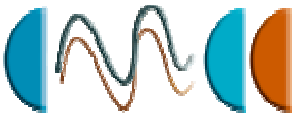
Report



CENTRO EURO-MEDITERRANEO PER I CAMBIAMENTI CLIMATICI

ISC - Impacts on ground and coast

Research activity on statistical downscaling for precipitation.

Tiberiu-Eugen Antofie

Centro Euro-Mediterraneo

per i Cambiamenti Climatici

www.cmcc.it

February 2009

Index

Abstract

This study is a result of the activity carried out at **ISC CMCC Division CIRA.** The activity was dedicated on understanding the different modalities used to manage the discrepancy between the coarse scale at which the COSMO LM[1] delivers output and the scale that is required for most impact studies.

This study reviews 8 methods of interpolation to be used on digital model output data from the COSMO LM. The model output data are predicted precipitation on a regular grid shaped surface with 305 point location and a resolution of 2,8 Km. Most of the geographical spatial analysis require a continuous data set and this study is designed to create that surface. This study identifies the best spatial interpolation method to use for the creation of continuous data for predicted precipitation. ArcGIS was employed as the software for this study. The following interpolated methods were developed in ArcGis: Inverse Distance Weight, Radial Basis Function (RBF), Kriging (Ordinary, Simple, Universal and Disjunctive), Local Polynomial interpolation and Global Polynomial Interpolation. A statistical measurement of the resultant continuous surfaces indicates that there is little difference between the estimating ability of the 8 interpolation methods with RBFs performing better overall.

1 Introduction

Precipitation is one of the most frequently used meteorological parameter in impact studies. The spatial variability of the precipitation depends not only on the nonperiodic or periodic behaviour of the general atmosphere but also on the sub-grid scale atmospheric processes such as cloud formation, turbulences, convection, evaporation etc. (Lorenz, E.N., 1966: ''Nonlinearity, Weather Prediction and climate deduction'' Final Report, Statistical forecasting project, 22 pg);

[1] The Lokal-Modell (LM) is a nonhydrostatic limited-area atmospheric prediction model. It has been designed for both operational numerical weather prediction (NWP) and various scientific applications on the meso- and meso- scale. The LM is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere. The model equations are formulated in rotated geographical coordinates and a generalized terrain following height coordinate. A variety of physical processes are taken into account by parameterization schemes. The basic version of LM has been developed at the Deutscher Wetterdienst (DWD). The subsequent developments related to LM have been organized within COSMO, the Consortium for Small-Scale Modelling. COSMO aims at the improvement, maintenance and operational application of a non-hydrostatic limited-area modelling system based on the LM. At present, different meteorological services partecipate to COSMO. (For other info see www.cosmo-model.org)

When these physical processes that are governing the atmosphere are integrated into digital atmosphere models, thank to parameterisation processing (Zorita and von Storch 1999), the output of these models needs only to fulfil the areal character that is needed in impact studies, which means a higher resolution.

Therefore, when the impact studies are constructed directly from a digital model output (in our case COSMO LM), they are unsuitable because the spatial resolution is too coarse (von Storch et al. 1993, Palutikof & Wigley 1996). So downscaling techniques were required to generate input data with a finer spatial/temporal resolution. The final result of these downscaling techniques is expected to have the characteristic of areal distribution which means not only data with finer spatial resolution but also data that preserve the characteristics of the conditions that generated them.

The large range of methods encountered in the studied bibliography, has imposed problems not only for finding the most suitable method but also because of the limited range of data that is available to use for these methods.

Therefore for this first period I have compressed my activity in finding a way in which the predicted output of the COSMO LM can be used as base data for the statistical operation that compose the downscaling techniques.

From the bibliography studied and due to the logistical and data availability, emerged the idea of concentrating the activity on the statistical downscaling.

Statistical Downscaling (SD): is a method of obtaining high-resolution climate/meteorological information from relatively coarse-resolution model.

SD methods establish statistical relations among large-scale variables (predictors) and the variables on a finer-grid scale (predictands).

This paper presents, therefore, an assessment of a regression-based SD method that has been widely used for constructing climate/meteorological scenarios for daily precipitation at local sites using digital model grid point information.

If, in most cases, as predictors were considered sea-level pressure (SLP), geopotential height (Z), temperature or relative (RH) and specific humidity (SH) etc. and as methods, re-sampling (analogue methods) or weather generators, this time we have decided on a simpler approach – spatial interpolation.

This method is based on the idea that terrain variables and geographical location are used as predictors of the meteorological/climate variables. This is a plausible idea if we consider that different morphology and geographical location of a region receive in an equal different way the general atmospheric circulation and these morphological and geographical characteristics of a region impose local characteristics to the general atmospheric circulation. So, having in mind that the radiation fluxes stands at the base of atmospheric circulation, what we are doing is using the very roots of the factors (surely not all of them) that produces variability of the radiation fluxes (and by so the variability of the atmospheric circulation at the local scale) – the geography of relief (morphology, location). In other words, the different orientation of the slopes or the vicinity of an aquatic basin (sea, ocean), for example, determine a different dynamic of the atmosphere due to their different degree of isolation (solar radiation), which activates or not elements of the boundary layer climate (turbulence and air movement, thermal convectivity etc.). As consequence this elements are responsible of creating diversity at a local level within the general circulation.

Though a high number of variables (latitude, longitude, elevation, distance from the nearest coast, slope, aspect, etc.) are used in the bibliographical studies to create the areal distribution of the precipitation, the position examined in this initial paper would, mostly, be that of underling methods, characteristics, limitations and accuracy (at least from statistical point of view) of the predicted areal precipitation interpolated from the grid shaped output of the Cosmo LM. So, a low number of variables – latitude an longitude – were used in order to keep the attention on the ‘’behaviour’’ of areal distribution of precipitation when interpolated from a regular grid Cosmo Lm output with a various amount of methods.

It remains for the future work to examine if the number of variable, alone, are inducing an evident positive or negative evolution in the representation of areal precipitation using similar data and spatial interpolation characteristics and methods.

2 Spatial Interpolation

Interpolation methods allow creating a surface on the base of sample points and predicting of values in all point of territory. The justification underlying spatial interpolation is the assumption that points closer together in space are more likely to have similar values than points more distant. This observation is known as Tobler’s First Law of Geography

There are many interpolation methods which produce different results using the same data. The methods can be divided in two main groups: Regular and Irregular.

In case of using the ***irregular methods*** the points are connected at triangles. Each triangle determines a surface which is used for assessment of values of each point. It is necessary to have many known points for irregular methods implementation. Because of this, these methods are not suitable for climate/meteo data interpolation. These methods are often used for relief modelling – creating Digital Elevation Models (DEM) .

The ***regular methods*** for interpolation make close initial points and create a net of identical rectangles with determined by user sides. In this case two main groups of methods can be used: deterministic and geostatistical.

- ***DETERMINISTIC*** interpolation techniques create surfaces from measured points, based on either the extent of similarity (e.g., Inverse Distance Weighted) or the degree of smoothing (e.g., Radial Basis Functions). These techniques do not use a model of random spatial processes.

  Deterministic interpolation techniques can be divided into two groups, *global* and *local.* Global techniques calculate predictions using the entire dataset. Local techniques calculate predictions from the measured points within neighbourhoods, which are smaller spatial areas within the larger study area.

  a. Global interpolation – uses every control point available to derive an equation or model, so a change in one input value affects the entire map. In other words global interpolators determine a single function which is mapped across the whole region.

  The global interpolation is a 2 step method:

  - identification of a statistical relation between the examined parameter and the potential explicative factors (latitude, altitude, longitude etc.)
  - computation of the values of the unknown points using the known values of the explicative factors.

  b. Local interpolation – uses a sample of control points in estimating an unknown value, so a change in an input value only affects the result within the neighbour points. In other words the local method is based on the idea that the values of the nearest points are similar and the variation of the values is increasing with the increasing of the distance between the points. The advantage of local

interpolation is that the values of the known points stay unchanged in the process of interpolation so it can display the spatial anomalies of a phenomenon. The disadvantage of local interpolation are that:

- Even if displays the spatial anomalies of a phenomenon this does not explains causal factors.
- It needs a dense network of points with known values.

Some example of ***deterministic*** Interpolation methods**:**

- Polynomials,
- Spatial join (point in polygon),
- Thiessen-Voronoi polygons,
- Triangular Irregular Networks (TIN) and linear interpolation,
- Bi-linear interpolation,
- Spline,
- Inverse Distance Weighting (IDW),
- Radial basis functions.

Whether one of this deterministic methods can be considered local or global depends on the parameterization of the search neighbourhood. In other words each of this interpolation methods becomes global if the function is settled to fit the entire surface or local if the function is settled to fit specified neighbourhoods. For example Global Polynomials fits a polynomial to the entire surface, Local Polynomial interpolation fits many polynomials, each within specified overlapping neighbourhoods.

- ***GEOSTATISTICAL*** interpolation techniques (kriging) utilize the statistical properties of the measured points. Geostatistical techniques quantify the spatial autocorrelation among measured points and account for the spatial configuration of the sample points around the prediction location.

Establishing the theoretical presentation, an amount of question has arisen:

1. Are the predicted data of the COSMO LM accurate enough to be considered reliable for areal distribution of precipitation?
2. Do the characteristics of the data influence the result of interpolation?
3. Can the interpolation be considered a downscaling method?
4. Which one of the interpolation method is more efficient in downscaling?

The objective of these questions is to underline in a simpler and communicative manner the characteristics of the data and the interpolation appropriate for the prediction of areal precipitation.

3 Data characteristics

3.1 How reliable are predicted data compared with the observed data?

The lack of direct measurement of areal precipitation it is the problem that have been faced when trying to find input precipitation data for all impact studies involving the ground phase of the water cycle. Reliable direct areal rainfall measurements can be obtained only at a very limited spatial scale, while the rainfall process is known to exhibit a high degree of variability both in space and time.

The areal precipitation has been obtained using 3 methods:

- traditional method, which estimates precipitation at ungauged sites through suitable interpolation method. These are based on the hypothesis that rainfall estimates at ungauged sites can be obtained as linear or non-linear combinations of the values measured at a number of instrumented locations.
- indirect estimates of areal rainfall based on the measurement provided since the late ’60s by ground-based meteorological RADARs and remote sensing devices borne on satellite platforms, such as RADARs and other sensors.
- physically-based numerical models of the atmosphere — though relying on various theoretical approximations — provide predictions of temporal accumulation values for areal rainfall over wide spatial scales.

Anyway, the accuracy of each and every one of this techniques is always questioned: the traditional method is based only on the assumption that rain gauge measurements can reliably account for the “true point rainfall”, the indirect method encounters problems of calibration and validation using historical data and the data from the physically-based

numerical models relays to much on the quality of the model accuracy and also the output of such models is not known of recreating in a reliable way the actual precipitation measurements.

So, our attention was drawn towards making a statistical comparison of the predicted and observed data in order to see how much the predicted data conserves the statistical characteristics of the observed data (which is assumed that they are “true point rainfall”).

For this, it was used a simple statistical analysis with the purpose of finding whether data from several groups have a common mean. That is, to determine whether the groups are actually different in characteristics.

Based on the idea that a valid statistical comparison between analogue variables can not only be made at the level of the points but also at the areal level, as revealed in the results of ISC research activity, has been found that 79% of the average values of observed precipitation it’s included between the area occupied by the maximum and the average are of the predicted values of precipitation (fig. 1).

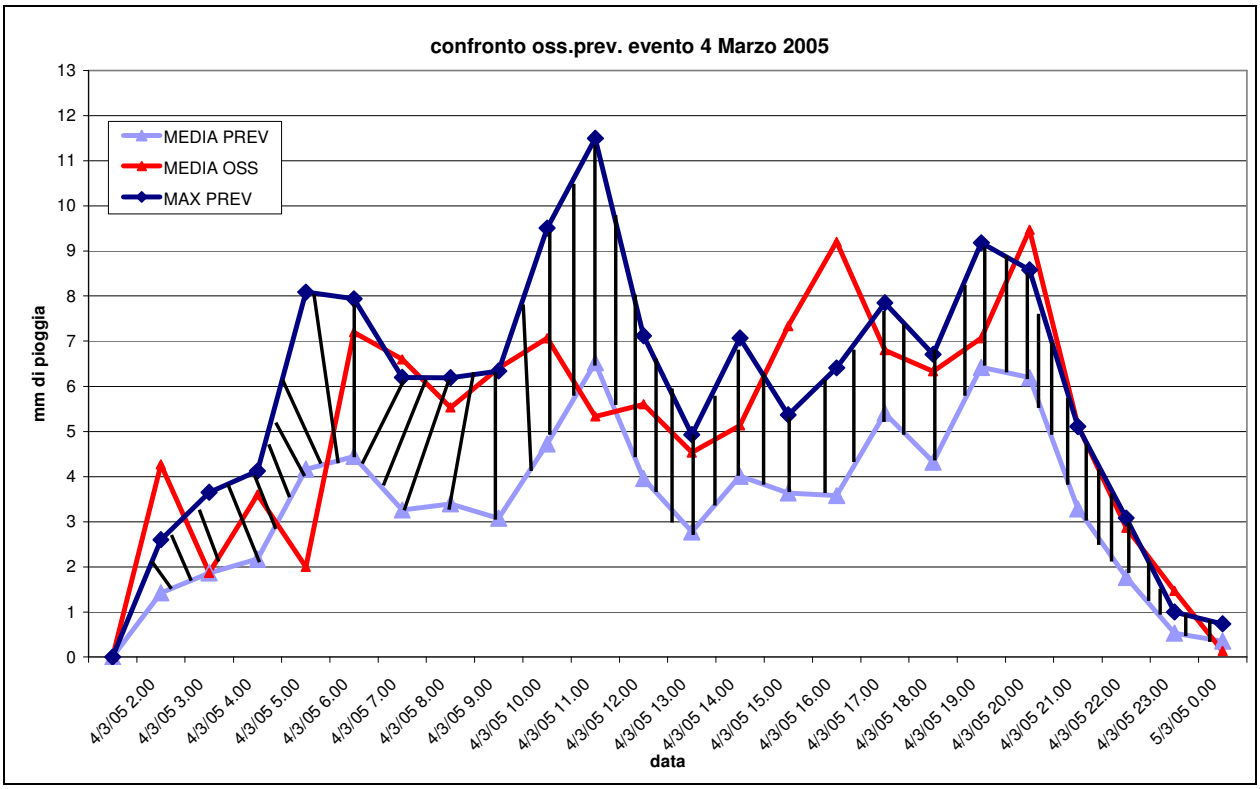


Figure 1. Comparison between observed and predicted precipitation values
(ISC research activity, 2008)

This establishes that development of the predicted values of precipitation follows in a satisfactory measure the evolution of the observed precipitation values, which means that,

to a great extent, the predicted data conserves the statistical characteristics of the observed data.

3.2 Do the characteristics of the data influence the result of interpolation?

Knowledge about the data that are to be interpolated is critical to decide on an appropriate interpolation method and to understanding the results produced by the interpolation. Characteristics of the data (spatial representativeness, measurement accuracy, and existence of spatial relationships) important to consider in interpolation (Shelly Eberly, et al. 2004) were concentrated in: the analysis of summaries, geometrical configuration, distance distribution and linearity of the space between grid points of the data.

3.2.1 Summaries of the available data.

It is important to generate some initial summaries of the available data prior to analysis in order to obtain a better understanding of its spatial characteristics. Reasonable summaries include, a histogram of the overall values distribution, and summary statistics such as the data's mean, standard deviation, and various percentiles (e.g., minimum, median, maximum, etc.).

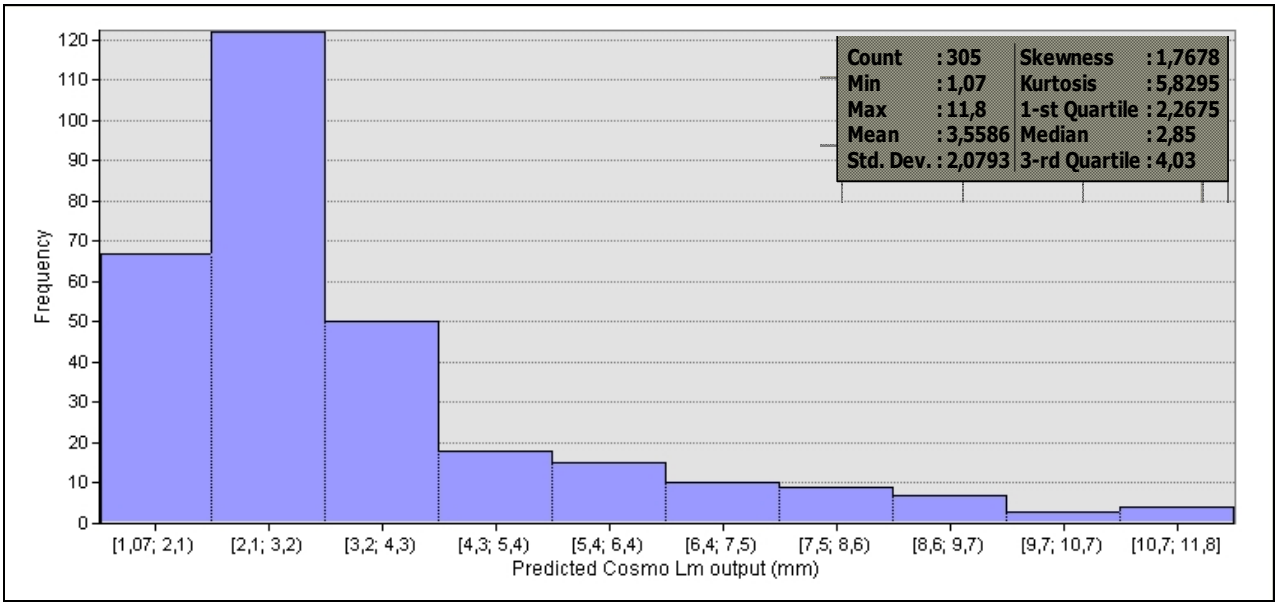


Figure 2. Summaries of the COSMO LM output data

Histogram shows that our precipitation values are not perfectly normally distributed. One of the crosscheck of normal distribution of data is that mean should be closer to the median. In our case mean is 3,5mm and median is 2,85mm. Also the data shows an asymmetric distribution with positive skewness and a leptokurtic character which means that a great number of observations cluster near the average and the rest of observations are skewed towards the right.

3.2.2 Geometrical configuration of the data

Some of the studies (Morrissey et al. 1995) have assumed that the standard error of the data disposed in a network, depends not only on the density (which in our case is almost 0.5 grid points/ Km^2) of the points with known values but even on the geometrical configuration. The overall conclusion of this studies concluded that the uniform network are the best in the terms of accuracy of the representation of the data in space. As the output of the Cosmo LM is represented on a regular uniform grid we can establish that in terms of spatial structure of measured field the error that might occur in the process of interpolation is minimal.

3.2.3 The distribution of the nearest-neighbour distance

The distribution of the nearest neighbour distance it is considered also important in many studies. It has been examined as mean of distribution (Smith et al. 1986) or as coefficient of skewness (Matthew Garcia et al 2008) which is said that constitutes a clustering factor (CF). In the case of the regular grid, the clustering factor remains undefined because of the singular value of nearest-neighbour distance (2,8Km).

It was considered though that a CF < 0 (a more distributed network) will produce interpolation errors by reduced resolution of the precipitation field and that CF > 0 (clustering in the network) will produce errors because of reduced areal representation of the precipitation field.

In the case of CF= 0, which is the case of a regular grid, it is considered that both the resolution of the precipitation field and the areal representation of the precipitation filed are characterized by reduced uncertainty and thus by lower errors of the prediction

3.2.4 Linearity of the space between grid points of the COSMO LM output

The data set required for any two-dimensional spatial interpolation exercise consists of three variables: the parameter of interest (Z), location in the first spatial dimension (x), and location in the second spatial dimension (y). The variables x and y are in the COSMO LM data output, longitude and latitude respectively. Such a coordinate system is subject of the curvature of the earth's surface. Statistical spatial interpolation techniques assume some sort of spatial correlation structure defined with respect to the linear distance between two points in space. (Yan Yu, Deepak Ganesan, Lewis Girod, Deborah Estrin, Ramesh Govindan, 2003). Therefore, it is not strictly accurate to calculate the distance between two points in a longitude by latitude coordinate system using a simple linear distance function.

But in the case of the output data of the COSMO LM, the spatial domain under study is small enough in geographic extent (2,8Km) in such the error of calculation will be minimal.

As a conclusion we can say that the output data of the COSMO LM presents the most desirable characteristics in terms of uniform, geometrical and density distribution of the data for the purposes of prediction of values trough out the studied territory.

3.3 Interpolation characteristics

3.3.1 Interpolation = downscaling?

Based on the available measurement and modelling approaches and the nature of application, the tendency adopted, in the bibliographical studies, was to replace the unavailable areal rainfall observations, at the required space-time scales, with suitable surrogates based on *interpolation* or *downscaling* techniques.

Interpolation is the process of predicting the values of a certain variable of interest at unsampled locations based on measured values at points within the area of interest (Burrough and McDonnell 1998).

"Downscaling" is based on the view that regional climate is conditioned by climate on larger, for instance continental or even planetary, scales. Information is cascaded "down" from larger to smaller scales (Hans von Storch, 2004). In other words, downscaling is any process where large (coarse) scale output of models is reduced or made finer.

Practically both are establishing the same relation of transfer function between coarse scale and finer scale, and the spatial and temporal distance is the variables that have to solve.

“In general, interpolation is applied when ground-based rain gauge and/ or radar networks are available, while downscaling is applied when using indirect measurements from other remote sensing devices, or predicted values from atmospheric models, all of which are usually available at much coarser scales than those required in most hydrological applications.’’ (L.G. Lanza, J.A. Ramírez and E. Todini - *Stochastic rainfall interpolation and downscaling,* Hydrology and Earth System Sciences, 5(2), *139–143* (2001)).

“However, a sharp distinction cannot be made since interpolation and downscaling can both be incorporated in one single approach, e.g. in order to exploit jointly the information content of both remotely sensed and rain gauge data (Fiorucci *et al.*, 2001; Todini, 2001)’’

3.3.2 Which one of the interpolation methods is more efficient in downscaling?

Several characteristics of spatial interpolation have been considered in order to facilitate the identification of the most efficient method of transferring the character of the coarse scale COSMO LM to a finer scale trough interpolation. These characteristics include point-based versus areal-based, global versus local, exact versus approximate, stochastic versus deterministic, gradual versus abrupt. A compressive description of this characteristics has been the subject of a subchapter of a study by Shelly Eberly, Jenise Swall, David Holland, Bill Cox, Ellen Baldridge,2004, study on which we will draw our conclusion in the following part of the report.

- **Point-based versus Areal-based:**

  Point-based interpolation methods predict values at specific points in space, based on the values and locations of other individual points in space. Areal interpolation methods estimate values for entire zones or areas based on data available for a different set of zones or areas. Even if areal interpolation methods seems to correspond with our task of representing precipitation on an areal level, a amount of downfall – like data shaped in a grid network and even limitation of equipment and analogue data (areal-based has the form of: surrounding area A, B, C, and D, estimate the values in area E) - are the subject of our reserve toward choosing this method.

- ***Global versus Local*:**

  Global interpolators develop and use a single function that estimates values for the entire sample area whether *local interpolators* break the full sample area into smaller pieces that are each evaluated individually by a particular function. Changing one input data point - in global interpolation - affects the predictions for the entire area, whether in local interpolator, affects only those areas that consider that point in the prediction algorithm. In order to choose one of these methods we have to establish if the function that is mapped has to use the entire area of concern or has to break up the area into smaller blocks that are evaluated individually. Due to the characteristics of precipitation that assumes abrupt spatial variance suitable to their complex morphology our attention was drawn on local interpolation. Global interpolation may be useful for interpolating surfaces with gradual variation over the area of interest.

- ***Exact versus Approximate*:**

These two methods are considered when we want to produce a surface that avoids or not sharp peaks and troughs in the estimated surface. Also these features are important when there is uncertainty about the accuracy of the measured values (i.e., measurement error).

These methods vary, based on whether the predicted surface must include the exact values of the measured data points (exact interpolation) or not (as in the case of approximate interpolation).

Even on these characteristics the choice of an interpolation method is driven by the phenomenon that we want to represent. For areal representation of precipitation an accurate data that reproduces the variation of the precipitation might request that the predicted surface has to replicate the measured values exactly - so it would be more appropriate to use an exact method of interpolation.

- ***Stochastic versus Deterministic*:**

Whether methods utilize the concept of randomness is another important characteristic to consider. Stochastic methods incorporate the idea of randomness into the interpolation process. These methods, which include kriging, allow the uncertainty of the predicted values to be calculated. Deterministic methods do not incorporate statistical probability theory into development of the predictions. Instead, these methods use mathematical formulas or other relationships to interpolate values.

An example of a deterministic method would be one that derives a predicted value by a simple averaging of nearby measured points. Inverse Distance Weighted (IDW) is a deterministic method that uses a weighted average of nearby points with distance being the only factor influencing calculation of the weight. The advantage of stochastic methods is the ability to provide estimates of uncertainty for the spatial interpolation model's output. Kriging is a stochastic method because it assigns weights based not only on the distance between surrounding points but also on the spatial autocorrelation among the measured points, which is determined by modelling the variability between points as a function of separation distance.

- ***Gradual versus Abrupt*:**

Another distinguishing characteristic of spatial interpolators is the smoothness of the predicted surface that is produced. A gradual interpolator produces a surface with

gradual, relatively smooth, changes. An abrupt interpolator produces a discontinuous surface with sharp changes. The proximal “nearest-neighbour” method, which sets unknown points equal to the nearest measured point, is an example of an abrupt interpolator.

As a conclusion drawn by the theoretical information, but yet to be sustained by the practical exercise, we can say that – considering the characteristic of our output data of the COSMO LM and the complex characteristics of precipitation – the most suitable method that can be used for downscaling has to be a point-based, local, exact (with reservation due to the accuracy of measurements), deterministic and abrupt method.

These characteristics of the interpolation method, at the end, are pointing on the spatial variability of the precipitation that the predicted areal distribution has to include. So, the characteristics of the interpolation has to use a function that evaluates individually the samples (local method) on detriment of the global which accentuates the smoothness of the distribution, also for the accuracy of the variability it has to use the exact and not the approximate method. Thus, the prediction has not to avoid sharp peaks and troughs in the estimated surface. Also for the accuracy of the spatial distribution of the precipitation the areal-based method was overlooked because it estimates values for entire zones or areas which might lead to a loss of detailed information reason for which the abrupt method has to be employed in order to realistically catch the gradual changes within the phenomenon. The deterministic method it might be the better method because in a regular grid the distances between the sampled points it has no major contribution. So, assigning weights based on the distance of the surrounding points or modelling the variability between points as a function of separation distance as the stochastic method does, do not bring contributions to the final result. In other words in a regular grid shaped data the spatial variability of a phenomenon can not be contained by the method that attributes weights to the neighbouring point based on any function of distance separation but by the method that gives weights taking into consideration mathematical/statistical relation between the magnitude of the values.

4 Spatial interpolation - methods

In the task of choosing the right interpolation method I've taking into consideration 3 factors:

- the objective of the activity;
- the availability and characteristics of the data (discussed in a previous part of the report);
- the software available.

The objective would be to find an appropriate method for transferring the characteristics from the coarse scale data at which the COSMO LM delivers output to finer scale data that preserves the nonlinear characteristics[2] of the conditions that generated them and retains a high degree of resemblance to the actual or true precipitation field.

The Data Available are:

- hourly predicted precipitation values from the COSMO LM digital model with the horizontal resolution of 2.8KM (fig.3) ;
- latitude and longitude of the grid points;
- hourly observed data from the meteorological stations: S. Mauro, Ponte Camerelle, meteorological station A3 highway ;

Both observed and predicted data were selected for a common period of 24h on 1h step from 4th March 2005.

The activity was developed in ArcGis software, which allows many techniques of interpolation. ArcGIS is an integrated collection of GIS software products for handling spatial data, developed by Environmental Systems Research Institute (ESRI). GIS is a system designed to capture, store, update, manipulate, analyze, and display the geographic information. For interpolation of meteorological/climatological parameters the Geostatistical Analyst extension of the version ArcGis 9.2. were used, which provides a set of tools to create a continuous surface using deterministic and geostatistical methods. Gis has been vastly used for areal representation of precipitation: Ahers, B., 2006; Hevesi, J.,

[2] "A time-space spectral analysis of the atmosphere certainly reveals pronounced periodically-varying large-scale motions, but the general nonperiodic behaviour and much of the small-scale structure'' known in digital models as sub-grid scale processes ''are direct results of nonlinearity''**. *Lorenz, E.N., 1966: ''Nonlinearity, Weather Prediction and climate deduction'' Final Report, Statistical forecasting project, 22 pg;***

Flint, A. & Istok, J. (1992); Fiorucci, P., La Barbera, P., Lanza, L.G. and Minciardi, R., 2001, D. Kastelec, M. Dolinar, 2000; Tveito,O.E 2002 etc.

5 Interpolation Methods Developed in ArcGis.

A full description of this characteristics has been the subject of the a studies by Shelly Eberly, Jenise Swall, David Holland, Bill Cox, Ellen Baldridge,2004;

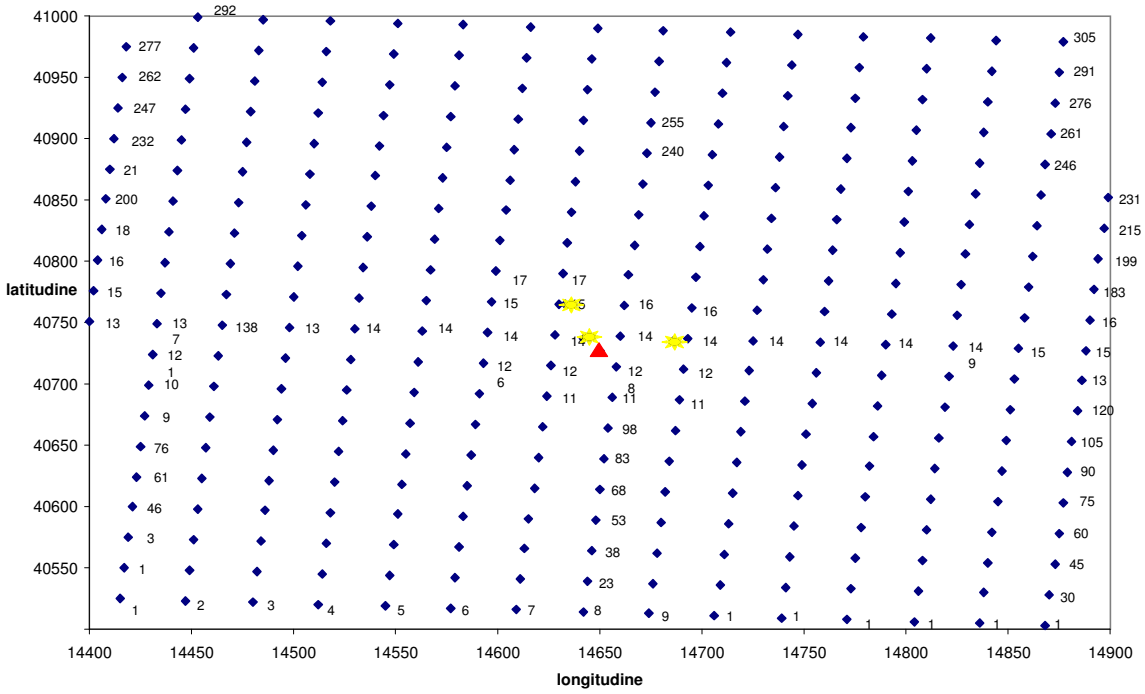


Figure 3. In blue the grid points of the COSMO LM model (resolution 2.8 km), in yellow meteo station in red the landslide position.

and *ArcGis desktop help,* on which we will draw our conclusion in the following part of the report.

5.1 Inverse Distance Weighted (IDW):

IDW is an exact local deterministic interpolation technique. IDW assumes that the value at an unsampled location is a distance-weighted average of values at sampled points within a defined neighbourhood surrounding the unsampled point (Burroughs and McDonnell 1998). In this sense, IDW considers that points closer to the prediction location will have more influence on the predicted value than points located farther away (Johnston et al. 2001). Specifying a higher power places more weight on the nearer points while a lower power increases the influence of points that are further away. Using a lower power will result in a smoother interpolated surface being generated (*ArcGis Desktop help*).

IDW uses:

$$\hat{z}(s_0) = \sum_{i=1}^{N} \lambda_i . z(s_i)$$

$\hat{z}(s_0)$ is the predicted value at the unsampled location s_0.

N is the number of measured sample points within the neighbourhood defined for s_0.

λ_i are the distance-dependent weights associated with each sample point.

$z(s_i)$ is the observed value at location *si*. Weights are calculated using:

$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{N} d_{i0}^{-p}}$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

Where:

d_{i0} is the distance between the prediction location *s0* and the measured location *si*.

P *is* the power parameter that defines the rate of reduction of the weights as distance increases.

IDW is forced to be an exact interpolator to avoid the division by zero that occurs when $d_{i0} = 0$ at the sampled points.

IDW is an extremely fast interpolation method, though it is very sensitive to the presence of outliers and data clustering. In addition, this method does not provide an implicit evaluation of the quality of the predictions (Burrough and McDonnell 1998, Johnston et al. 2001).

5.2 Radial Basis Functions (RBFs):

RBF are a series of exact deterministic interpolation techniques that include different basis functions like thin-plate spline, spline with tension, completely regularized spline, multiquadric function, and inverse multiquadric spline (*ArcGis Desktop help*). RBFs can be seen as the process of fitting a flexible membrane to the data points so the total curvature of the surface is minimized. Being also an exact interpolator, RBFs are different from IDW

because they allow the prediction of points above the maximum measured value and below the minimum measured value. In other words if IDW is based either on the extent of similarity the RBFs is based or the degree of smoothing.

The predictor defined by a RBF is a linear combination of *N* basis functions (one for each data point in the neighbourhood) of the form:

$$z(s_0) = \sum_{i=1}^{N} \omega_i \cdot \phi(\|s_i - s_0\|) + \omega_{n=1}$$

Where:

φ(r) is a radial basis function.

$r = ||s_i - s_o||$ is the distance between the prediction location s_0 and the measured location s_i.

$\{\omega_i: i = 1, 2, \ldots, n + 1\}$ are the weights to be estimated.

The vector of weights w=(ω1, ω2, … ωn) is calculated by solving the following system of equations:

$$\begin{pmatrix} \Phi & 1 \\ 1' & 0 \end{pmatrix} \begin{pmatrix} w \\ \omega_{n+1} \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}$$

Where:

Φ is a matrix with *i*, *j* th elements corresponding to $\Phi(||s_i - s_o||)$ for each pair of data points.

1 is a column vector of ones.

Z is a column vector containing the data points.

ω_{n+1} is a bias parameter.

5.3 Polynomial Interpolation (PI)

PI is an approximate, deterministic interpolation method that fits a mathematical function to the measured points. Options range from a first-order polynomial (linear) to a second-order polynomial (quadratic) to higher-order polynomials (ArcGis ranges from the first up to 10th order polynomial). The predictive surface is typically generated by using a least-squares regression fit that minimizes squared differences between the surface and measured

points. Because it is an approximate interpolator, the surface is not constrained to going through the measured points as with RBF interpolation. In addition, because the method generates the best fit (least squares criterion) between the measured points, it is unlikely that the fitted line will run outside the minimum or maximum measured value, except once it goes beyond the measured area (i.e., extrapolation).

There are two types of polynomial interpolation — global and local.

- *Global polynomial* interpolation fits a polynomial model to the entire surface based on all measured points.
- *Local polynomial* interpolation fits multiple polynomials using subsets of the measured points.

Global polynomial interpolation is more appropriate for a surface that varies slowly over the area of interest, while local polynomial interpolation captures more of the short-range variation in addition to the long-range trend. Global polynomial interpolation accounts for bends in the data — one bend with quadratic, two bends with cubic, and so forth. Surfaces that do not display a series of bends, however, such as one that increases, flattens out, and increases again, can be better represented using local polynomial interpolation. Both the global and local methods produce a gradual predicted surface.

5.4 Kriging

Krigging is an optimal interpolation based on regression against observed *z* values of surrounding data points, weighted according to spatial covariance values (Geoff Bohling, 2003).

Geostatistical interpolation methods are stochastic methods, with kriging being the most well-known representative of this category. Kriging methods are gradual, local, and may or may not be exact (perfectly reproduce the measured data). Also, they are not by definition set to constrain the predicted values to the range of the measured values. Similar to the IDW method, kriging calculates weights for measured points in deriving predicted values for unmeasured locations. With kriging, however, those weights are based not only on distance between points, but also the variation between measured points as a function of distance. The kriging process is composed of two parts — analysis of this spatial variation and calculation of predicted values.

Spatial variation is analyzed using variograms, which plot the variance of paired sample measurements as a function of distance between samples. An appropriate parametric

model is then typically fitted to the empirical variogram and utilized to calculate distance weights for interpolation. Kriging selects weights so that the estimates are unbiased and the estimation variance is minimized. This process is similar to regression analysis in that a continuous curve is fitted to the data points in the variogram.

Kriging creates a continuous surface for the entire study area using weights calculated based on the variogram model and the values and location of the measured points. The analyst has the ability to adjust the distance or number of measured points that are considered in making predictions for each point. A fixed search radius method will consider all measured points within a specified distance of each point being predicted, while a variable search radius method will utilize a specified number of measured points within varying distances for each prediction.

Because kriging employs a statistical model, there are certain assumptions that must be met. First, it is assumed that the spatial variation is homogenous across the study area and depends only on the distance between measured sites. There are different kriging methods and each has other assumptions that must be met.

Kriging methods are often classified as linear and nonlinear (Moyeed and Papritz, 2002; Papritz and Moyeed, 1999). There are no formal definitions for linear and nonlinear kriging.

5.4.1 Linear kriging (LK)

LK can be defined as kriging methods that derive the estimation using observed values by assuming a normal distribution of the samples. Linear kriging may include:

- *Simple Kriging,*
- *Ordinary Kriging and*
- *Universal Kriging.*

5.4.2 Non-linear kriging (NLK)

NLK are those methods that derive predictions based on the transformed values of the observed data. Nonlinear kriging methods consist of:

- *Disjunctive Kriging,*
- *Indicator Kriging,*
- *multiGaussian kriging,*
- *lognormal Ordinary Kriging*

- *Model-Based Kriging.*

Nonlinear kriging methods have two major advantages over linear kriging:

1) they were developed to model the conditional distribution of the primary variable (i.e., to give an estimate of its probability distribution conditional on the available information);
2) their estimations should theoretically be more precise when a Gaussian random process is inappropriate to model the observations.

The basic form of the kriging is (Geoff Bohling, 2003):

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

With

u, u_α: location vectors for estimation point and one of the neighbouring data points, indexed by **α**

n(u): number of data points in local neighbourhood used for estimation of *Z**(u)

m(u**),** m (u_α)**:** expected values (means) of *Z*(u) and (*Z* u_α)**;**

λ_α(u): kriging weight assigned to datum *Z*(**u_α)** for estimation location **u**; same datum will receive different weight for different estimation location

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

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

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

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

$$E\{R(u)\} = 0$$

$$Cov\{R(u), R(u+h)\} = E\{\mathrm{R(u)} \times \mathrm{R(u+h)}\} = \mathrm{C_R(h)}$$

The residual covariance function is generally derived from the input semivariogram model,

$$C_R(h) = C_R(0) - y(h) = Sill - y(h).$$

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

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

Resuming the mathematical expressions:

- *Simple kriging* assumes that there is a known constant mean, that there is no underlying trend, and that all variation is statistical (Wackernagel, 2003).
- *Ordinary kriging* is similar except it assumes that there is an unknown constant mean that must be estimated based on the data and the data have no trend (Clark and Harper, 2001; Goovaerts, 1997).
- *Universal kriging* differs from the other two methods in that it assumes that there is a trend in the surface that partly explains the data's variations. In other words it is incorporating the local trend within the neighbourhood search window as a smoothly varying function of the coordinates. Universal Kriging estimates the trend components within each search neighbourhood window and then performs Simple Kriging on the corresponding residuals. This should only be utilized when it is known that there is a trend in the data.
- *Disjunctive Kriging* is a nonlinear method that is more general than ordinary kriging. It considers functions of the data rather than using only the data. It assumes that all data pairs come from a bivariate normal distribution. The theory of disjunctive kriging and examples of its practical application are described by Armstrong and Matheron (1986a; 1986b), Rendu (1980) and Oliver *et al*. (1996).

6 Parameterization

Spatial interpolation methods in ArcGis have various decision parameters to choose from, no matter if we consider trend based methods or weight based methods. Based on the knowledge of the nature of data being sampled and processes involved (in our case rainfall on 305 locations) some parameters can be fixed before the calculations start. The choices that are made affect the results of the interpolation.

Interpolation methods such as IDW and RBF (weight based methods) require fewer decisions or parameters to manipulate in comparison to kriging methods. Parameters important for determining the validity of the surface model include (which may vary by model); surrounding point weight, neighbourhood search, and anisotropy. IDW and RBF both have similar parameters for determining the small scale variation involved within the dataset. Kriging interpolation methods utilize functions such as semivariogram and covariance to assess the weight given to surrounding data points, based on distance and direction.

Finding the most suitable weight for IDW or RBF is accomplished easily in GeoStatistical Analyst 9.2 through the “optimize power” feature or consecutively adjusting it during the procedure until the lowest RMS was obtained.

A curve is fit (quadratic local polynomial equation) to the points and from the curve, the power that provides the smallest RMS is determined as the optimal power (ESRI, 2004). When determining the influence of surrounding data points (i.e. neighbourhood) for weighing interpolation calculations, careful analysis of the involved parameters is essential.

The neighbourhood search is used to define the neighbourhood shape and the constraints of the points within the neighbourhood that will be used in the prediction of an unmeasured location. Neighbourhood search sizes should be large enough to capture the variability in the data, but small enough to avoid capturing distant points, which create reduces spatial autocorrelation with the prediction location, hence jeopardizing the appropriateness of stationarity (Isaaks and Srivastava, 1989).

Anisotropy is a characteristic of a random process that shows higher autocorrelation in one direction than another.

Although no measures are known that would or could be universally applied to choose the optimal set of parameters for kriging, cross-validation (a.k.a. "leaving-one-out" method) is often used to select an interpolator from different number of candidate (Davis, 1987).

Cross validation provides an array of statistical and graphical outputs for comparison of different parameters before surface model creation, allowing for manipulation of parameters if needed.

Among the prediction error output statistics for cross validation of *deterministic and stochastic* interpolation methods is the mean prediction error (MPE) and root-mean square

(RMS). The RMS statistic is a measurement of how close the predicted values are to the measured values, in which smaller values are preferred. The mean prediction error statistic (MPE) is a measure of the bias within the model, which will produce values centered on zero for unbiased models.

Stochastic methods provide additional statistics as an extra measure of uncertainty and potential error for the prediction model. The kriging standard error, a statistical measure of uncertainty in the prediction, is calculated by the square root of the kriging variance. RMS and MPE values can be "standardized" to account for scale dependence, by dividing the RMS and the MPE each by the standard prediction error to produce RMS standardized and MPE standardized. The RMS standardized is a measure of variability in addition to the kriging standard error, in which RMS standardized values will underestimate the variability when greater than one and overestimate variability where values are less than one (ESRI, 2004).

The use of cross validation prediction error statistics can be a beneficial tool for finding differences among interpolation methods, however may fall short of clear determination for finding the "optimal" interpolation method. In such situations, *Pearson's correlation* coefficient and *standard deviation* values calculations may be beneficial for interpolation model determination (Isaaks and Srivastava, 1989).

So our decision in choosing the best method of interpolation was highly influenced by the cross-validation techniques and in addition Pearson's correlation coefficient and standard deviation values.

7 Results and discussions of the interpolation methods developed in Arcgis

7.1 IDW

This spatial interpolation method has various parameters decision. The descriptions below include the options used in Geostatistical Analyst extend of ArcGis 9,2.

Parameters include: β - the weighting power (exponent); δ - the smoothing parameter; ρ - the anisotropy ratio; θ - the anisotropy angle.

The method used to consider best IDW interpolation was optimizing parameters via cross-validation. The parameters were simultaneously adjusted during the procedure until the lowest RMS was obtained.

At the end, the lowest RMS (0,8272) was obtained using a low power (two) in order to give influence to the points situated farther away too, and the number of points used for each cell's calculation was limited at 15 to reduce the risk of errors, because points far from the cell location where the prediction is being made, might have no spatial correlation. Also the smoothing parameter was not considered, on benefit of standard option, as it accentuated '' the bulls eye effect'' (concentric circles around the measured value at the locations).

The overall predicted contour map is shown in Figure 4 a. This surface shows low degree of smoothness, a increased spatial variance and even if the power has a low value, the prediction displays the effect of the concentration around the values of location still causes ''bulls eye effect''.

Figure 4 b. shows the predicted rainfall by the Cosmo LM plotted against interpolated values for the same locations. The linear correlation coefficient r=0.72 confirms relatively good overall agreement between independent (Cosmo LM output) and dependent (interpolated) values. The fact that the larger values of both variables (predicted COSMO LM and interpolated values) are associated gives a positive related character of correlation analysis.

The ability to predict extreme values is an important benchmark in evaluating the performance of an interpolator. Of the extreme values, the lowest values were predicted

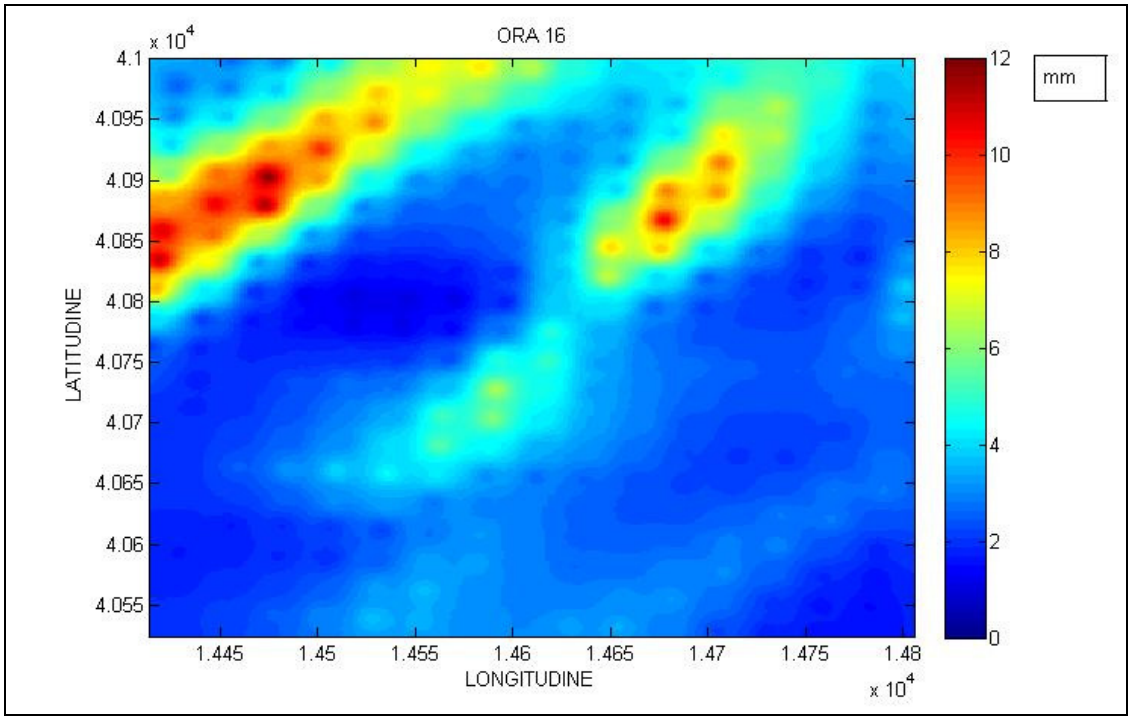


Figure 4 a. Interpolated surface using smooth *IDW*

better with dependent and independent low extremes being, up to some extent, similar but the highest values were inaccurately predicted.

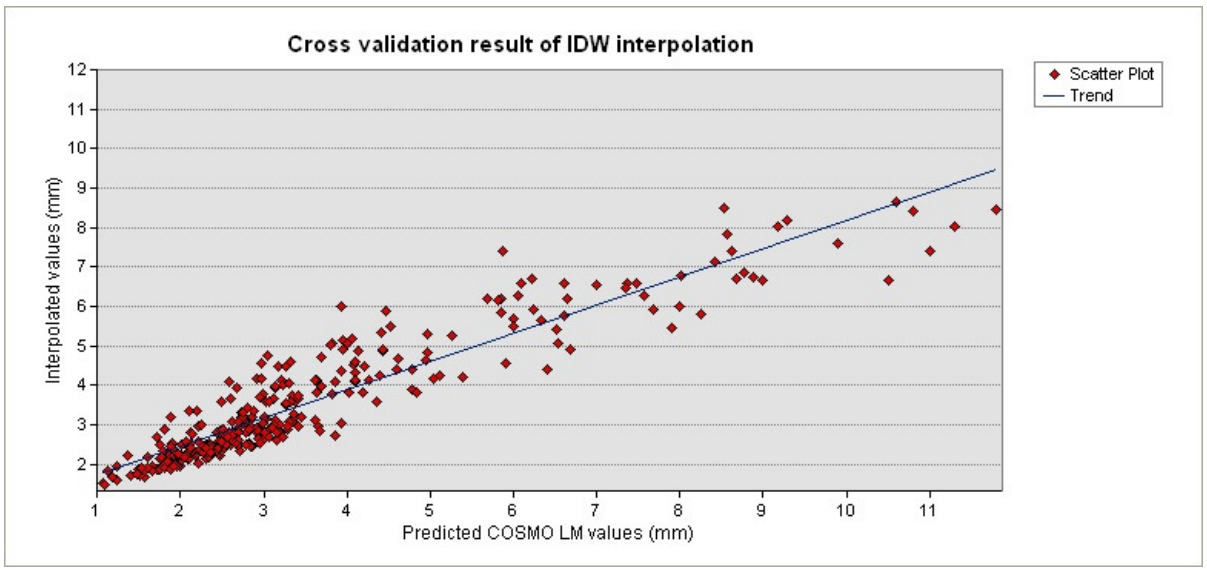


Figure 4 b. Cross-validation of the IDW interpolated prediction

Figure 4 c. shows the distribution of errors (value dependent minus value independent, or residual) as a function of the magnitude of independent values (Cosmo LM output).The residuals seems to have a weak tendency of under prediction of the values, with the increasing of the independent variables. Also the over predicted residuals of the dependent variable tend to have the outlier characteristics (residuals that fall far from the regression line) and have a greater variance compared with the under prediction of the smaller dependent values.

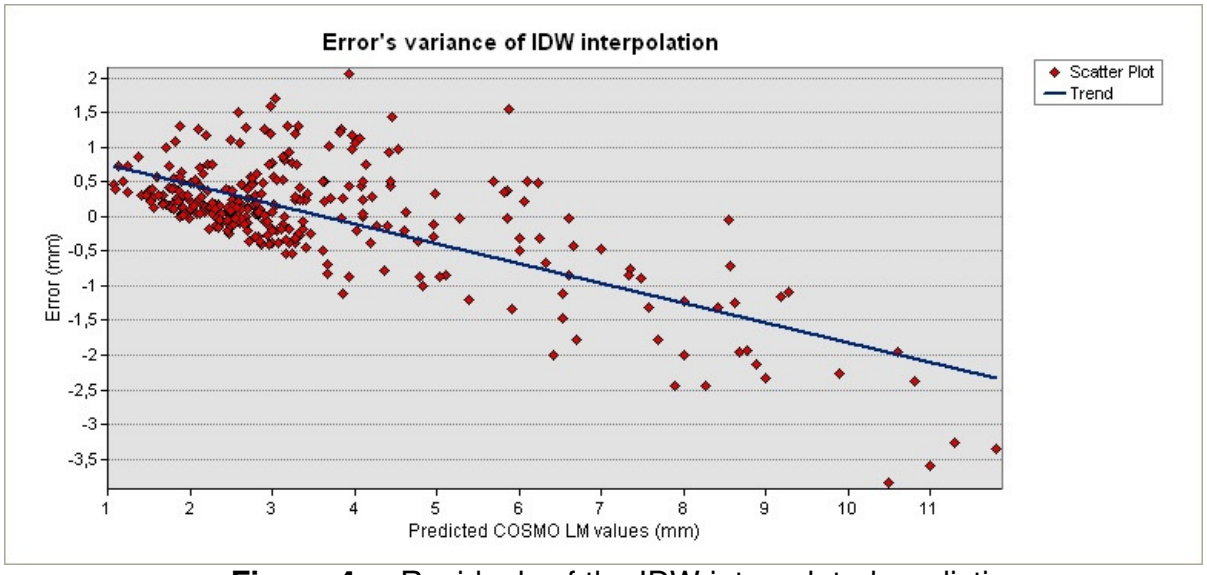


Figure 4 c. Residuals of the IDW interpolated prediction

7.2 Radial Basis functions

RBFs can be seen as the process of fitting a flexible membrane to the data points so the total curvature of the surface is minimized. Being also an exact interpolator, RBFs are different from IDW because they allow the prediction of points above the maximum measured value and below the minimum measured value.

There are five different basis functions:

- Thin-plate spline
- Spline with tension
- Completely regularized spline
- Multiquadric function
- Inverse multiquadric function

Each basis function has a different shape and results in a slightly different interpolation surface. Also the parameters selected for each of the five basis functions could bring an amount of changes in the final areal distribution of the independent variable. That is way the selection of the optimal interpolated prediction, for each of the five basis functions, was done by optimization of the parameters value until the smallest MPE and RMS was obtained for each of the interpolated predictions. For the optimization and selection of the parameter's values was used an identical approach, for the five different function in order to set a commune base to start from. In terms of the neighbourhood values ArcGis presents two functions: standard and smooth. The standard search neighbourhood is defined by the Ellipse parameters: Angle, Major Semiaxis, and Minor Semiaxis. The Smooth Interpolation option creates an outer ellipse and an inner ellipse at a distance equal to the Major Semiaxis multiplied by the Smooth Factor. Prediction to each point uses data inside each corresponding circle/ellipse (ArcGis Desktop help). For the *standard search neighbourhood* after the optimisation of the parameters (angle=0; neighbourhoods=15; major and minor semiaxis = 0,175), a good fitted surface with the smallest RMS (table 1) was predicted by the RBFs with *completed regularized spline* (RMS=0,177), followed by *multiquadric* (RMS=0,181) and *inverse multiquadric* RBFs (RMS=0,182). But when the smoothing effect was added (the same optimized parameters as the standard search neighbourhood were used and in addition a smooth factor of 0,5) the only RBFs available in ArcGis 9.2, the *inverse multiquadric* predicted the best interpolated surface, having a RMS equal to 0,1501. Changing the smooth factor by

increasing or decreasing it with a pass of 0,1 confirmed the fact that still RBFs with a smooth factor of 0,5 best predicts the interpolated surface.

Table 1 . Comparison of the RBFs (MPE = mean predicted error, RMS = root mean square of predicted surface):

Radial Basis Functions		MPE	RMS
Standard	Completely regularized Spline	0,00278	0,1778
	Spline with tension	0,0000841	0,2993
	Multiquadric	0,001143	0,1811
	Inverse multiquadric	0,001982	0,1824
	Thin plate spline	0,0006502	0,3018
Smooth	***Inverse multiquadric***	***0,0003563***	***0,1501***

Using the cross-validation of the data (fig. 5b and 5c) obtained by *inverse multiquadric* RBFs with smoothing factor and the graphical representation (fig. 6.a) an amount of characteristics could be underlined. The RBFs interpolation produces:

- a smooth surface prediction (fig.6a) due to the fitting of the interpolation curve trough the measured sample values while minimizing the total curvature of the surface which can be translated as - prediction above the maximum and below the minimum measured values;

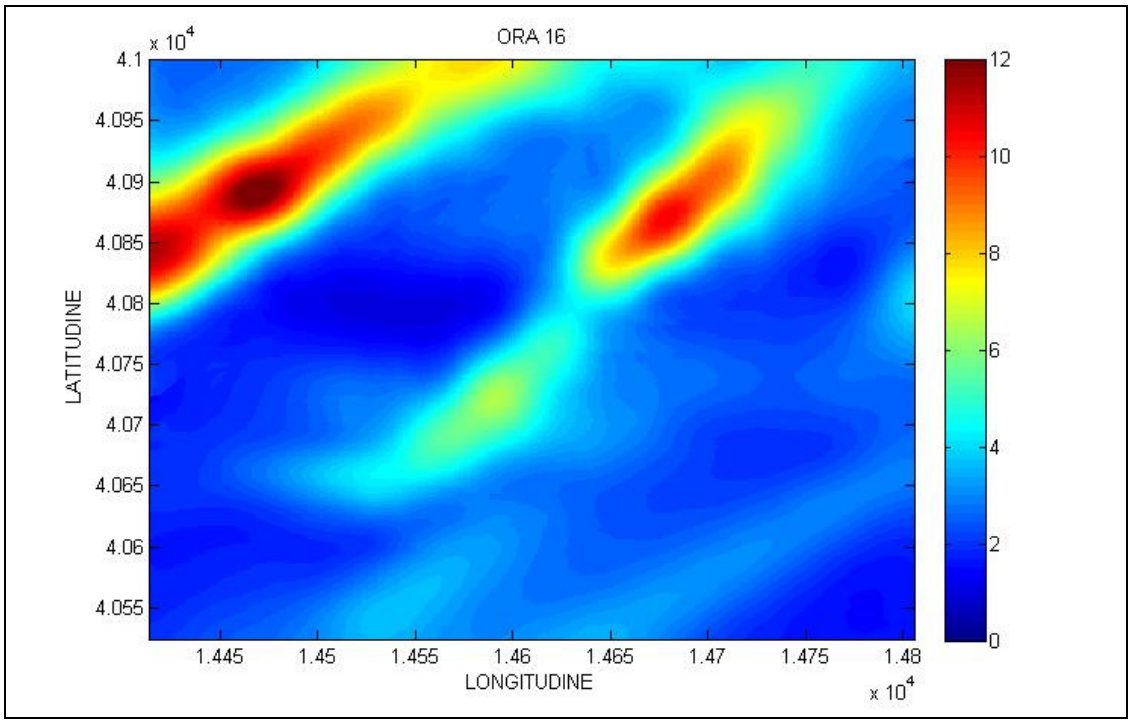


Figure 5 a. Interpolated surface using smooth *inverse multiquadric* RBFs

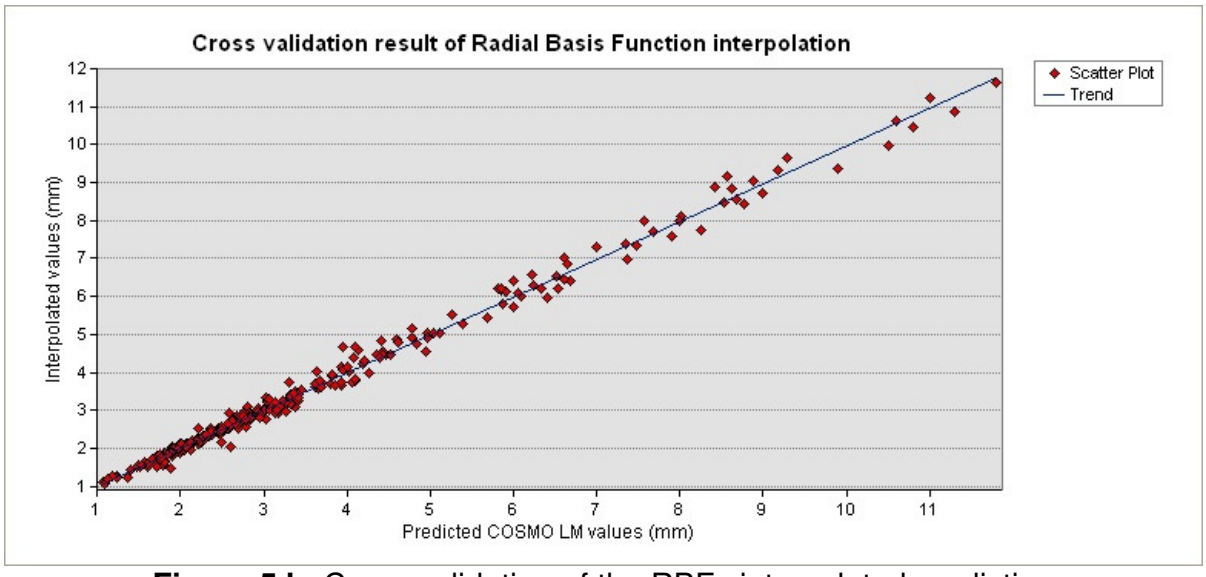


Figure 5 b. Cross-validation of the RBFs interpolated prediction

- an almost perfect correlation between the dependent and independent variables (r=0,99) due to the fact that RBF is an exact method of interpolation. The higher correlation of the RBFs, compared with another exact method IDW, might be caused by the higher smoothing factor.
- best prediction made in the area with small values of precipitation, poor prediction in the area of large and medium precipitation;

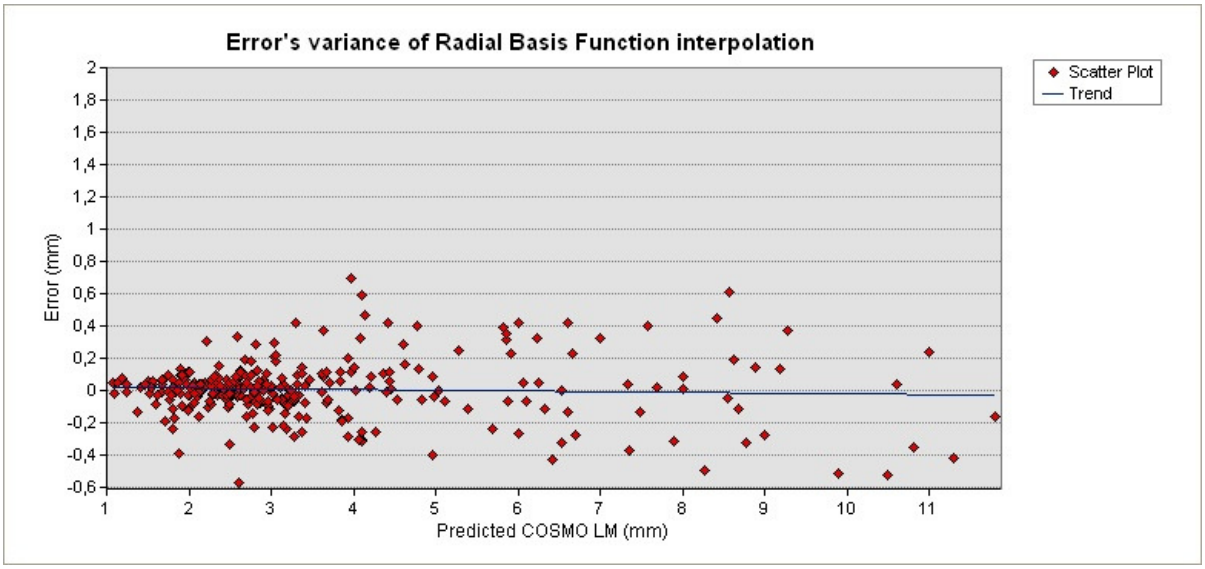


Figure 5 b. Residuals of the RBFs interpolated prediction

- a more accentuated underestimated prediction compared with the overestimation of the small precipitation, characteristic observed even in the area of largest precipitation;
- an overall slightly underestimate prediction with 155 underestimated values and 150 over estimated and maximum of 11,6 mm compared with the 11,8 mm of independent variable.

7.3 Kriging

If the other interpolation methods (inverse distance squared, splines, radial basis functions, triangulation, *etc.*) estimate the value at a given location as a weighted sum of data values at surrounding locations, Kriging assigns weights according to a (moderately) data-driven weighting function, rather than an arbitrary function, but it is still just an interpolation algorithm and will give very similar results to others in many cases (Isaaks and Srivastava, 1989).

Three kriging methods have been considerate for the analysis of the best geostatistic interpolation models that include autocorrelation — that is, the statistical relationships among the measured points.

Even in this case the cross-validation was used in order to select the most suitable of the methods. Also a statistical description was brought in, with the purpose of drawing a complete image of the efficiency of the methods (tab. 2).

Kriging	MPE	RMS	ASE	MS	RMSS	Std.dv	Mean.	Max.	Min.	Corel. Coef.(r)
Ordinary	0,001845	0,9897	1,583	0,000746	0,6261	1.9	3,56	10,2	1,12	0,66
Simple	***0,001118***	***0,4219***	***1,108***	***0,000620***	***0,3839***	**1.9**	**3,55**	***10,2***	***1,10***	0,92
Universal	0,004583	0,4037	0,427	0,007972	0,9516	1,97	3,56	10,6	1,03	0,95
Disjunctive	0,0119	0,4955	1,289	0,008148	0,3863	1,84	3,57	9,9	1,2	0,94
Cosmo LM output						2,05	3,55	11,8	1.07	

Table 2. Cross-validation and statistics of the predicted dependent variable

Cross-validation is used to determine "how good" the model is. The goal should be to have standardized mean prediction errors (MS) near 0, small root-mean-squared prediction errors (RMS), average standard error (ASE) near root-mean-squared prediction errors (RMS), and standardized root-mean-squared prediction errors (RMSS) near 1.

The selection of the optimal interpolated prediction, for each of the three kriging method, was done by comparing the parameters of the cross-validation characteristics in conformity with the description:

- a optimal prediction has to be unbiased (centered on the true values). If the prediction errors are unbiased, the mean prediction error (MPE) should be near zero.
- a optimal prediction has to have a valid assessment of uncertainty of the prediction standard errors (MS).

Each of the kriging methods gives the estimated ***prediction*** *kriging* ***standard errors***. Besides making predictions, it is estimated the ***variability of the predictions*** from the true values. It is important to get the correct variability. Examples of estimation of variability:

- If the average standard errors (ASE) are close to the root-mean-squared prediction errors (RMS), the variability in prediction It is correctly assessed.
- If the average standard errors (ASE) are greater than the root-mean-squared prediction errors (RMS), it is overestimated the variability of the predictions;
- if the average standard errors (ASE) are less than the root-mean-squared prediction errors (RMS), it is underestimating the variability in the predictions.

Another way to look at this is to divide each prediction error by its estimated prediction standard error. They should be similar, on average, so the root-mean-squared standardized errors should be close to 1 if the prediction standard errors are valid:

- If the root-mean-squared standardized errors (RMSS) are greater than 1, it is underestimated the variability in the predictions;

- if the root-mean-squared standardized errors (RMSS) are less than 1, it is overestimated the variability in the predictions.

In terms of fitted model, kriging was represented by the Spherical model which was chosen for all kriging methods. With this option, Kriging uses the mathematical function specified by the method to fit a line or curve to the semi-variance date in the semi-variogram. Spherical method seems to better fit the spatial variation of the data set compared with other methods: Circular, Gaussian, Exponential, Tetraspherical etc. The RMS was used to validate the best fitting model.

- **Ordinary Kriging**

  For ordinary kriging, rather than assuming that the mean is constant over the entire domain, it is assumed that it is constant in the local neighbourhood of each estimation point, that is that (m **u**) = m(**u**) a for each nearby data value, Z ($\mathbf{u}_{\alpha)}$, that we are using to estimate Z(**u**).

  The ordinary kriging prediction presents the following characteristics :

  - a level of the bias higher compared with the simple kriging's prediction but much better compared with the rest of the kriging predictions;
  - an overestimated level of variability in prediction (ASE > RMS) but closer to a correct assessing, compared with the simple kriging methods;
  - the prediction tends to under predict large values and over predict small values, as shown in the table 2 (max. 10,2 mm and min. 1, 12 mm);
  - a not so good correlation between the dependent and independent variables (r=0,66)
  - considerable extreme residual values settled between 1,2 mm for over prediction and - 2,3 mm for under prediction;
  - best prediction made in the area with small values of precipitation, poor prediction in the area of large and medium precipitation;

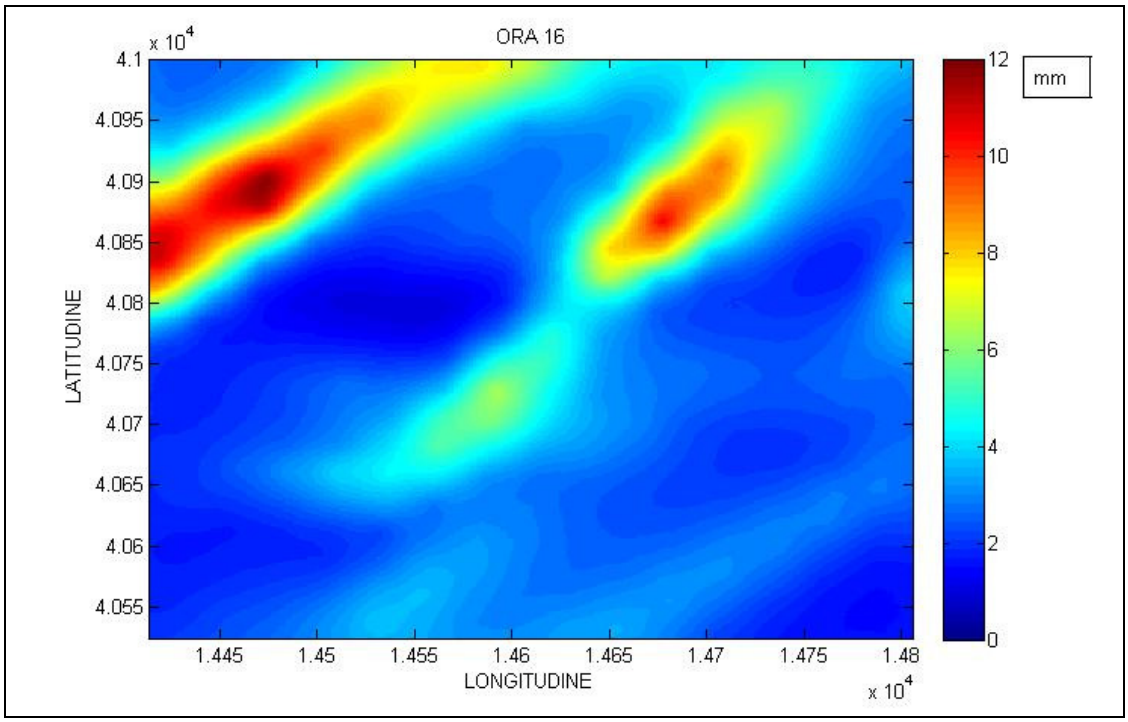


Figure 6 a. Interpolated surface using Ordinary Kriging

- a more accentuated underestimated prediction compared with the overestimation of the large precipitation;
- in the area of small precipitation the variability of the over estimated prediction is higher compared with the variability of the under estimated prediction;

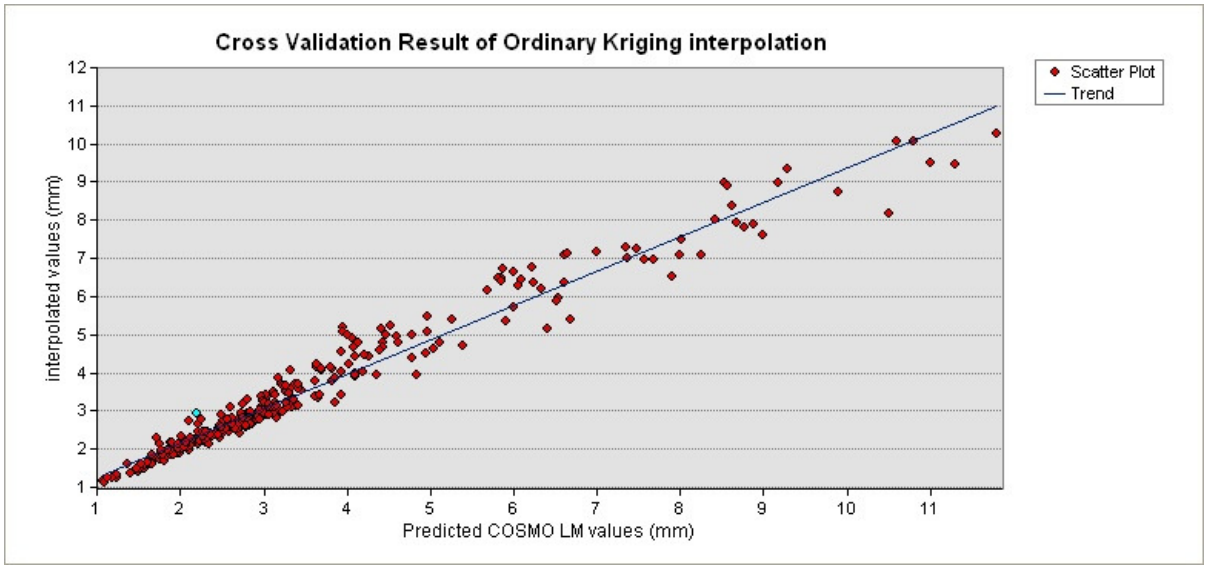


Figure 6 b. Cross-validation of the Ordinary kriging interpolated prediction

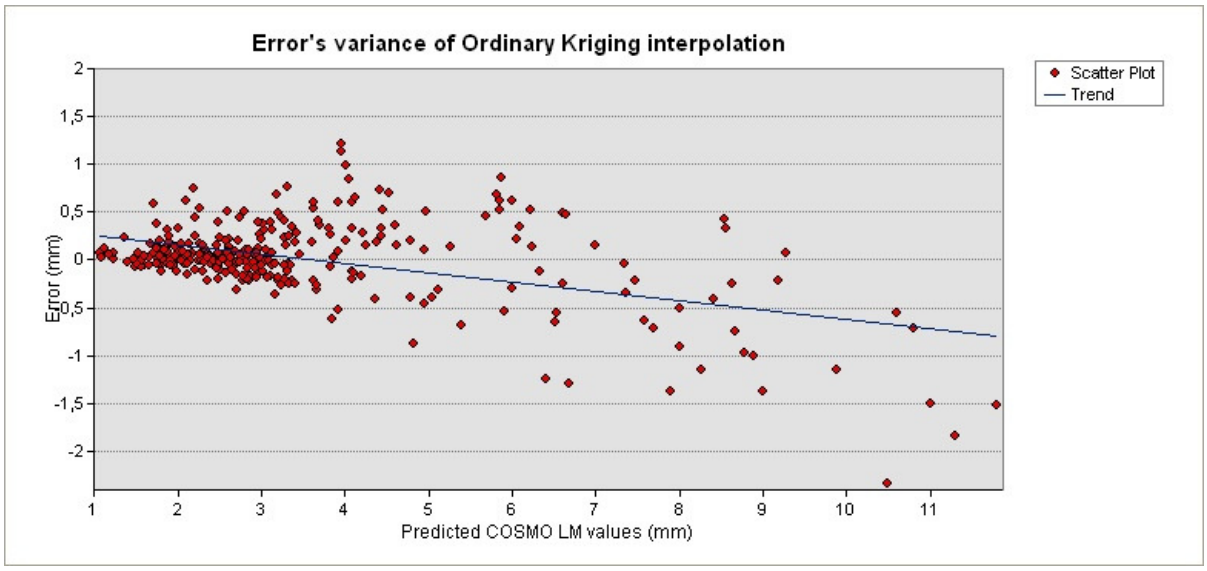


Figure 6 c. Residuals of the Ordinary kriging interpolated prediction

- **Simple kriging**

For simple kriging, it was assumed that the trend component is a constant and known mean, $m(u) = m$.

Characteristics (drawn from tabs. 2 and figs 7 a, b ,c) of prediction with simple kriging:

- Simple kriging prediction presents the lowest bias level, the MPE =0, 00118, and a standard deviation (1,9) close to that of the independent data ;

- the linear correlation coefficient r=0.92 confirms a very good overall agreement between independent and dependent variables;

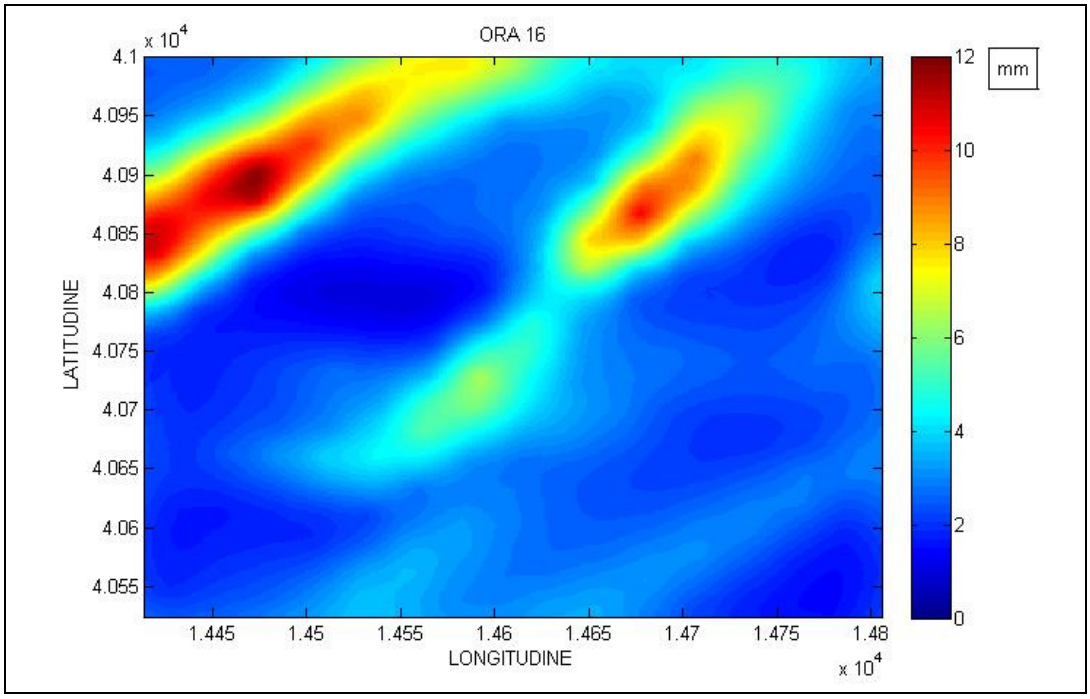


Figure 7 a. Interpolated surface using Simple Kriging

- poor assessment of the variability in prediction as the distance between the ASE and RMS is the highest ;
- the prediction tends to under predict large values and over predict small values, as shown in the table, with considerable extreme residual range (similar with the ordinary kriging residuals) ,values settled between 1,2 for over prediction and -2,3 for under prediction;
- simple kriging prediction presents an overall overestimated prediction with 166 overestimated values and 139 underestimated predicted values;

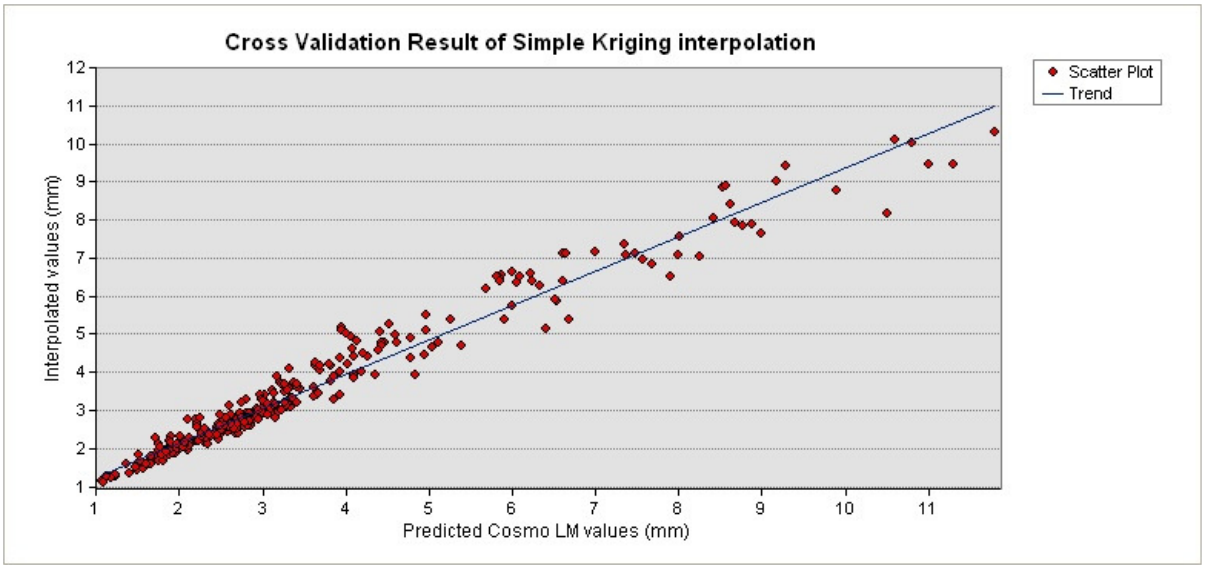


Figure 7 b. Cross-validation of the Simple kriging interpolated prediction

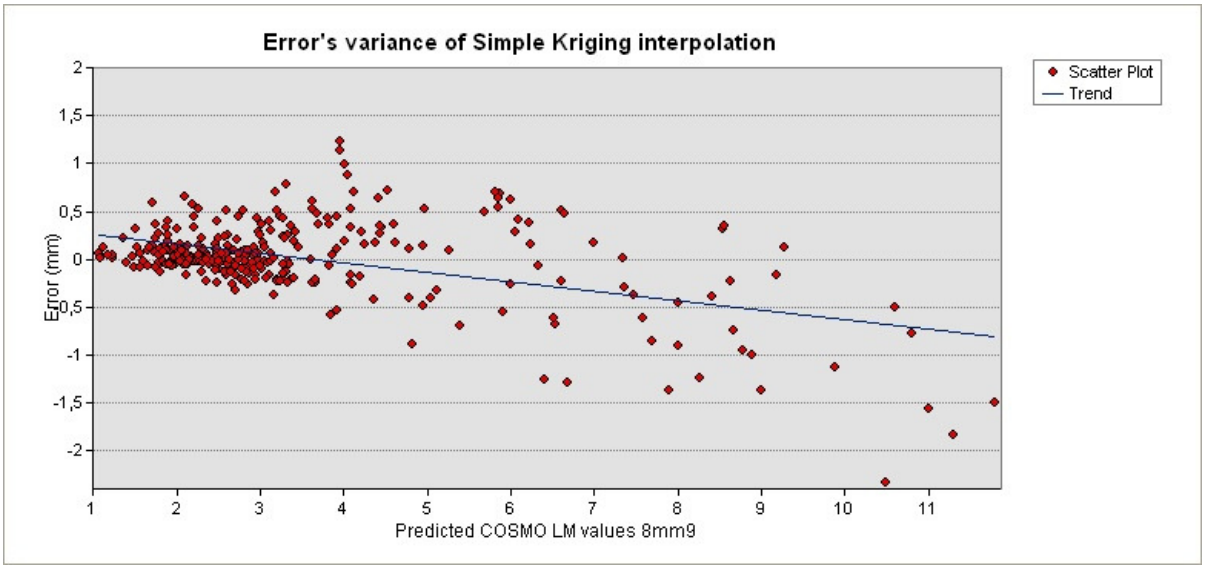


Figure 7 c. Residuals of the Simple Kriging interpolated prediction

➢ **Universal kriging** the method known as *kriging with a trend* is much like ordinary kriging, except that instead of fitting just a local mean in the neighbourhood of the estimation point, it fits a linear or higher-order trend in the (x,y) coordinates of the data points.

The cross validation of the Universal kriging prediction has underlined the following characteristics:

- in terms of bias, accuracy and variability of the prediction the Universal kriging used

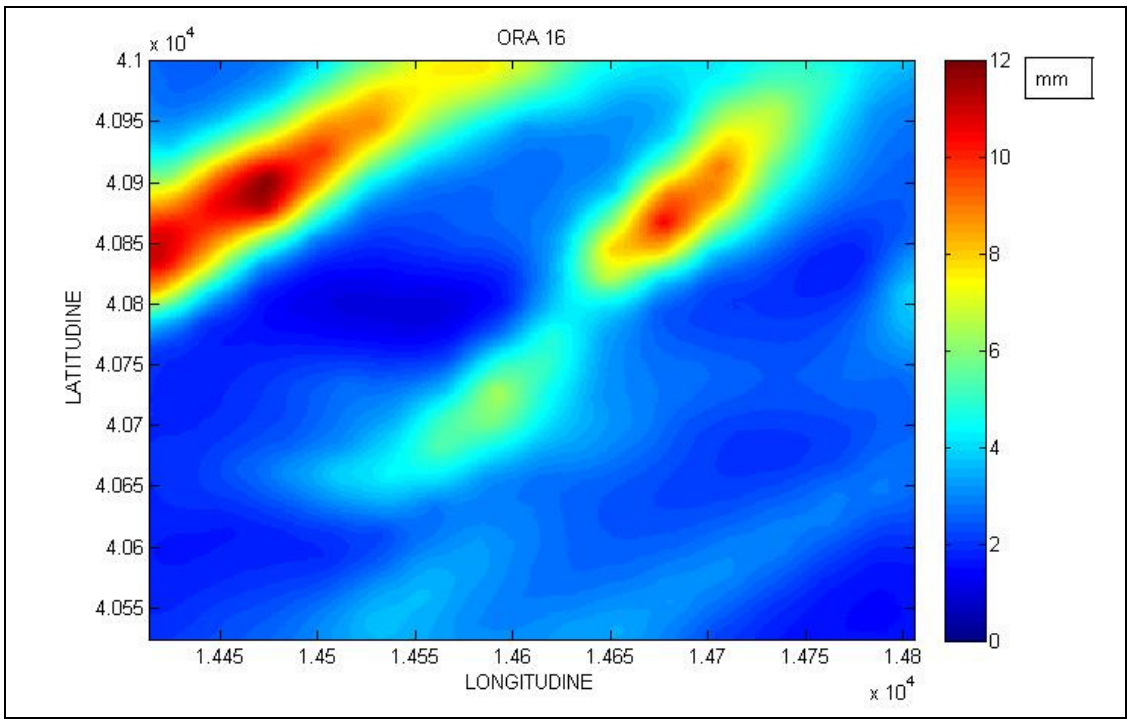


Figure 8 a. Interpolated surface using Universal Kriging

on the grided Cosmo LM data performs better than the ordinary kriging as it uses more the characteristics of the geometry of data;

- the relation of correlation between the dependent and independent variable is explained in a proportion of r= 0,95.

- Universal kriging presents an almost exact level of variability in prediction (ASE > RMS) compared with the other interpolation methods, being slightly overestimated ;

- the prediction tends to under predict large values and also the small one, as shown in the table 2 (max. 10,6 mm and min. 1, 03 mm);

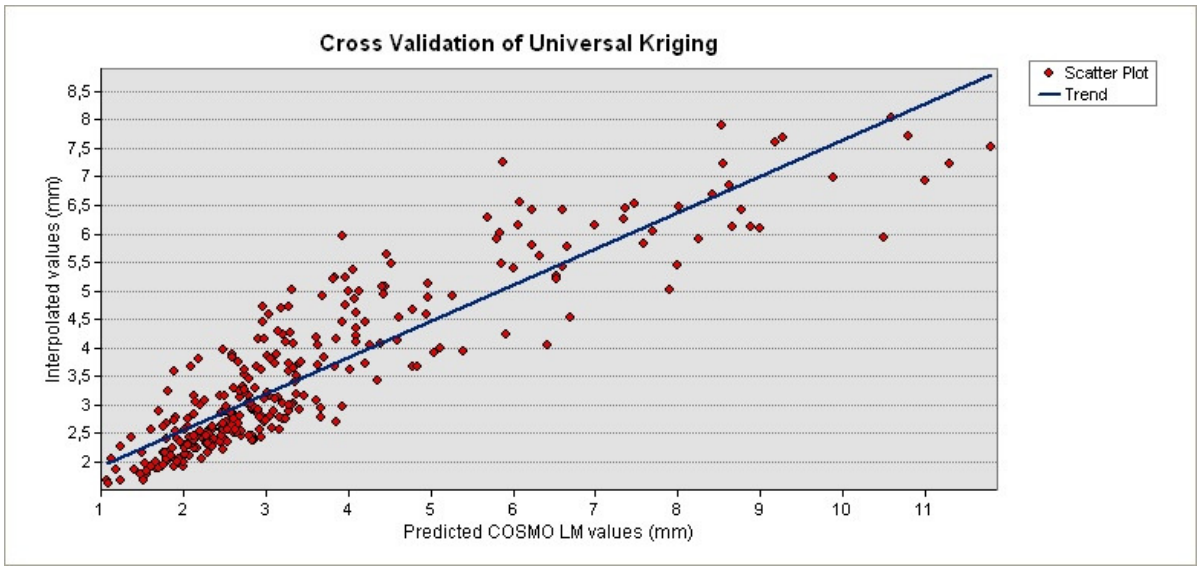


Figure 8 b. Cross-validation of the Universal kriging interpolated prediction

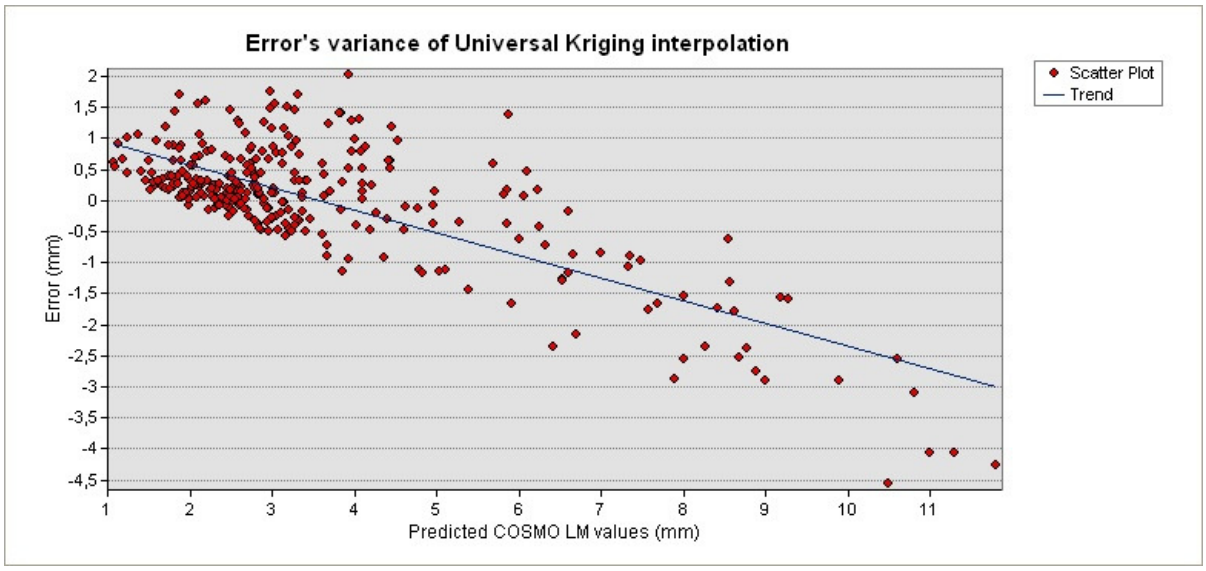


Figure 8 c. Residuals of the Universal Kriging interpolated prediction

- considerable extreme residual values settled between 1,3 mm for over prediction and - 2,1 mm for under prediction;

- universal kriging predicts better in the area with small values of precipitation and makes a more accurate (less variable) underestimation compared with the overestimation;

- in the area of small precipitation the variability of the over estimated prediction is higher compared with the variability of the under estimated prediction;

- **Disjunctive Kriging** produces a nonlinear unbiased, distribution-dependent estimator with the characteristics of minimum variance of errors (Burrough and McDonnell, 1998; Yates et al., 1986).

  Considering the functions of the data rather than using only the data the Disjunctive Kriging predicts, based on the characteristics of the Cosmo LM output:

  - a accurate prediction with a low bias (MPE = 0,0119) but a not so good accuracy of the variability wich is highly overestimated (ASE > RMS);

  - disjunctive kriging present also a low standard deviation 1,84 compared with the independent data (2,05) which shows that the prediction is not taken into consideration the actual data;

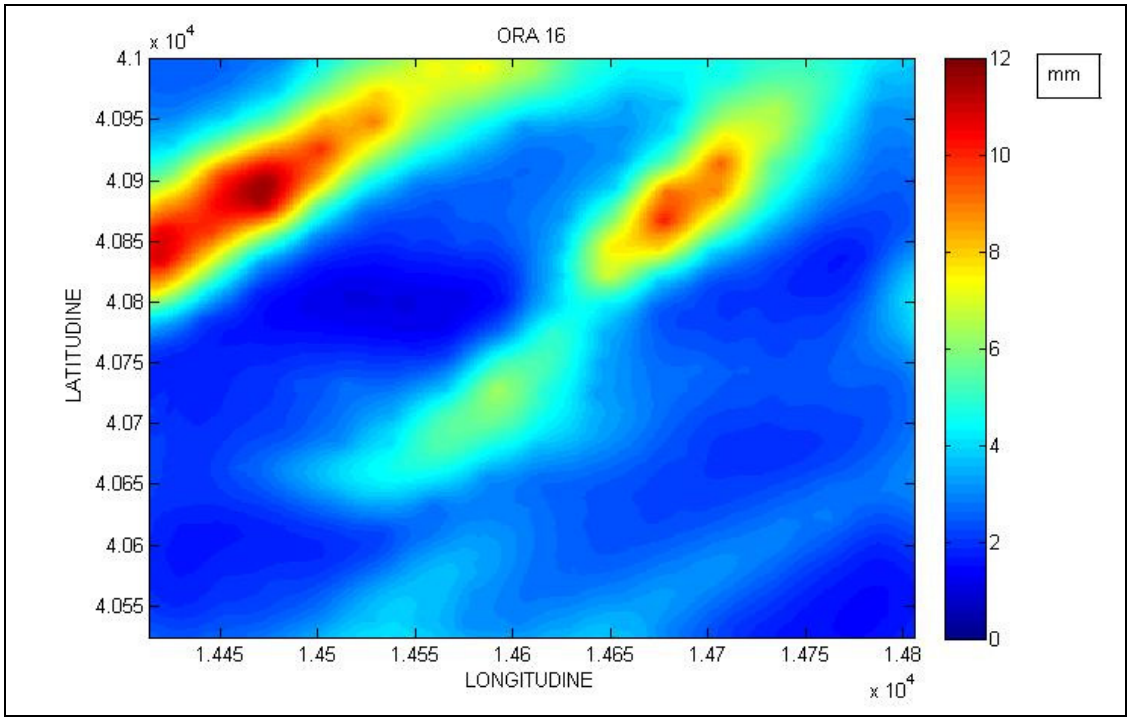


Figure 9 a. Interpolated surface using Disjunctive Kriging

- like all the interpolation method presented even DK under estimates the high and overestimates the low precipitations but in a more accentuated manner, also a characteristic of distribution-dependent estimator;

- creates considerable amplitude for residual values settled between 1,2 mm for over prediction and - 2,7 mm for under prediction;

- the prediction of DK shows a very high agreement with the independent data, the correlation coefficient being situated at 0,94;

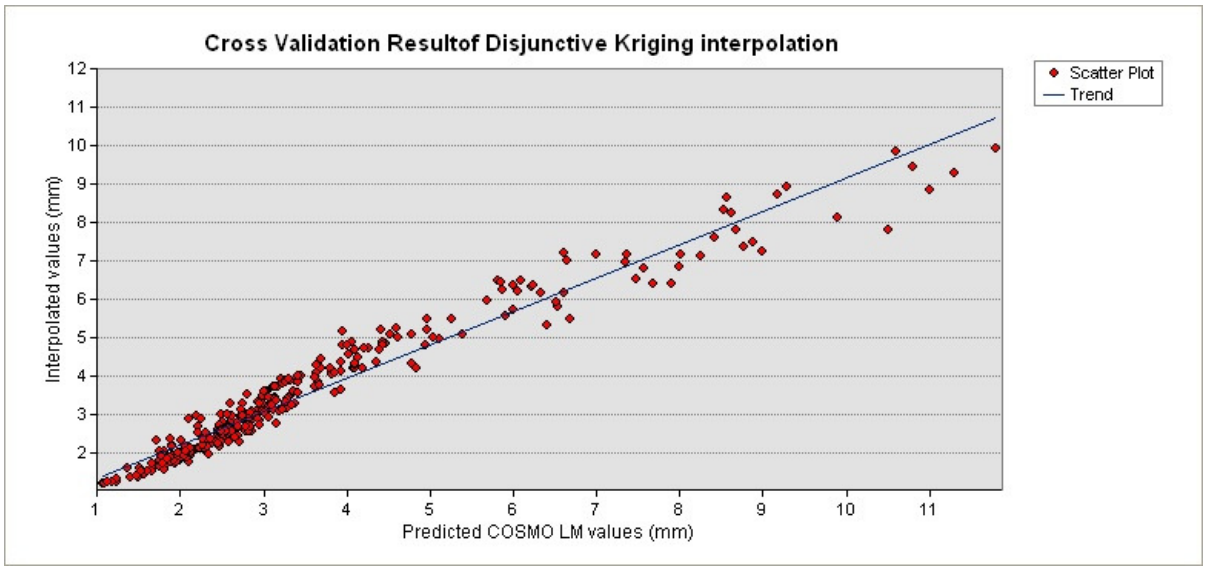


Figure 9 b. Cross-validation of the Disjunctive kriging interpolated prediction

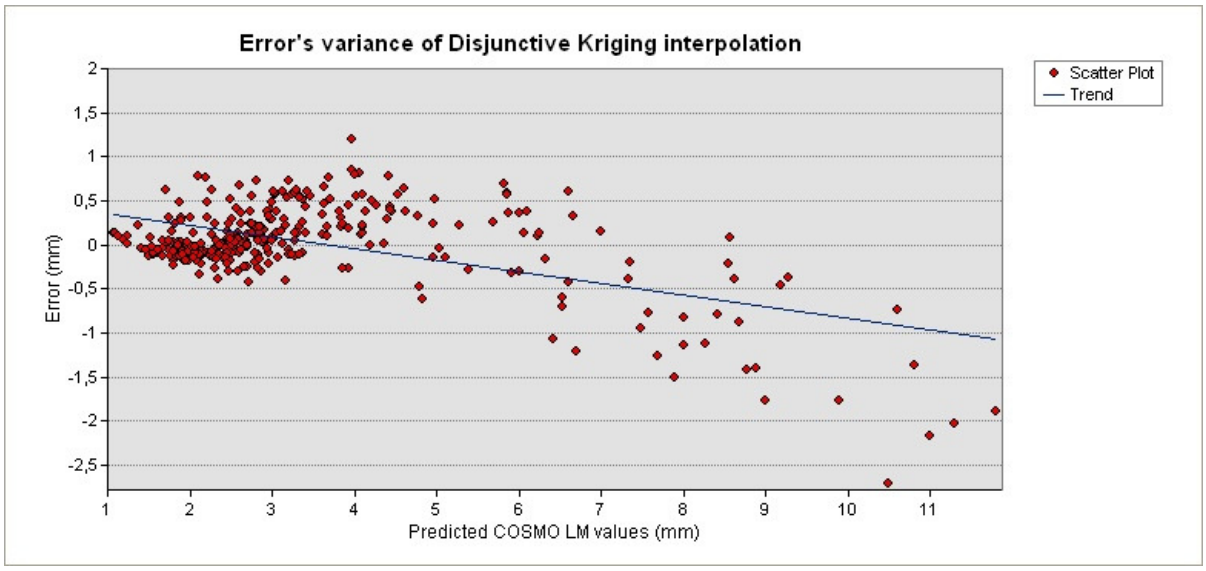


Figure 9 c. Residuals of the Disjunctive Kriging interpolated prediction

- DK predicts better in the area with small values of precipitation where it makes a more accurate (less variable) underestimation compared with the overestimation;

- In the area with high values of precipitation DK has a more accentuated tendency of under prediction.

- In the area with mean values of precipitation DK has the tendency of over predicting;

As a initial conclusion concerning the best prediction *among the kriging* models used in this study, we can say that the best prediction is obtained with simple kriging. On this assumption we have to take into consideration that the main factor that leaded our assumption towards this conclusion is the spatial characteristics of the data (geometry

of the distribution of the data, the distribution of the nearest-neighbour distance,). Many other interpolation studies that used kriging concluded different based on the different type of data characteristics:

- ''universal kriging results were encouraging and comparable with the subjectively obtained map'' (D. Kastelec, 2002), the study was concentrated on mean annual precipitation on 1 KM regular grid data;

- ''Indicator kriging gives better estimates than traditional kriging'' (X. Sun, M.J. Manton and E.E. Ebert, 2003), the study is combining rain gauge measurements with satellite infrared data ;

- ''ordinary kriging, the best approach to depict the unique variation within the data set'' (Julie Earls.2006); study is concentrated on data obtained on radar (NEXRAD);

The features that made the Simple kriging the best method of interpolation (within the stochastic methods) lays in the characteristics of the data both at the level of geometry and at the level of measurements and less on modelling the variability of the data as a function of separation distance.

From the studied bibliography (Oliver Schabenberger and Carol A. Gotway 2004) it was found that what brings these features in the interpolation computations as a decisive selection element is the nugget effect. It can be attributed to measurements errors or spatial source of variation at distance smaller than the sampling interval. In other words the nugget effect is simply the sum of the measurement error and microscale variation. Both the measurements error (defined as accuracy of the variability) and spatial source of variation (thinking at the regular grid) are at level of the Coasmo LM output minimal and thus it can be said that the nugget effect is also minimal.

This is mostly, important, if we consider a property of simply kriging which says that ***when a no nugget effect is encountered, then ˆU (si) =U(si). More clearly – without nugget effect, kriging interpolates, with nugget effect kriging it smooths. Interpolation means following the independent data (observed or predicted) closely*.**

Also It was noted , that a good predictor should be variable, in the sense that it follows the independent data closely. The simple kriging predictor has an this property.

Consider predicting at locations where data are actually independent, the predictor $p_{sk}(Z;s_0)$ becomes $p_{sk}(Z; [s_{1,....,}s_n]')$, and in (the simple kriging predictor equation):

$$p_{sk}(Z;s_0) = \lambda_0 + \lambda' Z(s) = \mu(s_0) + \sigma' \sum^{-1} (Z(s) - \mu(s)).$$

when replacing Cov $[Z(s_0),Z(s)]=\sigma'$ with $[Z(s),Z(s)]=\Sigma$ and $\mu(s_0)$ with $\mu(s)$ it is obtained

$$p_{sk}(Z,[s_1,\ldots,s_n]) = \mu(s) + \sum\sum^{-1} (Z(s) - \mu(s)) = Z(s).$$

Where:

z(s) is the predicted value at the unsampled location s_0.

λ are the distance-dependent weights

μ(s) mean of the random field

In this way, the simple kriging predictor when is interpolating is taking into consideration the independent data not only the variability of the data as a function of separation distance . '' It is an 'exact' interpolator''(Oliver Schabenberger and Carol A. Gotway 2004).

That is way, when cross-validation is performed, the prediction values of the simple kriging are most likely centered on the true value (low bias), the mean error ,the RMS and RMSS of the prediction have the smallest values, the correlation between the dependent and independent variables is most explained (92%), the residuals have a small fluctuation range but on the other side the variability of the prediction is less correctly assessed (simple Kriging is the method with the highest overestimated variability ASE > RMSPE).

7.4 Polynomial interpolation

- ***Local Polynomial*** interpolation is a sensitive to the neighbourhood distance that fits the specified order (zero, first, second, third, and so on) polynomial using all points

only within the defined neighbourhood. The neighbourhoods overlap and the value used for each prediction is the value of the fitted polynomial at the center of the neighbourhood. (ArcGis Desktop help).

Other theoretical characteristics are:

- Creates a surface from many different polynomial formulas.
- Each is optimized for a specified neighborhood.
- The neighborhood shape, maximum and minimum number of points, and a sector configuration can be specified.
- The sample points in a neighbourhood can be weighted by their distance from the prediction location.
- Local Polynomial interpolation maps can capture the short-range variation

Characteristics of the Local polynomial prediction underlined by the cross-validation method (Fig. 9 a, b, c):

- a smooth surface prediction (fig.9 a) due to the fitting of the interpolation curve trough the measured sample;
- a good coefficient gradient between the dependent and independent variables (r=0,83);
- a clear tendency of underestimation of the prediction MPE (-0,0056) and a low degree of bias;
- the prediction tends to under predict large values (max. of prediction 9,6mm) and over predict small values (min of prediction 1,3 mm);
- presents a high fluctuation of the residuals situated between -3,2 mm for under prediction and 1,28 mm for over prediction;
- a higher variability of the over predicted residuals along the trend line in the area of low precipitation.

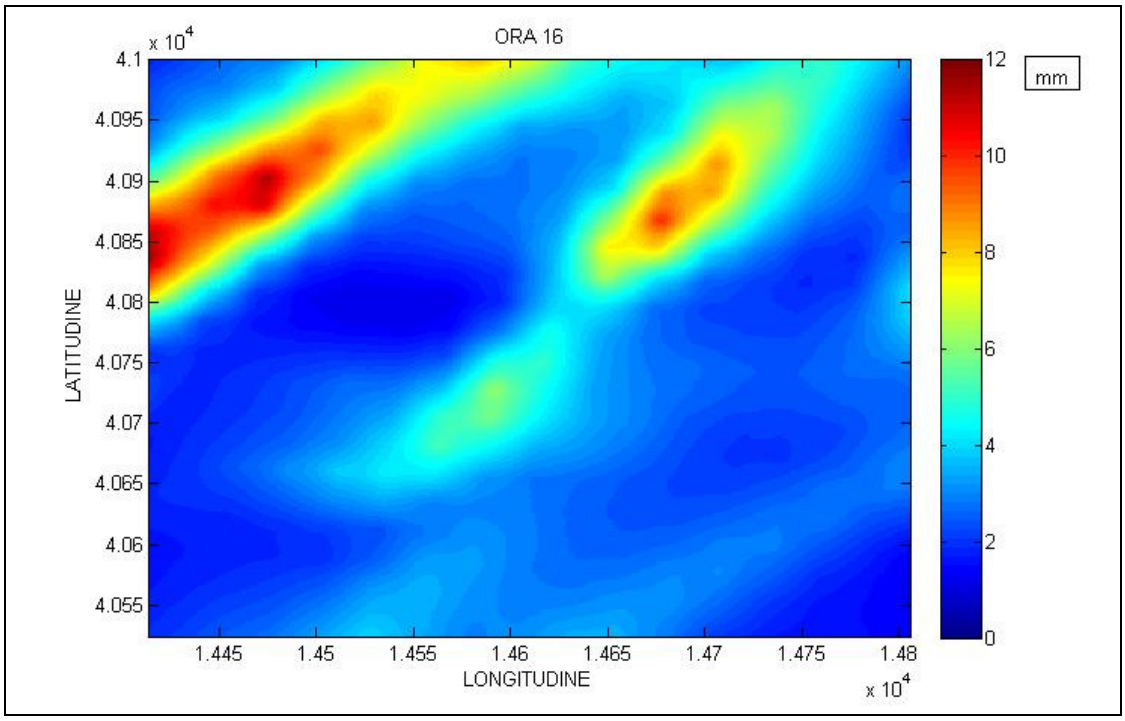


Figure 10 a. Interpolated surface using Local Polynomial interpolation

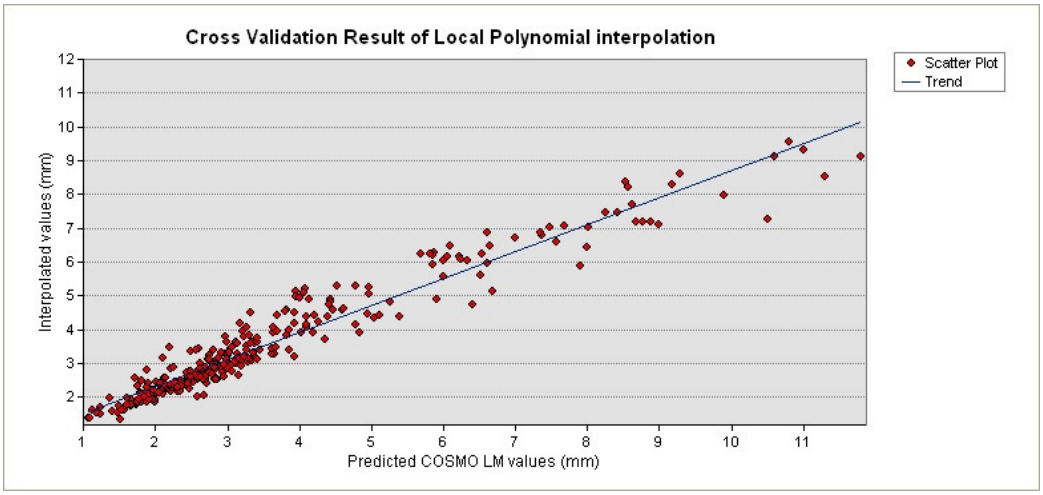


Figure 10 b. Cross-validation of the Local polynomial prediction

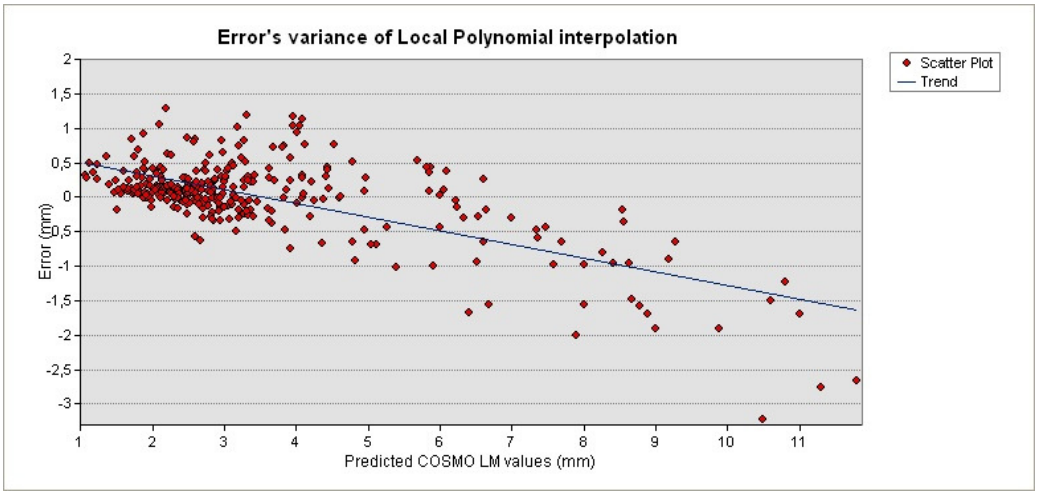


Figure 10 c. Residuals of the Local polynomial interpolated prediction

- ***Global Polynomial*** interpolation fits a smooth surface that is defined by a mathematical function (a polynomial) to the input sample points. The Global Polynomial surface changes gradually and captures coarse-scale pattern in the data. (ArcGis Desktop help)

  This method generates a smooth surface that does not have to fit to measured points and does not use a search neighbourhood. A polynomial is used to fit the surface to the data, so a first order polynomial would have no bends in it, a second order would have one bend, etc. This method is best used for data that varies slowly over a landscape or for looking at general trends. It is sensitive to outliers, especially at the edge of the area of interest.

  ArcGis 9,2 offers 10 different order polynomials that can be used from the Global Polynomial interpolation Set parameters dialog box.

  The selected global polynomial of different order brings changes in the final prediction. That is way the selection of the optimal order of interpolation, was done by optimization of the parameters value until the smallest RMS was obtained for each of the polynomial orders (table 3).

Global Polynomial interpolation (order)	1st	2nd	3rd	4th	5th	6th	7th	8th	**9th**	10th
MPE	-0,00051	0,001618	0,004909	-0,00016	0,00163	-0,00867	-0,02183	0,01316	**0,01244**	0,01712
RMS	1,815	1,795	1,778	1,706	1,587	1,435	1,495	1,111	**0,915**	0,953

Table 3. Mean error and root-mean-error of the 10 different order global polynomial prediction

Comparing the RMS of the polynomials with different order has been decided that for a regular grid data, the best prediction is offered by the 9^{th} order-polynomial.

Characteristics of this global polynomial prediction:

- creates smooth surfaces and identifying long-range trends in the dataset, therefore the prediction fails to represent the short range variability of the phenomenon.

- being sensitive to outliers, the edges of the predicted area are generating errors.

- presents a high fluctuation of the residuals situated between -4,1 mm for under prediction and 5,4 mm for over prediction;

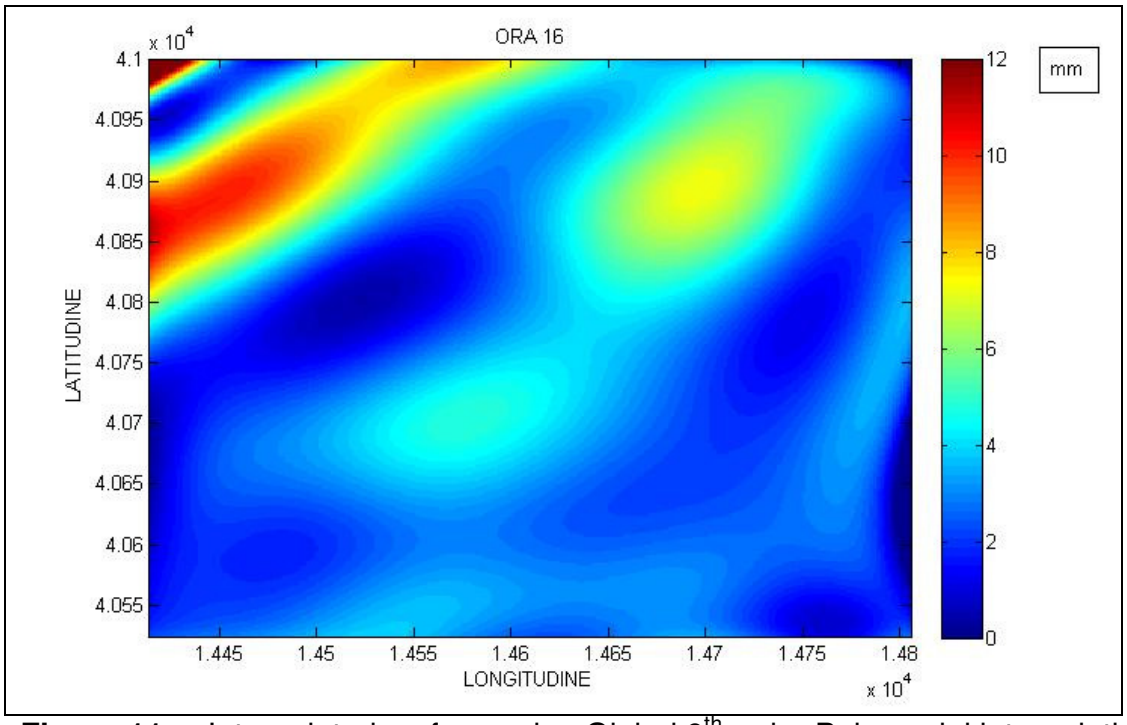


Figure 11 a. Interpolated surface using Global 9th order Polynomial interpolation

- presents a low accuracy of the variability of the prediction;

- also presents a good correlation (r=0,86) of the dependent and independent variables;

- the over and under prediction has a constant variability no matter of the small or large independent variable (precipitation);

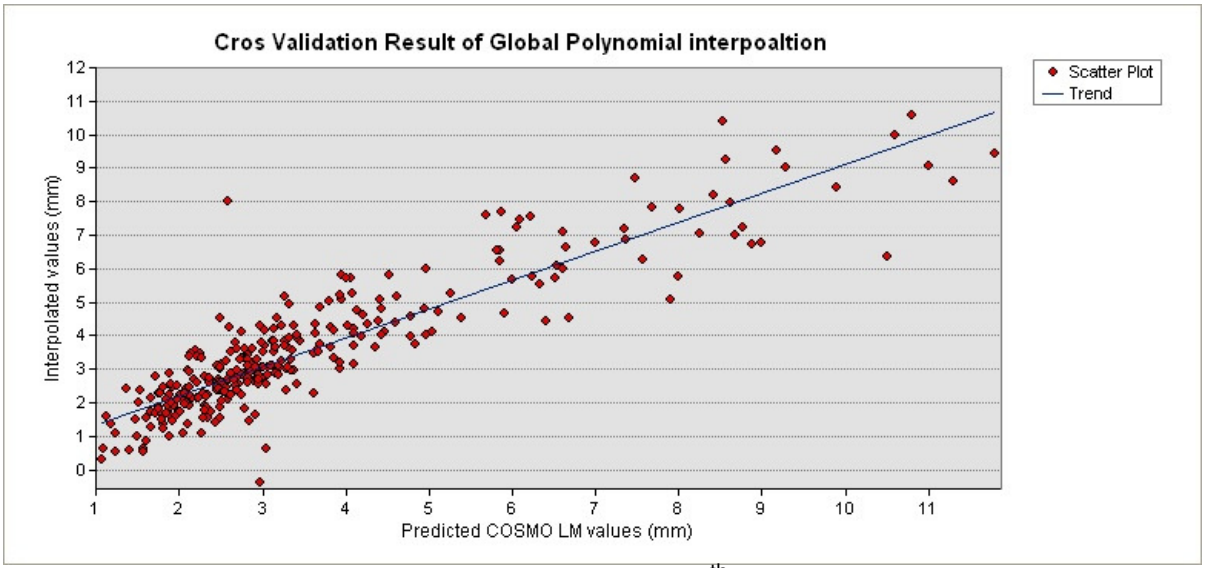


Figure 11 b. Cross-validation of the Global 9th order polynomial prediction

- the prediction tends to under predict not only the large values (max. of prediction 10,5mm) but even the small values (min of prediction -0.4 mm);

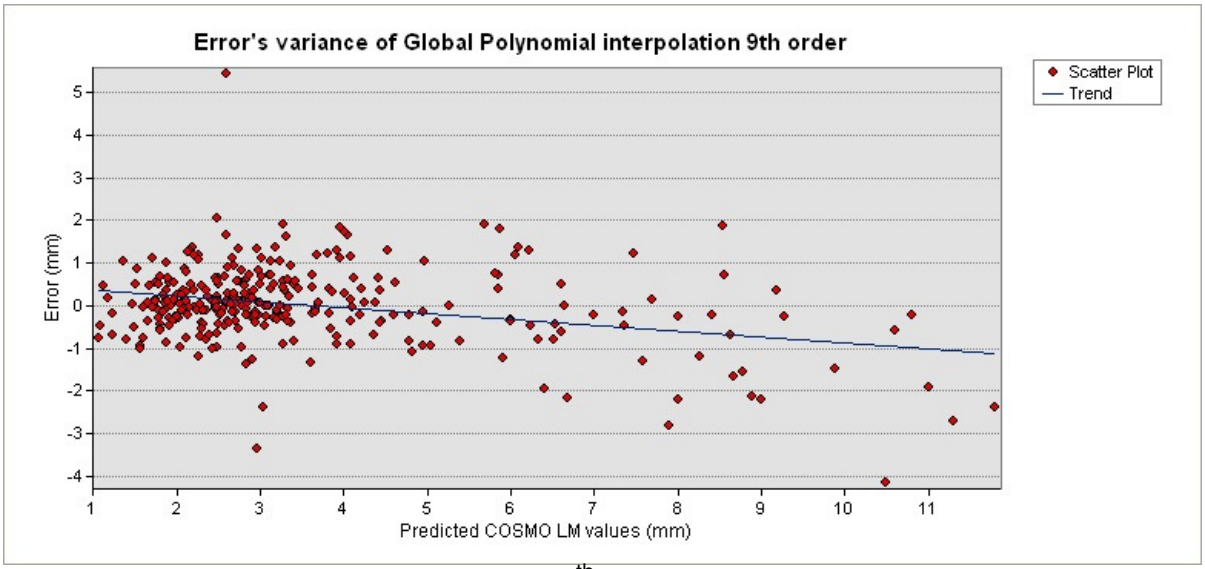


Figure 11 c. Residuals of the Global 9th order Polynomial interpolated prediction

8 Discussion

All interpolation methods create similar evolution and gradual increases and decreasing of the predicted area along the high and low independent values. The global similarities among methods were correlated to the similarity of generated statistics for each method. On a smaller scale, some qualitative variation did exist across all interpolation methods (table 4).

Interpolation method	MPE	RMS	Corel. coef.(r)	Std.dv	Mean.	Max.	Min.
IDW	0,01413	0,8272	0,72	1,5	3,57	8,6	1,4
RBF	**0,000356**	**0,1501**	**0,99**	2,06	**3,55**	**11,6**	**1,04**
Local Polynomial	-0,00565	0,595	0,82	1,7	3,55	9,5	1,3
Global Polynomial	0,01244	0,915	0,86	1,98	3,57	10,6	-0.3
Ordinary Kriging	0,001845	0,9897	0,66	1.9	3,56	10,2	1,12
Simple Kriging	0,001118	0,4219	0,92	1.4	3,55	10,2	1,10
Universal Kriging	0,004583	0,4037	0,95	1,97	3,56	10,6	1,03
Disjunctive kriging	0,0119	0,4955	0,94	1,8	3,57	9,9	1,2

Table 4. Statistical properties of the predictions.

The IDW surface model, based on the extent of smoothing, produced “bulls-eye” patterns, especially along higher values areas. Thus, even though IDW produced respectable statistics for both cross validation and validation, it was not considered a suitable final surface model choice.

The RBF methods, based on the degree of smoothing, are also affected by errors of prediction and under prediction and over prediction. RBF methods typically produce high

error or uncertainty in areas where values abruptly changes due to the rubber sheeting applied to the data but overall, RBFs, statistically, produced the highest quality prediction statistics compare with all other interpolation methods.

Overall the trend based methods (kriging) have had a good concentration of the predicted values along the centered values which made very smooth prediction maps.

Although, smoother surface models are more visually aesthetic, a smoother surface understates the variability and may be misleading from a qualitative point of view (Isaaks and Srivastava, 1989).

As a comparison, all kriging methods produced similar results (more than other interpolation models), both in prediction statistics and of output map features. Universal kriging compared with ordinary kriging appeared to contain less bias and lower RMS values in cross validation and validation statistics, thus the more qualified surface modeller.

In choosing a suitable optimal stochastic model for modelling a final surface from a Cosmo LM output, simple kriging did stand out when all aspects (cross-validation, validation, and output product) were examined.

All methods, statistically (RMS, MPE, and where applicable, RMS standardized) produced a optimal quality output surface. RMS values (a measure of the paired relationship between observed and measured values) indicated greater accuracy with lower RMS values, which varied from 0,15 with RBFs inverse multiquadric to 0.98 with Ordinary kriging, RMS values for the other interpolation methods used here fell between these RMS values for OK and RBFs methods.

Another measure of data variability is the mean prediction error (MPE), which is a measurement of data bias within the prediction surface model.

Since the MPE value is dependent on the scale of the data, kriging models provide a good representative of the MPE and the standardized MPE (the MPE divided by the kriging standard error (i.e. the square root of the variance prediction)) but still the RBFs performed better.

Examination of Pearson correlation coefficient and standard deviation values were useful enough for determining between interpolation prediction models. All interpolation methods displayed strong positive correlation between observed and predicted values, from a maximum of correlation by the RBFs to a minimal correlation by Ordinary kriging.

9 Conclusions

The comparison proposed in this study has allowed identifying the most accurate of 8 different interpolation methods used to derive areal precipitation from a COSMO LM output.

The spatial interpolation of daily rainfall output of the Cosmo LM was studied using the following interpolation methods: IDW, Radial Basis Functions and Ordinary kriging, Simple Kriging, Universal Kriging, Disjunctive kriging. The performance of each method was evaluated by cross-validation. There were no significant differences on the interpolations used but Radial Basis Function (Inverse Multiquadric) was the better efficient interpolation method of all, with the lowest errors (Tab.4) and with highest capability to accurately reproduce variability of the independent data.

Due to the high variability, normally associated with hourly precipitation records and the high density of the network, it was clear that techniques such as trend induced prediction (Ordinary kriging, Universal Kriging, Simple Kriging, Disjunctive kriging) would provide good estimates, but not the best. Trend surfaces are always smooth surfaces which do not normally pass through the original data points but performs a best fit for the entire surface. In other words it provides an approximate direction of the intensity of rain rather than an accurate description of the spatial variability of rain. On the other hand, surfaces generated using exact methods try to pass through the points which, in the case of hourly, complex with high variability precipitation, is not suitable because of the rapid changes in gradient/ slope in the vicinity of the data points with high values, that creates rain cells .

However, compared with the trend induced surfaces, RBFs is a more relaxed version of smooth fitting, that means a method that can fit a less smooth curves. Practically is a ”shared’’ method taking into consideration smoothness - allowing the prediction of points above the maximum measured value and below the minimum measured value - and in the same time is an exact method. All this associated with a regular grid data - with a not so high variability of the values of precipitation (the output of Cosmo LM is predicted) - has allowed a good prediction with a small bias and a good accuracy of reproducing the variability of the independent data.

From the polynomial methods the local method produced better prediction with smallest errors and a better accuracy of representing the variability, even if the underproduction of high and over prediction of the low is much more obvious.

10 Bibliography

1. Ahers, B., 2006. *Distance in spatial interpolation of daily rain gauge data*. Hydrol. Earth Syst. Sci. (10)197-208.
2. Armstrong, M. and Matheron, G., 1986a. Disjunctive kriging revisited: Part I.Mathematical Geology, 18(8): 711-728.
3. Armstrong, M. and Matheron, G., 1986b. Disjunctive kriging revisited: Part II. Mathematical Geology, 18(8): 729-741.
4. Burrough, P.A., and R.A. McDonnell. 1998. Principles of Geographical Information Systems. New York: Oxford University Press.
5. Bohling, G, 2005, Kriging, Kansas Geological Survey C&PE 940,
6. Davis, B. M. (1987) Uses and abuses of Cross-validation in geostatistics. Mathematical Geology, 19(3); 241-248.
7. Clark, I. and Harper, W.V., 2001. Practical Geostatistics 2000. Geostokos (Ecosse) Limited, 342 pp.
8. I.Dyras, H.Dobesch, E.Grueter, A.Perdigao, O.E.Tveito, J.E.Thornes, F.v.d. Wel, L.Bottai, The Use of Geographic Information Systems in Climatology and Meteorology, COST719, Geographical Information Systems, Interdisciplinary Aspects, Sosnowiec-Zagreb 2003, 99-107
9. Eduardo Zorita and Hans von Storch (1999)- The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods, Journal of Climate Article: pp. 2474–2489 ;
10. Fiorucci, P., La Barbera, P., Lanza, L.G. and Minciardi, R., 2001. A geostatistical approach to multisensor rain field reconstruction and downscaling, Hydrol. Earth. System Sci., 5, 201–213.
11. Hevesi, J., Flint, A. & Istok, J. (1992) Rainfall estimation in mountainous terrain using multivariate geostatistics. *J. Appl. Met*. **31**, 661–688.
12. Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.
13. Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 483 pp.
14. Julie Earls, 2006, - Spatial interpolation off rainfall data using ArcGis: A comparative study ,Geo-Spatial, Analytics,Lab
15. Johnston, K., J. M. Ver-Hoef, K. Krivoruchko, and N. Lucas. 2001. Using ArcGIS Geostatistical Analyst. ESRI. Redlands, 300pp.
16. D. Kastelec, T. Vrhovec, K. Košmelj, L. Kajfež-Bogataj: Objektivna prostorska interpolacija meteoroloških spremenljivk in njihovo kartiranje: disertacija; Objective spatial interpolation of meteorological variables and their mapping: Disertation;
17. D. Kastelec, T. Vrhovec: The spatial interpolation of the daily precipitations in the mountainous regions in Slovenia: presented at 26th International Conference on Alpine Meteorology, Innsbruck
18. D . Kastelec: The spatial interpolation of the daily precipitations in the mountainous regions in Slovenia: Geophysical research abstracts. – ISSN 1029-7006. - #Vol. #2 (2000), CD-ROM, OA27.
19. D. Kastelec, M. Dolinar: Spatial interpolation of precipitation in mountainous region of Slovenia with Kriging methods: Book of abstracts / MAP meeting 2000,Bohinjska Bistrica, 24. - 26. 5. 2000. - [Ljubljana : Faculty of Mathematics and Physics, 2000]. p.54.
20. D.,Kastelec, Košmelj K., 2002, Spatial Interpolation of Mean Yearly Precipitation using Universal Kriging Metodološki zvezki, 17, Ljubljana: FDV,

21. D. Kastelec, K. Košmelj: Comparison of different Kriging methods: a case study on daily precipitation, The XXth International Biometric Conference, July 1-7, 2000, Berkeley. - [Berkeley : University of California, 2000]. - Vol. 1, p. 77.
22. I. Hanssen-Bauer, E. J Førland, J.E. Haugen og O.E. Tveito (2003) Temperature and precipitation scenarios for Norway: Comparison of results from dynamical and empirical downscaling, met.no Rapport 06/03 Klima ;
23. I. Hanssen-Bauer, E. J Førland, J.E. Haugen og O.E. Tveito (2003): "Temperature and precipitation scenarios for Norway. Comparison of results from dynamical and empirical downscaling", Climate Research, vol 25, 15-27
24. Hans von Storch, 2004 - Statistical Downscaling Central Weather Bureau, Taipei, Taiwan;
25. Hanssen-Bauer,I., O.E. Tveito and E.J.Førland,(2002) Local precipitation and temperature scenarios for Norway, Proceedings of Nordic Hydrological Conference 2002, Røros, Norway 4-7.August 2002, NHP-report 47, pp. 641-648.
26. Hanssen-Bauer,I., O.E.Tveito and E.J.Førland (2001) Precipitation scenarios for Norway: Empirical Downscaling from the ECHAM/OPYC3 GSDIO integration (RegClim), DNMI Report 10/01 KLIMA
27. Dobesch,H., O.E.Tveito, P.Bessemoulin (2001) Geographical information systems in climatological application – Final report of project no. 5 in the framework of the climatological projects in the applications area of ECSN, DNMI Report 13/01 KLIMA;
28. G. Skok, T. Vrhovec: Interpolation considerations when comparing 24h precipitation derived from a model and measured by a raingauge network, Booklet of abstracts / II. Voltaire Workshop,Ljubljana, 6. - 8. Oct. 2004. - Ljubljana : Faculty of Mathematics and Physics, 2004. p. 22.
29. Shelly Eberly, Jenise Swall, David Holland, Bill Cox, Ellen Baldridge (2004) Developing Spatially Interpolated Surfaces and Estimating Uncertainty, Office of Air Quality Planning and Standards, Publication No. EPA-454/R-04-004
30. Smith, D. R., E. Pumphry, and J. T. Snow (1986), A comparison of errors in objectively analyzed fields for uniform and nonuniform station distributions, J. Atmos. Oceanic Technol., 3, 84 -97;
31. SILVA, Álvaro; SOUSA, A. Jorge and ESPÍRITO SANTO, Fátima (2004): “Mean air temperature estimation in Mainland Portugal: Test and comparison of spatial interpolation methods in GIS.” waiting for publication in Proceedings of Conference on the Spatial Interpolation Techniques in the Climatology and Meteorology, Budapest, Hungary, 25 a 28 de Outubro de 2004. 9p.
32. SILVA, Álvaro (2003): Neural Networks application to spatial interpolation of climate variables report of STSM – Short Term Scientific Mission in ZAMG - Vienna, Austria.
33. Tveito, O.E. and Wolfgang Schöner (eds) (2002) Applications of spatial interpolation of climatological and meteorological elements by the use of geographical information systems (GIS) (Cost 719 Report no 1) Met.no Report 28/02 Klima
34. Tveito, O.E. (2002) Gridding observed precipitation for verification of ECMWF-forecasts, met.no report 12/02 Klima
35. Tveito,O.E (2002) An objective comparison of observed and forecasted 24-hour precipitation, a spatial analysis, met.no Report 10/02 KLIMA
36. Todini, E., 2001a. A Bayesian technique for conditioning radar precipitation estimates to rain gauge measurements, Hydrol. Earth. System Sci., 5, 187–199.
37. Morrisey, M. L., J. A. Maliekal, J. S. Green, and J. Wang (1995), The uncertainty of simple spatial averages using rain gauges networks, Water Resour. Res. , 31, 2011-2017

38. Palutikof JP, Wigley TML (1996) Developing climate change scenarios for the Mediterranean region. In: Jeftic L, Keckes S, Pernetta JC (eds) Climate change and the Mediterranean, Vol 2. Arnold, London, p 27–56
39. Garcia, M., C. D. Peters-Lidard, and D. C. Goodrich (2008), Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States, Water Resour. Res., 44, W05S13, doi:10.1029/2006WR005788).
40. Lorenz, E.N., 1966, Nonlinearity, Weather Prediction and climate deduction; Final Report, Statistical forecasting project, 22 pg;
41. L.G. Lanza, J.A. Ramírez and E. Todini, Stochastic rainfall interpolation and downscaling, Hydrology and Earth System Sciences, 5(2), *139–143* (2001);
42. Moyeed, R.A. and Papritz, A., 2002. An empirical comparison of kriging methods for nonlinear spatial point prediction. Mathematical Geology, 34(4): 365-386.
43. Oliver, M.A., Webster, R. and McGrath, S.P., 1996. Disjunctive kriging for environmental management. Environmetrics, 7: 333-358.
44. Oliver Schabenberger, Carol A. Gotway, 2005 Statistical Methods for Spatial Data Analysis, CRC Press, 488pg''
45. Papritz, A. and Moyeed, R.A., 1999. Linear and non-linear kriging methods: tools for monitoring soil pollution. In: V. Barnett, A. Stein and K.F. Turkman (Editors), Statistical Aspects of Health and the Environment. Statistics for the Environment. John Wiley & Sons, Chichester, pp. 303-336.
46. Rendu, J.-M., 1980. Disjunctive kriging: Comparison of theory with actual results. Mathematical Geology, 12(4): 305-320.
47. Song, J. and J. V. DePinto (1995) A GIS-based Data Query System. Presented at the International Association for Great Lakes Research (IAGLR) Conference, Windsor, Ontario.
48. von Storch H, Zorita E, Cubash U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6: 1161–71
49. Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications. Springer, Berlin, 387 pp.
50. Yan Yu, Deepak Ganesan, Lewis Girod, Deborah Estrin, Ramesh Govindan (2003) – ''Synthetic Data Generation to Support Irregular Sampling in Sensor Networks''
51. Yates, S.R., Warrick, A.W. and Myers, D.E., 1986. Disjunctive kriging: 1. Overview of estimation and conditional probability Water Resources Research, 22(5): 615-621.
52. X. Sun, M.J. Manton and E.E. Ebert, 2003, Regional rainfall estimation using double-kriging of raingauge and satellite observations, BMRC Research Report no. 94,