BENCHMARKING RAINFALL INTERPOLATION OVER THE NETHERLANDS

MENGXI YANG
March, 2015

SUPERVISORS:
Dr. Y. Zeng (Yijian)
Dr. Ir. C.M.M.(Chris) Mannaerts

BENCHMARKING RAINFALL INTERPOLATION OVER THE NETHERLANDS

MENGXI YANG

Enschede, The Netherlands,

March, 2015

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Water Resources and Environmental Management

THESIS ASSESSMENT BOARD:

Prof. Dr. Ing. Wouter Verhoef (Chair)

Dr. R. Sluiter (External Examiner, Royal Netherlands Meteorological Institute, KNMI)

Dr. Y. Zeng (Yijian)

Dr. Ir. C.M.M.(Chris) Mannaerts

ABSTRACT

Different applications (e.g. urban flood early warning system or agriculture management) need rainfall information at different spatial resolutions. On the other hand, most of the meteorological departments only produce one set of precipitation data at fixed spatial resolution (e.g. KNMI provides rainfall data at 1km resolution). This study tries to derive rainfall data at different scales. One common method to obtain gridded rainfall data at different specific spatial resolutions is interpolation. The gridded rainfall data that are interpolated from in situ rain gauge varies depending on many factors. One of such impacting factors is the different interpolation methods that are used to produce gridded data, since each method has its own benefits and drawbacks. It is therefore needed to understand the uncertainty that may be caused by different interpolation methods. In this thesis, two geostatistical algorithms (ordinary kriging, thin plate spline) and one deterministic algorithm (inverse distance weighting) have been used to interpolate daily rainfall data at five different resolutions (1km, 3km, 8km, 12km, 25km) over the Netherlands from 2003 to 2013. Meanwhile, the grid data was resampled at 1km to other four resolutions and compared with interpolated data. Moreover, the scale factor may have influence on the interpolation method when interpolating rainfall measurement which means different interpolation algorithms may suit for specific scales. So there is also a need to understand the uncertainty may be caused by different spatial scales. As an extension of research, monthly data has been interpolated to see how the temporal scale may affect the interpolation results. In addition, the interpolated data were used to validate satellite data.

The main objective of this thesis is to identify the optimal interpolation method vs spatial-scale pair for generating reliable rainfall datasets over the Netherlands, meanwhile generate the relevant reference data which is prepared for generating a long term dataset.

Results contain interpolated rainfall data and resample rainfall data at 1km, 3km, 8km, 12km and 25km resolutions. Through comparing with the observed data from 32 automatic meteorological stations we found that for long-term daily rainfall interpolation, IDW interpolation is suitable at 1km, 3km and 8km and resampling method is suitable at 12km and 25km. Ordinary kriging is preferred on monthly rainfall interpolation.

Keywords: rainfall data, interpolation method, spatial resolution, resample

ACKNOWLEDGEMENTS

I would like to express my gratitude to all of the people who support me during the study and research in ITC as an MSc.

First of all, I would like to express my honest gratitude to my first supervisor Dr. Yijian Zeng and my second supervisor Dr C.M.M.Mannaerts, since they teaching me a lot with patience, supporting and encouraging me all the time and giving me many useful suggestion and excellent comments. Their encouragement and trust helped me to come across the difficult period during the research. This thesis would not be finished without their help. I also want to thank Dr. Raymond Sluiter and Tomislav Hengl for helping me with R and giving me useful suggestions.

I would like to extend my thanks to PHD student Vahid Rahimpour， Xu Yuan， Matthew and staff of Water Resources department, who shared insightful knowledge with me and helped me a lot during the research.

I am also grateful to all my classmates, Ms Sonam, Ms Gloria Afua, Ms Tsehay Amare Agegnehu, Ms Amie Elizabeth, Mr Henry Munyaka, Mr Hossein, Mr Damas Patrick, Mr David Ongo, for giving me help and happiness during my MSc study.

I would like to thank my Chinese friends in ITC: Lianghui Xing, TingTing Wei, Yannan Zhang, Ziwei Cao, Linyue Kong, Gaoyan Wu, Chao Mai, Yaojun Cao, Chunhui Shen, Huiming Cai, Xiaoxu Li, etc. The unforgettable nice time will stays forever.

Thanks to all ITC staff, especially to all staff in WREM during the lectures. I learned a lot and it will help me in the future.

My greatest appreciation goes to my family and my boyfriend for their permanent support.

TABLE OF CONTENTS

1. Introduction 1
   1.1. Background 1
   1.2. Problems statement 2
   1.3. Objectives and questions 2
2. State-of-the-art on interpolation 5
   2.1. Interpolation 5
   2.2. Deterministic interpolation methods 6
   2.3. Geo-statistical methods 7
   2.4. Summary and selection of interpolation method 8
   2.5. Resample 10
3. Datasets 11
   3.1. Study area 11
   3.2. Datasets 11
4. Methodology 13
   4.1. Flow chart 13
   4.2. R scripts 14
   4.3. Data distribution 14
   4.4. Interpolation 14
   4.5. Resample 15
   4.6. Temporal rescalling satellite data 16
   4.7. Result analysis 16
5. Implementation and result 17
   5.1. Data distribution and transformation 17
   5.2. Data histogram 18
   5.3. Variograms modeling for ordinary kriging interpolation 19
   5.4. Interpolation result at different resolutions 20
6. Validation and discussion 29
   6.1. Ordinary Kriging variance 29
   6.2. Comparation of interpolation methods 30
   6.3. Correlation coefficient and Root Mean Square Error 33
   6.4. Interpolation on monthly rainfall data 36
   6.5. Compare with satellite data 38
   6.6. Discussion 39
7. Conclusions 41

LIST OF FIGURES

Figure 1-1: the flooding caused by extreme rainfall....1
Figure 2-1: The theory of interpolation: yellow point is the unmeasured point, interpolated this value by a function using the around red points (From ArcGIS Help Meau) 5
Figure 3-1 locations of 325 precipitation stations and 32 automatic meteorological stations....12
Figure 4-1 flow chart of methodology....13
Figure 4-2 flow chart of resample 15
Figure 5-1 bubble rain on December 8 2006 and June 2 2004: stratiform rainfall, convective rainfall17
Figure 5-2 normal qq plot for December 8 2006 and June 2 2004: without transformation, with log transformation. 18
Figure 5-3 histogram of rainfall after log transformation: December 8 2006 represent stratiform rainfall and June 2 2004 which represent convective rainfall....19
Figure 5-4 semi-variogram of daily rainfall on December 8 2006 and June 2 2004....19
Figure 5-5 interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 1km resolution....20
Figure 5-6 interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 3km resolution....22
Figure 5-7 interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 8km resolution....24
Figure 5-8 interpolation result of ordinary kriging, IDW, TPS and resample on December 6 2006 at 12km resolution....25
Figure 5-9 interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 25km resolution....27
Figure 6-1 ordinary kriging variance on stratiform rainfall at 1km, 3km, 8km, 12km and 25km 29
Figure 6-2 the correlation coefficient for 32 stations with each method....31
Figure 6-3 RMSE of each method at specific resolution....35
Figure 6-4 RMSE and CC of ordinary kriging, IDW, TPS and resample at different resolutions 36
Figure 6-5 RMSE and CC of each method on monthly interpolation at different spatial resolutions 38

LIST OF TABLES

Table1-1: resolutions will be produce and related application 3
Table 2-1: summary of different interpolation methods 8
Table 5-1 CC and RMSE of each method at 1km resolution 21
Table 5-2 CC and RMSE of each method at 3km resolution 23
Table 5-3 CC and RMSE of each method at 8km resolution 24
Table 5-4 CC and RMSE of each method at 12km resolution 26
Table 5-5 CC and RMSE of each method at 25km resolution 28
Table 6-1 OK variance on December 8 2006 and June 2 2004 at different resolutions 30

1. INTRODUCTION

1.1. Background

Precipitation is liquid, solid or gaseous water that in the atmosphere falls to the surface (Lawford, 2014). As the most significant component in water cycle (Flato et al., 2000), precipitation plays an important role in various hydrological models (Tapiador et al., 2012). For instance, daily rainfall is major meteorological input for water resources and agricultural modelling system. Therefore, obtaining reliable and accurate rainfall data is important for local, regional and global hydrologic prediction (Jiang et al., 2012).

Moreover, precipitation also has a directly effect and important influence on human beings. Not only because it is used for drinking and for irrigation in agriculture, but also essential factor of the urban development. For example, the role of urban sewer system is to drain out the sewage and rainwater to reduce the vulnerability from flooding. A lot of factors have influence on the sewer system: rapid urbanization, complex infrastructure, human activities and changes in the precipitation patterns. Among these, extreme precipitation is a major threat to urban drainage system because it can cause overpressure to the drainage systems causing urban flooding and lead to loss to the society. Figure 1-1 shows urban flooding in the Netherlands. To prevent such risk and hazard, appropriate sewer systems should be built and some old systems should be replaced. It means extreme rainfall data as the dominant factor of the capacity of drainage design is crucially important. The full understanding of extreme rainfall data, such as intensity, frequency, and duration become necessary. For better understanding, high resolution precipitation data is needed. Therefore, one of the research directions of precipitation science is the areal interpolation of surface rain for urban development and agricultural application (Tapiador et al., 2012).

Figure 1-1: the flooding caused by extreme rainfall

There are various sources of rainfall products, such as rain gauges, satellite observations, precipitation radars and weather prediction models output, each having different benefits and drawbacks (Sapiano & Arkin, 2009). So methodologies that can capitalize on the strengths and minimize the disadvantages are needed. A number of models have been developed to combine satellite data with ground measurements. These models can provide estimations of rainfall for both missing pixels and time that are not covered (Alemohammad & Entekhabi, 2013). Alemohammad (2013) tried to combine different types of rainfall measurement through image fusion. Madsen (2012) compared different regional models and statistical downscaling methods for extreme rainfall estimation. Berndt (2014) investigated the performance of merging radar and rain gauge data with different high temporal resolution and rain gauge network densities. Willems (2011) used RCMs and urban drainage models to estimate the climate change influence on statistical precipitation downscaling for small-scale hydrological.

Precipitation is difficult to estimate, and it is spatially and temporally sensitive. The spatial distribution of precipitation is irregular and the characteristic is very dependent on the scale factor (Camarasa-Belmonte & Soriano, 2014). It is, therefore, needed different scales of precipitation datasets to understand how spatial variability influences the hydrological state. In general, to solve significant features at the suitable scales of urban drainage systems, small urban catchment scale is needed (Willems, Arnbjerg-Nielsen, Olsson, & Nguyen, 2012). Moreover, particular spatiotemporal resolution is dependent on the effective use of details. For example, agricultural application requires monthly data over areas of thousands of square kilometres; flood prediction needs hourly precipitation data in areas of various 100s of square kilometres (De Marchi, 2006). So there is a need to scale rainfall products to different local levels.

1.2. Problems statement

Different applications require rainfall data at specific spatial resolutions. For the time being, there are various sources of precipitation products. However, it is still difficult to use them directly at various local scales. For instance, satellite data has full spatial coverage but often discontinuous records, while the rain gauge data is continuous but limited to point observations. Moreover, most of meteorological departments produce only one set of rainfall data at a fixed spatial resolution. In the Netherlands, for example, the KNMI provides 1km grids precipitation data.

Nowadays, spatial interpolation as an effective method to adjust scale is widely used for creating continuous data when data is collected at discrete locations (Akkala, 2010). There are many studies comparing different interpolation methods. Some applications consider only monthly or annual time resolution for rainfall spatial interpolation (Goovaerts, 2000; Todini & Ferraresi, 1996; Lloyd, 2005). Some research have focused on different interpolation at one specific spatial resolution (Soenario, Plieger, & Sluiter, 2010; Hofstra, Haylock, New, Jones, & Frei, 2008; Taesombat & Sriwongsitanon, 2009), while some used comparisons restricted to one or two methods because it is very cumbersome in terms of computation time (S. Ly et al., 2011). However, little experience exists on comparing multiple interpolations at different spatial and temporal resolutions. There is a need to know whether the interpolation method preferred at certain resolution can still work well at other spatial resolutions, and whether the interpolation methods used for monthly and yearly would be suitable to daily rainfall interpolation. Comparing more techniques may provide some insights on particular strengths and constraints.

In addition, as we mentioned before, the meteorological departments often provide rainfall data at specific spatial resolution. Resampling datasets to other spatial resolution is more simple and convenient than interpolation. However, the accuracy of resampling to different spatial resolutions is still relatively unknown.

1.3. Objectives and questions

1.3.1. Objectives

Generally, reliable dataset is paramount for good research. As motioned, knowing the suitable interpolation method is critical in using rainfall data, with different research application requiring precipitation data at certain spatial resolution. Therefore, using different interpolation methods to generate a set of trusted data at different spatial resolutions is necessary. For example, merging high resolution of satellite data with in situ data can provide new opportunities to study regional variation in rainfall over the Netherlands. But previous to merging, satellite data need downscaling. Since spatial and temporal resolution of different satellites are not the same, reference data at different scales that can validate the scaling results are needed. In addition, the precipitation pattern at the diurnal level is different with monthly timescales (Johnson & Hanson, 1995). Therefore, there is a need to know whether the best interpolation for monthly and annual rainfall is also suitable for the daily scale. With the problem indicated,

this thesis is focus on the different interpolation method choose at different spatial resolution, so the objectives are:

Main objective:

The main objective is to benchmark rainfall interpolation over the Netherlands at various spatial resolutions. To address the discussed research problems, the current study will try to investigate what is the best interpolation method for rainfall observation over the Netherlands at different spatial resolutions. Particularly, according to the different applications, five different spatial resolutions will be test with three different interpolation methods as showed in Table 1-1, in which the reasons for the choice of each spatial resolution was explained.

Table1-1: resolutions will be produce and related application

Spatial resolution	Choose reason
1 km	Radar data in the Netherlands is at this resolution after 2008
3 km	MSG_CPP is viewed at this resolution (Meirink & Plieger, 2012)
8 km	Fine resolution for CMORPH data (Joyce, Janowiak, Arkin, & Xie, 2004)
12 km	The RCM modes at this resolution(Christensen, Christensen, & Guldberg, 1990)
25 km	All of the satellite data

Sub-objectives:

1. To estimate if different interpolation methods could get the same rainfall result at certain fixed resolution.
2. To analyse the result and determine which method is the most suitable at different spatial resolutions.
3. To understand the effect of different temporal scales on the rainfall interpolation results
4. To use the generated rainfall data at certain spatial resolution, as the reference data, to validate satellite rainfall data product.
5. To compare the interpolated rainfall data with simply resampled rainfall data.

1.3.2. Research questions

With the indicated objectives, the research questions can be asked as below:

1. Do the different interpolation methods provide the same rainfall result at certain resolution? If not, what are the possible reasons?
2. Which interpolation method is better for a specific spatial resolution?
3. Is the interpolation method used for daily is still suitable for monthly?
4. What is the added-value for generating reference rainfall data at different resolution?
5. Is the resampling method adequate to provide reliable rainfall data at different spatial resolution?

1.3.3. The innovate aim at

Identify the optimal rainfall interpolation method in the Netherlands and generate reference data at different spatial resolutions at daily timescales.

2. STATE-OF-THE-ART ON INTERPOLATION

2.1. Interpolation

Interpolation as a numerical analysis method is widely used in engineering and science. It obtains data at locations where there is no historical record, and generates data at a finer resolution than the historical record (Wey, 2006). When one has a number of data points, the un-sampled point can be obtained by interpolation (see figure 2-1). This process is usually achieved by curve fitting or regression analysis ("Interpolation - Wikipedia, the free encyclopedia,"). In geostatistics, data could be measured anywhere and typically comes from a limited number of observation locations (Eda, 2013). Spatial autocorrelation is the premise of any spatial interpolation which means that close points tend to be more similar than distant samples.

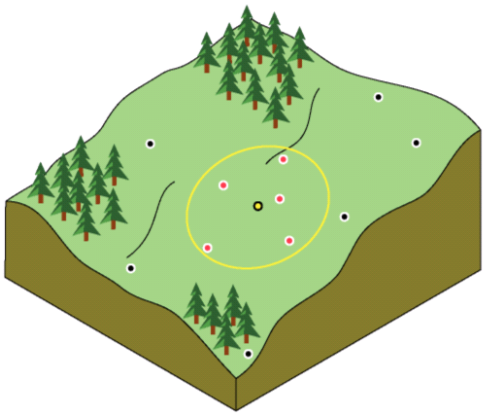


Figure 2-1: The theory of interpolation: yellow point is the unmeasured point, interpolated this value by a function using the around red points (From ArcGIS Help Meau)

Spatial interpolation allows the estimation of an attribute at any location by assuming that the attribute data is continuous over the study area, meanwhile the attribute is spatially dependent (Akkala, Devabhaktuni, & Kumar, 2010). The interpolated data is more close to the points which are nearby. Therefore, when datasets are collected at discrete and random locations, spatial interpolation can be effectively used for creating continuous data. The purpose of spatial interpolation is to create an empirical reality surface (Akkala et al., 2010). Spatial interpolation estimates values for cells in a raster from a limit number of known data points. It can be used to predict unknown values for any geographic point data, this study focus on rainfall interpolation.

A number of interpolation methods can be used to produce the spatial continuity of precipitation based on rain gauge measurement. There are a number of studies on analysing the difference between various interpolation methods. Goovaerts(2000) interpolated annual and monthly rainfall data by Thiessen polygon, IDW, ordinary kriging and cokriging. Kao & Hung (2004) used 5 meter DTMs as test data to compare twelve interpolation methods. Hofstra et al (2008) compared six interpolation methods of daily precipitation, temperature and sea level pressure. This part of literature review mainly based on (Sluiter, 2009), (Sterling, 2003), (Sarann Ly, Charles, & Degré, 2013) and (Akkala, 2010). A summary of 6 interpolation methods is shown in table 2-1.

Generally, the direct ground-based interpolation methods can be classified into two main types: deterministic methods and geostatistical methods.

2.2. Deterministic interpolation methods

Deterministic methods only use the geometric characteristics of point data to create a continuous surface.

Inverse distance weighting

Inverse distance weighting (IDW) is a simple and intuitive deterministic method for multivariate interpolation with a known scattered set of points. The un-sampled points are calculated with a weight function of the known points that includes more observations. So it is an advanced nearest neighbour theory that consider more points than only the nearest observation. It estimates values by weighted average using nearby observations. The weight decreases as distance increases (Sarann Ly et al., 2013). Therefore, the closer points have more influence to the predict point than the further distance point, which may cause "bulls eye" effect. IDW is a simpler interpolation technique in that it does not require pre-modelling like kriging (Tomczak, 1998).

The value at a certain grid cell is predicted by linear combination. When calculating a grid data, the sum of all the weights should be equal to 1.0, so the weights assigned to the data points are fractions (Kao & Hung, 2004). The formula for weighting determine are:

$$\lambda\, \mathrm{i} = d_{i0}^{-p} / \sum_{i=1}^{N} d_{i0}^{-p} \quad (1)$$

$$\sum_{i=1}^{N} \lambda \mathrm{i} = 1 \quad (2)$$

Where: $\lambda \mathrm{i}$ are the weights assigned to each known point, decrease with distance.

d_{i0} is the distance between the predicted points

p is the factor reduced weight

N is the number of surrounding points

The IDW formula is:

$$\bar{Z}(S_0) = \sum_{i=1}^{N} \lambda i Z(S_i) \quad (3)$$

Where: $\bar{Z}(S_0)$ is the value to be predicted at location of S_0.

$Z(S_i)$ is the observed point at location of Si.

This method is easy, fast and widely used in meteorology. Because having no extrapolation, all the interpolated data are within the original data range.

Nearest Neighbourhood (NN)

The nearest neighbourhood is a simple and fast method of multivariate interpolation. The theory of nearest neighbourhood method is to assign value to a certain grid cell from the nearest point (Sluiter, 2009). However, this method could not be success in all case due to its lack of success measures. So the method performs better when there are many data points.

Thiessen polygon (THI)

The Thiessen polygon method is also known as the nearest neighbour method. It assumes the predicted values are estimated from the closest observed values (Sarann Ly et al., 2013). The benefit of this method is its simplicity. This method creates a Thiessen polygon network formed by the segments in the nearby stations to the other related points. Each polygon surface is created to balance the rain quantity in the centre of the station, which means the polygon changes every time. This method is not suitable for the region which has many mountains due to the orographic rain. The disadvantage of this method is that the estimation is based only on one measurement and the other neighbour points are ignored. There may sudden jumps in two polygons(S. Ly et al., 2011).

Splines (Polynomial functions)

The splines interpolation methods are deterministic interpolators based on a polynomial functions for surface estimation that fits a minimum-curvature surface through the input points (Sarann Ly et al., 2013). Spline can generate sufficiently accurate surfaces by only a few known data and retain small features. There are five different spline functions (Spline with Tension, Multiquadratic Spline, Completely

Regularized Spline, Thin Plate Spline, Inverse Multiquadratic Spline) (Sterling, 2003). In this research, Thin Plate Spline has been used since there is little difference existing among spline equations.

In general, splines interpolation methods are global interpolators that ensure the result do not strongly oscillate between the sampled points (Sluiter, 2009). The polynomial functions perform well when the interpolation data is monthly and yearly. In addition, KNMI uses tension splines to interpolate all the climatological and meteorological data

Linear regression

Linear regression is a stochastic deterministic method that consider the probability distribution of the variable to expresses the relationship between a predicted variable and one or more explanatory variables (Sluiter, 2009). The form of linear regression is simple, using a straight line fitted through the data points. It is executed by a standard statistical program, using calculator functions to calculate maps. The linear interpolation models are commonly used as global interpolators due to its simple form. In this method, standard error, regression parameter and predicted values can be calculated.

The linear regression model assumes interpolated on the theory of physical reasons. In cases, such as random linear regression, the spatial independence and a normal distribution are commonly assumed. The multiple regression models may also include ancillary data.

Artificial neural networks (ANN)

The artificial neural network is a relatively new interpolation method for spatial interpolation. It is a non-linear statistical data modelling tool which is used to model complex relationships between inputs and outputs data (Sluiter, 2009). There are many types of ANN, most of them have little feedback about the data modelling and need powerful computation.

Karmakar (2009) applied the artificial neural network to the spatial interpolation of mean rainfall variable of 102 rain gauge stations in India.

2.3. Geo-statistical methods

Geo-statistics focus more on spatial statistical prediction than model fitting. The geo-statistics methods use the semi-variogram as main tool to analysis the spatial dependence of the datasets.

Kriging

Kriging is a geostatistical methodology which is based on a spatial correlation function. It was developed by Frech mathematician Georges Matheron in 1960s. The application was used in estimating gold deposited in a rock from some random core samples. Kriging was then used in earth sciences and other disciplines. As a geostatistical method, kriging interpolation widely used in various applications ranging from point measurement to continuous surfaces (S. Ly et al., 2011). It is also a kind of probabilistic method based on Kriging interpolation which works well in geosciences, when data is sparse (Sluiter, 2009). The kriging method assumes that the spatial variation of a continuous attribute is difficult to model by a simple function due to the irregular data distribution. The variation should be described by a regionalized variable in the stochastic surface.

Kriging interpolation is often regard as the optimum interpolation in geoscience. The basic tool of kriging is the semivariogram which captures the spatial dependence between points by plotting separation distance against semivariance. In addition, there are different types of that kriging can be used for spatial interpolate, each has their benefits and drawbacks.

Ordinary kriging

Ordinary kriging is the basic form of kriging interpolation. It measures values by linear combination, using variogram to determine the weight of data and describe the spatial correlation. As it is one of the three chosen methods, details will be described in methodology.

Cokriging

Cokriging using a multivariate variogram or covariance model and additional data (Sluiter, 2009). The theory of cokriging is based on the linear weighted sum of all the test data to estimate a location, so when there are two or more co-variable, the method may become more complex. Moreover, the result is better when both covariables and the spatial correlation are higher. Cokriging might improve the interpolation result when the primary variable is assumed under sampled and the variogram models has differ shapes.

Due to ancillary can be used in cokriging, this method often applied in meteorology. (J. M. Schuurmans, Bierkens, Pebesma, & Uijlenhoet, 2007) applied cokriging method to combine precipitation radar data with station data.

Universal kriging

Universal kriging uses a regression model as part of a process to calculate the mean value expressed as a linear or quadratic trend (Sluiter, 2009). It is a kriging with an external drift which often used in meteorology. For example, (J. M. Schuurmans et al., 2007) used universal kriging to combine precipitation radar data with station data.

Indicator kriging

Indicator kriging is a simple non-parametric interpolation method. The idea of indicator kriging is to estimates several models for different quantiles described by indicators. So indicator kriging is based on data transformed from continuous values to binary values. The indicator values are 0 or 1, so it is often used to interpolate a categorical variable like rainfall occurrence(Sluiter, 2009).

Residual kriging

Residual kriging is also called detrended kriging. The assumptions are the same with universal kriging. It is widely used in meteorology. Reyes et al. (2012) studied an approach based on the classical residual kriging method due to in practice the no stationary functional datasets are often exist. In Poland, it was found that the best method for monthly and seasonal averages of precipitation totals is residual kriging interpolation.

2.4. Summary and selection of interpolation method

In the Netherlands, precipitation data are being collected from more than 300 stations, so the density of sample data is adequate, while the spatial distribution is irregular. The planned highest spatial resolution for interpolation is 1km, so the method should be valid at local scale. Moreover, datasets used for interpolation are daily rainfall observations from 2003 to 2013 (as will be discussed in Chapter 3), which requires that the interpolation method should be fast. Table 2-1 is a summary of different interpolation methods and their characteristics. Based on advantages, disadvantages and suitable scenario, three different interpolation methods (ordinary kriging, Inverse Distance Weighting, Thin Plate Spline) are chosen. The details of the chosen interpolation methods will be introduced with more details in Chapter 4.

Table 2-1: summary of different interpolation methods

Interpolation method	Principle	Case	Advantage	Disadvantage	Best-suited scenario	Choose or not ?why
Inverse distance weighting (IDW)	Linear combination of the surrounding locations, weighted inversely by distance.	REGNIE (2008)	Exact interpolator. Fast, easy to use and tailored for specific needs. Preform will	Weights are not affected by spatial arrangement.	Middle dense sampling in local area	Yes. Ease to use and suitable dense sampling. Well performed in local area.

			with noisy data.			
Nearest Neighbor (NN)	Assigns the un-sample value from the nearest point.	Dense measurement network.	Fast and simple	Not work well in all case	Densely sampled data	No. Application in meteorology is limited
Thiessen polygon (THI)	Similar as the nearest neighborhood method		Fast and simple	Limited in mountainous regions	Uniform distribution data	No. The rain gauge data in the Netherlands is random
Polynomial functions (splines)	Fit trend functions through the observations by x-order polynomials.	Ancillary data can be included. eg. ANUSPLIN (2008)	Visually appealing curves or contour lines	May mask uncertainty present in the data.	Irregularly-spaced data	No. Monthly and yearly climate elements but are less suitable on higher temporal resolutions like days and hours.
Linear regression	A straight line is fitted through the data points.	Monthly gridded datasets in combination with IDW (2005)	Most are global interpolators.	stochastic	Expresses the relation between variables	No. It usually used as global interpolators. For spatial data, it is difficult to process in R.
Artificial neural networks (ANN)	Model complex relationships between inputs and outputs or to find patterns in data.	Calculating mean monthly temperature	Ability to learn and generalize data; works well with sparse data distributions; and extrapolation capability.	Requires good coverage of the input space; Risk of poor interpolation caused by over-learning or under-learning.	Regions ranging from sparse irregularly distributed data to well-distributed data	Yes. The use of ANN is a relative new methodology for interpolation, is highly black box and requires excessive computing power.
Kriging	Similar to the principle of IDW; however additionally accounts for the spatial arrangement		Best linear unbiased spatial predictor; and no edge-effects resulting from trying to force a polynomial to fit the data.	Sophisticated programming requited; and problems of nonstationarity in real-world datasets.	Well-distributed data, with no discontinuities.	Yes. Most promising techniques

The three interpolation methods will be accomplished in R script. Much of interpolation approach is included in the package 'gstat'. It offers widely functions in the geostatistics curriculum. This package

contains variogram modelling, everything from global simple kriging to local universal cokriging (Eda, 2013). Other R packages such as maptools, rgdal provide additional geostatistical functions.

2.5. Resample

Resample is a raster process in ArcGIS. During this process, the cell size will change while the extent of the raster dataset will not change. Generally, the accuracy of resample is higher when resample process is from high resolution to a coarser resolution.

In ArcGIS, four options resampling technique are provided: nearest, majority, bilinear, cubic. Among these methods, the bilinear and cubic will make the cell values altered when dealing with categorical data. At this time, the nearest method can be used since it does not crates new values.

Nearest

Nearest is the fastest interpolation method which performs as a nearest neighbour assignment ("ArcGIS Help 10.1 - Resample (Data Management),"). The maximum spatial error will be non-half the cell size due to it not changed the values in the cells.

Majority

Majority determine the new value by a majority algorithm, the new value is based on the most popular values in the filter window. It is also suitable for discrete data and smoother than nearest.

Bilinear

The bilinear interpolation calculated the predicted value based on a weighted distance average of the four nearest data centers. It works well with continuous data.

Cubic

The theory of cubic is based on cubic convolution. It predicts values by fitting a smooth curve of the sixty nearest data centers. Its predict value may out of the range of input raster data but it is still appropriate for continues data. The geometrical is less distorted, so the process of cubic requires more time.

3. DATASETS

3.1. Study area

The study area is the Netherlands. It covers an area of 41526 square kilometres, east longitude from 3 to 7 and north latitude of 49 to 53. The Netherlands can be defined as the urbanized area inside the river delta. It is a small and flat country, so climatological differences are small. Many parts are situated below the sea level. The topography is very flat, only a few hill in the east and south. Due to the proximity of the ocean and the effect of the north Atlantic Gulf Stream, it belongs to the temperate zone climate with small climatological variations. The mean annual rainfall changes from 725mm to 925mm. Because of the coastal effects, the amounts of precipitation are smaller in the coastal zone when spring and lager in late autumn (Attema & Lenderink, 2011).

3.2. Datasets

3.2.1. Interpolation source data

The source datasets of the interpolation part are the rain gauge data which was provided by The Royal Netherlands Meteorological Institute (KNMI). There are two types of KNMI stations, the manual voluntary precipitation stations and the automatic meteorological stations. These rain gauge data are stored in the Klimaat Informatie Systeem (KIS) database and are quality controlled. In this study, interpolation source data were daily rainfall data measured by precipitation stations. The rainfall datasets and coordinates of each station have been downloaded from KNMI website.

The observation of the manual voluntary network contains more than 300 stations at different locations and measures the amount of rainfall (snow) once a day at 08.00 UT. This implies that a measurement date Jan 1st contains the interval from Dec 31th 08.00 to Jan 1st 08.00. The locations of stations do not change significantly since 1946, but the stations have different start and stop time, for example, some stations started from 1951 and stopped in 1989 and some start from 1966 stopped in 1990. With these restrictions, the time range selected for this study was 11 years from 2003 to 2013, with 325 stations. The locations of the 325 stations are shown in figure 3-1.

For each observation point contains:

- Station number.
- Station name.
- Actual data of measurement.
- Amount of rainfall (mm).

The dataset is grouped by monthly that already contains 325 stations, and then combine the coordinates with rainfall data according to station name and number.

3.2.2. Validation data

As motioned before, KNMI has two kinds of stations. In this study, data observed from automatic meteorological stations were used as the reference data for validation purpose. There are 32 synoptic stations available which include data of temperature, sun hours, clouds and visibility, barometric pressure, wind and precipitation. The automatic stations provide daily data and hourly data. However, the daily data are calculated from 12.30 UT is not the same with interpolation data, so the hourly data are downloaded.

The hourly data is provided by individual station. Filtered precipitation data in combination with all of the 32 stations are needed before validation. The hourly data were then calculated into daily data, with the start and end times being the same with interpolation source data. In addition, since the units of

coordinates are longitude and latitude, the coordinates system was changed into Dutch system in ILWIS. The locations of the 32 stations are shown in figure 3-1.

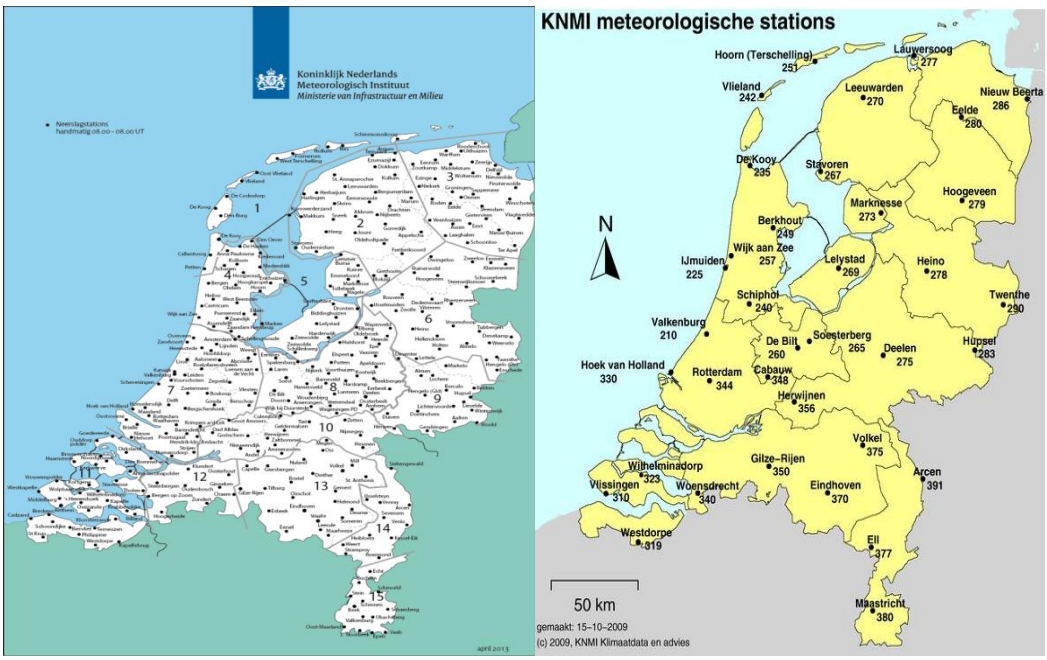


Figure 3-1 locations of 325 precipitation stations and 32 automatic meteorological stations

3.2.3. Resample data

The resample data are gridded files of a sum of daily precipitation in the Netherlands. Grids are based on invalidated data which have only been automatically pre-validated but near real time. The grids are measured on 100-300 locations of the voluntary network from 08.00-08.00UT. The file name contains the starting and end time, but the preceding day on the FTP server as 2/3 of the measurement period, so the measure time falls on the preceding day. ("KNMI Data Centre,")

The grids data is interpolated by ordinary kriging method and the observations are square root transformed and back-transformed after interpolation using quantiles calculation. In ordinary kriging method, the variogram is automatically fitted. So according to the best fit to determine the variogram model is spherical or exponential automatically and the nugget is zero.

3.2.4. Grids map and mask map

Grid maps were downloaded from WorldClim project page. WorldClim is a set of global climate layers with a spatial resolution of about 1 km. The precipitation map were then resampled it to 3km, 8km, 12km and 25km through ArcGIS software.

The Netherlands mask map was downloaded from CBS website. As grids map, they were resampled to 3km, 8km, 12km and 25km in ArcGIS.

3.2.5. CMORPH precipitation data

The National Oceanic and Atmospheric Administration (NOAA), the Climate Prediction Centre (CPC) and Morphing Technique (CMORPH) makes use of motions vectors based on 30 minutes and 8 kilometre (at equator) geo stationary satellite IR imagery to broadcast the estimations of precipitation produced by means of passive microwave information, and makes three hourly product in 0.25°×0.25° resolution for latitude between 60°N and 60°S since the last month of 2002. The high time and space resolution estimates of rainfall are required for many applications(Joyce et al., 2004).

The COMRPH data were downloaded from NOAA Climate Prediction Centre which is 30 minute estimates at 8km resolution. Temporal scaling has been done using ILWIS.

4. METHODOLOGY

4.1. Flow chart

Spatial interpolation projects the unknown point by estimating a regionalized value at some observed points by distance weighted. The main objective of the study is benchmarking rainfall interpolation over the Netherlands that at different spatial resolutions. To achieve this objective, the methodology is as follows:

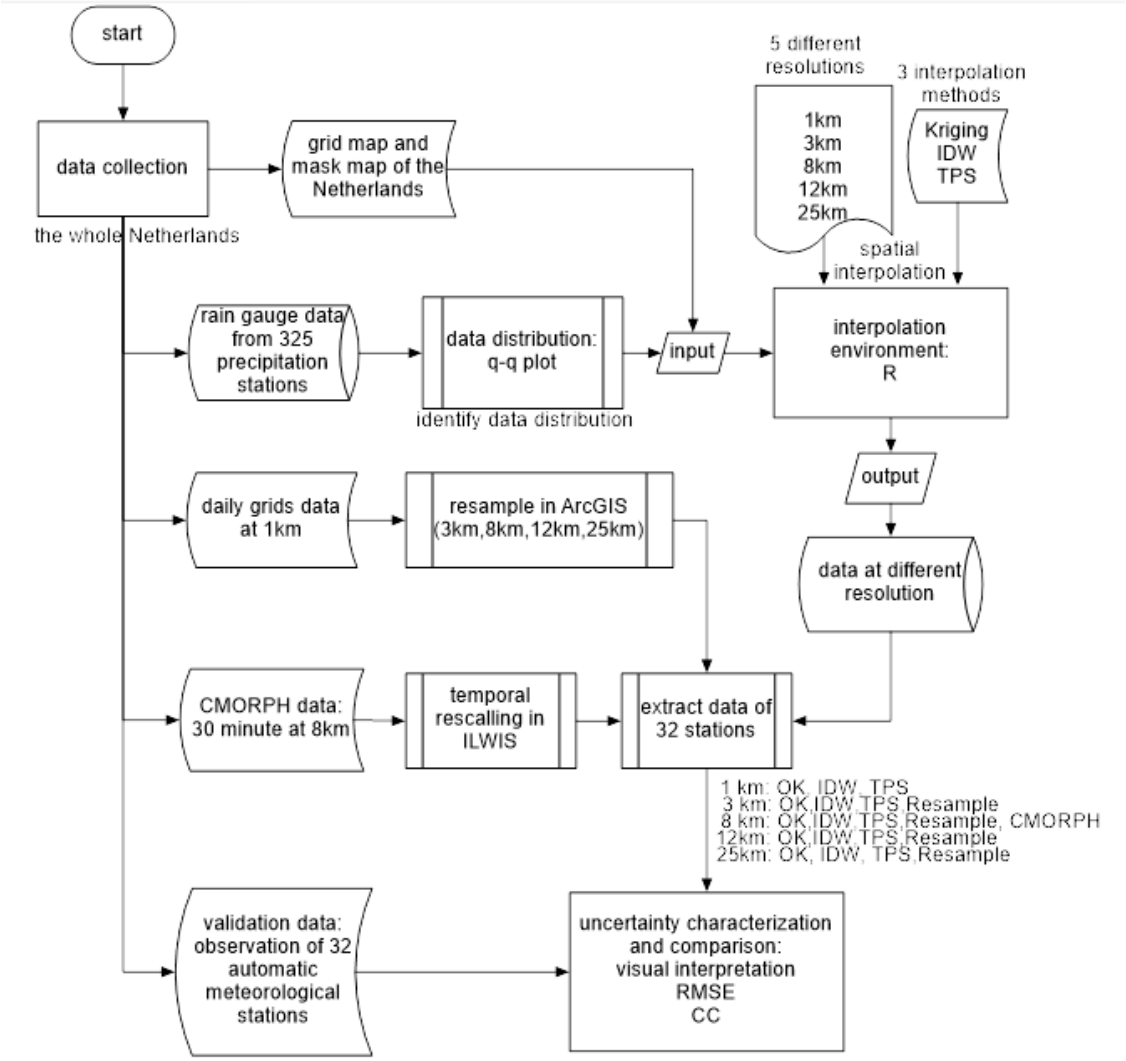


Figure 4-1 flow chart of methodology

The flow chart of methodology is shown in figure 4-1. Data collection included rain gauge data, grids data and satellite data. The q-q plot was used to estimate the data distribution. R script was the main interpolation environment. Three interpolation methods (ordinary kriging, IDW, TPS) were used to interpolate rainfall data at five different spatial resolutions. At the same time, grid data at 1km spatial resolution were resampled to 3km, 8km, 12km and 25km resolution in ArcGIS software. According the coordinates of 32 automatic meteorological stations, predicted rainfall data were extracted from each station at different spatial resolutions. The interpolated data and resample data were compared with the in situ data by visual interpretation and two statistical metrics (correlation, root mean square error). At 8km

resolution, satellite data from CMORPH were compared with the interpolated rainfall data for validation purpose.

4.2. R scripts

R is a free software environment for statistical computing and graphics, which uses packages for geospatial analysis (R-Project). The interpolation was done with R studio. Usually, interpolation is done on a regular grid, so the grid map and mask map of the Netherlands will be used.

The following packages have been applied:

- sp: a package that provides classes and methods to deal with spatial data. The classes document contains spatial location information resides and provides utility functions (Bivand, Rowlingson, & Gomez-rubio, 2014).
- maptools: a set of tools for manipulating and reading geographic data (Roger et al., 2014).
- gstat: containing variogram modelling; simple, ordinary kriging (Suggests et al., 2014). This package provides a number of functions for univariate and multivariate geostatistics and larger datasets.
- Fields: it is for the curve, surface and function fitting with an emphasis on thin plate splines (Douglas, Furrer, Sain, & Nychka, 2014).
- rgdal: bindings for the geospatial data and assess to projection/transformation operations

4.3. Data distribution

Prior to interpolation, the data distribution is needed because precipitation is irregular. Rainfall usually shows non-normal distribution, especially on days with convective rainfall. Point data were imported in R, with each station containing unique station number, coordinates and precipitation value. A normal quantile-quantile plot (q-q plot) has been used to show the data distribution. The plot would show whether the data is normal distributed or not. If the data is non-distributed, a log transformations or square root transformation could be used to meet normal distribution. Moreover, for kriging interpolation, Gaussian distribution is preferred. The log transformation could exaggerate the non-normal distribution.

Usually, quantile-quantile plot draws a quantile to compare two graphic methods of probability distribution. If the two distributions are linear correlation, the data points in the plot approximation to fall on a straight line. Moreover, the plot is also a kind data nonparametric method that compares the distribution of random variables.

4.4. Interpolation

4.4.1. Ordinary kriging

Kriging characterized spatial correlation through a variogram model. It is an optimal interpolation based on regression against unknown z values of surrounding data points, weighted according to spatial distance. There are many types of kriging (Sluiter, 2012). In this study, ordinary (point) kriging was used.

Ordinary kriging is the basic form of kriging. It assumes that mean is constant in the local neighbourhood of each estimation point, prediction using weighted linear combinations. The variogram describe the spatial correlation between the stations and determines the weight.

In R, the spatial correlation is modelled by the variogram instead of a correlogram or covariogram. Therefore, the first step of ordinary kriging interpolation is the semi-variogram modelling that determines the suitable variogram for all moment data. Usually, the variogram choose by cross validation. In this thesis, semi-variogram chose through considered most rainfall pattern in the Netherlands and refer to a spatial interpolation exercise in Netherlands did by Tomislav Hengl.

Following the q-q plot, a log-transformation was preferred before spatial prediction. The transformation of rainfall data could avoid higher estimation. In addition, because the input rainfall data is organized as

monthly, some details in the code should be change before running. The output files include precipitation map and a variance maps, grids data in ASCLL format and txt format.

4.4.2. Inverse Distance Weighting (IDW)

Inverse distance weighting is an advanced method of spatial interpolation. It is based on the theory that the value of an attribute z at an un-sampled point is a distance-weighted average of data points which nearly the point. That means values at unknown points can be calculated by using linear combination of values at known points. The inverse distance power determines the degree to which the nearer points are preferred over more distant points (Eda, 2013). However, the main disadvantage of IDW is 'bulls-eye' influence. It occurs when value of certain individual point is much higher than surroundings. Code of IDW is similar to Kriging, but IDW using data directly that without log transformation.

4.4.3. Thin Plate Splines (TPS)

The thin plate spline is the two-dimensional analogy of the cubic spline in one dimension. It works by fitting a surface to the data with some allowed error at each station (Tait, Henderson, Turner, and Zheng, 2006).

To fit irregularly space data using a thin plate spline, smoothing parameter should be chosen by generalized cross-validation. The assumed model in R is

$$Y = f(x) + e \tag{4}$$

Where: f(x) is the dimensional surface.

The residual sum of squares subject is minimizing in a thin plate spline. The amount of smooth data is controlled by the smoothing parameter.

In this study, the first step is to define the smooth function according to the coordinates and amount of rainfall.

4.5. Resample

Image resampling process is used to interpolate the new cell values raster image during a resizing operation. The raster dataset resampled by changing the cell size and resampling method. Actually, ArcGIS software resampling is widely used in geostatistical analysis. In the thesis, the nearest interpolation method has been used for resample.

The following is the model used to resample:

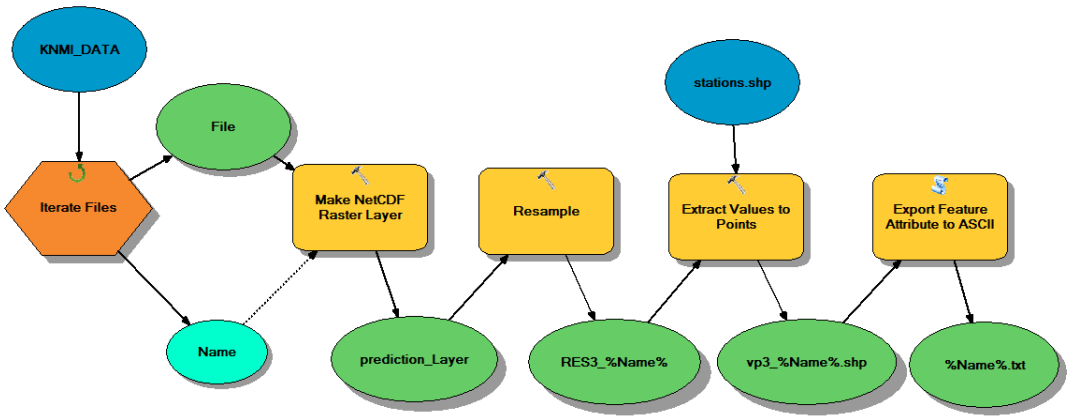


Figure 4-2 flow chart of resample

From the figure 4-2 we can see, the format of raster data is NetCDF, so the first step is to get the prediction layer which contains observations of precipitation by Make NetCDF Raster Layer function. Than using nearest function resample the raster 1km grids to 3km, 8km, 12km, and 25km. After resampling, put in the station map that includes the location of 32 automatic meteorological stations. The last step is to exact the resample value according to the coordinates and export these predict values into txt.

4.6. Temporal rescalling satellite data

The CMORPH data is 30 minutes at 8km, so the temporal rescaling did in ILWIS. Create map list which include the images for each from 08.00 AM to 08.00 AM. Each day contains total 48 images. Then don summation of these 48 images.

4.7. Result analysis

According to the objectives, three different interpolation methods (kriging, IDW, TPS) have been done at five different spatial resolutions (1km, 3km, 8km, 12km, and 25km). Each of them is daily data from 2003 to 2013 contains 4018 days in each resolution. In order to compare these results, the first step is to extract the predicted/interpolated rainfall values of 32 automatic metrological stations according to their locations. As a matter of fact, this is a repeat time-consuming work due to the result is daily, so a small produce is needed to extract each day in ASCII file. In this research, using R to extract point that location is near the automatic stations or the same with the reference stations, then extract the predicted values from daily to yearly into excel. Each resolution includes the observations of the most reference stations and interpolated values from three different interpolation methods. There are three ways of validation in this study.

4.7.1. Visual compare

The characteristic of visual interpretation is simple and directly. While visual compare may seem subjective, but the importance should not be underestimated or overlooked. Via reviewing a lot of interpolation results, the credibility of an interpolation method will be better guaranteed.

4.7.2. Correlation coefficient

Correlation coefficient is a statistical indicator that shows the strength and the direction of a linear relationship between two variables. CC is high if the two dataset have a strong positive linear correlation. As a matter of fact, the validation data we used is the existing observations that are validated by KNMI. There automatic stations are spread over the Netherlands. The interpolated data were compared with these observations that are not attends interpolation. They are two independent datasets.

4.7.3. Root mean square error

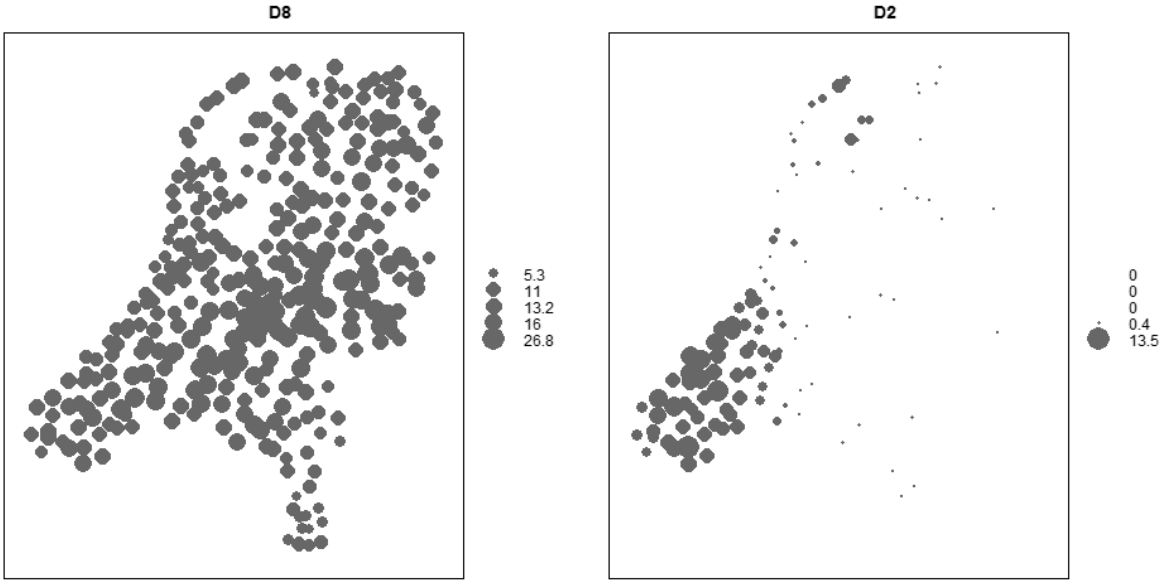
The Root Mean Square Error is a frequently used measure of the difference between values predicted by a model. The individual difference is also called residuals. In this thesis, RMSE will be calculated by taking the root of the sum of all squares of the differences between each sample of grid A and grid B, divided by the total number of samples:

$$\mathrm{RMSE} = \sqrt{\frac{\sum(A_i - B_i)^2}{n}} \qquad (4)$$

Where i stand for each individual pixel and n is the total number of pixels.
If the spatial distribution of the datasets is same, the values will be zero. The lower root-mean-squared error (RMSE) indicates the interpolator is likely to give reliable predict values for the un-sampled area.

5. IMPLEMENTATION AND RESULT

Three different interpolation methods and resampling method in ArcGIS are used to find the optimum method at different spatial resolutions over the Netherlands. Considering the type of precipitation may change every day, two kinds of different rainfall scenarios (stratiform rainfall and convective rainfall) (Schuurmans & Bierkens 2007) have been chosen to demonstrate the comparison results. The different spatial distribution of these two forms of rainfall can be seen in figure 5-1. They are bubble rainfall produced by 325 precipitation stations. Figure 5-1a is rainfall on December 8 2006 that presents stratiform rainfall. The intensity of stratiform rainfall is continuous but relatively low which means a fair amount of rainfall in one day. So it shows evenly spread spatial distribution. Figure 5-1b is precipitation on June 2 2004, a typical example of convective rainfall. The character of convective rainfall is high intensity with short duration in a local area, extreme rainfall event may occur in this situation. For some days when all the observations are zero, supposed that there was no rainfall in the Netherlands, so these days have no interpolation.



a b

Figure 5-1 bubble rain on December 8 2006 and June 2 2004: stratiform rainfall, convective rainfall

5.1. Data distribution and transformation

Because of the irregularity of precipitation, the observed rainfall data often shows a non-normal distribution. So q-q plot is needed to visualize the distribution of sample data. Result of data distribution analysis includes each day from 2003 to 2013 which is totally 4018 days. Figure 5-2 shows the originally data distribution and the distribution after log transformation, here the two examples are December 8 2006 which is stratiform rainfall and June 2 2004 which represent the convective rainfall (as shown in Figure 5-1). From these two pictures we can see, the non-normal distribution is particular obvious on convective rainfall.

As mentioned before, choosing appropriate variogram for ordinary kriging interpolation is depending on the spatial data distribution. According to the q-q plot, a log transformation is needed in ordinary kriging method. IDW and TPS interpolations need linear relationship to build spatial autocorrelation.

Comparing the origin data distribution with the distribution after a log transformation we can see, the log transformation will exaggerate the non-normal distribution but not heavily distort the distribution which is already good.

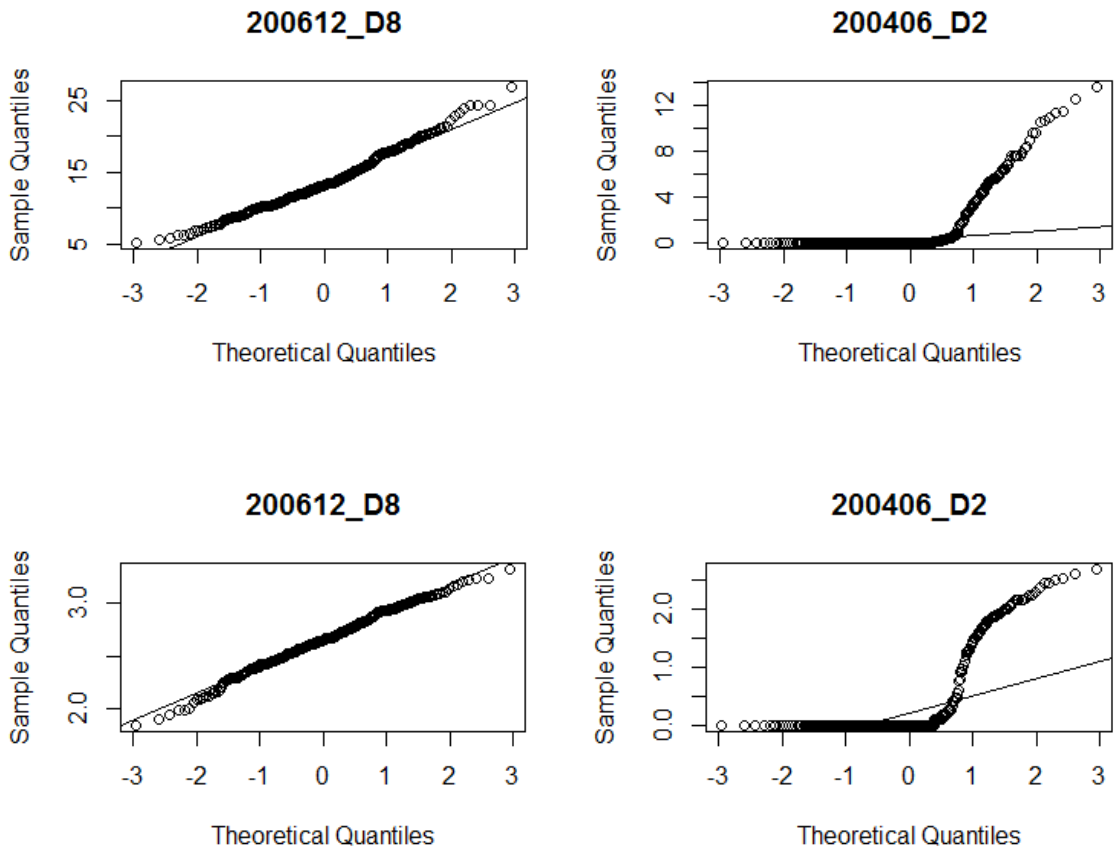


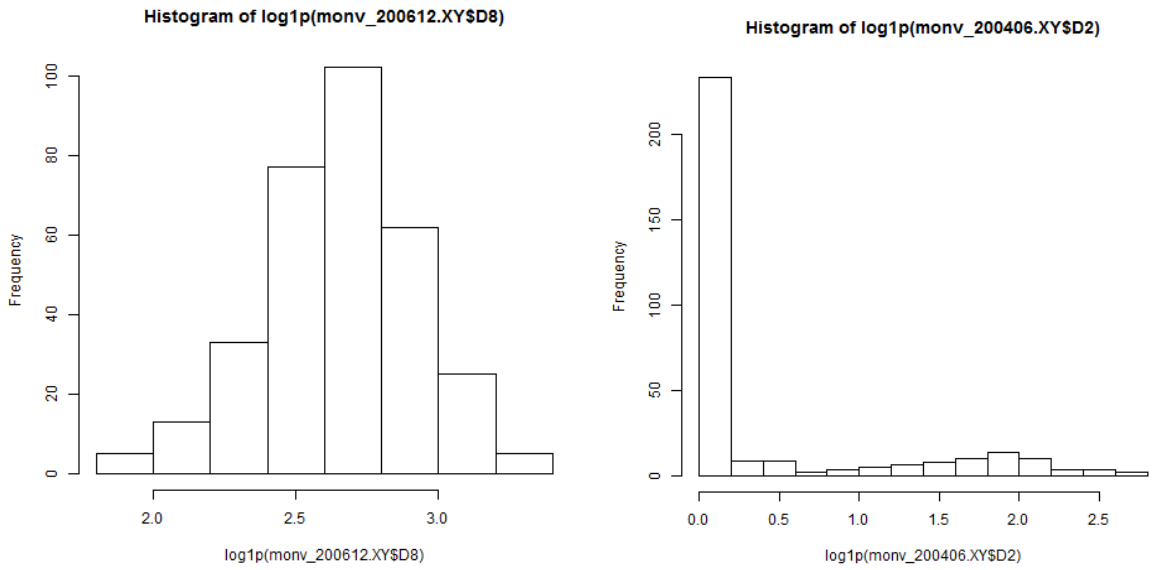
Figure 5-2 normal qq plot for December 8 2006 and June 2 2004: without transformation, with log transformation.

To identify whether the data obeyed the normal distribution by qq plot is to see if the data points is approximate a straight line. From the figure 5-2 we can see, stratiform rainfall is more close to normal distribution. In the qq plot, y coordinate is a standardized residual values and x coordinate is the theory interval.

5.2. Data histogram

Histogram is another kind of statistical graphical that represents the distribution of data and also shows the quality distribution. It consists of a series of high longitudinal stripes or lines represent data distribution.

From the histogram, the histogram of some days (like to stratiform rainfall) is close to normal distribution, some days (convective rainfall) are not very good, the histogram is various. The data distribution is consistent when compare the results of histogram with qq plots.



a b

Figure 5-3 histogram of rainfall after log transformation: December 8 2006 represent stratiform rainfall and June 2 2004 which represent convective rainfall.

After visual compare the histogram with bubble rain and qq-plots of all the days we found that when most stations can detect the rainfall (see picture 5-3a), histogram close to the normal distribution. So in the day that only some parts have rainfall (see picture 5-3b), the histogram is uneven distribution which will reduce the accuracy of the interpolation.

5.3. Variograms modeling for ordinary kriging interpolation

The first step of ordinary kriging interpolation is determining the semi-variograms. In fact, the semi-variogram is used to look the spatial structure of the data which produce the kriging weights by plotting semivariance against separation distance. The semivariance equals to one-half squared difference between the distances of two points. The good result is the Semi variance increased with the increase of distance between two samples. Therefore, choosing the appropriate variogram to ordinary kriging interpolation is very important. The prediction values were calculated based on the semi-variogram and the spatial arrangement of the surrounding data. Though the fit graph, the suitable model can be found when the trend line match the data. Generally, the semi-variogram determines the model which will be used to all the moment data. So the variogram can just meet majority of data.

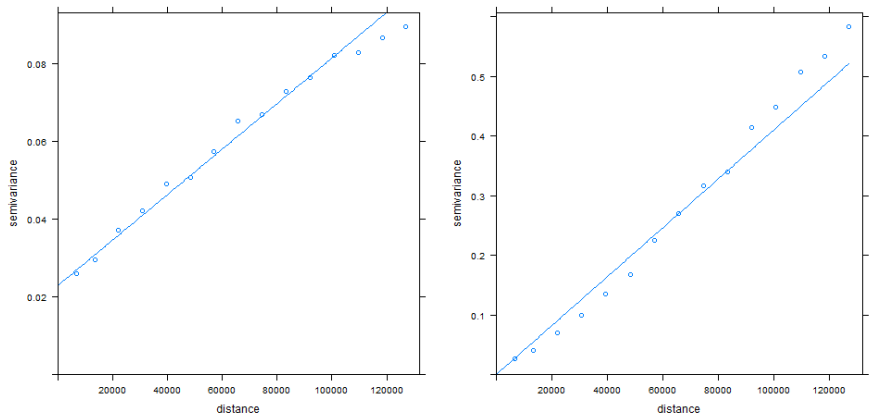


Figure 5-4 semi-variogram of daily rainfall on December 8 2006 and June 2 2004

The two examples in figure 5-4 are December 8 2006 and June 2 2004. The variogram includes 4018 pictures from 2003 to 2013. From those figures we can see that the data match majority of the trend line which ensure the accuracy of the interpolated data. Actually, the appropriate variogram could avoid negative daily rainfall in ordinary kriging interpolation (S. Ly, Charles, & Degré, 2011). After analysis most variogram, the negative estimates of kriging were observed for convective more than stratiform rain. So the spatial pattern is an important factor for interpolation.

5.4. Interpolation result at different resolutions

5.4.1. Ordinary Kriging, IDW, TPS and Resample at 1km

Figure 5-5 shows the results of three different interpolation results and resampling map from KNMI gridded file on December 8 2006 and June 2 2004 at 1km resolution. From left to right are ordinary kriging, IDW, TPS and resample. Using standard deviation stretching to enhance the visualisation of the patterns and each image is individually scaled, so the absolute value is different.

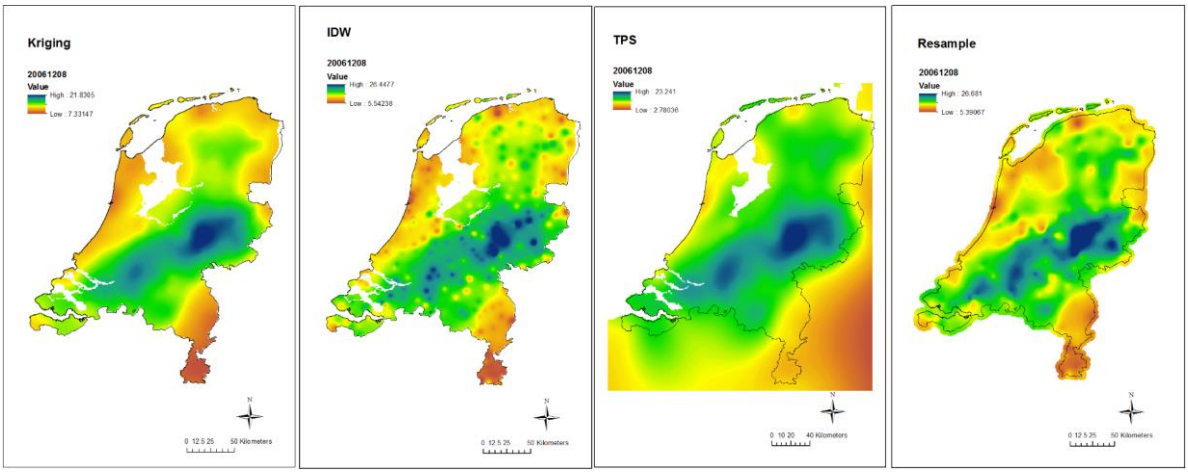


Figure 5-5a interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 1km resolution

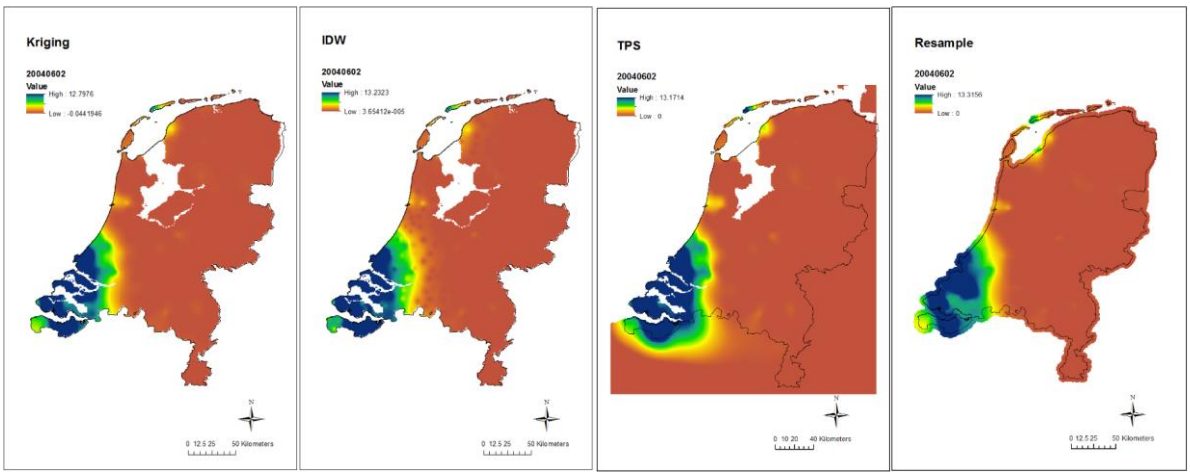


Figure 5-5b interpolation result of ordinary kriging, IDW, TPS and resample on June 2 2004 at 1km resolution

Figure 5-5a is the stratiform rainfall which is equally distributed. Comparing these graphs we can see, they show great overlap for the whole rainfall distribution. In fact, each of these methods has advantages and disadvantages. The range of precipitation on December 8, 2006 is about from 5.3mm to 26.8mm (see bubble rain in 5-1). For ordinary kriging which is more complex than other interpolation method, the range of interpolated rainfall is from 7.33mm to 21.83mm that is narrower than real situation. But it has a smooth representation of the spatial patterns. The main difference between kriging and IDW is the tendency of a bulls-eye pattern in the IDW results. Ordinary kriging gives a smooth surface than IDW interpolation. The rainfall range of IDW interpolation is from 5.54mm to 26.44mm which are closer with in-situ observed rainfall data than other interpolation methods. TPS shows the worse spatial patterns compared to the other three methods because some details are missing. For example, rainfall in the yellow range (on the North area) is not as clear as other three maps. The interpolated rainfall range of TPS is from 2.78mm to 23.24mm that is lower than actually rainfall. The resample map using grids data provide by KNMI website, here we just produce the map by ArcGIS. The spatial pattern of rainfall is between kriging and IDW. Rainfall range is 5.39mm to 26.6mm that is best matched with bubble rain result, so it ensures the future resampling is established on a very good basis.

Figure 5-5b shows the convective rainfall (the example is June 2 2004) interpolation results from three different interpolation methods. As can be seen in figure 5-5b, the spatial pattern in four maps shows a strong similarity. The practical range of rainfall is from 0mm to 13.5mm (see bubble rain in 5-1). All the range of IDW, TPS and resample are very close to this scope. However, the ordinary kriging interpolation result has negative values that may be explained by the negative kriging weights to extreme values (S. Ly et al., 2011). Moreover, this kind of situation often occurs only in convective rainfall.

After visual interpretation, predicted values at locations of 32 automatic meteorological stations are extracted to compare with the measured data. Correlation coefficient is calculated between each interpolated values and the in situ data for 11years at each station. Also the calculated Root Mean Square Error to analyse the interpolated data. Table 5-1 shows the results of correlation coefficient and Root Mean Square Error for each station. From the table we can see, all the interpolation method gives a good result, among them IDW interpolation has a better result.

Table 5-1 CC and RMSE of each method at 1km resolution

CC				RMSE			
STN	kriging	idw	tps	STN	kriging	IDW	TPS
210	0.958262	0.967733	0.962137	210	1.418096	1.314803	1.400875
235	0.949902	0.959456	0.955631	235	1.369492	1.291664	1.331927
240	0.932489	0.935653	0.937168	240	1.782776	1.771008	1.775146
249	0.942491	0.94861	0.950615	249	1.550034	1.473745	1.465095
251	0.948206	0.954555	0.951824	251	1.329792	1.279308	1.371551
257	0.654314	0.651004	0.630616	257	3.609583	3.676204	3.851932
260	0.969896	0.988536	0.978832	260	1.183208	0.788476	1.018241
265	0.619768	0.635123	0.63067	265	3.946452	3.874427	4.054212
267	0.945879	0.954453	0.948666	267	1.337065	1.230284	1.308721
269	0.944649	0.964147	0.956091	269	1.517895	1.225346	1.342165
270	0.949438	0.962314	0.953984	270	1.457397	1.302883	1.400708
273	0.935914	0.976553	0.955782	273	1.739326	1.046446	1.427768
275	0.945608	0.959476	0.950704	275	1.53195	1.359254	1.480814
277	0.940374	0.945303	0.949661	277	1.484237	1.430991	1.383027
278	0.869853	0.884533	0.848034	278	2.163948	2.034028	2.343883
279	0.930027	0.929352	0.933331	279	1.625627	1.625187	1.590159
280	0.950842	0.976537	0.965276	280	1.335379	0.950108	1.141351
283	0.964831	0.971385	0.973218	283	1.311655	1.214579	1.141679
286	0.930093	0.942421	0.939392	286	1.525499	1.44914	1.473879
290	0.94647	0.955864	0.951939	290	1.461259	1.354323	1.389779
310	0.946483	0.954428	0.957764	310	1.512787	1.419866	1.39726
319	0.835331	0.841588	0.810232	319	2.597545	2.549281	2.86526
323	0.956086	0.982078	0.963664	323	1.338295	0.957029	1.245608
330	0.949432	0.958655	0.951666	330	1.515278	1.385226	1.510606
344	0.951047	0.954816	0.956607	344	1.512444	1.464754	1.464683
350	0.953105	0.968173	0.964452	350	1.40171	1.213596	1.232537
356	0.931836	0.945544	0.948271	356	1.715882	1.544438	1.510883
370	0.947181	0.952976	0.955804	370	1.348198	1.319348	1.270773
375	0.93776	0.96765	0.967065	375	1.504805	1.21854	1.14294
377	0.949768	0.962362	0.96128	377	1.25344	1.108858	1.121975
380	0.949076	0.957011	0.955204	380	1.311369	1.250235	1.277879
391	0.943926	0.961279	0.96175	391	1.456999	1.230153	1.242698
mean	0.921261	0.933424	0.927417	mean	1.660919	1.511048	1.593

5.4.2. Ordinary Kriging, IDW, TPS and Resample at 3km

Figure 5-6 shows results of three different interpolation methods and resampling on December 8 2006 and June 2 2004 at 3km resolution. From left to right are ordinary kriging, IDW, TPS and resample. Each image has different scales and using standard deviation stretching to enhance the visualisation of the patterns. So the absolute value is different. Like results in 1km, all of the images show great overlap on spatial pattern.

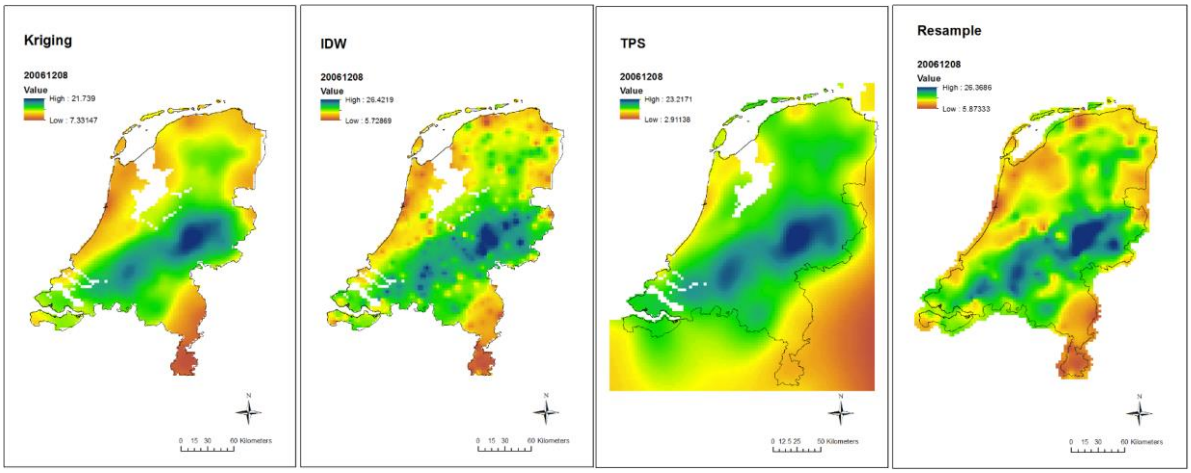


Figure 5-6a interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 3km resolution

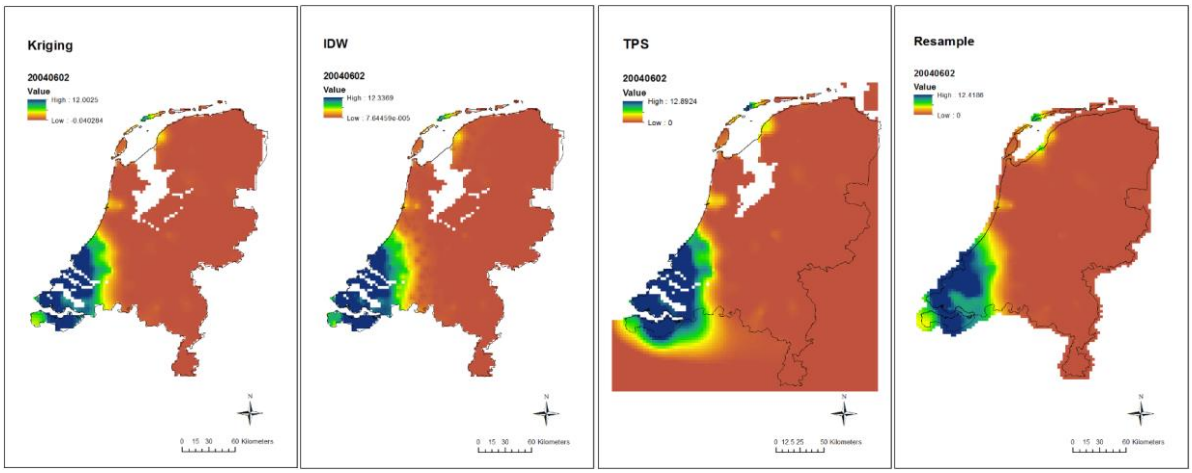


Figure 5-6b interpolation result of ordinary kriging, IDW, TPS and resample on June 2 2004 at 3km resolution

Figure 5-6a is interpolation results of each method on stratiform rainfall. From the pictures we can see, the range of rainfall in IDW and resample is more close to practical rainfall, ordinary kriging is narrow and TPS is lower.

Figure 5-6b is interpolation results of convective precipitation at 3km resolution. Here the example is June 2 2004. The spatial patterns in these images produced by different methods are very similar. However, the extreme values in the images are all decreased. So the result of convective rainfall interpolation is not as good as stratiform precipitation interpolation.

Exacted interpolated values at 32 automatic meteorological stations for 11 years were compared with the observation data. Correlation coefficient and Root Mean Square Error are calculated between them for each station. The following table shows the result. According to the table 5-2, IDW interpolation shows a better result both in CC and RMSE.

Table 5-2 CC and RMSE of each method at 3km resolution

CC					RMSE				
STN	kriging	IDW	TPS	resample	STN	kriging	IDW	TPS	resample
210	0.954769	0.967742	0.962137	0.966949	210	1.479145	1.314644	1.400875	1.402591
235	0.94383	0.958113	0.955233	0.9591	235	1.457535	1.336765	1.340095	1.310391
240	0.926319	0.935102	0.935612	0.936076	240	1.866308	1.768611	1.795946	1.834104
249	0.936201	0.948425	0.950169	0.953352	249	1.631457	1.475854	1.469268	1.421364
251	0.941696	0.955007	0.951936	0.95654	251	1.41467	1.282488	1.361851	1.277897
257	0.649087	0.641154	0.633725	0.624922	257	3.642579	3.742767	3.827099	3.852462
260	0.96397	0.985647	0.977444	0.986471	260	1.29295	0.838402	1.043593	0.824066
265	0.61735	0.635228	0.630803	0.636862	265	3.98326	3.935273	4.05258	3.966597
267	0.935797	0.954452	0.948666	0.955448	267	1.451128	1.230298	1.308721	1.216665
269	0.936268	0.964583	0.955744	0.964787	269	1.61915	1.21248	1.34792	1.212694
270	0.942712	0.962311	0.953984	0.959774	270	1.547965	1.302976	1.400708	1.365359
273	0.92676	0.976431	0.956393	0.975389	273	1.839745	1.04913	1.418452	1.07485
275	0.940466	0.959738	0.951573	0.959541	275	1.603117	1.355195	1.465637	1.364966
277	0.931488	0.939823	0.945064	0.942213	277	1.586967	1.493389	1.440856	1.473505
278	0.861853	0.885788	0.848034	0.85387	278	2.224972	2.023467	2.343883	2.290046
279	0.920472	0.929354	0.933407	0.940948	279	1.724668	1.62542	1.591971	1.492976
280	0.942056	0.974628	0.963285	0.974763	280	1.448329	0.976861	1.17097	0.976169
283	0.955985	0.97449	0.972842	0.961863	283	1.462948	1.125004	1.148369	1.35162
286	0.923421	0.939689	0.938992	0.940134	286	1.601746	1.50274	1.477042	1.485745
290	0.939942	0.95701	0.952768	0.959134	290	1.548329	1.330604	1.382858	1.308534
310	0.941351	0.951699	0.954657	0.956438	310	1.58681	1.466475	1.447403	1.413083
319	0.833873	0.842319	0.808851	0.807928	319	2.614984	2.547607	2.874993	2.857905
323	0.951787	0.981904	0.962453	0.980627	323	1.400164	0.920678	1.26327	0.954299
330	0.939432	0.945822	0.944365	0.956445	330	1.657017	1.575327	1.617811	1.432847
344	0.947728	0.955354	0.957269	0.958643	344	1.562584	1.455444	1.450481	1.445486
350	0.946742	0.969427	0.963921	0.969902	350	1.492104	1.167677	1.240387	1.165233
356	0.929445	0.948973	0.94972	0.955231	356	1.745272	1.496689	1.487953	1.410839
370	0.939799	0.953387	0.955929	0.957675	370	1.439756	1.313357	1.266767	1.243675
375	0.930871	0.969729	0.966503	0.970828	375	1.586076	1.140494	1.152558	1.130024
377	0.944067	0.962497	0.96128	0.967543	377	1.32107	1.107041	1.121975	1.030134
380	0.942119	0.953782	0.952149	0.954621	380	1.39657	1.293677	1.318867	1.288446
391	0.939558	0.961335	0.960782	0.908853	391	1.5089	1.237638	1.257905	1.856402
mean	0.914913	0.93253	0.92674	0.929777	mean	1.741821	1.52014	1.602783	1.554093

5.4.3. Ordinary Kriging, IDW, TPS and Resample at 8km

The following pictures are results of ordinary kriging, IDW, TPS interpolations and resample at 8km on December 8 2006 and June 2 2004. Each image has different scales and using standard deviation stretching to enhance the visualisation of the patterns. The absolute value is different. Like results before, the spatial pattern of all the images are similar. The minimum rainfall value in ordinary kriging, IDW and resample are higher than 1km and 3km. All the images in 8km are not as smooth as before, the small grids can been seen.

Figure 5-7a is December 8 2006 which represents stratiform rainfall. From these images we can see, rainfall on ordinary kriging interpolation map and TPS interpolation map have a uniform trend of decreased from middle part to boundary. However, rainfall on IDW and resampling maps is complex and irregular. Except resample, all the rainfall range is narrowing during the increase of spatial resolution, which means some extreme value are missing. The theory of resample is to alter the raster dataset by changing the cell size, so the range of grid value will not change which ensure the real precipitation scale at a certain extent.

Figure 5-7b is June 2 2004 that stands for convective rainfall. See maps of ordinary kriging and IDW, some rainfall values around rainfall centre are missing. This situation also occurs on the boundary.

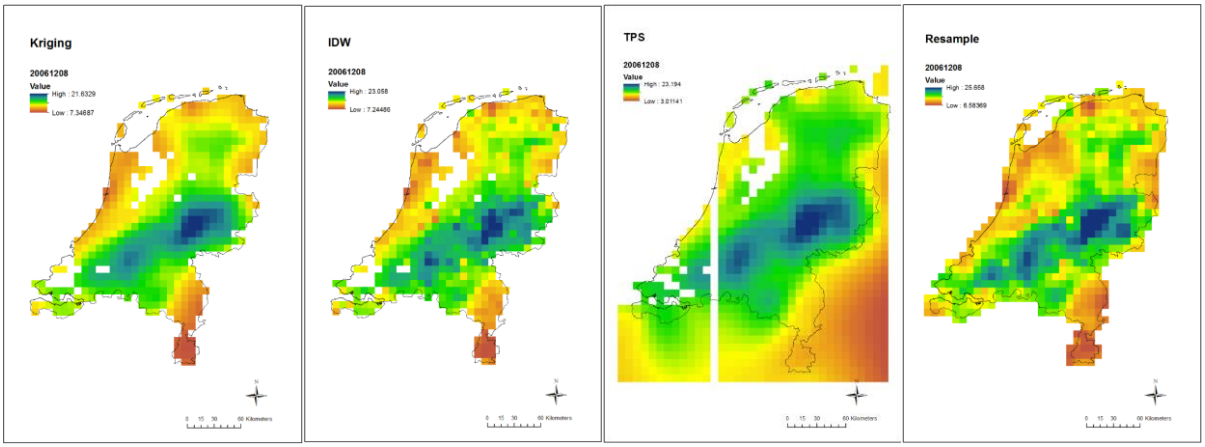


Figure 5-7a interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 8km resolution

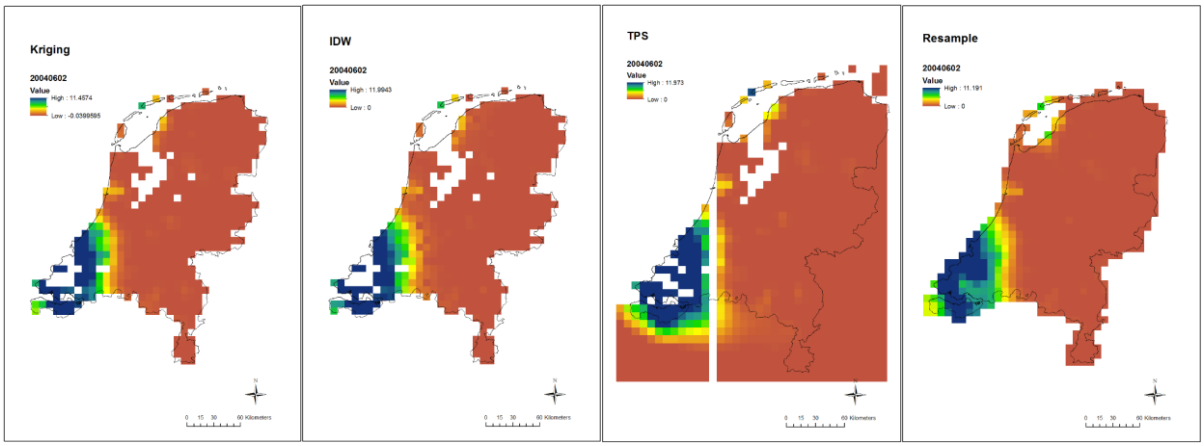


Figure 5-7b interpolation result of ordinary kriging, IDW, TPS and resample on June 2 2004 at 8km resolution

Table 5-3 CC and RMSE of each method at 8km resolution

CC					RMSE				
STN	kriging	IDW	TPS	resample	STN	kriging	IDW	TPS	resample
210	0.930195	0.938145	0.949894	0.956777	210	1.830869	1.87724	1.640909	1.488174
235				0.958844	235				1.306717
240	0.907807	0.926899	0.92517	0.928755	240	2.056543	1.872993	1.927765	1.930951
249	0.916059	0.937035	0.943217	0.943588	249	1.872714	1.634004	1.577768	1.566007
251				0.954066	251				1.310816
257					257				
260	0.945062	0.973045	0.973849	0.983999	260	1.589742	1.131376	1.120072	0.898229
265	0.613517	0.634608	0.630039	0.635437	265	3.928216	3.910878	4.067066	3.962749
267	0.920468	0.94058	0.94443	0.954552	267	1.615798	1.413034	1.366676	1.230806
269	0.928369	0.950557	0.953144	0.964968	269	1.720214	1.426882	1.386705	1.199597
270	0.938021	0.960064	0.956399	0.963568	270	1.607644	1.369016	1.366073	1.271485
273	0.920407	0.972349	0.954534	0.975336	273	1.916309	1.141446	1.451656	1.07526
275	0.931639	0.959243	0.952142	0.959021	275	1.708421	1.383076	1.460554	1.383847
277				0.935422	277				1.54521
278	0.846602	0.860921	0.829316	0.821333	278	2.328145	2.240982	2.499245	2.567513
279	0.907842	0.925084	0.924596	0.939278	279	1.855649	1.674668	1.683829	1.515021
280	0.935368	0.970202	0.961804	0.974503	280	1.521302	1.054985	1.195493	0.980386
283			0.972296	0.960588	283			1.162967	1.374963
286	0.917356	0.941877	0.939938	0.942231	286	1.648633	1.465735	1.483633	1.429461
290	0.935042	0.954159	0.949201	0.95314	290	1.597485	1.402999	1.430148	1.412925
310	0.924837	0.945541	0.949572	0.955295	310	1.793111	1.594028	1.532826	1.43534
319			0.810943	0.801254	319			2.863067	2.900281
323				0.980627	323				0.954299
330	0.930056	0.959077	0.951818	0.958315	330	1.78259	1.38559	1.512167	1.406511
344	0.927288	0.954284	0.954611	0.952094	344	1.84832	1.495914	1.511127	1.567554
350	0.936528	0.968826	0.964084	0.968586	350	1.624518	1.185849	1.24165	1.173418
356	0.917366	0.947673	0.95035	0.95527	356	1.882216	1.518434	1.482624	1.408846
370	0.919331	0.957825	0.954348	0.955328	370	1.670739	1.228156	1.292306	1.264409
375	0.909915	0.957274	0.958863	0.952596	375	1.79997	1.276339	1.267234	1.351767
377	0.929835	0.965468	0.961803	0.968263	377	1.482107	1.072707	1.120886	1.028531
380	0.902307	0.951406	0.951099	0.957839	380	1.846703	1.327436	1.337983	1.24229
391	0.922208	0.958391	0.959751	0.904847	391	1.715215	1.28424	1.278517	1.898609
mean	0.908537	0.936421	0.930638	0.935991	mean	1.849727	1.53472	1.602257	1.518773

Exacted interpolated values at 32 automatic meteorological stations for 11 years were compared with the observation data. Table 5-3 shows correlation coefficient and Root Mean Square Error that calculated

between predicted values and observations in each station. The same with the result of visual compare, rainfall data at some station are missing at 8km resolution. Correlation coefficient of IDW, TPS and resample are almost the same. Correlation coefficient of Ordinary kriging is less than other three methods. For Root Mean Square Error, resample gives lowest value, and then is IDW and TPS. Ordinary kriging has highest RMSE at 8km resolution.

5.4.4. Ordinary Kriging, IDW, TPS and Resample at 12km

Figure 5-8 are results for ordinary kriging interpolation, IDW interpolation, TPS interpolation and resampling at 12km on December 8 2006 and June 2 2004. In order to compare changes in different interpolation methods better, each image using linear minimum-maximum stretching to enhance spatial pattern. All the images in this spatial resolution are not smooth, the grids are obviously seen.

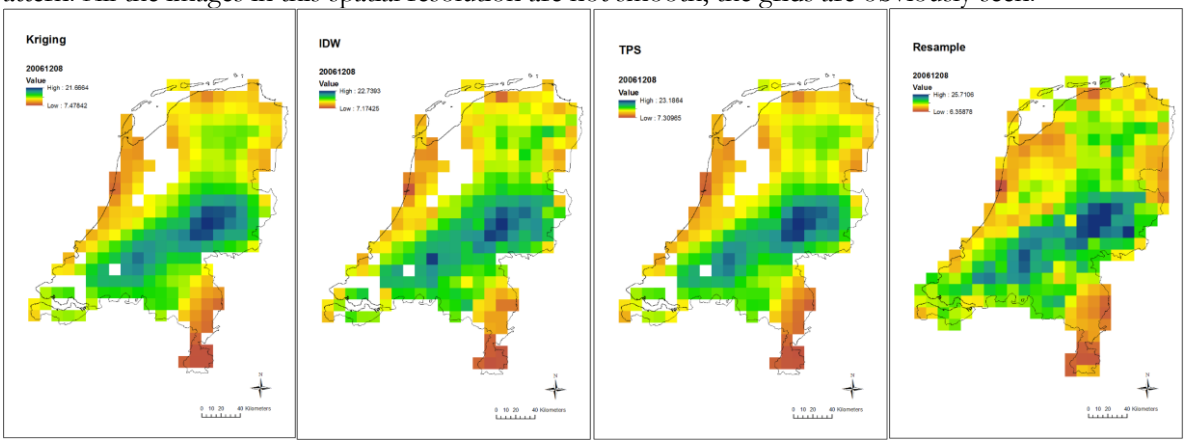


Figure 5-8a interpolation result of ordinary kriging, IDW, TPS and resample on December 6 2006 at 12km resolution

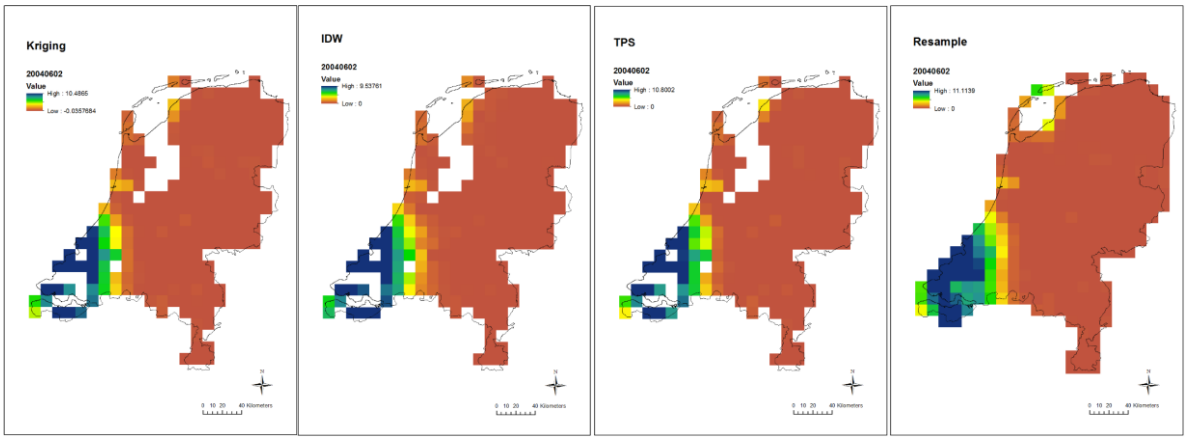


Figure 5-8b interpolation result of ordinary kriging, IDW, TPS and resample on June 2 2004 at 12km resolution

Figure 5-8a is interpolated rainfall results from three different interpolation methods and resampling method. The example is December 8 2006 which represents stratiform rainfall. Comparing these figures, it is easy to see that some points are uncovered in ordinary kriging, IDW and TPS interpolation methods. In addition, the similarity of these images is relatively reduced. Ordinary kriging and TPS are almost the same. IDW and resample are more dispersed. The predicted ranges of rainfall by ordinary kriging, IDW and TPS are further narrow than before.

Figure 5-8b is convective rainfall results from three different interpolation methods and resampling method. The example here is June 2 2004. For convective rainfall, the extreme rainfall information may miss when the spatial resolution is too big. The highest rainfall value is almost 4mm lower than real situation.

Extracted predicted values of 32 automatic meteorological stations for 11 years were compared with the practical observations. Correlation coefficient and Root Mean Square Error are calculated for each station, the results can be seen in table 5-4. From the table, interpolated values of some stations are missing due to the coarse spatial resolution. The mean CC of IDW is highest and the mean RMSE is lowest.

Table 5-4 CC and RMSE of each method at 12km resolution

CC					RMSE				
STN	kriging	IDW	TPS	resample	STN	kriging	IDW	TPS	resample
210	0.954385	0.957329	0.958844	0.965649	210	1.485983	1.492801	1.477084	1.399179
235	0.949169	0.959217	0.953114	0.959479	235	1.380873	1.297529	1.366702	1.293389
240	0.923989	0.930231	0.926717	0.927769	240	1.8793	1.838687	1.907196	1.860534
249	0.9315	0.930823	0.93886	0.943926	249	1.700279	1.720863	1.653567	1.567747
251				0.955611	251				1.278719
257				0.636556	257				3.776466
260	0.954845	0.957809	0.966201	0.979909	260	1.447967	1.403013	1.258684	0.977779
265	0.618869	0.62873	0.629976	0.634268	265	4.000373	4.176109	4.051832	4.005014
267	0.932816	0.926252	0.926338	0.946272	267	1.487921	1.588615	1.608383	1.333175
269	0.940215	0.955627	0.953547	0.959206	269	1.578717	1.349764	1.379255	1.29288
270	0.935635	0.944308	0.93599	0.918817	270	1.639729	1.520251	1.650604	1.898247
273	0.935666	0.976925	0.956863	0.969373	273	1.745222	1.044723	1.411396	1.194488
275	0.943458	0.948954	0.946377	0.958253	275	1.55792	1.49269	1.539367	1.372766
277				0.930926	277				1.632211
278	0.85877	0.880591	0.824663	0.79745	278	2.248692	2.066362	2.509763	2.768455
279	0.919703	0.927904	0.922142	0.933795	279	1.737277	1.640111	1.70721	1.574743
280	0.949436	0.976165	0.96391	0.968971	280	1.35422	0.95427	1.166545	1.07993
283	0.959116	0.963605	0.96491	0.962481	283	1.419133	1.334366	1.303957	1.339706
286	0.92722	0.941073	0.93662	0.935002	286	1.557198	1.472826	1.498782	1.513219
290	0.94858	0.955286	0.951486	0.956436	290	1.434882	1.383245	1.406949	1.346954
310	0.933996	0.937755	0.940527	0.955249	310	1.700118	1.738557	1.662638	1.414343
319				0.802508	319				2.891698
323	0.946831	0.949962	0.954597	0.972072	323	1.4793	1.493238	1.399877	1.134112
330	0.934198	0.926687	0.933356	0.950239	330	1.730445	1.852461	1.770428	1.539083
344	0.948839	0.950523	0.954456	0.95925	344	1.54555	1.52947	1.493278	1.43665
350	0.94157	0.949461	0.951695	0.962744	350	1.5619	1.459002	1.430675	1.261465
356	0.931226	0.948055	0.944466	0.95323	356	1.723216	1.51704	1.562955	1.441424
370	0.937477	0.944348	0.942379	0.918757	370	1.468854	1.405027	1.465672	1.729701
375	0.922851	0.942112	0.949729	0.961485	375	1.66125	1.462405	1.381622	1.23726
377	0.94147	0.941638	0.948068	0.952067	377	1.350298	1.363479	1.289717	1.236085
380	0.933946	0.913348	0.936713	0.941667	380	1.516013	1.889014	1.545945	1.503315
391	0.943494	0.961217	0.962569	0.911738	391	1.462602	1.229807	1.224674	1.828274
mean	0.924974	0.933069	0.931254	0.921286	mean	1.673401	1.59699	1.611598	1.661219

5.4.5. Ordinary Kriging, IDW, TPS and Resample at 25km

Images below are interpolation results of ordinary kriging, IDW, TPS interpolations and resampling at 25km. The two examples are December 8 2006 and June 2 2004. Linear minimum-maximum stretching has been used to enhance spatial pattern in each image. As a matter of fact, images on 25km resolution are crude, the grids are very obvious.

Figure 5-9a is interpolation results of different interpolation methods and resample on December 8 2006. From these maps we can see, the spatial pattern in ordinary kriging and IDW are almost the same. Value changes of TPS map are more simply. To the contrary, resample contains various rainfall changes. Moreover, many stations are uncovered in ordinary kriging and IDW.

Interpolation results of three different interpolation methods and resample method can be seen in figure 5-9b. Here the example represents the convective rainfall. Ordinary kriging interpolation gives less value changes than other three methods. In addition, the predicted data are missing which provides difficulty for study convective rainfall in local or small area. The predicted rainfall range of kriging and IDW is nearly

6mm which is half less than the observed rainfall (13mm). The predicted range is relatively higher in TPS and resample.

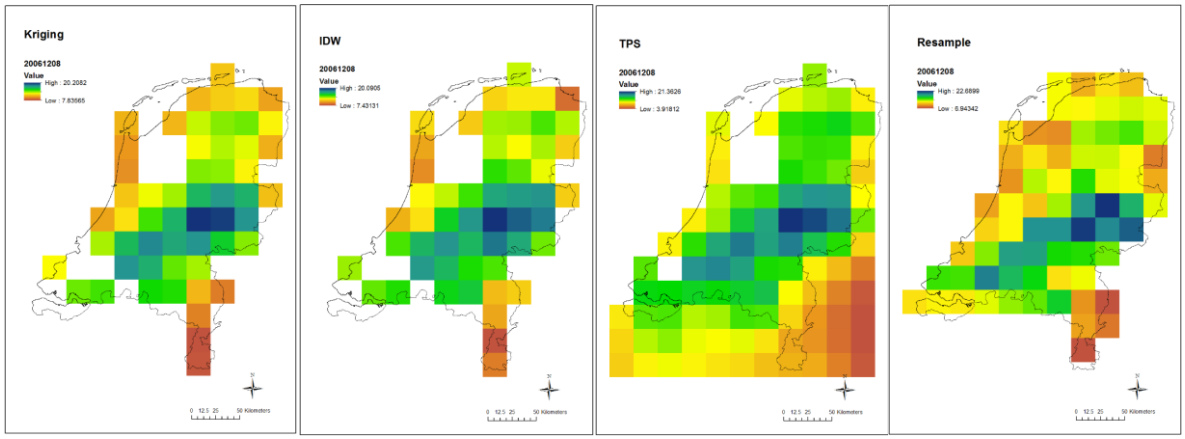


Figure 5-9a interpolation result of ordinary kriging, IDW, TPS and resample on December 8 2006 at 25km resolution

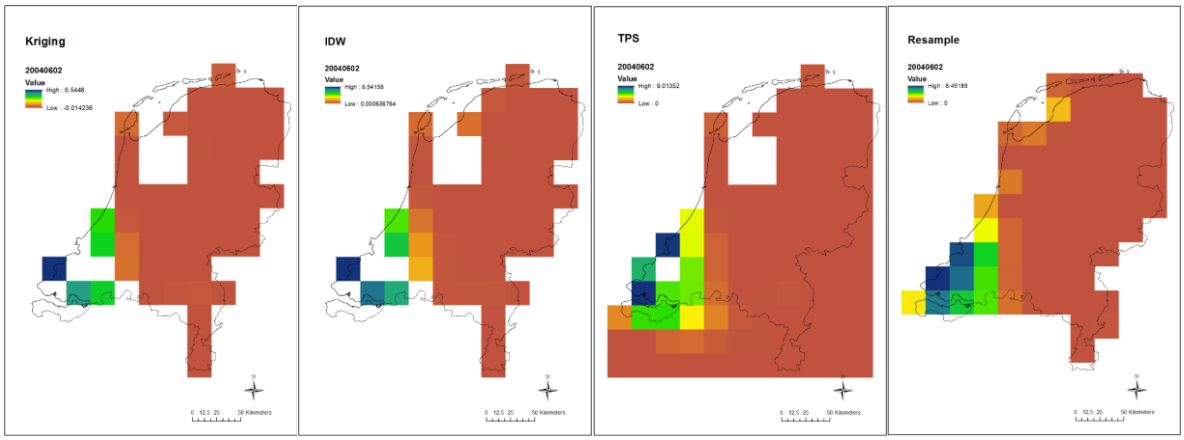


Figure 5-9b interpolation results of ordinary kriging, IDW, TPS and resample on June 2 2004 at 25km resolution

The interpolated values in 32 automatic meteorological stations are extracted according to the coordinates. Correlation coefficient and Root Mean Square Error are calculated in each station. From the table below we can see that many stations cannot be interpolated in 25km because of the coarse resolution. At this spatial resolution, resample gives a better result due to the higher correlation coefficient and lower Root Mean Square Error. Moreover, this method will contain almost all of the stations.

Table 5-5 CC and RMSE of each method at 25km resolution

CC					RMSE				
STN	kriging	idw	tps	resample	STN	kriging	idw	tps	resample
210	0.937745	0.955965	0.959496	0.916164	210	1.695779	1.49547	1.453565	2.044377
235	0.888146	0.908697	0.907166	0.955059	235	2.014961	1.853644	1.883698	1.330811
240	0.908003	0.929923	0.924404	0.930624	240	2.044827	1.852207	1.935393	1.838992
249				0.914443	249				1.9055
251				0.925575	251				1.677089
257				0.6491	257				3.736678
260	0.936499	0.952083	0.957778	0.974208	260	1.700196	1.491223	1.420065	1.10015
265	0.612363	0.626527	0.617977	0.615762	265	3.81984	3.835575	4.050397	3.990401
267				0.922665	267				1.642928
269	0.907185	0.915619	0.930341	0.955487	269	1.927269	1.852288	1.697412	1.350348
270	0.930576	0.94013	0.9472	0.943354	270	1.700848	1.580337	1.487216	1.537361
273	0.860884	0.876749	0.893211	0.950508	273	2.469078	2.331268	2.181217	1.5128
275	0.907006	0.909228	0.919481	0.959541	275	1.98026	1.981688	1.889195	1.364966
277	0.881203	0.901494	0.892238	0.935863	277	2.067734	1.899069	2.022567	1.542358
278	0.852766	0.872426	0.838175	0.863762	278	2.279657	2.144657	2.421756	2.212883
279	0.908906	0.91932	0.917089	0.932863	279	1.836907	1.73537	1.807057	1.584206
280	0.912066	0.921482	0.926627	0.927537	280	1.770053	1.694011	1.675192	1.638626
283	0.929128	0.933446	0.943306	0.947242	283	1.852939	1.795952	1.642653	1.585691
286	0.885562	0.866871	0.88188	0.92976	286	1.930397	2.176231	2.024756	1.647224
290	0.905453	0.914381	0.912337	0.951217	290	1.922805	1.824856	1.842459	1.402517
310			0.79234	0.915167	310			3.086467	1.964895
319	0.799715	0.828629	0.800141	0.820446	319	2.848446	2.652786	2.927798	2.758893
323	0.897695	0.915446	0.906462	0.903328	323	2.021667	1.858175	1.990468	1.986539
330			0.888733		330			2.310879	
344	0.920684	0.944754	0.947823	0.915588	344	1.922811	1.617851	1.592057	2.055569
350			0.913302	0.911392	350			1.911624	1.951346
356	0.898792	0.921356	0.927131	0.948001	356	2.070128	1.846589	1.800262	1.522803
370	0.913993	0.946607	0.947763	0.911881	370	1.719911	1.369637	1.371435	1.776366
375	0.896047	0.934267	0.927224	0.897225	375	1.928686	1.566334	1.68355	1.987818
377	0.91133	0.922798	0.925536	0.937656	377	1.657752	1.569525	1.556799	1.406129
380	0.869556	0.913341	0.905009	0.940194	380	2.145771	1.798258	1.906519	1.506084
391	0.921062	0.961339	0.956531		391	1.722117	1.232651	1.323623	
mean	0.887695	0.905315	0.900239	0.90672	mean	2.042034	1.882226	1.960574	1.852078

6. VALIDATION AND DISCUSSION

Results of the different interpolation algorithms provided some insights in terms of advantage and disadvantages. All the deterministic and geo-statistical methods to interpolate daily rainfall data over the Netherlands were able to produce 11 year daily precipitation data at different spatial resolutions.
However, ordinary kriging can lead to negative estimates (Deutsch, 1996). Negative weights occurs when applied to high rainfall may lead to negative and non-physical estimates (S. Ly et al., 2011). In this research, negative weights usually occurred in convective rainfall situation. All the negative values were replaced with zero value in this thesis.

6.1. Ordinary Kriging variance

All ordinary kriging methods create both a prediction map and a variance map. The kriging variances at different spatial resolutions are shown in figure 6-1. From left to right are 1km, 3km, 8km, 12km, and 25km. Here examples are December 8 2006 and June 2 2004. From the pictures, the variance on stratiform rainfall is lower than convective rainfall. Some places are nearly zero which ensured the accuracy of ordinary kriging.

Figure 6-1a ordinary kriging variance on stratiform rainfall at 1km, 3km, 8km, 12km and 25km

Figure 6-1b the ordinary kriging variance on convective rainfall at 1km, 3km, 8km, 12km and 25km

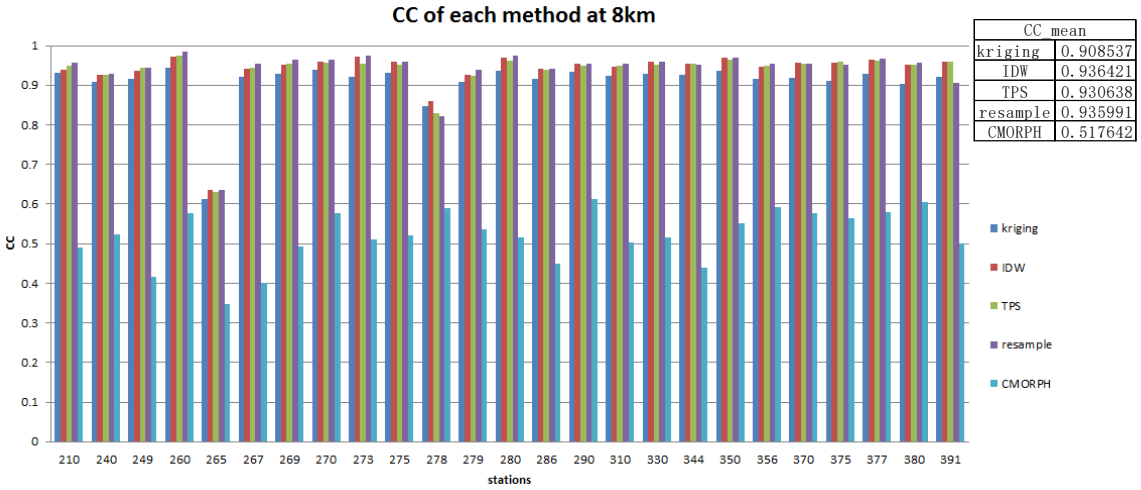
Table 6-1 is details about OK variance on December 8 2006 and June 2 2004. As table shown, the OK variance of convective rainfall is higher than stratiform rainfall. Therefore, ordinary kriging is good at interpolation on the precipitation that has an evenly distribution.

Table 6-1 OK variance on December 8 2006 and June 2 2004 at different resolutions

OK variance	1km	3km	8km	12km	25km
December 8 2006	0.0027-0.0040	0.0027-0.0039	0-0.035	0-0.0035	0.028-0.034
June 2 2004	0.0029-0.088	0.0046-0.084	0-0.053	0-0.050	0.01-0.04

6.2. Comparation of interpolation methods

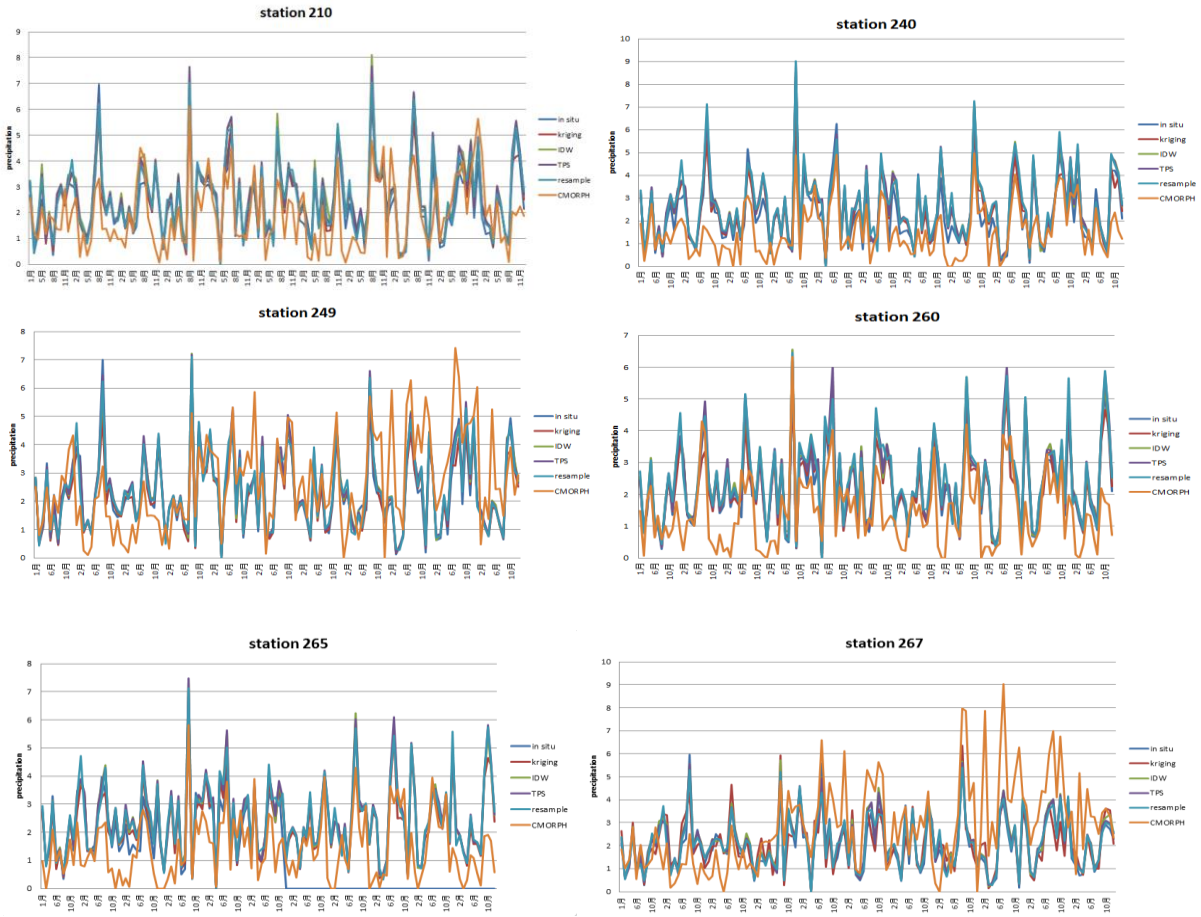
In order to find the optimal interpolation at each spatial resolution, the predicted values of 32 automatic meteorological stations were extracted and compared with in-situ data. Figures below is the results at 8km. Considering the CMORPH data has a spatial resolution of 8km, the comparison implemented at 8km with in situ precipitation data, interpolated data from ordinary kriging, IDW, TPS interpolation and resampling data. The comparison process at other four resolutions is the similar, without satellite data.

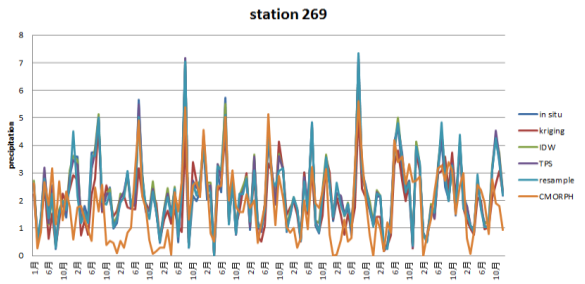


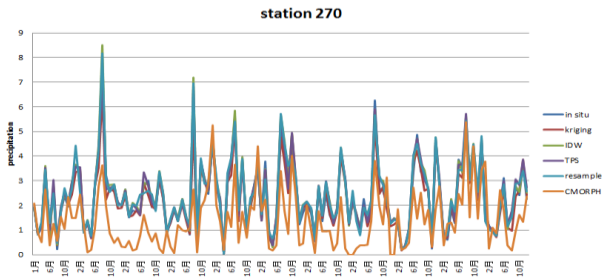
CC_mean	
kriging	0.908537
IDW	0.936421
TPS	0.930638
resample	0.935991
CMORPH	0.517642

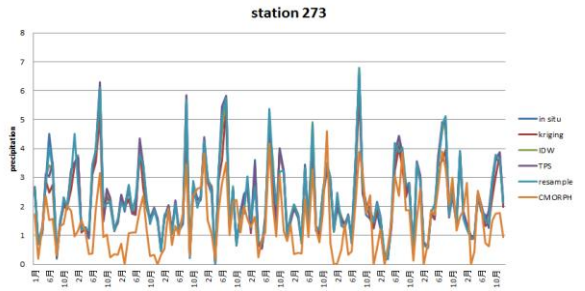
Figure 6-2a the correlation coefficient for 32 stations with each method

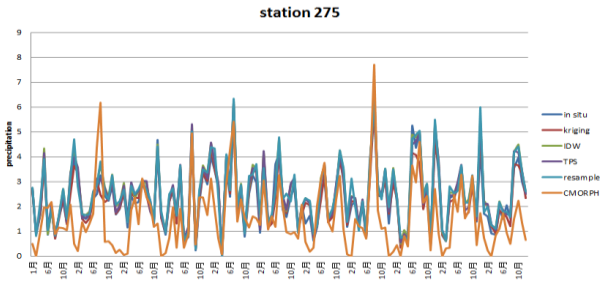
Figure 6-2a is the correlation coefficient of ordinary kriging, IDW, TPS, resample and CMORPH in each station. As graph shown, correlation coefficient of most stations are more than 0.9 which means the result is acceptable. IDW and resample are nearly 0.94. The satellite data is relatively lower. However, from the following graphs below, trends of CMORPH data is very similar with interpolated data. So the interpolated data can be a good reference data for CMORPH downscaling.

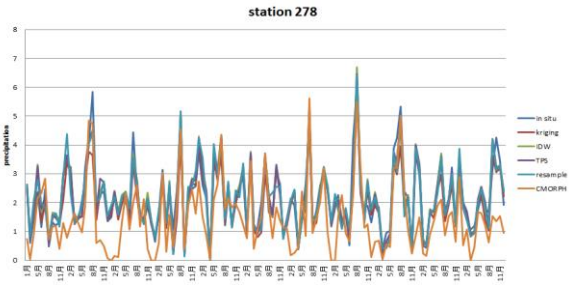


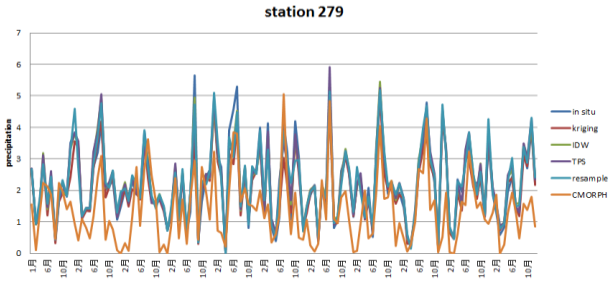

station 269
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

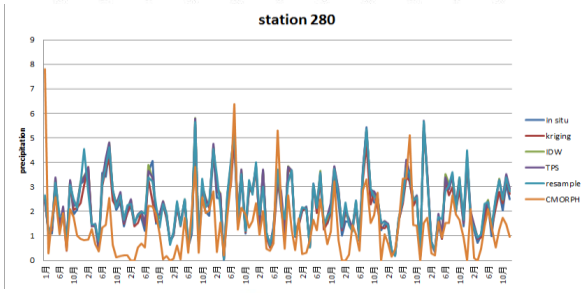

station 270
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

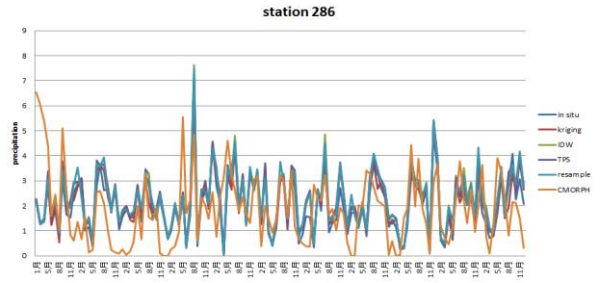

station 273
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

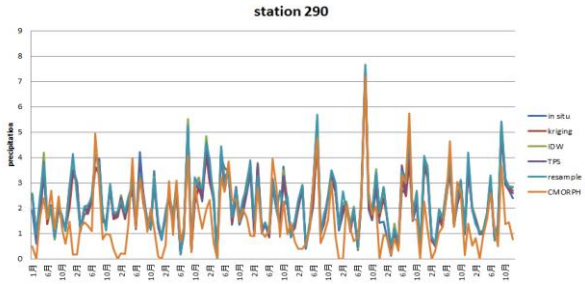

station 275
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

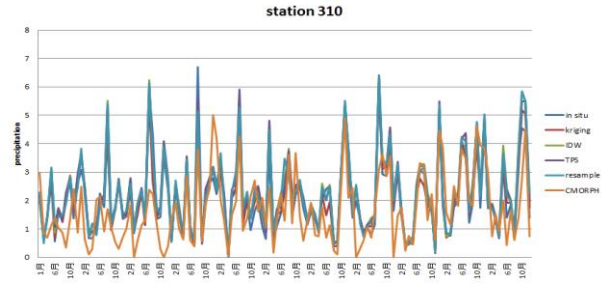

station 278
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

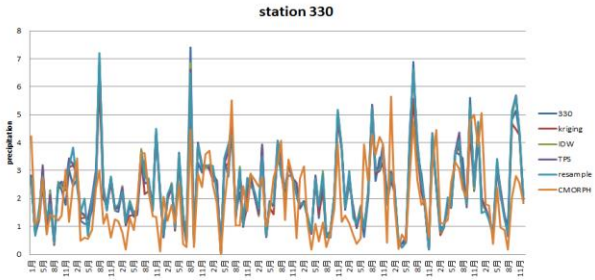

station 279
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

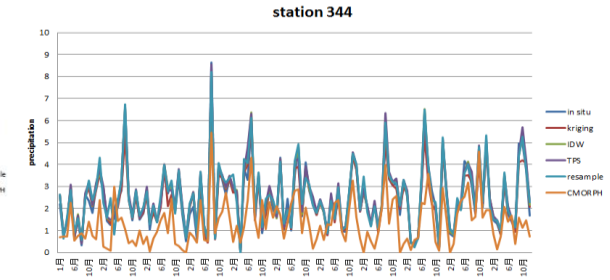

station 280
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH


station 286
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH


station 290
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH


station 310
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH


station 330
precipitation
330
kriging
IDW
TPS
resample
CMORPH


station 344
precipitation
in situ
kriging
IDW
TPS
resample
CMORPH

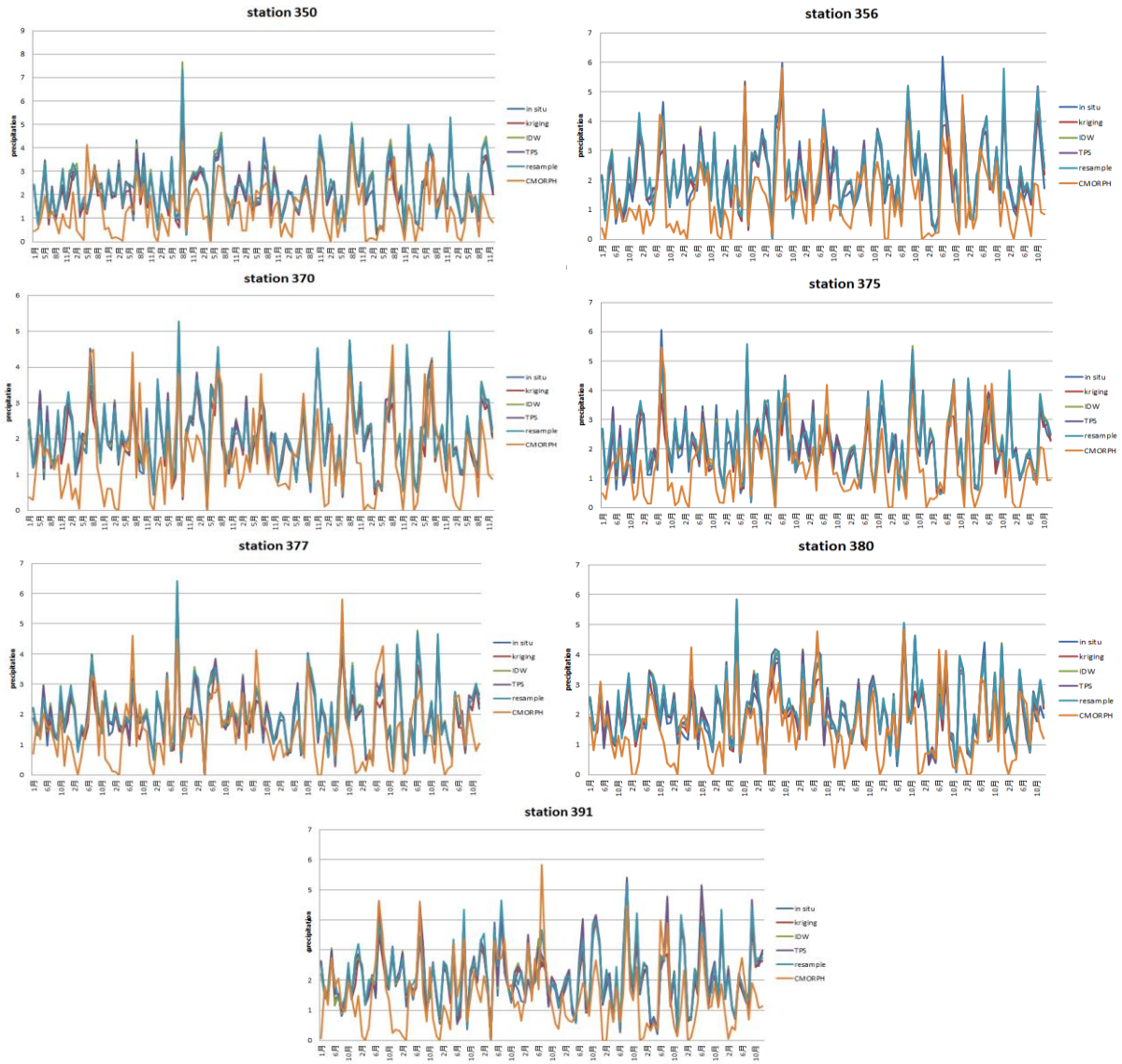
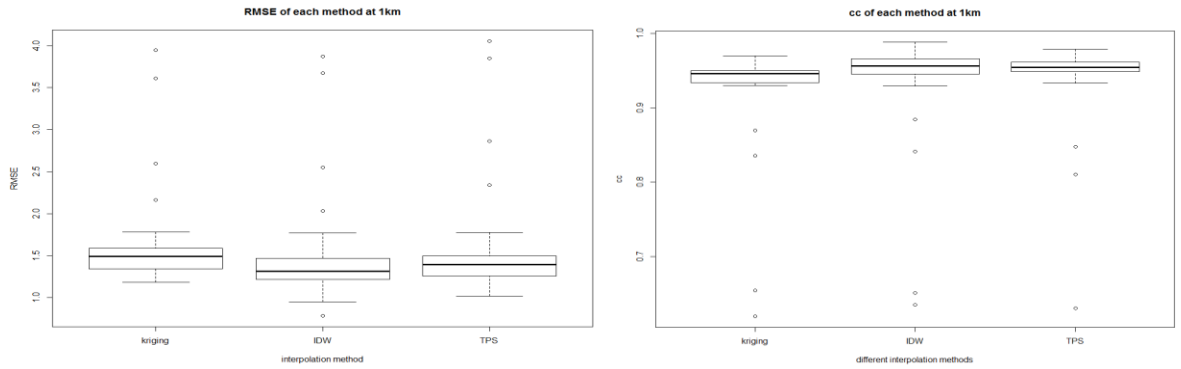


Figure 6-2b the curve of the forecast value in each station with different method

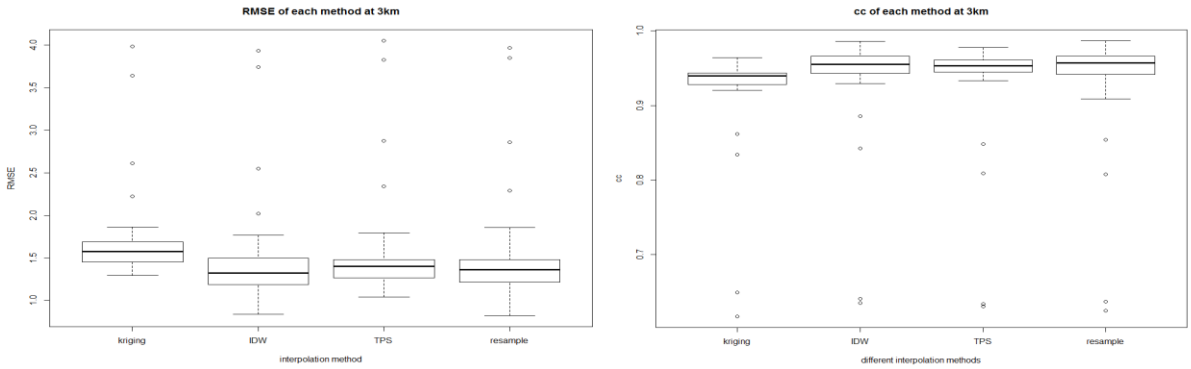
The graphs above (figure 6-2b) are precipitation data produced by each method on validation stations. These pictures show the trend of precipitation change in the past 11 years. For each year, June usually has more rainfall according to the plots. Moreover, interpolated data of three different interpolation methods and resampling data almost have the same variation trend as the in situ data. For the satellite data, although the value looks not accurate enough (e.g. in generally lower than observation), but the change in trend is regular and similar with in situ data.

6.3. Correlation coefficient and Root Mean Square Error

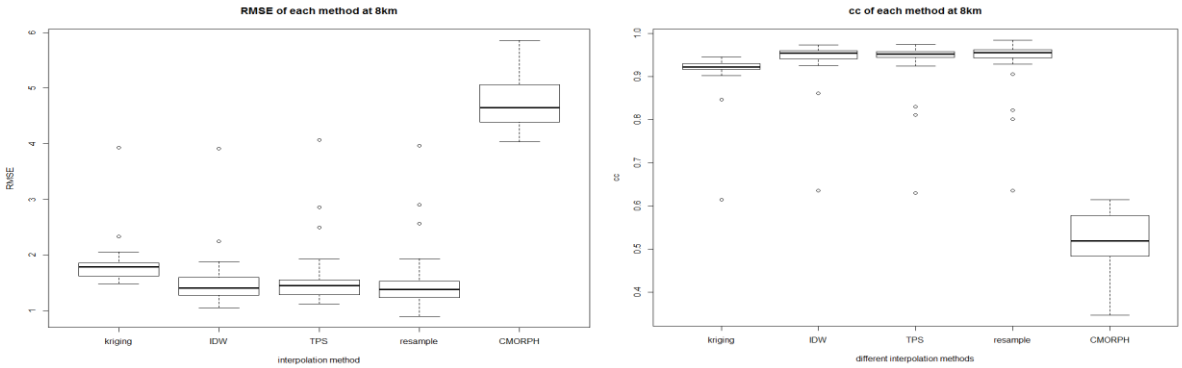
Correlation coefficient and Root Mean Square Error is calculated between the in situ data and interpolated and resampling data, at different spatial resolutions. The boxplots below are RMSE and CC of each method at different spatial resolutions. Summary of all the boxplot can be seen in appendix 1 and 2.



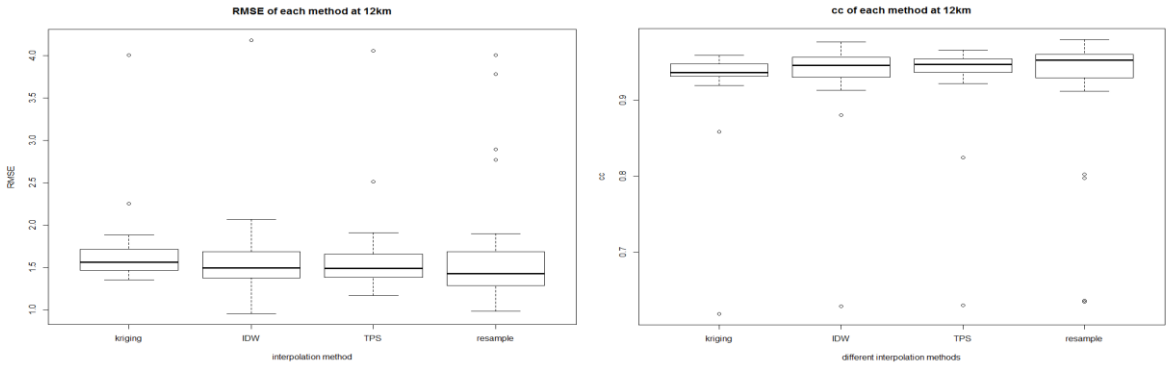
a. 1km



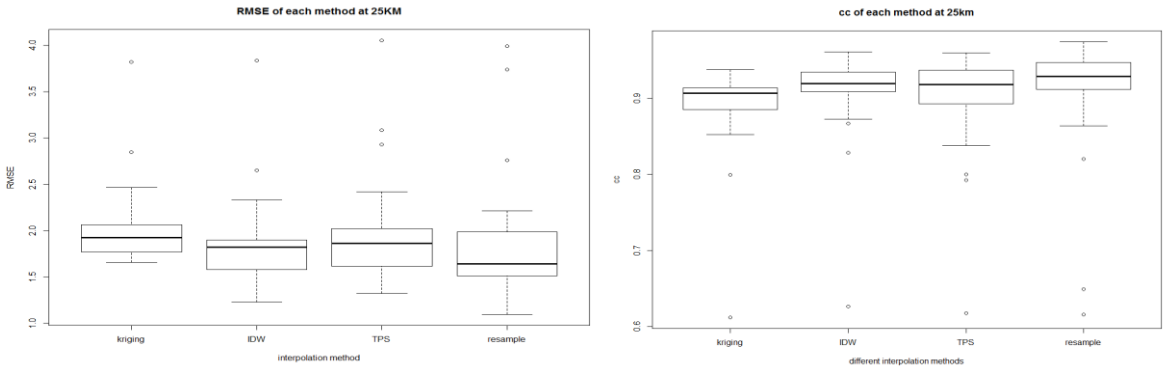
b. 3km



c. 8km



d. 12km



e. 25km

Figure 6-3 RMSE of each method at specific resolution

Comparing different methods at certain spatial resolution

Figure 6-3a shows RMSE and CC of ordinary kriging, IDW and TPS interpolation at 1km resolution. Form the figures we can see, RMSE of IDW is lowest and ordinary kriging is highest. RMSE of TPS is between them. So the sequence of RMSE is IDW<TPS<ordinary kriging. For correlation coefficient, ordinary kriging is lowest. CC of IDW and TPS are very close. Mean CC of IDW is 0.933 and TPS is 0.927. Therefore, sequences of correlation coefficient at 1km is IDW>TPS>ordinary kriging. Through comparison, IDW has lowest RMSE and highest CC which is better than ordinary and TPS at 1km resolution.

Figure 6-3b is boxplot comparing RMSE and CC for ordinary kriging, IDW, TPS and resample at 3km resolution. Form RMSE boxplot, ordinary kriging is higher than the other three methods; TPS is a little bigger than IDW and Resample; IDW and resample are very close. Mean RMSE of IDW is 1.52 and resample is 1.55. So for RMSE, IDW<resample<TPS<ordinary kriging. Compare CC of each method, IDW (0.9325) > resample (0.9298) >TPS (0.9267) >ordinary kriging (0.9149). Therefore, according to RMSE and CC of each method, IDW is more suitable at 3km with the lower RMSE and high CC. In addition, results of TPS and resample are similar, but interquartile range of resample in both RMSE and CC boxplot is larger than TPS. So TPS is stable than resample at this resolution.

RMSE and CC of ordinary kriging, IDW, TPS, resample and CMORPH at 8km are shown in figure 6-3d. It is clearly see that CMORPH data has very big RMSE and low CC. RMSE result from low to high is resample (1.51), IDW (1.53), TPS (1.60) and ordinary kriging (1.85). Result of CC is IDW (0.9364)>resample (0.9360)>TPS (0.9306)>ordinary kriging (0.9085). From comparison of RMSE and CC we can see, result of IDW, TPS and resample are very close.

Figure 6-3d shows the boxplot comparing RMSE and CC for each method at 12km. From the figure we can see, for RMSE, resample< IDW < TPS <ordinary kriging. Compare CC, resample> IDW> TPS> ordinary kriging. So resample is preferred at this resolution.

RMSE and CC of each method at 25km can be seen in figure 6-3e. CC of all the method decreased obviously, and RMSE is higher as well. RMSE of resample is lower than any other methods, and CC is highest. So resample is appropriate at this resolution. In addition, RMSE and CC of IDW and TPS are very similar.

Comparing each method under different spatial resolutions

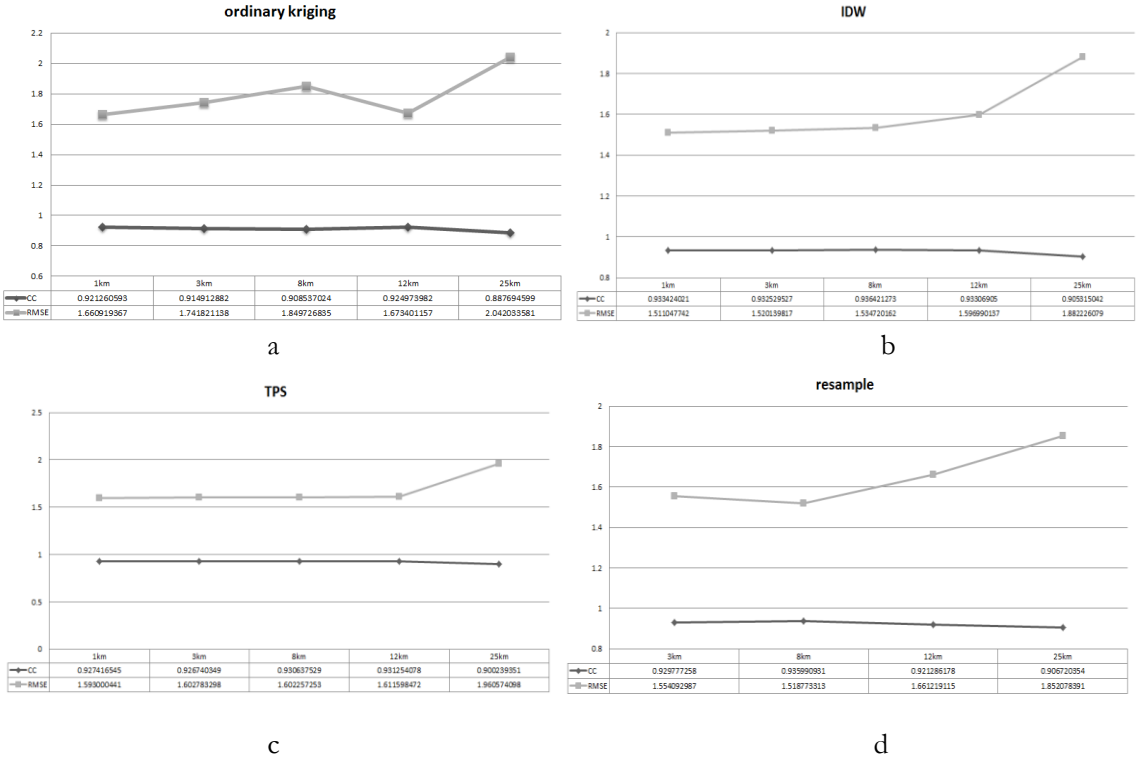


Figure 6-4 RMSE and CC of ordinary kriging, IDW, TPS and resample at different resolutions

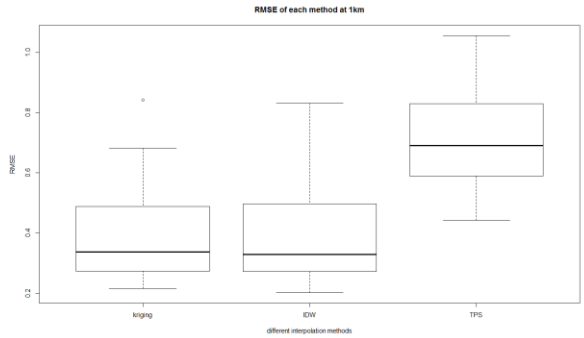
RMSE and CC of ordinary kriging for different resolutions can be seen in figure 6-4a. High correlation coefficient occurs at 1km and 12km that is around 0.92 and lowest is at 25km. RMSE at 1km and 12km are lower than the other three resolution. Therefore, ordinary kriging is suitable at 1km and 12km. From figure 6-4b we can see, CC of IDW interpolation is stable around 0.93 at 1km, 3km, 8km and 12km, RMSE is around 1.5. However, the RMSE obviously increased at 25km resolution. CC decrease at 25km as well. Figure 6-4c is RMSE and CC of TPS interpolation at different resolutions. RMSE and CC is stable at 1km, 3km 8km and 12km, CC is around 0.92 and RMSE is about 1.6. While at 25km, RMSE suddenly increased and CC decreased. From figure 6-4d we can see, resample is good at 8km. At this resolution, RMSE is lowest and CC is highest. Range of RMSE is from 1.5 to 2 that are bigger than the other three interpolation methods. Highest CC of resampling is 0.935 and lowest is about 0.9.

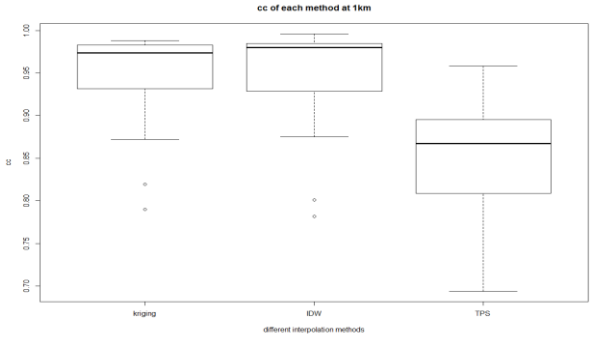
6.4. Interpolation on monthly rainfall data

6.4.1. Interpolation result

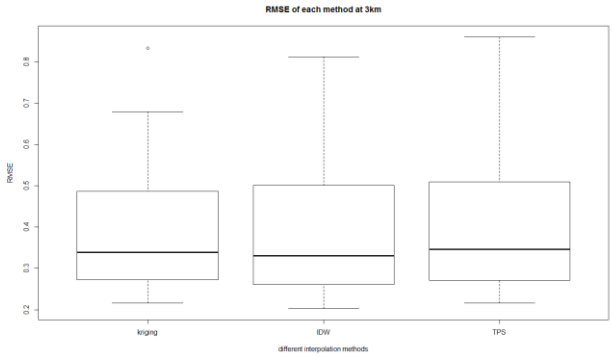
In order to know whether the interpolation method used for daily is still suitable for monthly interpolation. We also use ordinary kriging, IDW, TPS interpolation method on monthly data at different spatial resolutions. The monthly data is the mean of every month from 2003 to 2013. The methodology used is the same as before.

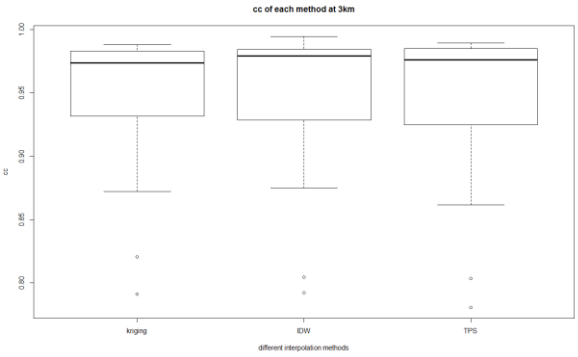
The following plots are RMSE and CC of each method interpolated on monthly data at different spatial resolutions. Summary of boxplot can be found in appendix 3 and 4.



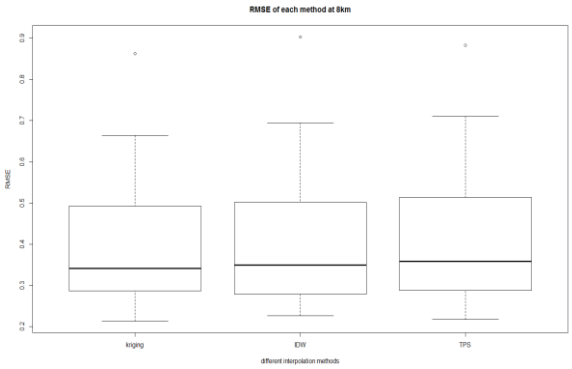


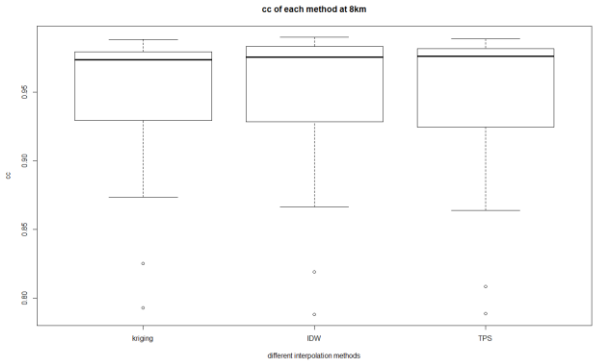
1km



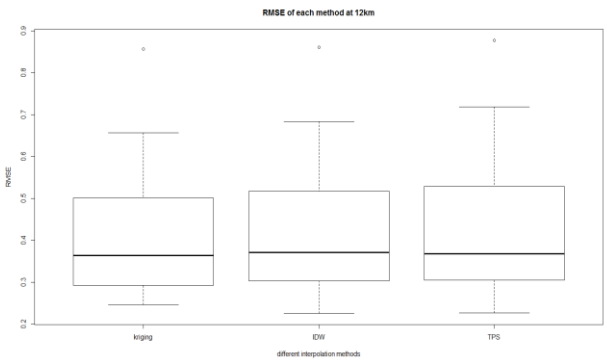


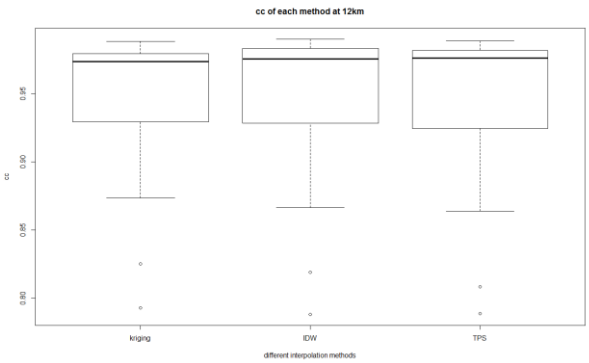
3km



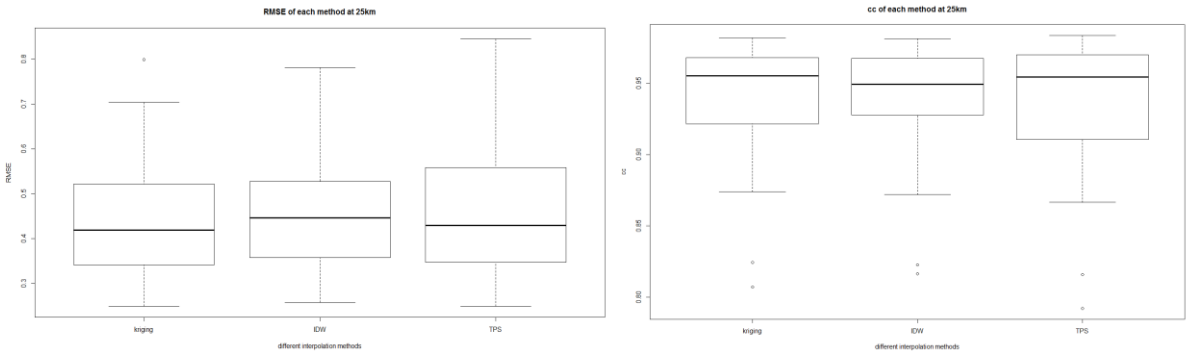


8km





12km



25km

Figure 6-5 RMSE and CC of each method on monthly interpolation at different spatial resolutions

From figure 6-8 we can see, at 1km resolution, RMSE of these three methods are IDW< ordinary kriging < TPS. CC of the three method are IDW> ordinary kriging > TPS. So IDW is more suitable at this resolution. At 3km resolution, RMSE and CC of each method are very similar. Result of IDW is a little bit better than ordinary and kriging. At 8km and 12km, CC of ordinary kriging and IDW are almost the same. However, RMSE of ordinary kriging is lower than IDW. In addition, considered the image of ordinary kriging is often smoother than IDW and TPS. Therefore, ordinary kriging interpolation is preferred at these two resolutions. At 25km resolution, ordinary kriging is appreciated with lower RMSE and higher CC.

6.4.2. Comparison between monthly interpolation and daily interpolation

Overall comparing daily interpolation and monthly interpolation we found, ordinary kriging shows better result than IDW and TPS when temporal scale is monthly. The difference may be caused by the spatial pattern and rainfall type that vary dramatically on daily rainfall. And for ordinary kriging, variogram will be used for all moment data, so it is difficult for one variogram to fit all the situations. In addition, we found that ordinary kriging usually shows better interpolation on stratiform rainfall than convective rainfall. However, in this thesis, daily interpolation starts from 2003 to 2013 that contains 4018 days, which means there are a lot of convective rainfall during this time range. So accuracy of ordinary kriging may decrease. Moreover, the spatial pattern of monthly rainfall is like stratiform rainfall. So it is easier to model one semi variogram that could fit majority rainfall patterns.

6.5. Compare with satellite data

With advances in technology, satellite data are widely used in all kinds of discipline. For example, study climate change and applications on various hydrological models. So the accuracy of satellite data is important to know. However, the spatial resolution for each satellite is different which is to say reference data at different spatial resolution are need due to these datasets can be used to analyse the uncertainty of satellite data.

From all the graphs in section 6-3 we can see, the satellite data usually have the similar trend with in situ data, also the interpolated data. Actually, in situ data are often point data that may not enough to correct satellite data. Moreover, in situ data are uneven distributed, so sometimes one station cannot represent one area. At this time, interpolated data will have more advantages. It is grid data that covered more area and is very close to actual situations as well. To be reference data for scaling merging can effectively bridge the gap between the satellite data and the reality.

In addition, corrected satellite data may be used to extract information on other places. For instance, in the thesis the validation data are observations of 32 stations, compare the result we can see, the accuracy of station 265 is not ideal no matter for correlation coefficient or Root Mean Square Error. When the

study area is around this station, using interpolated data to bias correct satellite data will improve its match with in-situ observations.

6.6. Discussion

Comparing different interpolations result at different spatial resolution, IDW outperforms ordinary kriging, TPS at all spatial resolutions on daily rainfall interpolation. Except for 12km and 25km resolutions, resample is better than IDW, while IDW shows better result than resample at other resolutions. The correlation coefficient of IDW is higher and Root Mean Square Error is lower.

The overall spatial pattern for ordinary kriging, IDW, TPS and resample matches each other very well at different spatial resolutions. The drawback of interpolation is the loss of interpolated values at some stations when the spatial resolution is increased. In addition, the bulls-eye pattern in IDW is obvious and is not as smooth as ordinary kriging.

If we check the predicted range of rainfall for a certain specific method, the range decreased with the increase of spatial resolution. Usually, the range will be narrowing than actual situation (Costa, Oliveira, Dias, & Painho, 2009). In this respect, resample is a good way to keep the scope. In theory, resample just changes the cell size and the extent of raster data will not change.

As a matter of fact, when the rain gauge data is enough, the result of different interpolation is similar. In the Netherlands, there are more than 300 stations can be used, so the correlation coefficient is close to 0.95 which is a good result. Compared with ordinary kriging and IDW, the accuracy of TPS is relatively low.

In section 5, rainfall is simply divided into two stratiform rainfall and convective rainfall. They have different spatial distributions and characteristics. During the research we found that ordinary kriging shows better result in stratiform rainfall and is not good at interpolating convective rainfall. However, there are many days are convective rainfall during 11 years. It is difficult to model a semi-variogram to adjust all kinds of situations. So this may be the main reason that the ordinary kriging results are different at daily and monthly scale. The spatial pattern of monthly data is like that of stratiform rainfall. So when interpolating monthly data, ordinary kriging is preferred. In addition, IDW method is simpler to programme and provides a measure of uncertainty of the estimates which is related to the estimated values directly (Tomczak, 1998). That is to say IDW method is appropriate to the small area where the semi-variogram is difficult to fit, such as high density rainfall in local area. So when interpolate long-term daily rainfall data which means the interpolation contains a lot of convective rainfall, IDW interpolation as a relatively simple and fast method is preferred, and the result also has high quality.

7. CONCLUSIONS

Based on this research we can conclude the followings:
1. For long-term daily rainfall interpolation, via visual interpretation, correlation coefficient and Root Mean Square Error, IDW interpolation method outperforms ordinary kriging, TPS and resample at 1km, 3km, 8km. Resample shows better result at 12km and 25km.
2. The overall predicted spatial patterns of ordinary kriging, IDW, TPS and resample are quite similar. But ordinary kriging image is smoother than IDW, TPS and resample.
3. When the rain gauge data is enough, there is little different in ordinary kriging, IDW, TPS and resample. The correlation coefficient of each method is quite good which is close to 0.92. Root Mean Square Error is low.
4. The overall accuracy of resample is acceptable in all spatial resolutions, especially at coarse resolution like 12km and 25km.
5. For monthly rainfall data, ordinary kriging is preferred. Since the spatial distribution of monthly rainfall is stable, it is easy to model a semi variogram that could match most rainfall spatial patterns.

The research could be useful on interpolating long-term daily rainfall data in the Netherlands. However, there are some limitations in this research. Followings are some aspects might improve the interpolation result:
1. Ordinary kriging interpolation result has negative values that may due to negative kriging weights to extreme values (S. Ly et al., 2011). This often happened in convective rainfall. In this research the negative value is simply replaced by zero. So an ideal solution to solve this problem is needed.
2. Usually, a cross validation technique is used for evaluating monthly or annual time steps. However, it would be time consuming to use it for daily data for 11 years with different interpolation methods at different spatial resolutions. Therefore, adding cross validation to validation could have a multiple evaluations on interpolation method.
3. Interpolation usually overestimates small values and underestimates large one. Therefore, some geostatistical stochastic simulation that can obtain equiprobable variables from natural phenomena and also be consistent with the statistical characteristics of rainfall data can overcome this limitation (Costa et al., 2009).
4. Analysing the influence of rain gauge density on different interpolation methods is needed. In the Netherlands, precipitation stations start at different times, for instance, some start from 1951 and some start from 1990. So understanding such impact of interpolation is helpful if we want to do interpolation for 30 years or longer.
5. A suitable transformation for each individual day instead of log transformation for all the dataset can make rainfall more normally distributed.

LIST OF REFERENCES

Akkala, A. (2010). *Development of artificial neural networks based interpolation techniques for the modeling and estimation of radon concentrations in Ohio*. Retrieved from http://utdr.utoledo.edu/cgi/viewcontent.cgi?article=1795&context=theses-dissertations

Akkala, A., Devabhaktuni, V., & Kumar, A. (2010). Interpolation techniques and associated software for environmental data. *Environmental Progress and Sustainable Energy*. doi:10.1002/ep.10455

Alemohammad, S. H., & Entekhabi, D. (2013). Using Multi-scale Imagery Technique. Retrieved from http://arxiv.org/pdf/1304.3406.pdf

ArcGIS Help 10.1 - Resample (Data Management). (n.d.). Retrieved January 23, 2015, from http://resources.arcgis.com/en/help/main/10.1/index.html#//00170000009t000000

Attema, J., & Lenderink, G. (2011). Mean precipitation changes and coastal effects in the ENSEMBLES regional climate model simulations. *KNMI Scientific Report 2011-03*, 42 pp.

Berndt, C., Rabiei, E., & Haberlandt, U. (2014). Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. *Journal of Hydrology*, *508*, 88–101. doi:10.1016/j.jhydrol.2013.10.028

Bivand, R., Rowlingson, B., & Gomez-rubio, V. (2014). Package " sp ." Retrieved from https://r-forge.r-project.org/projects/rspatial/

Camarasa-Belmonte, a. M., & Soriano, J. (2014). Empirical study of extreme rainfall intensity in a semi-arid environment at different time scales. *Journal of Arid Environments*, *100-101*, 63–71. doi:10.1016/j.jaridenv.2013.10.008

Christensen, O. B., Christensen, J. H., & Guldberg, A. (1990). *Effects of resolution in an RCM : From 50 to 12 km*. Retrieved from http://www.wcrp-climate.org/WGNE/BlueBook/2006/individual-articles/07_Christensen_Ole_B_ssing_hirham12km.pdf

Costa, A. C., Oliveira, T., Dias, S., & Painho, M. (2009). Spatial Patterns of Daily Precipitation Intensity in Southern, 2007–2010. Retrieved from http://www.agile-online.org/Conference_Paper/CDs/agile_2009/AGILE_CD/pdfs/95.pdf

De Marchi, C. (2006). Probabilistic Estimation of Precipitation Combining Geostationary and TRMM Satellite Data. Retrieved from https://smartech.gatech.edu/handle/1853/14068

Deutsch, C. V. (1996). Correcting for negative weights in ordinary kriging. *Computers and Geosciences*, *22*(7), 765–773. doi:10.1016/0098-3004(96)00005-2

Douglas, A., Furrer, R., Sain, S., & Nychka, M. D. (2014). Package "fields." Retrieved from http://www.image.ucar.edu/Software/Fields

Eda, N. B. (2013). *An introduction to geostatistics with R / gstat*. Retrieved from http://www.css.cornell.edu/faculty/dgr2/teach/R/gs_short_ex.pdf

Flato, G. M., Boer, G. J., Lee, W. G., McFarlane, N. a., Ramsden, D., Reader, M. C., & Weaver, a. J. (2000). The Canadian Centre for Climate Modelling and Analysis global coupled model and its climate. *Climate Dynamics*, *16*, 451–467. doi:10.1007/s003820050339

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, *228*, 113–129. doi:10.1016/S0022-1694(00)00144-X

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres*, *113*(November). doi:10.1029/2008JD010100

Interpolation - Wikipedia, the free encyclopedia. (n.d.). Retrieved January 21, 2015, from http://en.wikipedia.org/wiki/Interpolation

Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., & Ma, M. (2012). Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *Journal of Hydrology*, *452-453*, 213–225. doi:10.1016/j.jhydrol.2012.05.055

Johnson, G. L., & Hanson, C. L. (1995). Topographic and Atmospheric Influences on Precipitation Variability over a Mountainous Watershed. *Journal of Applied Meteorology*. doi:10.1175/1520-0450-34.1.68

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*. doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Kao, C. Y. S., & Hung, F. L. P. (2004). Twelve Different Interpolation Methods : a Case Study. *Proceedings of the XXth ISPRS Congress*, *35*, 778–785. Retrieved from http://www.isprs.org/proceedings/XXXV/congress/comm2/papers/231.pdf

Karmakar, S. (2009). Spatial Interpolation of Rainfall Variables using Artificial Neural Network. *Test*, 547–552. Retrieved from http://delivery.acm.org/10.1145/1530000/1523212/p547-karmakar.pdf?ip=130.89.227.140&id=1523212&acc=ACTIVE SERVICE&key=0C390721DC3021FF.7DEDEACE9AC2380A.4D4702B0C3E38B35.4D4702B0C3E38B35&CFID=625549061&CFTOKEN=92181398&__acm__=1423582933_3d17e0982e046ae8ac

KNMI Data Centre. (n.d.). Retrieved January 23, 2015, from https://data.knmi.nl/portal/KNMI-DataCentre.html#

Lawford, R. (2014). *the Geoss Water Strategy*. Retrieved from ftp://ftp.earthobservations.org/TEMP/Water/GEOSS_WSR_Full_Report.pdf

Lloyd, C. D. (2005). Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology*, *308*, 128–150. doi:10.1016/j.jhydrol.2004.10.026

Ly, S., Charles, C., & Degré, a. (2011). Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, *15*(7), 2259–2274. doi:10.5194/hess-15-2259-2011

Ly, S., Charles, C., & Degré, A. (2013). Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. *Biotechnol. Agron. Soc. Environ.*, *17*(2), 392–406.

Meirink, J. F., & Plieger, M. (2012). *The MSG-CPP Project*. Retrieved from http://msgcpp.knmi.nl/documents/MSG-CPP_2012EUMETSATSatConf_20120907.pdf

Reyes, A., Giraldo, R., Mateu, J., Reyes, A., & Mateu, J. (2012). Residual kriging for functional data . Application to the spatial prediction of salinity curves. *CrossMark*, (20). Retrieved from http://www.estadistica.unal.edu.co/unciencias/data-file/user_45/RI20.pdf

Roger, A., Archer, E., Baddeley, A., Brey, S., Callahan, J., Forrest, D., … Rogerbivandnhhno, M. R. B. (2014). Package " maptools ." Retrieved from http://r-forge.r-project.org/projects/maptools/

Sapiano, M. R. P., & Arkin, P. a. (2009). An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data. *Journal of Hydrometeorology*, *10*(1), 149–166. doi:10.1175/2008JHM1052.1

Schuurmans, J. M., & Bierkens, M. F. P. (2006). Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. *Hydrology and Earth System Sciences Discussions*. doi:10.5194/hessd-3-2175-2006

Schuurmans, J. M., Bierkens, M. F. P., Pebesma, E. J., & Uijlenhoet, R. (2007). Automatic Prediction of High-Resolution Daily Rainfall Fields for Multiple Extents: The Potential of Operational Radar. *Journal of Hydrometeorology*, *8*, 1204–1224. doi:10.1175/2007JHM792.1

Sluiter, R. (2009). *Interpolation methods for climate data: literature review. KNMI, R&D Information and Observation Technology,* … (pp. 1–28). Retrieved from https://www.snap.uaf.edu/sites/default/files/files/Interpolation_methods_for_climate_data.pdf

Sluiter, R. (2012). Interpolation Methods for the Climate Atlas, (July). Retrieved from http://www.knmi.nl/knmi-library/knmipubTR/TR335.pdf

Soenario, I., Plieger, M., & Sluiter, R. (2010). *Optimization of Rainfall Interpolation* (p. 31). Retrieved from http://www.knmi.nl/bibliotheek/knmipubIR/IR2010-01.pdf

Sterling, D. L. (2003). A Comparison of Spatial Interpolation Techniques For Determining Shoaling Rates of The Atlantic Ocean Channel Approved : A Comparison of Spatial Interpolation Techniques For Determining Shoaling Rates of The Atlantic Ocean Channel. *Virginia Polytechnic Institute and State University*, *C*, 1–94.

Suggests, F. N. N., Gpl, L., Url, L., Pebesma, A. E., Graeler, B., Pebesma, M. E., & Date, R. C. (2014). Package " gstat ." Retrieved from https://r-forge.r-project.org/projects/gstat/

Sunyer, M. a., Madsen, H., & Ang, P. H. (2012). A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change. *Atmospheric Research*, *103*, 119–128. doi:10.1016/j.atmosres.2011.06.011

Taesombat, W., & Sriwongsitanon, N. (2009). Areal rainfall estimation using spatial interpolation techniques. *ScienceAsia*, *35*, 268–275. doi:10.2306/scienceasia1513-1874.2009.35.268

Tait, A., Henderson, R., Turner, R., & Zheng, X. (2006). Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *International Journal of Climatology*, *26*(14), 2097–2115. doi:10.1002/joc.1350

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. a. T., … de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104-105*, 70–97. doi:10.1016/j.atmosres.2011.10.021

Todini, E., & Ferraresi, M. (1996). Influence of parameter estimation uncertainty in Kriging. *Journal of Hydrology*, *175*, 555–566. doi:10.1016/S0022-1694(96)80024-2

Tomczak, M. (1998). Spatial Interpolation and its Uncertainity Using Automated Anisotrophic Inverse Distance Weighting (IDW) - Cross - Validation /Jackknife Approach. *Journal of Geographic Information and DEcision Analysis*, *2*(2), 18–30. Retrieved from http://www.ai-geostats.org/pub/AI_GEOSTATS/Papers20100623104140/Tomczak.pdf

Wey, K. (2006). *Temporal disaggregation of daily precipitation data in a changing climate*. university of Waterloo. Retrieved from http://hdl.handle.net/10012/2859

Willems, P., Arnbjerg-Nielsen, K., Olsson, J., & Nguyen, V. T. V. (2012). Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings. *Atmospheric Research*, *103*, 106–118. doi:10.1016/j.atmosres.2011.04.003

Willems, P., & Vrac, M. (2011). Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology*, *402*(3-4), 193–205. doi:10.1016/j.jhydrol.2011.02.030

APPENDIX

1. Summary Root Mean Square Error at each spatial resolution on daily rainfall interpolation

Resolution	kriging	IDW	TPS	Resample
1km	Min. :1.183	Min. :0.7885	Min. :1.018	
	1st Qu.:1.346	1st Qu.:1.2236	1st Qu.:1.264	
	Median :1.495	Median :1.3171	Median :1.394	
	Mean :1.661	Mean :1.5110	Mean :1.593	
	3rd Qu.:1.569	3rd Qu.:1.4670	3rd Qu.:1.488	
	Max. :3.946	Max. :3.8744	Max. :4.054	
3km				
	Min. :1.293	Min. :0.8384	Min. :1.044	Min. :0.8241
	1st Qu.:1.456	1st Qu.:1.2013	1st Qu.:1.266	1st Qu.:1.2157
	Median :1.574	Median :1.3226	Median :1.401	Median :1.3652
	Mean :1.742	Mean :1.5201	Mean :1.603	Mean :1.5541
	3rd Qu.:1.674	3rd Qu.:1.4982	3rd Qu.:1.480	3rd Qu.:1.4766
	Max. :3.983	Max. :3.9353	Max. :4.053	Max. :3.9666
8km				
	Min. :1.482	Min. :1.055	Min. :1.120	Min. :0.8982
	1st Qu.:1.625	1st Qu.:1.276	1st Qu.:1.285	1st Qu.:1.2365
	Median :1.783	Median :1.403	Median :1.452	Median :1.3838
	Mean :1.850	Mean :1.535	Mean :1.602	Mean :1.5188
	3rd Qu.:1.856	3rd Qu.:1.594	3rd Qu.:1.555	3rd Qu.:1.5301
	Max. :3.928	Max. :3.911	Max. :4.067	Max. :3.9627
12km				
	Min. :1.350	Min. :0.9543	Min. :1.167	Min. :0.9778
	1st Qu.:1.467	1st Qu.:1.3783	1st Qu.:1.381	1st Qu.:1.2893
	Median :1.558	Median :1.4927	Median :1.485	Median :1.4255
	Mean :1.673	Mean :1.5970	Mean :1.612	Mean :1.6612
	3rd Qu.:1.706	3rd Qu.:1.6603	3rd Qu.:1.651	3rd Qu.:1.6566
	Max. :4.000	Max. :4.1761	Max. :4.052	Max. :4.0050
25km				
	Min. :1.658	Min. :1.233	Min. :1.324	Min. :1.100
	1st Qu.:1.770	1st Qu.:1.580	1st Qu.:1.630	1st Qu.:1.515
	Median :1.929	Median :1.825	Median :1.863	Median :1.645
	Mean :2.042	Mean :1.882	Mean :1.961	Mean :1.852
	3rd Qu.:2.068	3rd Qu.:1.899	3rd Qu.:2.023	3rd Qu.:1.981
	Max. :3.820	Max. :3.836	Max. :4.050	Max. :3.990

2. Summary correlation coefficient at each spatial resolution on daily rainfall interpolation

Resolution	kriging	IDW	TPS	resample
1km	Min. :0.6198	Min. :0.6351	Min. :0.6306	
	1st Qu.:0.9351	1st Qu.:0.9455	1st Qu.:0.9486	
	Median :0.9462	Median :0.9564	Median :0.9546	
	Mean :0.9213	Mean :0.9334	Mean :0.9274	
	3rd Qu.:0.9498	3rd Qu.:0.9650	3rd Qu.:0.9614	
	Max. :0.9699	Max. :0.9885	Max. :0.9788	
3km				
	Min. :0.6173	Min. :0.6352	Min. :0.6308	Min. :0.6249
	1st Qu.:0.9288	1st Qu.:0.9443	1st Qu.:0.9449	1st Qu.:0.9419
	Median :0.9397	Median :0.9552	Median :0.9534	Median :0.9571
	Mean :0.9149	Mean :0.9325	Mean :0.9267	Mean :0.9298
	3rd Qu.:0.9430	3rd Qu.:0.9654	3rd Qu.:0.9609	3rd Qu.:0.9653
	Max. :0.9640	Max. :0.9856	Max. :0.9774	Max. :0.9865
8km				
	Min. :0.6135	Min. :0.6346	Min. :0.6300	Min. :0.6354
	1st Qu.:0.9161	1st Qu.:0.9406	1st Qu.:0.9438	1st Qu.:0.9429
	Median :0.9222	Median :0.9542	Median :0.9518	Median :0.9553
	Mean :0.9085	Mean :0.9364	Mean :0.9306	Mean :0.9360
	3rd Qu.:0.9302	3rd Qu.:0.9592	3rd Qu.:0.9576	3rd Qu.:0.9621
	Max. :0.9451	Max. :0.9730	Max. :0.9738	Max. :0.9840
12km				
	Min. :0.6189	Min. :0.6287	Min. :0.6300	Min. :0.6343
	1st Qu.:0.9314	1st Qu.:0.9307	1st Qu.:0.9365	1st Qu.:0.9301
	Median :0.9366	Median :0.9462	Median :0.9472	Median :0.9526
	Mean :0.9250	Mean :0.9331	Mean :0.9313	Mean :0.9213
	3rd Qu.:0.9473	3rd Qu.:0.9561	3rd Qu.:0.9545	3rd Qu.:0.9600
	Max. :0.9591	Max. :0.9769	Max. :0.9662	Max. :0.9799
25km				
	Min. :0.6124	Min. :0.6265	Min. :0.6180	Min. :0.6158
	1st Qu.:0.8856	1st Qu.:0.9087	1st Qu.:0.8930	1st Qu.:0.9125
	Median :0.9070	Median :0.9193	Median :0.9183	Median :0.9286
	Mean :0.8877	Mean :0.9053	Mean :0.9002	Mean :0.9067
	3rd Qu.:0.9140	3rd Qu.:0.9343	3rd Qu.:0.9336	3rd Qu.:0.9463
	Max. :0.9377	Max. :0.9613	Max. :0.9595	Max. :0.9742

3. Summary of RMSE at each spatial resolution on monthly rainfall interpolation

	kriging	IDW	TPS
1km	Min. :0.2160	Min. :0.2020	Min. :0.4420
	1st Qu.:0.2763	1st Qu.:0.2727	1st Qu.:0.6028
	Median :0.3379	Median :0.3295	Median :0.6900
	Mean :0.3827	Mean :0.3830	Mean :0.7138
	3rd Qu.:0.4863	3rd Qu.:0.4916	3rd Qu.:0.8247
	Max. :0.8415	Max. :0.8302	Max. :1.0554
3km			
	Min. :0.2160	Min. :0.2024	Min. :0.2164
	1st Qu.:0.2748	1st Qu.:0.2616	1st Qu.:0.2713
	Median :0.3394	Median :0.3302	Median :0.3462
	Mean :0.3823	Mean :0.3811	Mean :0.3914
	3rd Qu.:0.4850	3rd Qu.:0.5008	3rd Qu.:0.5079
	Max. :0.8334	Max. :0.8123	Max. :0.8606
8km			
	Min. :0.2134	Min. :0.2268	Min. :0.2176
	1st Qu.:0.2894	1st Qu.:0.2840	1st Qu.:0.2891
	Median :0.3411	Median :0.3491	Median :0.3587
	Mean :0.3952	Mean :0.4033	Mean :0.4049
	3rd Qu.:0.4903	3rd Qu.:0.5006	3rd Qu.:0.5134
	Max. :0.8626	Max. :0.9032	Max. :0.8826
12km			
	Min. :0.2461	Min. :0.2255	Min. :0.2270
	1st Qu.:0.2935	1st Qu.:0.3074	1st Qu.:0.3058
	Median :0.3636	Median :0.3711	Median :0.3679
	Mean :0.4044	Mean :0.4146	Mean :0.4187
	3rd Qu.:0.5006	3rd Qu.:0.5165	3rd Qu.:0.5288
	Max. :0.8573	Max. :0.8611	Max. :0.8783
25km			
	Min. :0.2487	Min. :0.2572	Min. :0.2492
	1st Qu.:0.3489	1st Qu.:0.3605	1st Qu.:0.3579
	Median :0.4192	Median :0.4463	Median :0.4290
	Mean :0.4478	Mean :0.4574	Mean :0.4688
	3rd Qu.:0.5209	3rd Qu.:0.5246	3rd Qu.:0.5513
	Max. :0.7993	Max. :0.7812	Max. :0.8456

4. Summary of correlation coefficient of each method on monthly rainfall interpolation

	kriging	IDW	TPS
1km	Min. :0.7899	Min. :0.7817	Min. :0.6939
	1st Qu.:0.9322	1st Qu.:0.9308	1st Qu.:0.8120
	Median :0.9736	Median :0.9800	Median :0.8675
	Mean :0.9523	Mean :0.9539	Mean :0.8533
	3rd Qu.:0.9829	3rd Qu.:0.9841	3rd Qu.:0.8922
	Max. :0.9881	Max. :0.9958	Max. :0.9580
3km			
	Min. :0.7911	Min. :0.7920	Min. :0.7806
	1st Qu.:0.9322	1st Qu.:0.9308	1st Qu.:0.9259
	Median :0.9739	Median :0.9793	Median :0.9763
	Mean :0.9524	Mean :0.9544	Mean :0.9511
	3rd Qu.:0.9831	3rd Qu.:0.9840	3rd Qu.:0.9847
	Max. :0.9881	Max. :0.9946	Max. :0.9896
8km			
	Min. :0.7928	Min. :0.7881	Min. :0.7887
	1st Qu.:0.9303	1st Qu.:0.9301	1st Qu.:0.9255
	Median :0.9737	Median :0.9756	Median :0.9761
	Mean :0.9504	Mean :0.9517	Mean :0.9493
	3rd Qu.:0.9792	3rd Qu.:0.9831	3rd Qu.:0.9812
	Max. :0.9886	Max. :0.9900	Max. :0.9887
12km			
	Min. :0.7928	Min. :0.7881	Min. :0.7887
	1st Qu.:0.9303	1st Qu.:0.9301	1st Qu.:0.9255
	Median :0.9737	Median :0.9756	Median :0.9761
	Mean :0.9504	Mean :0.9517	Mean :0.9493
	3rd Qu.:0.9792	3rd Qu.:0.9831	3rd Qu.:0.9812
	Max. :0.9886	Max. :0.9900	Max. :0.9887
25km			
	Min. :0.8071	Min. :0.8163	Min. :0.7922
	1st Qu.:0.9213	1st Qu.:0.9284	1st Qu.:0.9129
	Median :0.9549	Median :0.9492	Median :0.9542
	Mean :0.9395	Mean :0.9387	Mean :0.9359
	3rd Qu.:0.9666	3rd Qu.:0.9663	3rd Qu.:0.9674
	Max. :0.9816	Max. :0.9810	Max. :0.9832