

## RANKING SPATIAL INTERPOLATION TECHNIQUES USING A GIS-BASED DSS

S. NAOUM<sup>1</sup>  
I.K. TSANIS<sup>2, \*</sup>

<sup>1</sup> Department of Civil Engineering,  
McMaster University,  
1280 Main Street West,  
Hamilton, Ontario, Canada, L8S 4L7  
<sup>2</sup> Laboratory of Water Resources  
Management & Coastal Engineering  
Technical University of Crete  
Polytechniopolis, Chania,  
73100, Crete, Greece

\*to whom all correspondence should be addressed:  
Tel: + (30) 2821037799  
Fax: + (30) 2821037849  
e-mail: tsanis@enveng.tuc.gr

Received: 15/05/02  
Accepted: 17/12/02

---

### ABSTRACT

A GIS-based Decision Support System (DSS) was developed to select the appropriate interpolation technique used in studying rainfall spatial variability. The DSS used the ArcView GIS platform by incorporating its spatial analysis capabilities, the programming language "AVENUE", and simple statistical methods. The system consists of a series of modules and can be applied in spatial studies of other hydrological parameters. A test case from the country of Switzerland is used to demonstrate the applicability of the system. This should aid in better input to hydrological models.

---

**KEYWORDS:** ArcView GIS, AVENUE, Mean, Raingages, Spatial Interpolation Techniques, Standard Deviation.

### INTRODUCTION

The analysis and interpretation of spatial data sets forms an important part of geostatistics and is, unfortunately, highly human dependent (Genton and Furrer, 1998). For instance, it is well known that different individuals will take different approaches, yielding a large assortment of distinct solutions. It is often the case where judgement and experience play a key role in selecting the proper spatial interpolation technique for

each individual case (Englund, 1990). This is partly due to the variety of available spatial interpolation methods, which range from simple intuitive predictions to more sophisticated and complex procedures (Cressie, 1991). Estimating both rainfall at unaged locations and mean areal rainfall over an area (e.g. a catchment) based on the results of meteorological observations, motivated the development of gridded estimates of precipitation to provide inputs to spatially distributed

hydrologic and management models.

Although there are numerous articles have been written that are concerned with spatial interpolation, there is little or no agreement among the authors on the superiority of some techniques over others. Additionally, the increasing interest in Geographic Information Systems (GIS) with their broad usage and popularity, made it crucial to simply investigate the credibility and applicability of the different ready-to-use spatial interpolation techniques that are embedded in those systems. Generated with that in mind, this work has also been inspired by the Journal of Geographic Information and Decision Analysis initiative's special edition on spatial interpolation (Spatial Interpolation Comparison SIC97).

### APPROACH AND PROCEDURE

Variability is often a result of changes in conditions under which observations are made, differences in the way people do the work, difference in process variables, difference in environmental factors, the measurement system, or sampling. Statistical techniques are used to describe and understand variability. To provide a basis of comparison between the different techniques/models in this work, simple statistical methods are adopted. Since the method is data-driven and fully automated, it does not require preprocessing. This could be of value in an emergency situation where rapid, yet justifiable, results are required. At first, the concept of an objective function has to be established. This is normally followed by defining the constrained optimization of that function. The process initiates by randomly eliminating some of the available gages. The different interpolation techniques are then applied to estimate the "unobserved/missing" values on the basis of the "observed/remaining" ones. The purpose of the random selection of gages, which in this case takes the form of twenty tries, is intended to overcome the problem of outliers, if they exist. Experience shows that measured data contains between 10 to 15 per cent of outlying values due to gross errors, measurement mistakes, and faulty recording. Identifying and rejecting, or removing, outliers is highly opinion dependent and is not normally recommended since they, being extremes, represent critical cases or worst case scenarios. The errors/residuals at the unobserved/ungaged locations are then calculated as the difference between observed and estimated

values and categorized as positive and negative residuals. If the absolute value of the sum of the positive residuals is greater than the absolute value of the sum of the negative residuals, it implies that the observed values are greater than estimated ones. The model is then said to be *underestimating*. If the absolute value of the sum of the positive residuals is less than the absolute value of the sum of the negative residuals, it implies that the estimated values are greater than observed ones. The model is then said to be *overestimating*. At the second phase, three values are calculated for each technique. The errors at the ungaged (unobserved) locations are grouped in one column as absolute values, where the mean and the standard deviation are calculated as follows:

$$\bar{e} = \frac{\sum_{i=1}^n |x_{\text{obs}}| - |x_{\text{est}}|}{n} \quad (1)$$

$$S = \sqrt{\frac{\sum_{i=1}^n (|e_i| - \bar{e})^2}{n-1}} \quad (2)$$

where:  $\bar{e}$ : the mean of absolute residuals,  $x_{\text{obs}}$ : the observed value of rainfall,  $x_{\text{est}}$ : the estimated value of rainfall,  $n$ : the number of observed/estimated value (sample size),  $S$ : the standard deviation, and  $|e_i|$ : the absolute value of the individual residual. Generally, for a model to be considered satisfactory, the mean ( $\bar{e}$ : MeanAbsErr) and standard deviation ( $S$ : StDevAbsErr) of the absolute values of residuals are expected to be as low as possible among the other techniques. From the rain surface (grid), which is generated using the observed values, the average value is calculated as the sum of the rain values at each grid cell divided by the total number of cells in the grid. This represents the mean areal precipitation (MeanEst). The average of the observed values at all locations (observed and unobserved) is then calculated (MeanObs). A good model should generate a value (MeanEst) that matches (or be as close as possible to) the (MeanObs) and the difference between these two values is minimal. The difference is then calculated (Diff). The main criterion for judging the best model for each run is based on the minimum value obtained by averaging (Avg) the values (Diff, MeanAbsErr, StDevAbsErr),

assuming equal weights for all. The twelve techniques are ranked accordingly from best (MinAvg) to the worst (MaxAvg) in addition to many other statistics in a report format.

### **SPATIAL INTERPOLATION METHODS IN A GIS**

The following is a description of the interpolation techniques available in ArcView GIS 3.2.

#### **Spline (Regularized & Tension):**

Spline interpolation consists of the approximation of a function by means of series of polynomials over adjacent intervals with continuous derivatives at the end-point of the intervals. Smoothing spline interpolation enables to control the variance of the residuals over the data set. The solution is estimated by an iterative process. It is also referred to as the basic minimum curvature technique or thin plate interpolation as it possesses two main features: (a) the surface must pass exactly through the data points, and (b) the surface must have minimum curvature. The reader is referred to Franke (1982) and Mitas and Mitasova (1988) for further reading about the technique.

#### **Inverse Distance Weighting (IDW)**

Inverse Distance Weighting (IDW) is an interpolation technique in which interpolated estimates are made based on values at nearby locations weighted only by distance from the interpolation location. IDW does not make assumptions about spatial relationships except the basic assumption that nearby points ought to be more closely related than distant points to the value at the interpolate location. This technique determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. IDW allows the user to control the significance of known points upon the interpolated values, based upon their distance from the output point. The reader is referred to Tung (1983) and Watson and Philip (1985) for further reading about the technique.

#### **Kriging**

Kriging provides a means of interpolating values for points not physically sampled using knowledge about the underlying spatial relationships in a data set to do so. Variograms provide this knowledge. Kriging is based on regionalized variable theory which provides an optimal interpolation estimate

for a given coordinate location, as well as a variance estimate for the interpolation value. It involves an interactive investigation of the spatial behavior of the phenomenon before generating the output surface. It is based on the regionalized variable theory, which assumes that the spatial variation in the phenomenon is statistically homogeneous throughout the surface; that is, the same pattern of variation can be observed at all locations on the surface. This hypothesis of spatial homogeneity is fundamental to the regionalized variable theory. Data sets known to have spikes or abrupt changes are not appropriate for the Kriging technique. In some cases, the data can be pre-stratified into regions of uniform surface behavior for separate analysis. The reader is referred to Burrough (1986); Heine (1986); McBratney and Webster (1986); Oliver (1990); Press (1988); and Royle *et al.* (1981) for further reading about the technique.

#### **Trend Surface**

The linear trend surface interpolator creates a floating-point grid. It uses a polynomial regression to fit a least-squares surface to the input points. It allows the user to control the order of the polynomial used to fit the surface. Trend interpolation is easy to understand by considering a first-order polynomial. A first-order linear trend surface interpolation simply performs a least-squares fit of a plane to the set of input points. Trend surface interpolation creates smooth surfaces. The surface generated will seldom pass through the original data points since it performs a best fit for the entire surface. When an order higher than 1 is used, the interpolator may generate a grid whose minimum and maximum might exceed the minimum and maximum of the input points. The most common order of polynomials is 1 through 3. The reader is referred to Chidley and Keys (1970); Shaw and Lynn (1972); Lee *et al.* (1974); and Kruizinga and Yperlaan (1978) for further reading about the technique.

#### **Theissen Polygons**

Another choice given in the software for coding based on the value of a chosen attribute of the seed feature. This is appropriate if we wish to define the "region of influence" of a point or line. The region of influence is based on "nearest neighbours" to the point or line. The region of influence for a series of points is represented by a

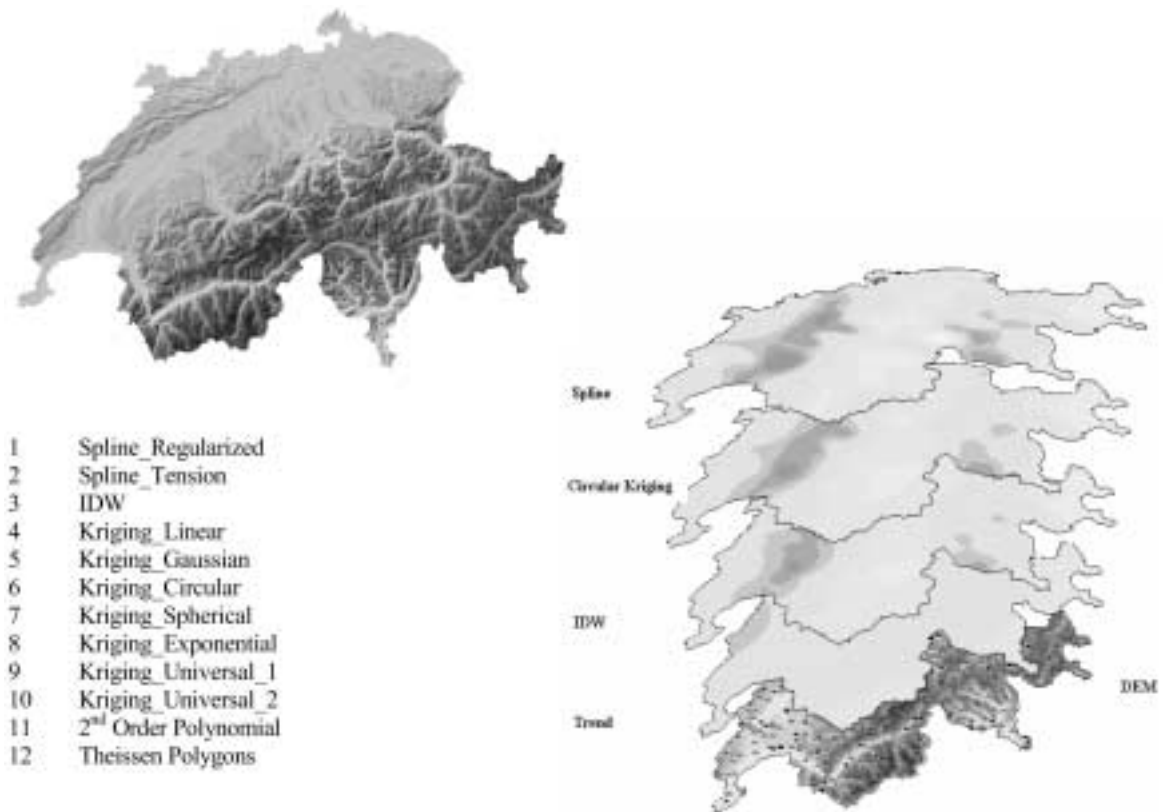


Figure 1. Digital Elevation Model and Interpolated Rain Surfaces in Switzerland

set of polygons encoded with the nominal value for each point. These polygons are referred to as Thiessen Polygons or collectively as a proximal map. Thiessen (1911) came up with the first technique to estimate areal average precipitation. Thiessen polygons are probably the most common approach for modeling the spatial distribution of rainfall. The approach is based on defining the area closer to a gage than any alternate gage and the assumption that the best estimate of rainfall on that area is represented by the point measurement at the gage. Because the basis of the model is geometry and gage location, implementation of Thiessen polygons in a GIS environment is not difficult. However, one impact of the use of Thiessen polygons is the development of discontinuous surfaces defining the rainfall depth over the area under study. This effect arises at the boundaries of the polygons where a discrete change in rainfall depth occurs (Ball and Luk, 1998). The reader is referred to Whitemore *et al.* (1961); Rainbird (1967); Hutchinson (1969); Diskin (1969); and Diskin (1970) for further reading about the technique.

As a result, and to summarize, the twelve spatial interpolation techniques employed in this study are listed in Figure 1.

### TEST CASE

The module is applied to a group of raingages in Switzerland to illustrate its applicability. Switzerland lies at the heart of Western Europe and covers an area of 41,284 km<sup>2</sup>. A Digital Elevation Model and some interpolated rain surfaces (interpolated from observed values using different techniques) are shown in Figure 1. The data set being used is related to the period of Chernobyl Nuclear Power Plant accident (April, 26th 1986) (Dubois, 1998). During the days following the accident, a radioactive plume, led by the action of atmospheric flow, was crossing many European countries. Radioactive deposition on the ground was mainly a function of the rainfall. The primary set of data includes 467 daily rainfall records made in Switzerland on May 8th 1986. The collection of the data was carried out by the Air Pollution Group at Imperial College in London under financial support of JRC-Ispra.

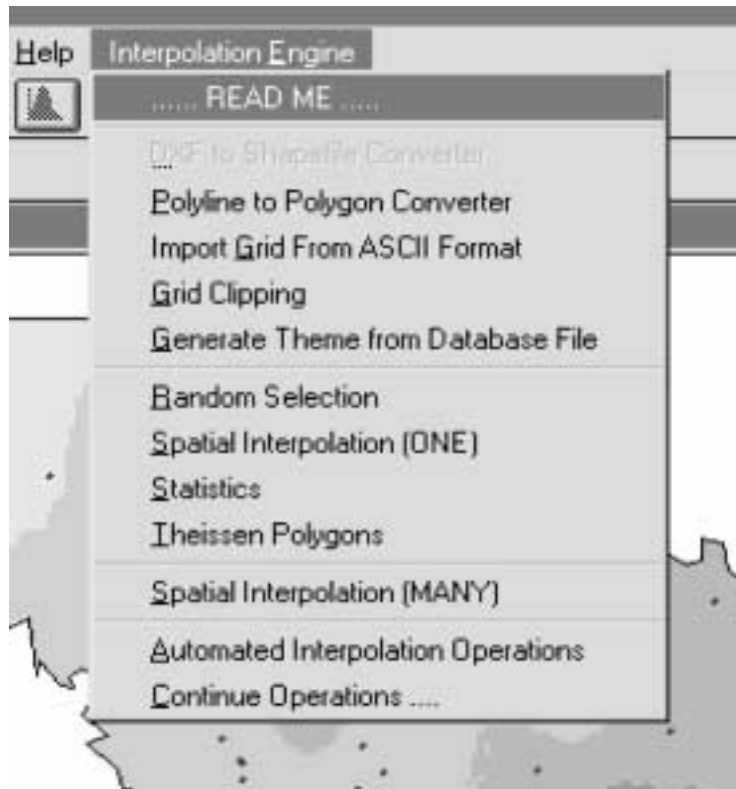


Figure 2. The Main Module of the Interpolation Engine Project

The location of the data points was also provided. Both Digital Elevation Model (DEM), with a resolution of around 1 km \* 1 km, and the country border, used to define the area under study, were provided as secondary information. The rainfall measurements were in the form of text files, the DEM in the form of a normal *ASCII ARC/INFO* format, and country borders were available as an AutoCad Interchange Drawing file.

### STRUCTURE AND DESCRIPTION OF THE PROJECT

A Geographic Information System can be more powerful when some added features, represented in this case by statistical methods, are combined with its many capabilities, as described above, resulting in the generation of a good decision support system. This GIS module, as shown in Figure 2, was developed in the ArcView GIS environment using AVENUE (the ArcView programming language). The programming language AVENUE provides a well-defined mechanism for allowing user-written routines to be called from within the normal user interface of the GIS pack-

age. In addition, this language also provides a menu-driven graphic interface, which makes it possible to guide a user with prompts and explanations throughout the application. The Dialog Designer and Spatial Analyst extensions are loaded to the ArcView project.

The project is composed of AVENUE scripts, dynamic link libraries (dlls), and designed dialogs. The scripts were mainly created by the authors. In some cases, however, they were modified versions of scripts that existed in the ArcView on-line help. As the name implies, the "makedir.dll" and "deldir.dll" dynamic link libraries were built to create and delete directories without the use of batch files. A dialog, created for the convenience of the user, included all spatial interpolation techniques available in ArcView but not available in the normal interface of the program. The developed module allows for one-/multi-process interpolation. The main module uses four main pieces of information to perform its task and generate the new network: the location of gages, rainfall data, region boundary, and a Digital Elevation Model and it consists of five sub-modules:

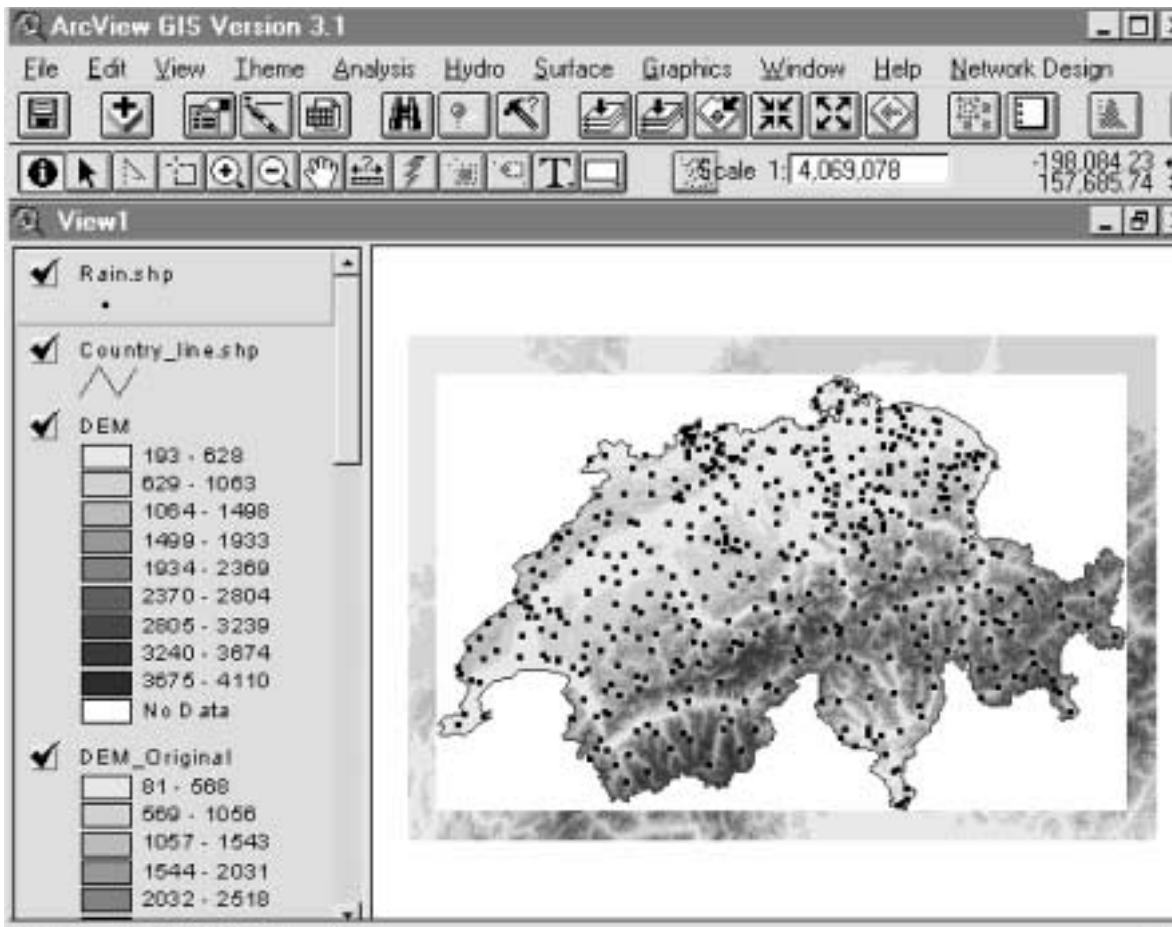


Figure 3. Clipped DEM Using the Menu Item "Grid Clipping"

### Module 1: ReadMe

This module documents the project by informing the user of the different parts of the main module using many graphic illustrations. It can be viewed as either a Word Perfect Document (help.wpd) or an HTML (help.htm). The advantage of the HTML is that it provides links to all AVENUE scripts and examples of output tables and text files.

### Module 2: Data Preparation

This is a five-step module that prepares the project for modules to follow. The data input requirements for this module include information on the boundary of the region, the Digital Elevation Model, and the data and location of raingages.

#### *DXF to Shapefile Converter*

In many cases, data is provided as AutoCad drawing files (for example: dxf files) which must then be converted to ArcView Shapefiles.

#### *Polyline to Polygon Converter*

After converting the AutoCad drawing file to an ArcView shapefile, the user should convert the resultant "polyline" shapefile into a "polygon" shapefile, which will be used in a later step.

#### *Import Grid from ASCII Format*

A grid, DEM in many cases, will be saved in the *ASCII ARC/INFO* format. In this case, it has to be imported into an ArcView grid format.

#### *Grid Clipping*

An area has to be extracted (clipped) from the original imported grid from the previous step in order to calculate mean areal rainfall or mean elevation over a specific region.

The clipping process (as shown in Figure 3) is accomplished by using the "polygon" shapefile generated in step 2 and the imported grid from step 3.

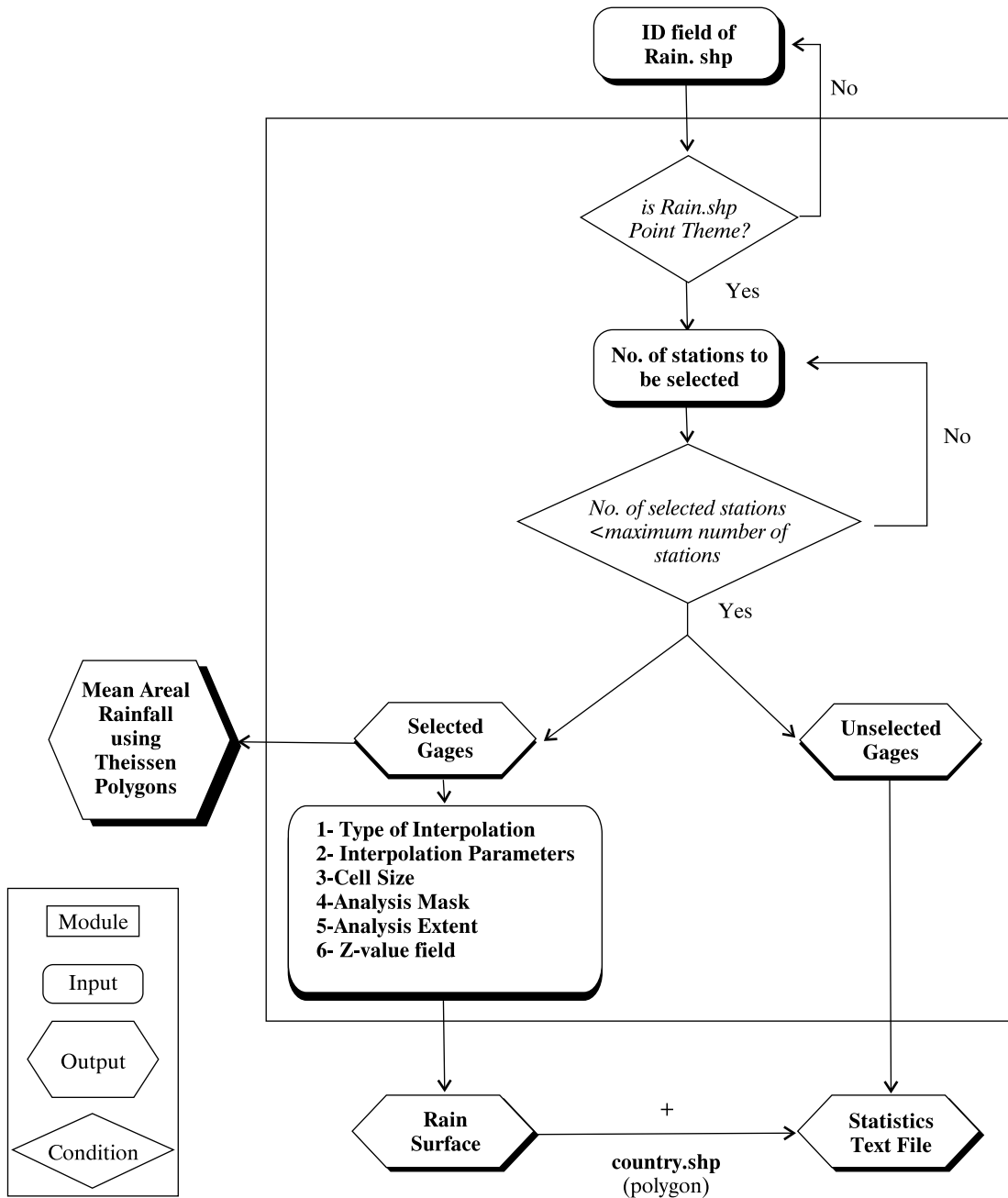


Figure 4. Module 3 (Interpolators)

**Generate Shapefile from a Database File**

A database file that contains information about location of raingages and amounts of rainfall can be converted to a "point" shapefile and added to the project.

It should be noted that this module was created specifically for the purpose of this study. If all input requirements are satisfied and the data is in the proper format, there is no need to go through the five steps.

**Module 3: Interpolators**

This is a key four-step module that is responsible for executing many tasks as shown in Figure 4, where a number of gages are selected randomly and a rain surface (grid) is generated. The remaining number of gages (the unselected ones) is then projected on the generated grid, and estimated values for rainfall at those locations are then extracted. A comparison between the observed and estimated precipitation values is held and the residu-

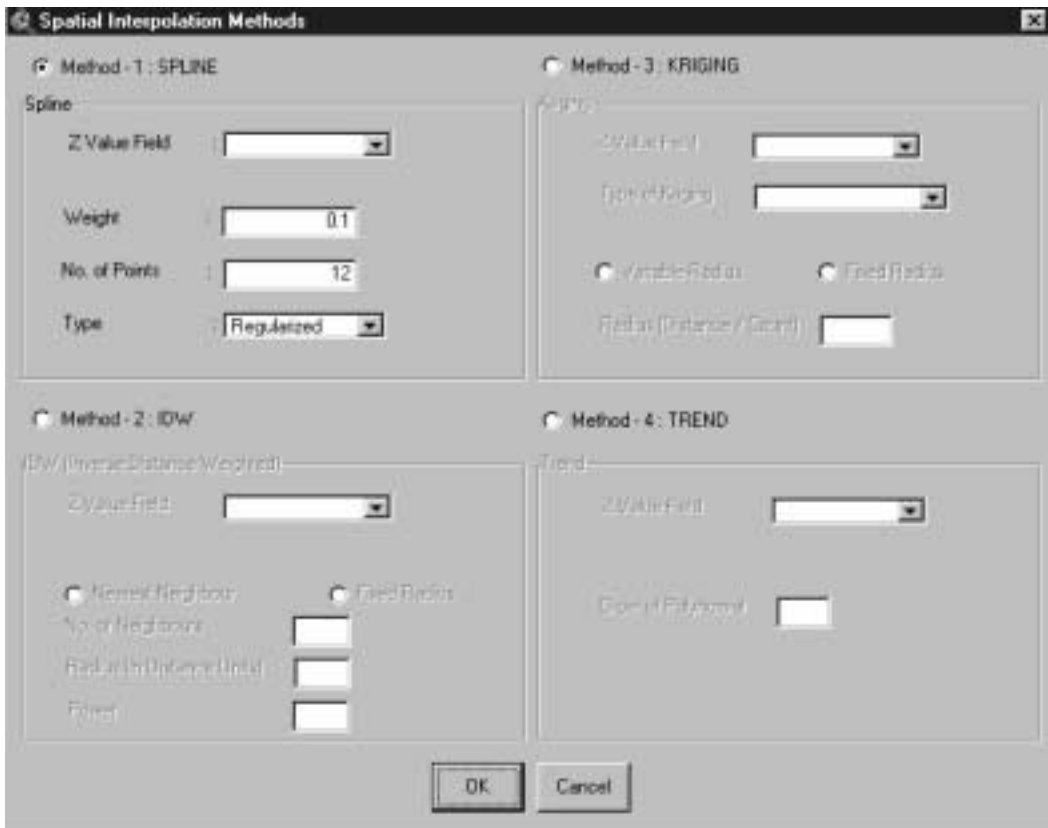


Figure 5. The Developed Interpolation Dialog as Part of Module (3)

als/errors are calculated. The results of the comparison are summarized in a text file and a database file from which illustrative charts can be generated if desired. The last item of the module generates Thiessen polygons for the selected gages.

#### *Random Selection*

In this step, the user is prompted to enter the required number of gages that should be randomly selected. The result is two shapefiles. One represents the selected gages and the other represents the unselected ones.

#### *Spatial Interpolation (ONE)*

The user is presented with the interpolation dialog (Figure 5) described earlier so that they may select the type of interpolation technique they will be using to generate the rain surface from the selected gages (as shown in Figure 6).

#### *Statistics*

This task is executed on the shapefile of the unselected gages. There is some interaction between

the user and the program. The user is prompted by some messages and a text file and a database file are then generated, as shown in Figure 7, in the working directory.

#### *Thiessen Polygons*

This task is executed on the shapefile of the selected gages. The mean areal rainfall will be calculated and presented to the user.

#### **Module 4: Spatial Interpolation (MANY)**

This module is an extended (advanced) version of module (3), where the user interference is minimized. It simply generates grids using all the spatial interpolation techniques available (including Thiessen polygons) by using the shapefile of the selected gages. Twelve grids are then generated using the interpolation techniques and the estimated values are compared to the observed values of the unselected gages. The result is a text file and twelve database tables. Each table represents the results of each interpolation technique in the same sequence as in the text file. At the end



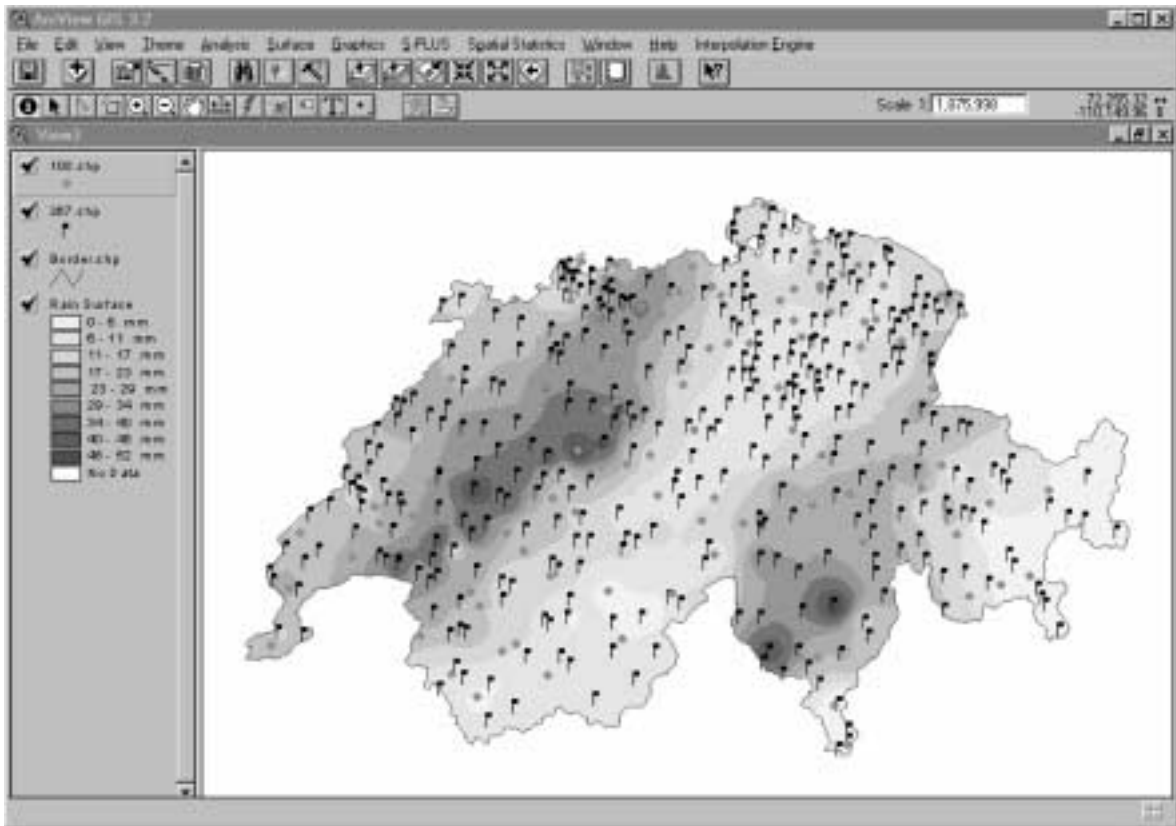


Figure 6. Module 3 - Menu Item: Spatial Interpolation [ONE] (Output)

of the text file, the different interpolation techniques are ranked from the best to the worst according to their performance.

#### Module 5: Automated Interpolation Operations

This module is an advanced version of module (4) with the least interference from the user. Three inputs are required from the user at the start of the operation: the number of iterations, the cell size, and the number of gages to be randomly selected, and the files DEM, Rain.shp, and Country.shp as shown in Figure 8. The main script, when executed, opens new views that are equal to the number of iterations specified by the user. It also generates new working directories for each one of the views inside the main working directory "C:\Inter", sets the properties of the views (map units and distance units), sets the analysis properties (extent, cell size, and mask), and copies the necessary themes (Rain.shp, Country.shp, and DEM) from the main view to the other views. The main script, then, triggers another script which, in turn, opens each of the

views and performs the random selection of the gages according to the number that was initially specified by the user. A third script is run from within the second, which is responsible for opening each of the previously generated views and performing the spatial interpolation task using the different methods and generate the text file, as shown in Figure 9, and the 12 database files. All output files are located in the respective working directory of each view, each of which is named after the view (i.e. all work done in view1 is stored in the sub-directory "C:\Inter\View1"). In many cases, the user will choose to terminate the run. For example, if the technique is not suitable for a certain data set or the user would like to run the same case with different cell size, the run may be terminated. The user will then choose to resume execution without generating new views or new shapefiles. At this point, the menu item "Continue Operations ..." may be used to work with the existing files. This menu item is attached to the script "Inter\_Continue" which will run the script "Inter\_Continue1" or it will run the script

```

Stat1 - WordPad
File Edit View Insert Format Help

INTERPOLATION RESULTS:

Number of Columns:      380.00
Number of Rows:        256.00
Left X-Coordinate:     -185556.38
Right X-Coordinate:    194443.63
Lower Y-Coordinate:    -127261.52
Upper Y-Coordinate:    128738.48
Cell Size:             1000.00
Center:                4443.63, 738.48
MODATA Value:         -9999

(1)  Interpolation Technique: SPLINE_REGULARIZED

Statistics:  Model Underestimated !!

Mean Precipitation (Observed):      18.345 mm
Mean Precipitation (Estimated):     17.593 mm
Maximum Observation:                 51.700 mm
Minimum Observation:                 0.00 mm
Maximum Estimation:                 100.000 mm
Minimum Estimation:                 -9.192 mm
Positive Residuals (Observed - Estimated): 1391.390 mm
Negative Residuals (Observed - Estimated): -1115.405 mm
Absolute Residuals (Observed - Estimated): 2506.795 mm
Residuals/Observed Precipitation:    4 %
Absolute Residuals/Observed Precipitation: 37 %
Mean of Absolute Residuals:          6.631 mm
Standard Deviation of the Residuals:  8.276 mm
Variance of the Residuals:           68.4841 mm^2
Coefficient of Variation of Residuals: 121.16 %

For Help, press F1
NUM

```

Figure 7a. Module 3 - Menu Item: Statistics (Output)

"Inter\_DifferentCellSize" if the user would like to run the "Automated Operations" using different cell size on multiple views that had been previously generated.

### Project Accessories

The project is equipped with more scripts that are automatically executed upon opening and closing the project. They perform additional functions which are intended to facilitate the project/user interaction and results presentation.

### EXECUTION

Speculating that the cell size and the number of available gages may influence the ranking of the interpolation techniques, a total of 12 runs were

performed. The first 4 runs were done using only 40% of the available gages (187 out of 467 gages) as observed records with a cell size of 500m, 1000m, 5000m, and 10000m. The second set of 4 runs used 60% of the available gages (280 out of 467 gages) as observed records with a cell size of 500m, 1000m, 5000m, and 10000m. The third and final set of 4 runs used 80% of the available gages (374 out of 467 gages) as observed records with a cell size of 500m, 1000m, 5000m, and 10000m. Each of the 12 techniques was then evaluated based on the average value of the 20 tries within each of the 12 runs. The evaluation was done on a scale of "0" to "10" with "10" being a perfect technique which means the smallest average value for the 20 (Avg) values; while a "0" is the worst tech-

	A	B	C	D	E	F	G	H
1	ID	LONG	LAT	OBS_RAINEST_RAIN	POS_RES	NEG_RES	RESID	
2	1	-159812	-39393	21.50	35.99	0.00	-14.49	-14.49
3	2	-159806	-68316	16.70	32.27	0.00	-15.57	-15.57
4	3	-158368	-42768	27.10	32.46	0.00	-5.36	-5.36
5	4	-156586	-62638	16.60	29.58	0.00	-12.98	-12.98
6	7	-153090	-46242	14.90	24.74	0.00	-9.84	-9.84
7	8	-148708	-24105	19.00	21.08	0.00	-2.08	-2.08
8	11	-145922	-35296	14.90	14.60	0.30	0.00	0.30
9	12	-143939	-16435	22.80	17.55	5.25	0.00	5.25
10	13	-140463	-30977	15.10	8.55	6.55	0.00	6.55
11	14	-136211	-12166	25.50	12.47	13.03	0.00	13.03
12	15	-129848	-72	21.00	5.29	15.71	0.00	15.71
13	16	-126276	-33474	21.00	10.58	10.42	0.00	10.42
14	17	-127684	-5680	23.40	6.55	16.95	0.00	16.95
15	19	-125363	-3505	21.70	3.82	17.98	0.00	17.98
16	20	-124685	-7968	10.00	6.95	3.05	0.00	3.05
17	23	-121276	9758	19.10	4.12	14.98	0.00	14.98
18	24	-120674	1960	19.40	0.54	18.86	0.00	18.86
19	26	-118064	17479	23.40	12.04	11.36	0.00	11.36
20	28	-116742	8555	18.30	3.13	15.17	0.00	15.17
21	29	-116780	-32596	33.40	29.19	4.21	0.00	4.21
22	30	-115480	-4816	10.70	6.36	4.34	0.00	4.34
23	31	-114528	-34864	37.10	35.25	1.85	0.00	1.85
24	32	-113301	-10418	23.00	15.89	7.12	0.00	7.12
25	33	-111242	-22691	29.00	31.46	0.00	-2.46	-2.46
26	35	-109680	20655	29.60	17.86	11.74	0.00	11.74
27	37	-109172	-34963	39.40	38.59	0.81	0.00	0.81
28	38	-108729	-10503	30.00	23.29	6.71	0.00	6.71
29	39	-108027	-13952	27.80	29.18	0.00	-1.38	-1.38

Figure 7b. Module 3 - Menu Item: Statistics (Output)

nique which means the largest average value for the 20 (Avg) values.

## RESULTS

Relying on the multiple random selection of gages to eliminate the effect of any errors or outliers, no statistical data preparation or preliminary analysis was done. Descriptive statistics were employed to provide inferences for the different models. The module output takes various forms:

1. Visual Grids (Surfaces): as shown in Figure 10, where the user can see the distribution of selected (solid dots) and unselected gages (x-marked locations). In addition to this, the different techniques can be visually compared to each other.

2. Text Files (Report Format): as shown in Figures 7 and 9, the user is able to obtain permanent records of the various runs for comparison purposes with helpful statistics listed for each technique. In addition, the techniques are listed in performance sequence.
3. Database Files: as shown in Figure 7, database files are permanently stored and from which plots can be generated (as shown in Figure 11). This confirms the results previously obtained in the text (report) format in Figure 9.

Generally, the results are shown in Figure 12, where the horizontal axis represents the interpolation techniques in the same order as in Figure 1 and the vertical axis represents the 0-to-10 scale.

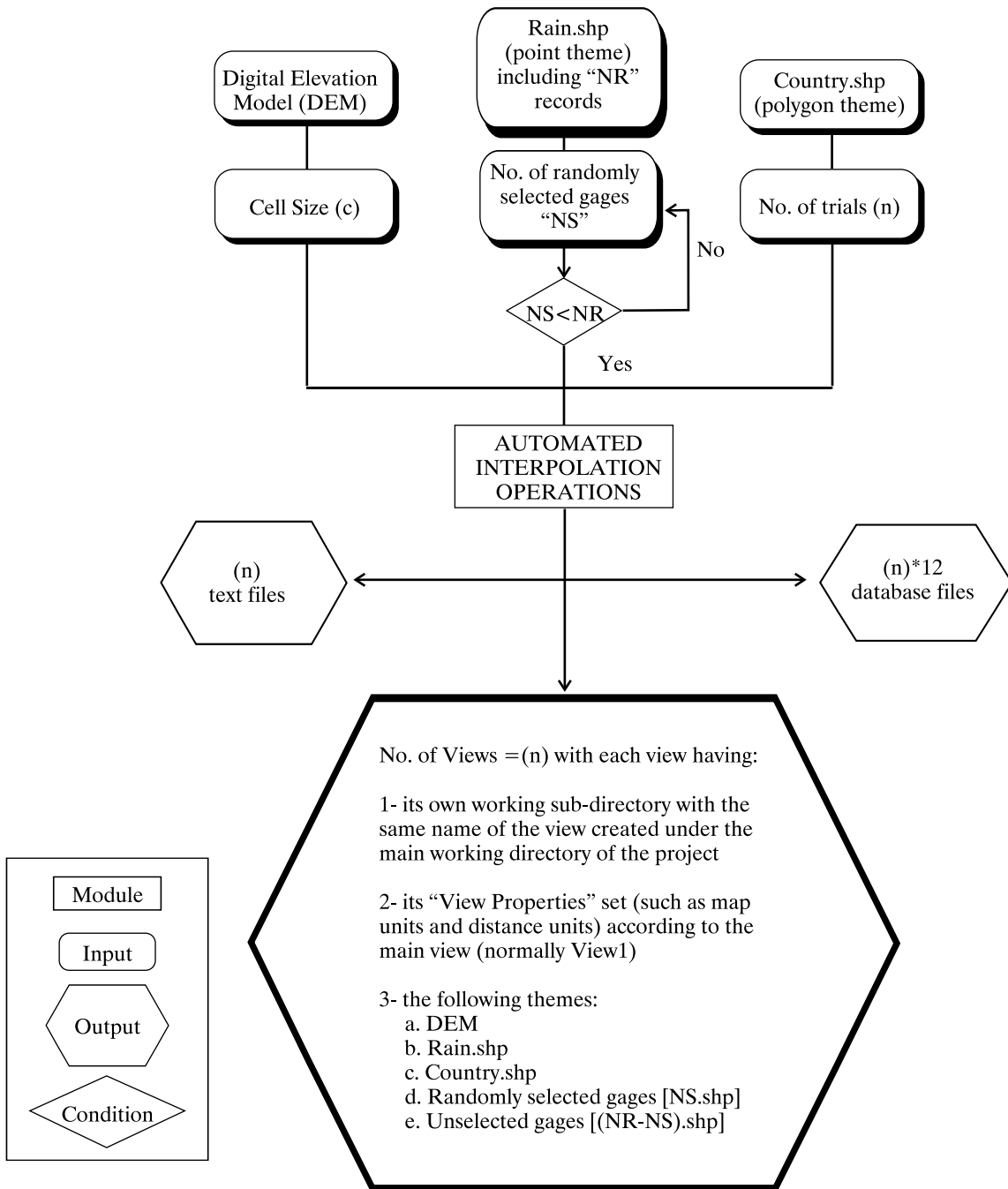


Figure 8. Module 5 - Automated Interpolation Operations

```

INTERPOLATION RESULTS..... C:\DSS\I\VIEWC

Number of Columns:      100.00
Number of Rows:         100.00
Left X-Coordinate:     -188884.38
Right X-Coordinate:    194443.01
Lower Y-Coordinate:    -117361.54
Upper Y-Coordinate:    118798.46
Cell Size:              1000.00
Center:                 8437.83,718.46
NODATA Value:          -9999

(1) Interpolation Technique: SPLINE REGULARIZED
Statistics: Model Underestimated !!

Mean Precipitation (Observed):      18.492 mm
Mean Precipitation (Estimated):     18.315 mm
Maximum Observation:                 81.700 mm
Minimum Observation:                 0.00 mm
Maximum Estimation:                  55.518 mm
Minimum Estimation:                  -18.211 mm
Positive Residuals (Observed - Estimated): 389.121 mm
Negative Residuals (Observed - Estimated): -552.001 mm
Absolute Residuals (Observed - Estimated): 1142.102 mm
Residuals/Observed Precipitation:    1 %
Absolute Residuals/Observed Precipitation: 28 %
Mean of Absolute Residuals:          5.263 mm
Standard Deviation of the Residuals:  5.847 mm
Variance of the Residuals:           31.8841 mm2
Coefficient of Variation of Residuals: 107.29 %

(2) Interpolation Technique: SPLINE TENSION
Statistics: Model Underestimated !!
EstimateError

Coefficient of Variation of Residuals:  71.65 %

(12) Interpolation Technique: THIESSEN POLYGONS
Statistics: Model Overestimated !!

Mean Precipitation (Observed):      18.492 mm
Mean Precipitation (Estimated):     18.497 mm
Maximum Observation:                 81.700 mm
Minimum Observation:                 0.00 mm
Maximum Estimation:                  50.160 mm
Minimum Estimation:                  0.00 mm
Positive Residuals (Observed - Estimated): 536.400 mm
Negative Residuals (Observed - Estimated): -537.500 mm
Absolute Residuals (Observed - Estimated): 1073.900 mm
Residuals/Observed Precipitation:    0 %
Absolute Residuals/Observed Precipitation: 29 %
Mean of Absolute Residuals:          4.949 mm
Standard Deviation of the Residuals:  5.548 mm
Variance of the Residuals:           30.7843 mm2
Coefficient of Variation of Residuals: 112.06 %

A Ranked List of The Spatial Interpolation Techniques
KRIGING_CIRCULAR
KRIGING_EXPONENTIAL
KRIGING_UNIVERSAL_1
KRIGING_LINEAR
KRIGING_SPHERICAL
IDW
KRIGING_GAUSSIAN
SPLINE_TENSION
KRIGING_UNIVERSAL_2
THIESSEN_POLYGONS
SPLINE_REGULARIZED
INF_ORDER_POLYNOMIAL

```

Figure 9. Module 5 - Automated Interpolation Operations (Output: Interp1.txt)

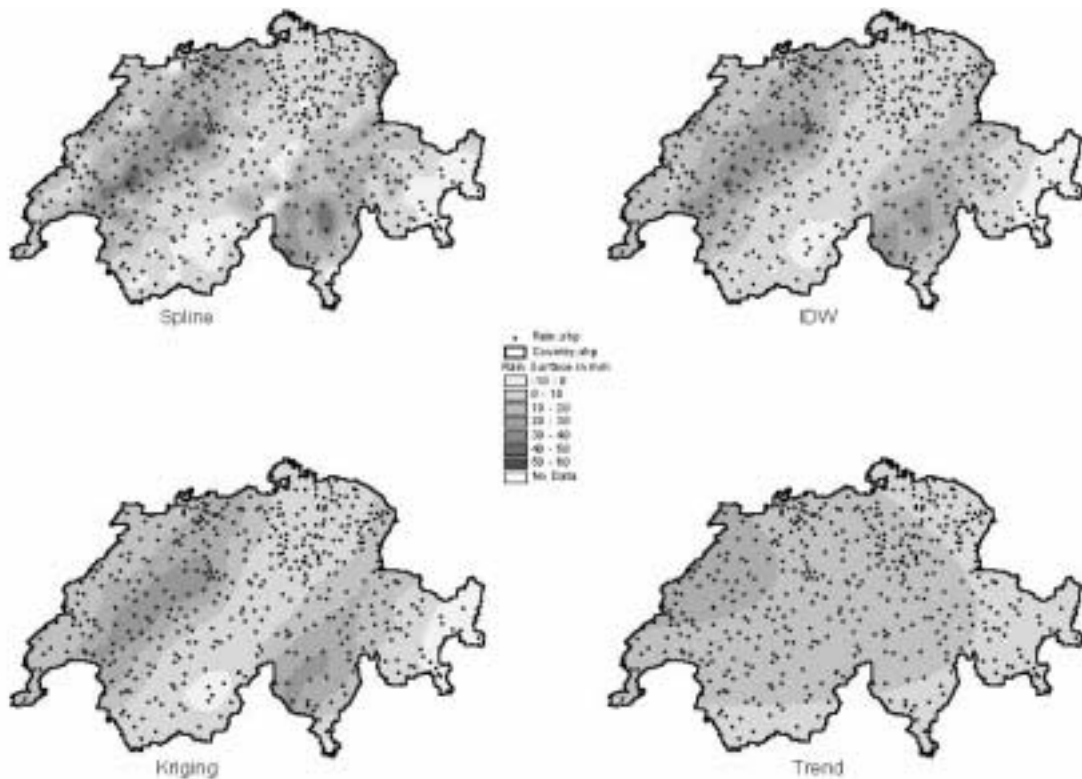


Figure 10. A Visual Comparison between Four Interpolation Techniques

The following can be concluded:

- the Spline-Regularized and the 2nd Order Polynomial techniques showed poor performance in almost all cases.
- Theissen Polygons and Kriging (Linear; Gaussian; Circular; Universal\_2) techniques fluctuated from one case to the other.
- the Spline-Tension, IDW, and Kriging (Spherical; Exponential; Universal\_1) techniques were able to provide reliable estimates. The Kriging-Exponential and Kriging-Universal\_1 models are recommended.

Results show that changing the cell size of the interpolated grid did not significantly affect the classification/rank of the interpolation techniques when using small number of gages (187 gages), as shown in Figure 13a, except for the Theissen polygons method which dropped on the scale significantly when a cell size of 10000m was used. However, by increasing the number of gages, the cell size started to show a more noticeable influence as some techniques show higher performance while the others show lower performance.

It should be noted that the 2<sup>nd</sup> order polynomial (technique 11) did not respond to any changes throughout the analysis. It was always ranked 12. Figure 13b shows that changing the number of gages used in the interpolation when using a cell size of 10000m did not have any effect on the performance of the different techniques. It is clear that increasing the number of gages available for interpolation enhanced the performance of the techniques except for three techniques: Spline-Regularized, Spline-Tension, and Theissen Polygons. Because Spline tries to fit a smooth surface that passes through the points, increasing the number of gages does not help the technique especially if there is abrupt changes in rain records, resulting in erratic estimated values. For Theissen polygons, as the number of gages increases, the size of the polygons decreases and their count increases resulting in erratic estimated rain values.

## DISCUSSION

This network is a relatively dense network with a density of one gage/88.4 km<sup>2</sup> and it records daily

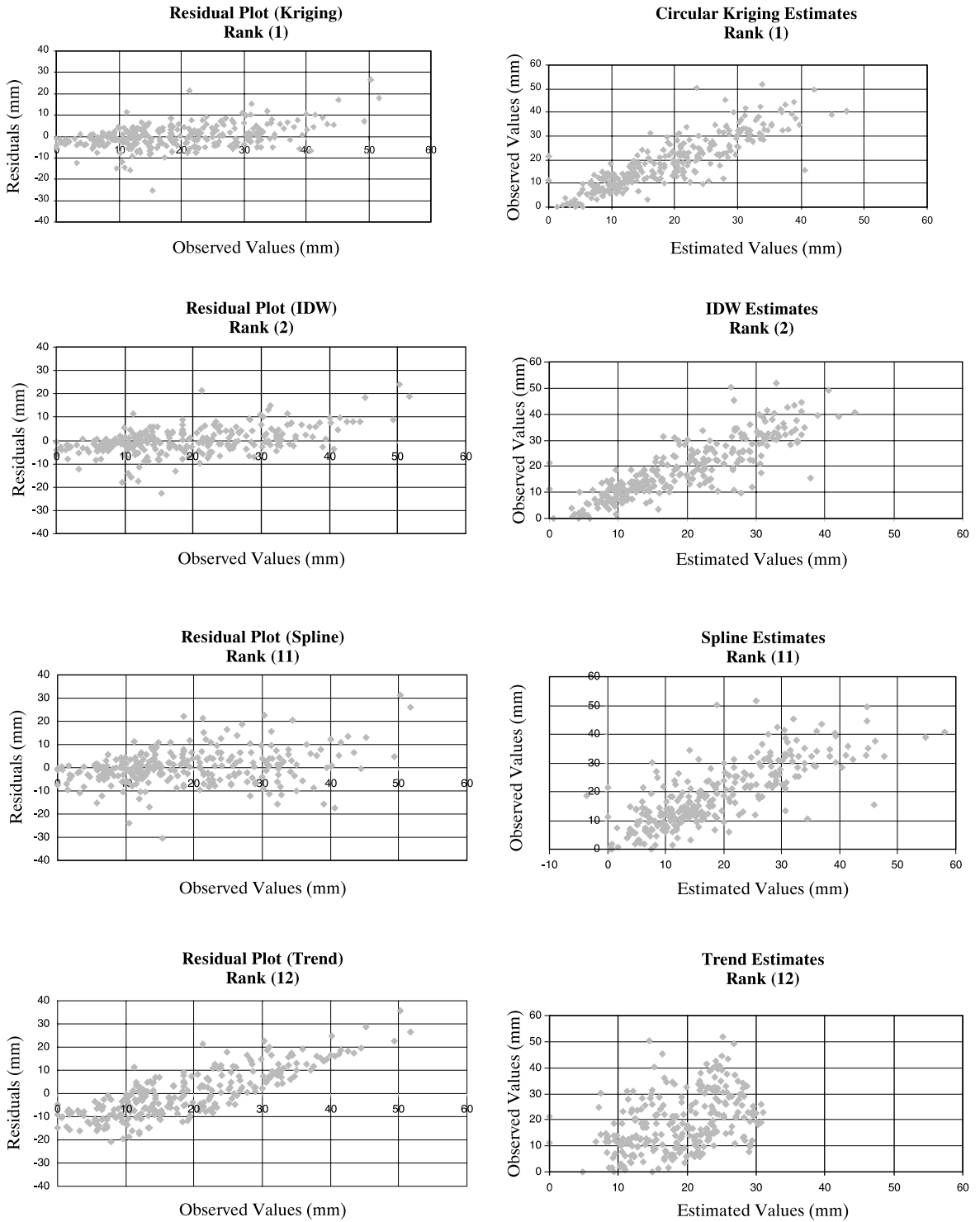


Figure 11. Residual Plots as well as Observed vs Estimated Rainfall Values for Four Interpolation Techniques

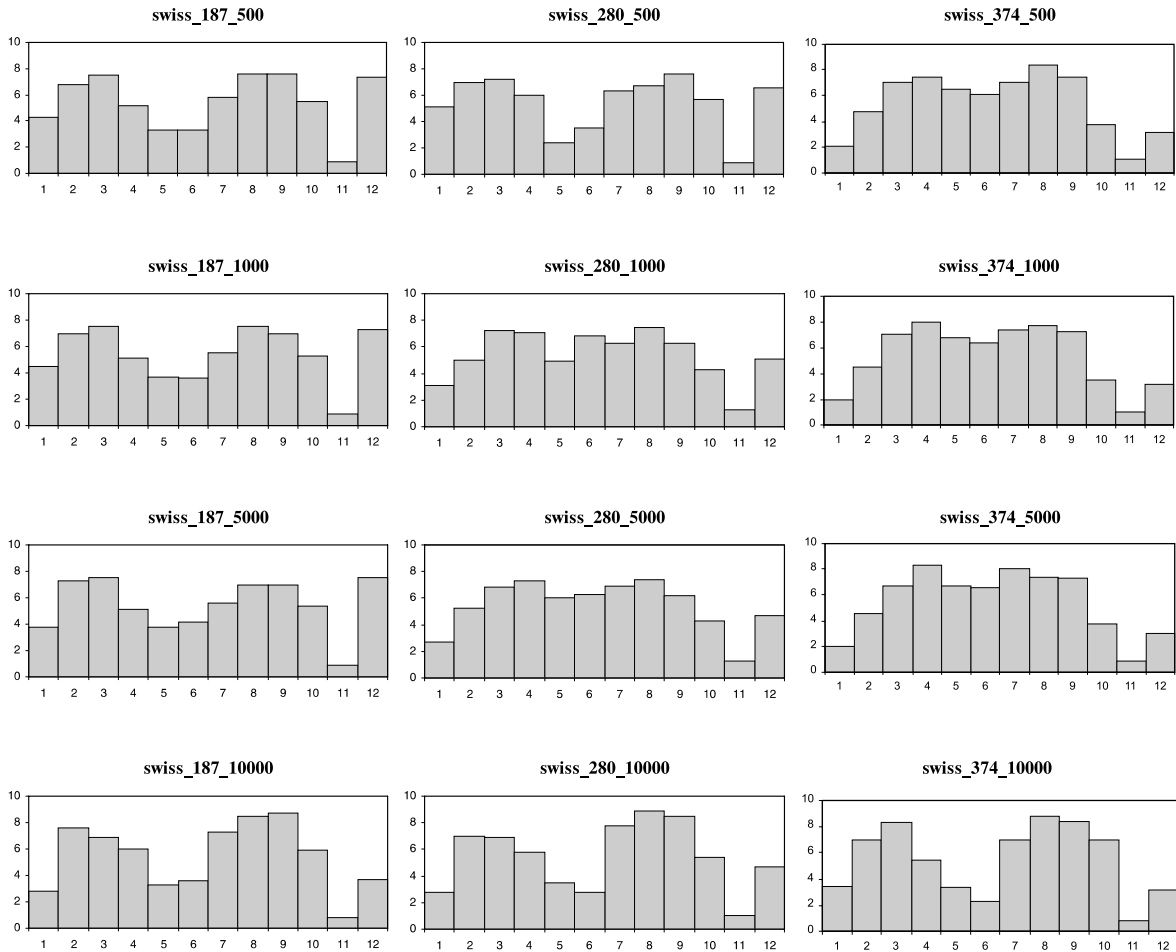


Figure 12. Performance of All Interpolation Techniques.

precipitation. Due to the high variability normally associated with daily precipitation records and the high density of the network, it is likely that techniques such as Trend and Spline\_Regularized would not provide nice estimates. 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 Spline\_Regularized try to pass through the points which, in this case, is not suitable because of the rapid changes in gradient/slope in the vicinity of the data points. However, Spline\_Tension is a more relaxed version of Spline which could fit a less smooth curves. Kriging is generally a good interpolator. The Ordinary Kriging is represented in this case by the Spherical, Circular, Exponential, Gaussian, and

Linear methods. With these options, Kriging uses the mathematical function specified by the method to fit a line or curve to the semi-variance data in the semi-variogram. These five models are provided to ensure that the necessary conditions of the variogram model are satisfied. The Exponential and Spherical methods seem to better fit the spatial variation of this data set. The Universal Kriging, represented by the Universal1 and Universal2 methods, assumes that the spatial variation across the surface has a structural component (drift). Drift is a systematic change in the cell values at a particular scale. This scale is related to the radius of the search area. The goal is to change the search radius to find the scale at which the drift can be detected and the variance is lowest. Universal1 uses a first order polynomial to approximate the drift and Universal2 uses a second order. The first derivative (Universal1) was appropriate for this specific application.



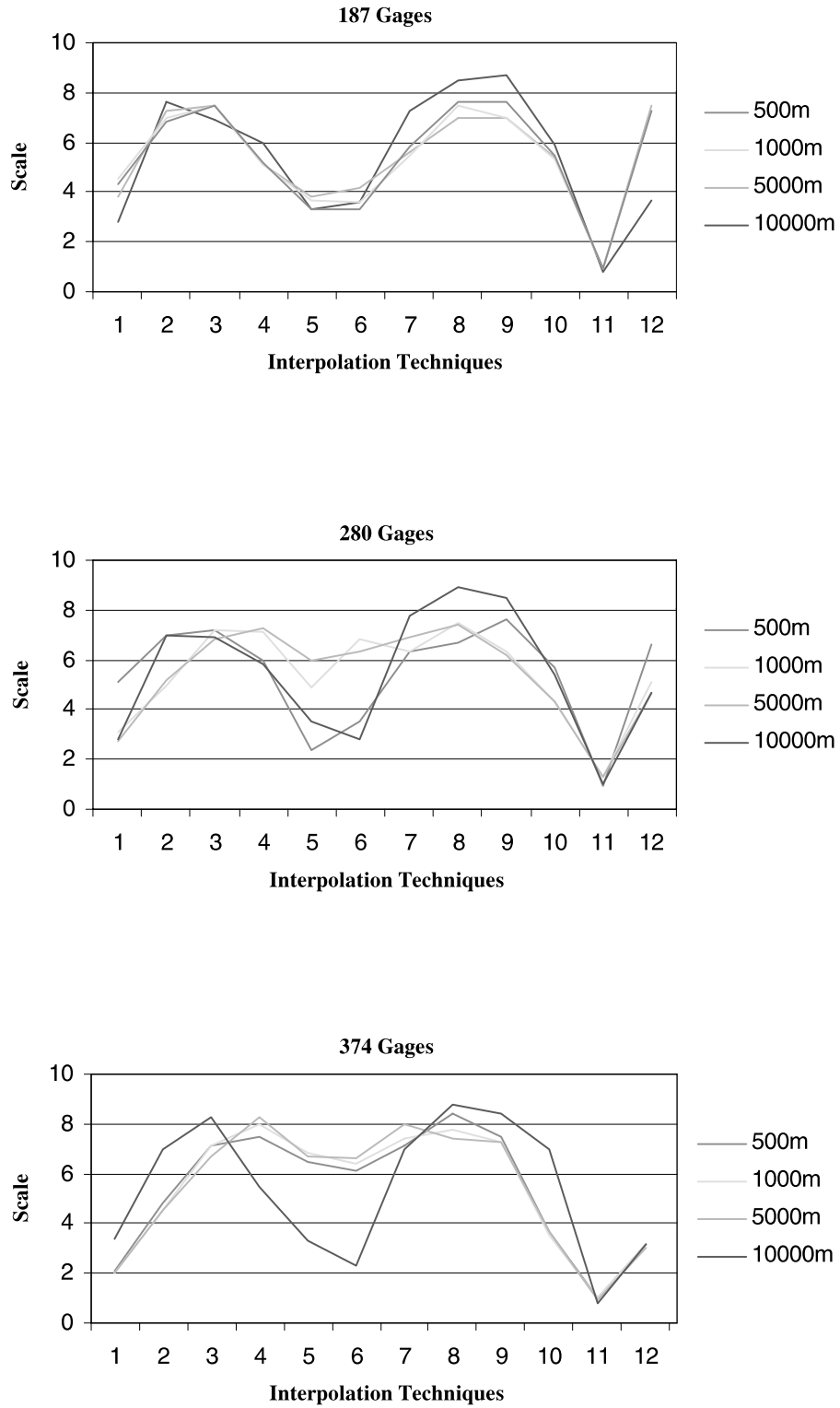


Figure 13a. The Effect of Changing the Cell Size on the Performance of the Interpolation Techniques.

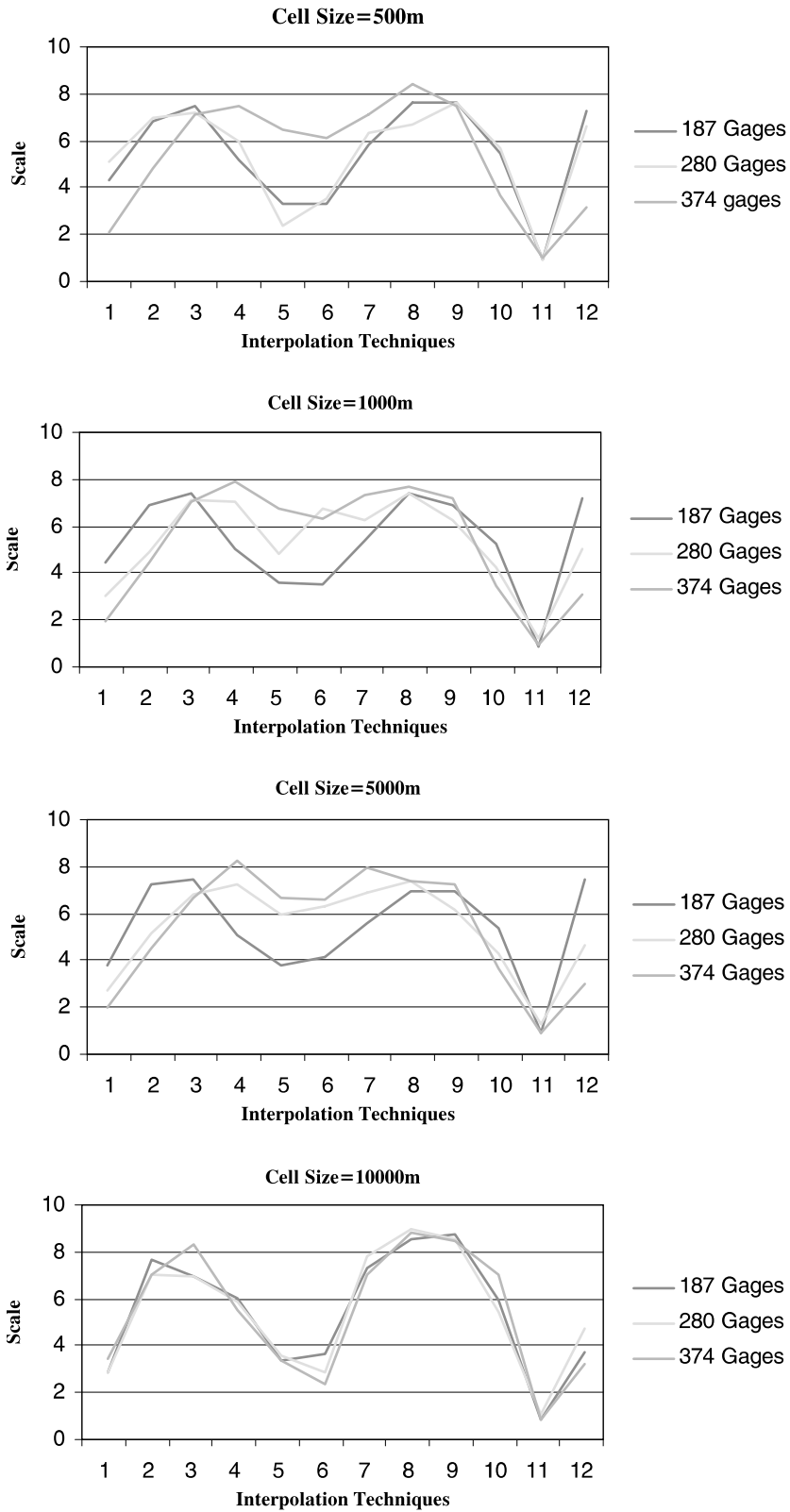


Figure 13b. The Effect of Changing the Number of Gages Used for Interpolation on the Performance of the Interpolation Techniques

Theissen polygons and IDW techniques are known to provide good results when used for relatively dense networks as in this case. However, increasing the number of gages can be problematic for the Theissen polygons technique.

It should be noted that repeated runs for different data sets is required to verify the results obtained. For example, wet, moderate, and dry conditions; hourly, daily, monthly, and yearly data; short and long term average; ...etc. The one available data set used as a test case in this study does not provide enough evidence that certain techniques are better than others.

## CONCLUSION

No interpolation technique, no matter how sophisticated, can accurately predict rainfall amounts at ungaged locations and, subsequently, estimate mean areal rainfall. This work establishes an approach by using GIS and historical data to locate the best technique. It should be noted that the selection of a given method to be used in

cases of emergency should be based on comparisons dealing with more than one data set. Rainfall data from different days, months, or years are likely to behave differently under various meteorological conditions. Performances of a method applied to a partial data set are likely to be different from performances of the same method with a larger data set. By taking a sample of the data set, the problem scale is changed, and the boundary between large and small scale variation is modified. For that reason, data sets with different sizes are recommended. Finally, it is highly advisable for any country or region to obtain data from the closest parts of neighboring countries or at least to have as many data as possible from locations within the country/region. Based on the one data set available for this study, it was clear that the Kriging\_Exponential and Kriging\_Universal\_1 models showed consistent performance and provided reliable estimates regardless of the number of gages or the cell size used in the interpolation.

## REFERENCES

- Ball J.E. and Luk K.C. (1998), Modeling Spatial Variability of Rainfall Over A Catchment, *Journal of Hydrologic Engineering*, **3**, 122-130.
- Burrough P.A. (1986), Principles of Geographic Information Systems for Land Assessment, Oxford University Press, New York.
- Chidley T.R.E. and Keys, K.M. (1970), A Rapid Method of Computing Areal Rainfall, *Journal of Hydrology*, **12**, 15-24.
- Cressie N. (1991), Statistics for Spatial Data, New York, Wiley.
- Diskin M.H. (1969), Thiessen Coefficients by Monte Carlo Procedures, *Journal of Hydrology*, **8**, 323-335.
- Diskin M.H. (1970), On the Computer Evaluation of Thiessen Weights, *Journal of Hydrology*, **11**, 69-78.
- Dubois G. (1998), Spatial Interpolation Comparison 97: Forward and Introduction, *Journal of Geographic Information and Decision Analysis*, **2**, 1-10.
- Englund E. J. (1990), A Variance of Geostatisticianism, *Journal of Mathematical Geology*, **22**, 417-455.
- Franke R. (1982), Smooth Interpolation of Scattered Data by Local Thin Plate Splines, *Journal of Computation and Mathematics with Applications*, **8**, 273-281.
- Genton M.G. and Furrer R. (1998), Analysis of Rainfall Data by Simple Good Sense: Is Spatial Statistics Worth The Trouble?, *Journal of Geographic Information and Decision Analysis*, **2**, 11-17.
- Heine G.W.(1986), A Controlled Study of Some Two-Dimensional Interpolation Method, *COGS Computer Contributions*, **3**, 60-67.
- Hutchinson P. (1969), Estimation of Rainfall in Sparsely Gaged Areas, *International Association of Hydrologic Sciences Bulletin*, **14**, 101-120.
- Kruizinga S. and Yperlaan G. J. (1978), Spatial Interpolation of Daily Total of Rainfall, *Journal of Hydrology*, **36**, 65-73.
- Lee P.S., Lynn P.S. and Shaw E.M. (1974), Comparison of Multiquadratic Surfaces for the Estimation of Areal Rainfall, *International Association of Scientific Hydrology Bulletin*, **19**, 303-317.
- McBratney A.B. and Webster R. (1986), Choosing Functions For Semi-Variograms of Soil Properties and Fitting Them to Sampling Estimates, *Journal of Soil Science*, **37**, 617-639.
- Mitas L. and Mitasova H. (1988), General Variational Approach to the Interpolation Problem, *Journal of Computation and Mathematics with Applications*, **16**, 983-992.

- Oliver M.A. (1990), Kriging: A Method of Interpolation for Geographical Information Systems, *International Journal of Geographic Information Systems*, **4**, 313-332.
- Press W.H. (1988). Numerical Recipes in C, The Art of Scientific Computing, New York, Cambridge University Press.
- Rainbird A.F. (1967), Methods of Estimating Areal Average Precipitation, Reports on WMO/IHD Projects, Report No. 3, pp. 45.
- Royle A.G., Clausen F.L. and Frederiksen P. (1981), Practical Universal Kriging and Automatic Contouring, *Geoprocessing*, **1**, 377-394.
- Shaw E.M. and Lynn P.O. (1972), Areal Rainfall Evaluation Using Two Surface Fitting Techniques, *International Association of Scientific Hydrology Bulletin*, **17**, 419-433.
- Thiessen A.H. (1911), Precipitation for Large Areas, *Monthly Weather Review*, **39**, 1082-1084.
- Tung Y.K. (1983), Point Rainfall Estimation for a Mountainous Region, *Journal of Hydraulic Engineering*, **109**, 1386-1393.
- Watson D.F. and Philip G.M. (1985), A Refinement of Inverse Distance Weighted Interpolation, *Geo-Processing*, **2**, 315-327.
- Whitemore J. S., Van Efden F. J. and Harvey K. J. (1961), Assessment of Average Annual Rainfall Over Large Catchments, In: *Inter-African Conference on Hydrology*, C.C.T.A. Publication No. 66, 100-107.