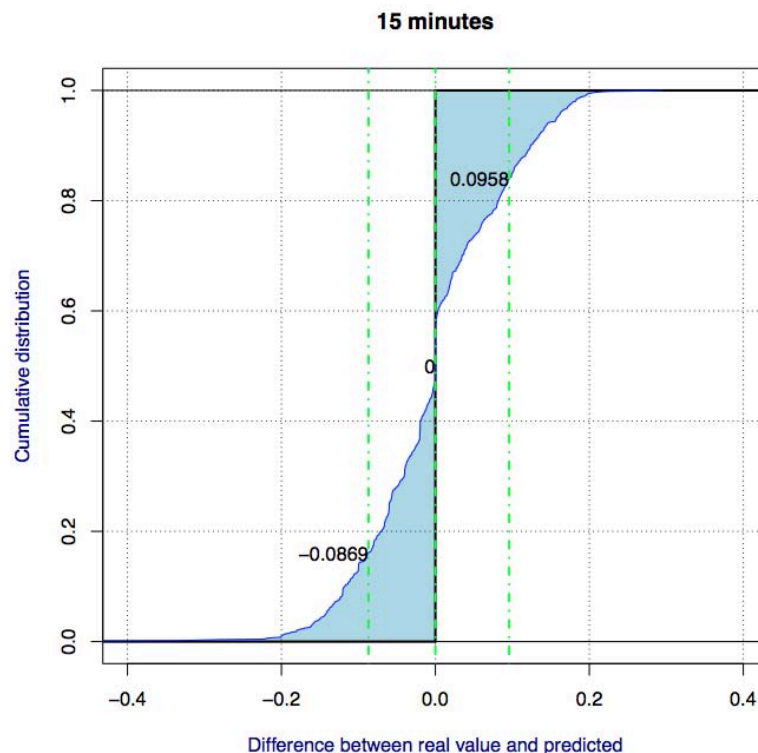
Figure 74: Trend of kriging estimation to provide the rainfall data for the area analysed. It is about the event 2, for a time scale of 15 minutes using as data input the single rain gauge.

89
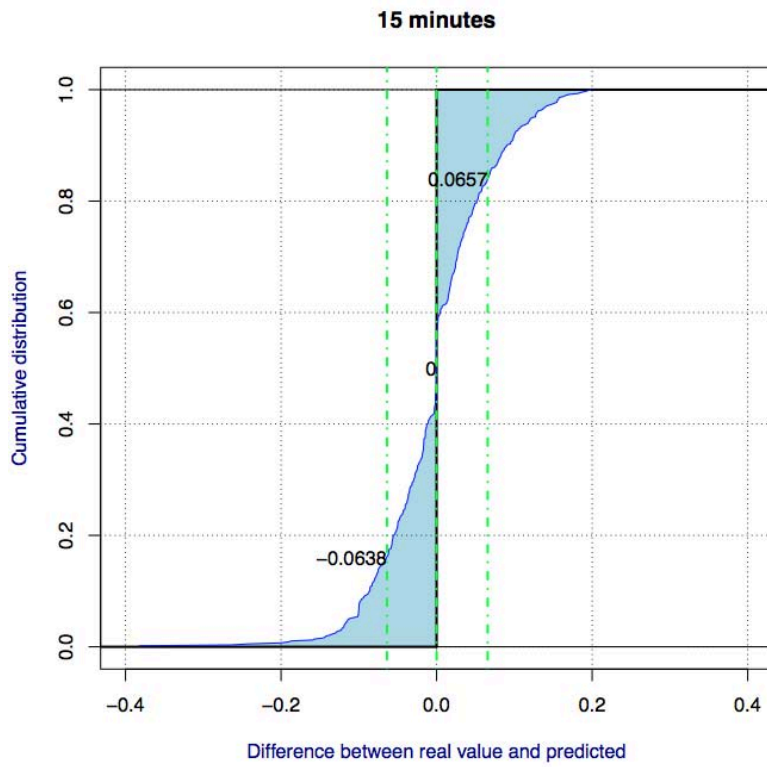
15 minutes

Figure 75: Trend of kriging estimation to provide the rainfall data for the area analysed. It is about the event 2, for a time scale of 15 minutes using as data input the average of 2 rain gauges.
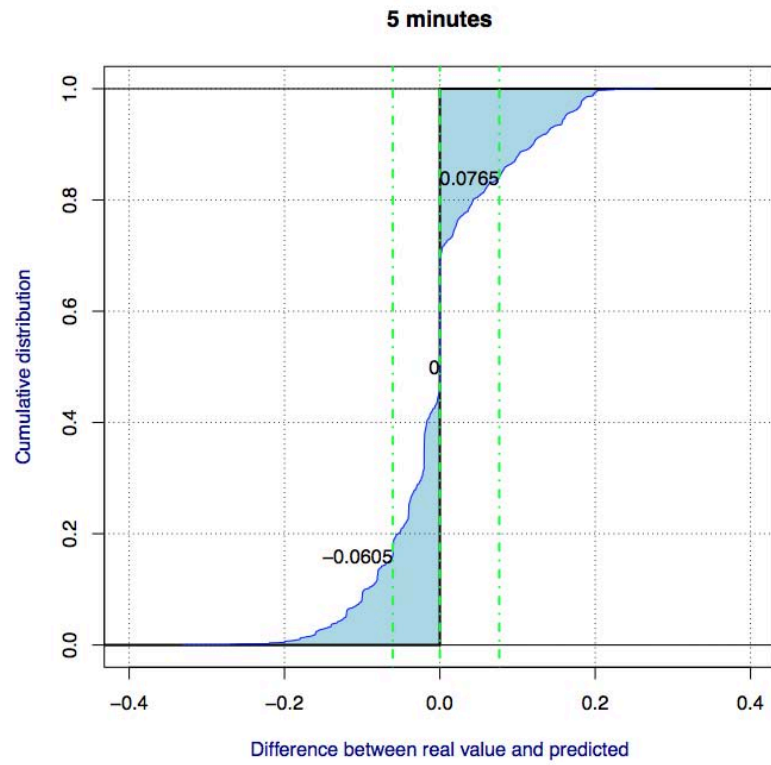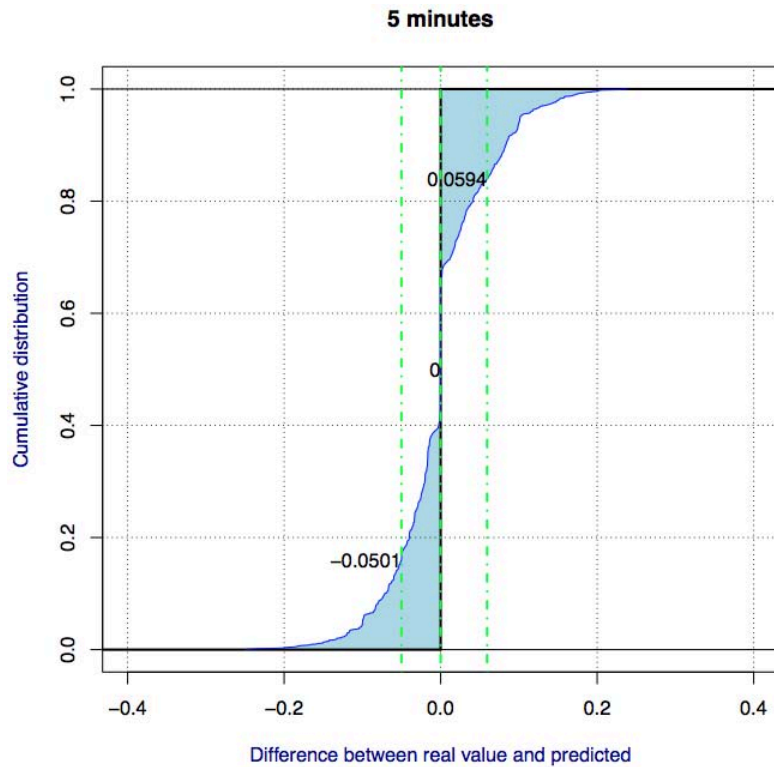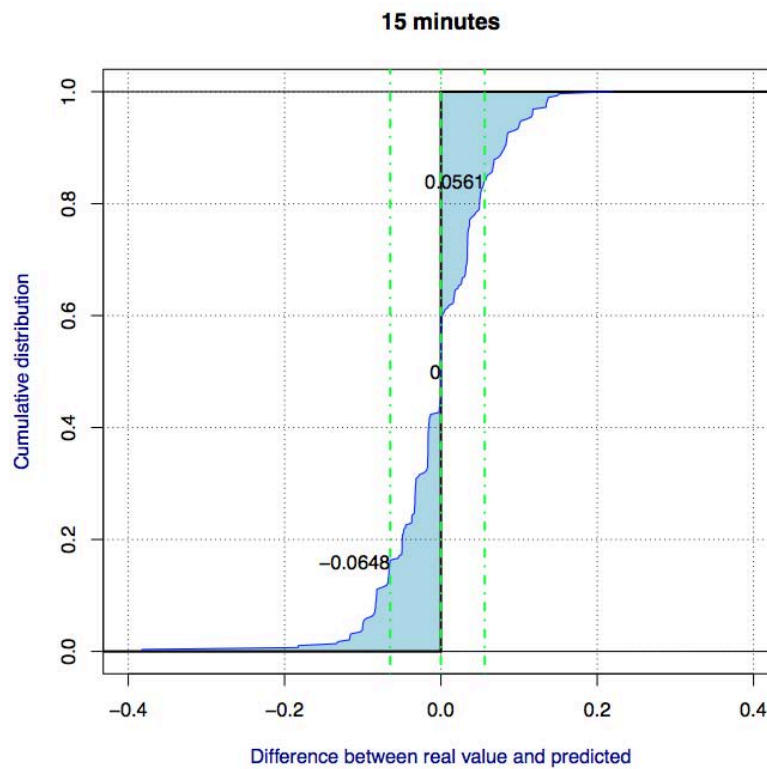

15 minutes

Figure 76: Trend of kriging estimation to provide the rainfall data for the area analysed. It is about the event 2, for a time scale of 15 minutes using as data input the average of 4 rain gauges.

6.4. Final consideration

In the end it is possible to conclude that for a small urban catchment, like the area that we have analysed, the kriging interpolation method is a good method to estimate the spatial variability of the rainfall. The real point is that using this interpolative method we need the more accurate value as data input, and for the events belonging to group C it is enough to use the data from the single rain gauges; but if the events belongs group B, to have a more accurate data input it is important to have couples of rain gauges to be sure to measure the real rainfall data with the smallest error. In the end, the worst case, if we have an event belonging to group A, we need at least a cluster of four rain gauges to have an accurate value.

Chapter 7: Conclusion

A detailed study into the differences between the correlations of the rain gauges on a small urban catchment, in the end gives us some results that allow us to know how the rain gauges behave, and to have a good estimation data as a function of the different events.

Starting from the studies of the correlation using a single rain gauge and evaluating how it changes for the different typologies of events, we noticed that there are three groups where we can put each event. For each one of these, there is a different behaviour of the correlation that indicates which is the best, in other words, the lowest temporal scale to use in the hydrological models. After a second analysis focused on the evaluation of the best trend of correlation, we noticed that using a group of rain gauges in every place is better than using only one. The reason is that by using this method, it is possible to check and verify if the data measured are accurate or if there are some errors.

After this examination, and the removal of the data with errors, we can carry out the same studies made above, but using an average of the rainfall data, and we will obtain a better trend of correlation for all the groups of the events.

In fact the correlation trend lines are higher than in the other graphs, so it is possible to use a lower temporal scale to estimate the data to be used on the hydrological model.

This study has also shown the accuracy of the kriging interpolation method in providing the values of the data calculated everywhere inside the boundaries of the studied area. If we are able to categorize the different events, knowing their behaviour, which allows us to establish the lowest cumulative temporal scale of reference with a few rain gauges well positioned, the kriging interpolation method can provide all the data estimated for the area and it allows using the hydrological models to study the surface runoff.

To conclude, in this small area of studies of approximately 500 m2 we have noticed that for this distance the correlation has a rather linear trend, in fact, the correlation degree is not affected from the different distance between the rain gauges, the low correlation degree of single rain gauge is due to some errors of the tipping bucket and for other external problems like the wind. To resolve these

problems it is necessary to use a group of rain gauges in the same location to cut the measurement errors.

If we use together with the rain gauges the data from the radar, which offers a general overview of the events to know how big the examined area is, it will be possible to have more accurate and specific data in real time. The next step of this study should be to find a good method to take the radar data under the pixel of 1 km square to have a good comparison between the radar and rain gauge, and so using the radar data where there are not rain gauges.

Appendix A

In this appendix there are some codes of the programs used to analyse the correlation and the interpolation by the kriging method. We have used a program called R. Most of them have been used to provide the data, using the interpolation method of Kriging, and also to obtain some graphs.

Program 1:

```
library(gstat)
library(lattice)
library(graphics)
source("/Users/Antonio/Desktop/Lavoro di tesi/programmi R/draw.cdf.r")
number_ev = 1
rain_yy <- read.table("C:/Users/Patania/Desktop/TESI/stazioniMed.csv",header=T,sep=";",dec=",")
rain_yy.grid <- read.table("C:/Users/Patania/Desktop/TESI/griglia.txt",header=T,sep=",",dec=".")
file_name = paste("C:/Users/Patania/Desktop/TESI/file di lettura/pm_med_ev",number_ev,".csv",sep="")
yy <- as.matrix(read.table(file_name,header=F,sep=";",dec=","))
durata = 5
prev_tot <- c()
oss_tot <- c()
scarti_tot <- c()
for(cn in 0:(((length(yy[,1])-1)/durata)-1))
{
    m0 <- (durata*cn)+2
    mD <- (durata*(cn+1)+1)
    yh <- as.vector(colSums(yy[m0:mD,]))
    rain_yy[,4]<-yh
#########ANALISI VARIOGRAMMA
    v1 <- variogram(rr~1,locations=~X+Y, data=rain_yy)
    model.1 <- fit.variogram(v1,model=vgm(1,"Exp",range=10,nugget=0))
    #CROSS VALIDATION
    x <- krige.cv(formula=rr~1,locations=~X+Y, data=rain_yy,model=model.1, nmax = 10, nfold=5)
    prev_tot <- c(prev_tot,x$var1.pred)
    oss_tot <- c(oss_tot,x$observed)
    scarti_tot <- c(scarti_tot,x$residual)
}
pdf(paste("C:/Users/Patania/Desktop/TESI/ScatterPlot/Evento_",number_ev,"_DT_",durata,".pdf",sep=""))
plot(oss_tot,prev_tot,pch=16,cex=0.5,col="black",main="Comparison",xlab="Empirical",ylab="Theoretical",
    xlim=c(round(min(c(oss_tot,prev_tot),digits=1)),round(max(c(oss_tot,prev_tot),digits=1))),
    ylim=c(round(min(c(oss_tot,prev_tot),digits=1)),round(max(c(oss_tot,prev_tot),digits=1))))
lines(c(0,1000),c(0,1000),col='red')
draw.cdf(scarti_tot,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),
            ylab_cdf=list('Cumulative distribution',col='darkblue'),
            quantils_lines=c(16,50,84),fill='lightblue',col_grid='black',col_quant='green')
draw.cdf(scarti_tot,oss_tot,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),
            ylab_cdf=list('Weighted cumulative distribution',col='darkblue'),
            quantils_lines=c(16,50,84),fill='lightblue',col_grid='black',col_quant='green')
dev.off()
```

Program 2:

```
# Kriging
library(gstat)
library(lattice)
library(graphics)
source("/Users/Antonio/Desktop/Lavoro di tesi/programmi R/draw.cdf.r")
rain_yy <- read.table("/Users/Antonio/Desktop/Lavoro di tesi/stazioni.csv",header=T,sep=";",dec=",")
rain_yy.grid <- read.table("/Users/Antonio/Desktop/Lavoro di tesi/griglia.txt",header=T,sep=",",dec=".")
```

```
yy <- as.matrix(read.table("/Users/Antonio/Desktop/Lavoro di tesi/file di
lettura/pm_tot_ev1.csv",header=F,sep=";",dec=","))
prev_tot <- c()
oss_tot <- c()
scarti_tot <- c()
dir.create("/Users/Antonio/Desktop/25_04_E1", showWarnings = TRUE, recursive = FALSE, mode = "0777")
############ INIZIO CICLO ###############################
cn <- 0
   m0 <- (60*cn)+2
   m60 <- (60*(cn+1)+1)
   yh <- as.vector(colSums(yy[m0:m60,]))
   rain_yy[,4]<-yh
   pdf(paste('/Users/Antonio/Desktop/25_04_E1/istant_',round(m60/60),'.pdf',sep=''))
aa <- c(seq(0,4.7,length=200))
bb <- c(seq(0,4.7,length=500)
   v1 <- variogram(rr~1,locations=~X+Y, data=rain_yy)
   v2 <- variogram(log(rr)~1,locations=~X+Y, data=rain_yy)
   model.1 <- fit.variogram(v1,model=vgm(1,"Exp",range=10,nugget=0))
   model.1
   z <- krige(formula=rr~1,locations=~X+Y, data=rain_yy,newdata=rain_yy.grid,model=model.1,nmax=10)
   levelplot(var1.pred ~ X+Y, z, aspect = "mapasp(z)",main="Ordinary kriging
prediction",col.regions=colorRampPalette(c("white","blanchedalmond","azure3","cadetblue3","cyan2","cornflowerbl
ue","dodgerblue4","blue3","chartreuse4","darkolivegreen","darkgreen","goldenrod","darkgoldenrod","darkorange","f
irebrick1","deeppink3","darkred","chocolate4","bisque4","black")), at =aa)
#errore standard di kriging
   levelplot(sqrt(var1.var) ~ X+Y, z, aspect = "mapasp(x)",main = "Ordinary kriging prediction standard errors")
   #CROSS VALIDATION
   x <- krige.cv(formula=rr~1,locations=~X+Y, data=rain_yy,model=model.1, nmax = 10, nfold=5)
   cros_val <- array( ,dim=c(length(x[ ,1]),3))
   cros_val[ ,1] <- x$var1.pred
   cros_val[ ,2] <- x$observed
   cros_val[ ,3] <- x$residual
   write.table(cros_val,paste("/Users/Antonio/Desktop/25_04_E1/CV_istant_",
         round(m60/60,digits=1),".csv",sep=''),sep=';',dec=',',row.names=F,
         col.names=c('Previsione','Dati osservati','Residui'),quote=F)
   write.table(x,paste("/Users/Antonio/Desktop/25_04_E1/CV_tot_istant_",
         round(m60/60,digits=1),".csv",sep=''),sep=';',dec=',',row.names=F,quote=F)
   model1<-lm(rain_yy$rr ~ x$var1.pred)
########################GRAFICI###########################
   prev <- cros_val[,1]
   oss <- cros_val[,2]
   scarti <- cros_val[,3]
   # aggiorno i vettori totali di CV
   prev_tot <- c(prev_tot,prev)
   oss_tot <- c(oss_tot,oss)
   scarti_tot <- c(scarti_tot,scarti)
   plot(oss,prev,pch=16,cex=0.5,col="black",main="Comparison",xlab="Empirical",ylab="Theoretical",
       xlim=c(round(min(c(oss,prev),digits=1)),round(max(c(oss,prev),digits=1))),
       ylim=c(round(min(c(oss,prev),digits=1)),round(max(c(oss,prev),digits=1))))
   lines(c(0,1000),c(0,1000),col='red')
draw.cdf(scarti,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),ylab_cdf=list('Cumulative
distribution',col='darkblue'),fill='lightblue',col_grid='black',col_quant='green')
draw.cdf(scarti,oss,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),ylab_cdf=list('Weighted
cumulative distribution',col='darkblue'),fill='lightblue',col_grid='black',col_quant='green')
dev.off()
############ FINE CICLO ####################################################
# analisi sulla cumulata totale di pioggia
ytot <- as.vector(colSums(yy))
rain_yy[,4]<-ytot
pdf('/Users/Antonio/Desktop/25_04_E1/cum_tot.pdf')
aa <- c(seq(0,22,0.1),seq(22.5,24.0,0.05),seq(24.5,28.5,0.01),seq(28.6,30,0.2))
rg_grid <- read.table("/Users/Antonio/Desktop/Lavoro di tesi/coord_stazioni.csv",header=T,sep=";",dec=",")
plot(rg_grid$X,rg_grid$Y,pch=16,cex=1, col=terrain.colors(n=50), main= "Stations", xlab="Coordiantes
GB",ylab="Coordiantes GB")
#########PESO INV PROP ALLA DISTANZA
```

```
z <- krige(formula=rr~1,locations=~X+Y, data=rain_yy,newdata=rain_yy.grid,model=NULL)
levelplot(var1.pred ~ X+Y, z, main="Inverse distance
interpolation",col.regions=colorRampPalette(c("white","gray","black")),contour='T')
#########PESO INV PROP AL QUADRATO DELLA DISTANZA
z <- krige(formula=rr~1,locations=~X+Y, data=rain_yy,newdata=rain_yy.grid,model=NULL,set=list(idp=2))
levelplot(var1.pred ~ X+Y, z, main="Inverse square distance
interpolation",col.regions=colorRampPalette(c("white","gray","black")),at=aa)
#########ANALISI VARIOGRAMMA
v1 <- variogram(rr~1,locations=~X+Y, data=rain_yy)
#plot(v1)
v2 <- variogram(log(rr)~1,locations=~X+Y, data=rain_yy)
#plot(v2)
model.1 <- fit.variogram(v1,model=vgm(1,"Exp",range=10,nugget=0))
#########KRIGING ORDINARIO
z <- krige(formula=rr~1,locations=~X+Y, data=rain_yy,newdata=rain_yy.grid,model=model.1,nmax=10)
levelplot(var1.pred ~ X+Y, z, aspect = "mapasp(z)",main="Ordinary kriging
prediction",col.regions=colorRampPalette(c("white","gray","black")),contour='T')
levelplot(var1.pred ~ X+Y, z, aspect = "mapasp(z)",main="Ordinary kriging
prediction",col.regions=colorRampPalette(c("white","blanchedalmond","azure3","cadetblue3","cyan2","cornflowerbl
ue","dodgerblue4","blue3","chartreuse4","darkolivegreen","darkgreen","goldenrod","darkgoldenrod","darkorange","f
irebrick1","deeppink3","darkred","chocolate4","bisque4","black")))
#errore standard di kriging
levelplot(sqrt(var1.var) ~ X+Y, z, aspect = "mapasp(x)",main = "Ordinary kriging prediction standard
errors",contour='T')
#CROSS VALIDATION
# five-fold cross validation:
x_tot <- krige.cv(formula=rr~1,locations=~X+Y, data=rain_yy,model=model.1, nmax = 10, nfold=5)
write.table(x_tot,"/Users/Antonio/Desktop/25_04_E1/CV_cum_tot.csv",sep=';',dec=',',row.names=F,quote=F)
os <- x_tot$var1.pred
prv <- x_tot$observed
res <- x_tot$residual
plot(os,prv,pch=16,cex=0.5,col="black",main="Comparison",xlab="Empirical",ylab="Theoretical",
    xlim=c(round(min(c(os,prv)),digits=1)),round(max(c(os,prv)),digits=1))),
    ylim=c(round(min(c(os,prv)),digits=1)),round(max(c(os,prv)),digits=1))))
lines(c(0,1000),c(0,1000),col='red')
draw.cdf(res,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),ylab_cdf=list('Cumulative
distribution',col='darkblue'),fill='lightblue',col_grid='black',col_quant='green')
draw.cdf(res,os,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),ylab_cdf=list('Weighted
cumulative distribution',col='darkblue'),fill='lightblue',col_grid='black',col_quant='green')
dev.off()
###############################################################################
pdf('/Users/Antonio/Desktop/25_04_E1/tot_istanti.pdf')
# aggiorno i vettori totali di CV
prev_tot <- c(prev_tot,prev)
oss_tot <- c(oss_tot,oss)
scarti_tot <- c(scarti_tot,scarti)
plot(oss_tot,prev_tot,pch=16,cex=0.5,col="black",main="Comparison",xlab="Empirical",ylab="Theoretical",
    xlim=c(round(min(c(oss_tot,prev_tot)),digits=1)),round(max(c(oss_tot,prev_tot)),digits=1))),
    ylim=c(round(min(c(oss_tot,prev_tot)),digits=1)),round(max(c(oss_tot,prev_tot)),digits=1))))
lines(c(0,1000),c(0,1000),col='red')
draw.cdf(scarti_tot,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),ylab_cdf=list('Cumulative
distribution',col='darkblue'),fill='lightblue',col_grid='black',col_quant='green')
draw.cdf(scarti_tot,oss_tot,main_cdf='Residuals', col_line_cdf='blue',label_quant=T,
            xlab_cdf=list('Diffrences between Real and Predicted',col='darkblue'),ylab_cdf=list('Weighted
cumulative distribution',col='darkblue'),fill='lightblue',col_grid='black',col_quant='green')
dev.off()
xx_tot <- as.matrix(array( ,dim=c(length(prev_tot),3)))
xx_tot[,1] <- prev_tot
xx_tot[,2] <- oss_tot
xx_tot[,3] <- scarti_tot
write.table(xx_tot,"/Users/Antonio/Desktop/25_04_E1/CV_tot_ist.csv",sep=';',dec=',',row.names=F,quote=F)
```

Bibliography:

[1.] Gires, I. Tchiguirinskaia, D. Schertzer, S. Lovejoy, A. Schellart, (2012) Definition and implementation of innovative comparison tools between radar and rain gauge rainfall measurement;

[2.] A.K. Gayen, (1951) The frequency distribution of the product moment correlation coefficient in random samples of any size draw from non-normal universes;

[3.] A.N.A. Schellart, L. Wang, C. Onof, (2012) High resolution rainfall measurement and analysis in a small urban catchment;

[4.] A.N.A. Schellart, W.J. Shepherd, A.J. Saul, (2012) Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale;

[5.] E. Habib, W.F. Krajewski, A Kruger, (2001) Sampling errors of tipping-bucket rain gauge measurements;

[6.] E. Habib, W.F. Krajewski, G.J. Ciach, (2001) Estimation of rainfall inter-station correlation;

[7.] G.J. Ciach, (2002) Local random errors in tipping-bucket rain gauge measurements;

[8.] G.J. Ciach, W.F. Krajewski, (2005) Analysis and modeling of spatial correlation structure in small-scale rainfall in central Oklahoma;

[9.] H. Zang, Y. Wang, (2009) Kriging and Cross-validation for Massive Spatial Data;

[10.] J.M. Faurès, D.C. Goodrich, D.A. Woolhiser, S. Sorooshian, (1995) Impact of small-scale spatial rainfall variability on runoff modeling;

[11.] J. Niemczynowicz, (1986) The dynamic calibration of tipping-bucket raingauges;

[12.] M. Borga, A. Vizzaccaro, (1997) On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging;

[13.] M. Colli, L.G. Lanza and P. La Barbera, (2012) Weighing gauges measurement errors and the design rainfall for urban scale applications;

[14.] P. Fiener1 and K. Auerswald, (2008) Spatial variability of rainfall on a sub-kilometre scale;

[15.]P. La Barbera, L.G. Lanza, L. Stagi, (2002) Tipping bucket mechanical errors and their influence on rainfall statistics and extremes;

[16.]P.V. Mandapaka, P. Lewandowski, W.E. Eichinger and W.F. Krajewski, (2009) Multiscaling analysis of high resolution space-time rainfall;

[17.]R. Fankhauser,(1997) Influence of systematic errors from tipping bucket gauges on urban runoff simulation. Use of historical rainfall series for hydrological modelling.