Application of Spatial Interpolation Method for Estimating the Spatial Variability of Rainfall in Semiarid New Mexico, USA.

Lam K.C.[1]

Bryant R.G.[2]

Wainright, J.[3]

[1] *Senior Lecturer, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia*
lam@ukm.edu.my
[2] *Reader, Department of Geography, University of Sheffield, UK*
R.G.Bryant@Sheffield.ac.uk
[3] *Professor, Department of Geography, Durham University, UK*
john.wainwright@durham.ac.uk

Abstract

Spatial interpolation of rain gauge data is important for ecohydrology study or modelling of land degradation. The monthly rainfall data recorded at 68 stations in Jornada basin was analysed to study the spatial patterns of rainfall. The inverse distance weighting spatial interpolation method was applied to model the spatial variability of rainfall for a wet (1992) and dry (1994) years. The rainfall interpolation was tested for the basin wide region and by constraining the rainfall-interpolation within the study area boundary. The accuracy of the interpolation result was measured by adopting the leave-one-out cross validation method. The result indicates that the rainfall displayed a strong spatial variability trend from the southwest to the northeast. The result from the CV analysis of the total data points for both year showed that the IDW interpolation method produced from data points within the study area boundary produced better fits compared to the CV result from all the data points within the Jornada basin. The result from constraining the rainfall-interpolation within the study-area boundary showed that the interpolation error from the edge effects can be minimised and result of the predicted values showed a closer fit to the measured values.

Keywords: spatial variability, rainfall interpolation, geo-statistics, Inverse Distance Weighting

1. Introduction

Spatial distribution of rainfall is important to many ecohydrological studies. In semiarid environments, rainfall is the main source of water which determines the availability of drinking water and the level of soil moisture (Justiniano and Fredericksen, 2000; Schwinning *et al.*, 2004). In semiarid desert, plants depend on rainfall for growth and survival (Noir Meir, 1973; Peñuelas *et al.*, 2004). Desert plants are very sensitive to the fluctuation of rainfall and are suitable for study of effects of climate change on the semiarid environment. However it is not possible to have meteorological stations covering large areas, particularly in areas which are not of interest to human activities such as in a semiarid desert. It is therefore important to choose a suitable method to interpolate rainfall data from existing meteorological station points to large areas.

Spatial interpolation methods use point measured data to estimate unknown values at unsampled locations. There are a large number of interpolation methods, from simple to complex, which are available for use. Much research has been done on the issue of spatial interpolation using different interpolation methods on various climatic variables. (e.g. Hevesi *et al.*, 1992; Goodale *et al.*, 1998; Goovaerts *et al.*, 2000; Vicente-Serrano *et al.*, 2003; Lloyd, 2005; Hofstra *et al.*, 2008). Accordingly Blöschl and Grayson (2000) suggested that, in choosing a particular interpolation model, the user must be aware that the quality of the interpolated surface is dependent on the accuracy of the original point data and how well the selected method reflects the underlining spatial structure. Some of the commonly used interpolation methods include Thiessen polygon, Inverse Distance Weighted (IDW), Spline, and Kriging.

One of the most commonly used spatial interpolation method for interpolating rainfall distribution is inverse distance weighting. IDW method was used to interpolate rainfall data by many researchers, for example Goodale *et al.* (1998) in Ireland, Dirks *et al.* (1998) in Norfolk Island and Price *et al.* (2000) in Canada.

Based on the research cited, the deterministic IDW interpolation method was chosen to be applied to generate the interpolated rainfall for the Jornada basin from the existing rainfall dataset. Areal mean rainfall, coefficient of variability, and accuracy of rainfall fields predicted were calculated and reported. The study site's simple topography and low relief negates the application of complex interpolation methods such as Kriging.

2. Study Area

The study area is located in the Jornada Basin, within the larger Jornada del Muerto basin and is typical of the Basin and Range physiographic of the American Southwest and the Chihuahuan Desert. The Jornada Basin (32° 31′ N, 106° 47′ W) is located in the northern Chihuahuan desert of New Mexico in the Jornada del Muerto Basin as shown in Figure 1. The topography of the Jornada Basin is of closed-basin, typical of many arid regions of the southwestern United States (US), with no exterior drainage where runoff generated by high intensity rainstorm events drains into the playas in the central part of the basin. It sit on the Rio Grande floodplains (1,186 m), bordered by the San Andreas Mountains (2,833 m) to the east, the Doña Ana Mountains and Summerford Mountains (1,780 m) to the southwest and Point of Rocks to the northwest. The landscapes in the Jornada Basin exemplify ecological conditions and desertification problems found in the southwest of US where it is characterised by the typical transition from perennial grassland to shrubland. Semiarid areas such as Jornada basin is characterized by complex rainfall patterns which highly influenced by seasonality. Phenomena like the thunderstorm occurred frequently during summer months which bring relatively large rainfall and during the winter months rainfall are lower due to the its frontal characteristic and rain shadow effect (Wainwright, 2006).

3. Data and Method

3.1 Rainfall Data

Monthly rainfall data from 78 rain-gauges distributed throughout the Jornada Basin (Figure 2) were summarised into hydrological year rainfall. The hydrological year used in this study starts runs from October to September. Two set of rainfall data representing a wet and a dry year were selected and tested in order to evaluate the effectiveness of the IDW technique in interpolating and capturing the vast difference in spatial heterogeneity of rainfall that fell during a dry and a wet year. A preliminary analysis to examine rainfall frequency distribution and spatial autocorrelation patterns over different locations of the study area was made. The hydrological year rainfall of each station was calculated for the 1992 (wet year) and 1994 (dry year). Only 68 out of the 78 rain-gauges provide data for both of the two selected years so that the rain-gauges with no data for both years were removed from analysis.
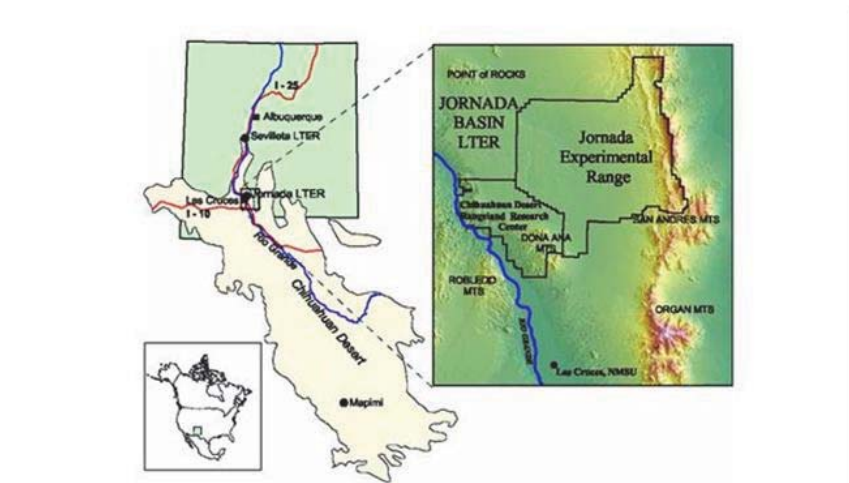

Figure 1. Location of Jornada Basin, New Mexico, USA (Modified from Monger *et al.* 2006).
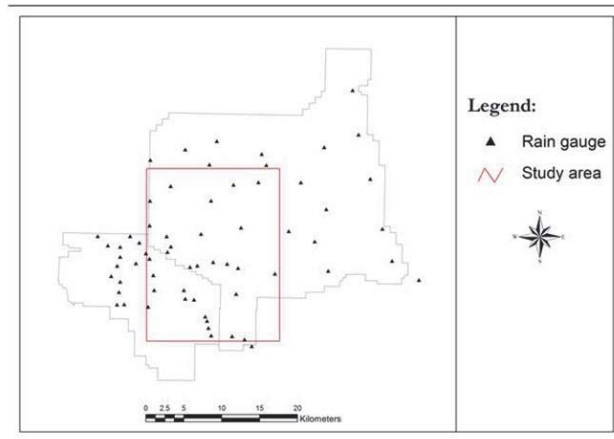
Figure 2. The study area and distribution of the 68 rainfall-monitoring stations within the Jornada Basin.

The rainfall dataset was subjected to a set of statistical exploration to determine the appropriate parameters for the interpolation model. The statistical tests applied included the following:

i) Using frequency-distribution test to explore the univariate distribution of the data
ii) Checking for the Normality of the dataset using Normal QQ plots
iii) Identify the existence of any global trend in the dataset. Indentifying global trends in a dataset is an important step towards creating an appropriate model which was then applied to produce a smooth continuous surface. For achieving better accuracy in geostatistical method, trend should be removed before modeling the semivariogram/covariance. When the trend is removed, only the random short-range variation in the residuals is modelled and the trend will be automatically added back before making an actual prediction, so that a reliable result is obtained.
iv) Checking for spatial dependencies in the dataset and exploring the spatial autocorrelation through the Semivariogram/Covariance cloud. The data were subjected to autocorrelation tests where different pairs of sample location were examined by measuring the distance between two locations and the semivariogram cloud was created by plotting half the difference squared between the values at the locations.

3.2 Inverse Distance Weighting method for spatial rainfall interpolation

Inverse Distance Weighting (IDW) is a simple deterministic interpolator that is exact with very few decisions to make regarding model parameters. It assumes that things that are closer are more similar than those farther apart and use measured values surrounding an unmeasured location to predict its value. Those measured values closest to the prediction location will have more influence on the predicted value and were given more weight than those farther apart. There is no assessment of prediction errors and no assumptions required of the data. The interpolation function can be written as (Isaaks and Srivastava, 1989):

$$Z^* = \frac{\sum_{i=1}^{N}\left(\frac{1}{dP}Z_i\right)}{\sum_{i=1}^{N}\left(\frac{1}{dP}\right)}$$ **Equation 1**

where:
Z^* = estimated value,
Z_i = a neighbouring data point value,
N = the number of neighbouring point,
d = the distance between the data point and the point being interpolated,
P = a positive-power parameter.

The weights are calculated based on the function of Euclidean distance, and the size of surrounding points or the search radius and the value of P are often chosen subjectively. As P approaches zero, an estimate approaches a simple average of the neighbouring data values, and as P approaches infinity, an estimate approaches the value of the closest data point. The accuracy of the model can be assessed by examining the cross-validation results for the root mean squared error (RMSE), mean relative error and correlation between the estimated and observed data values.

4. Validation

The accuracy of an interpolation method can be measured by the results of cross validation. The cross validation computes the residuals for all data points, and the self-consistency of the variogram is tested by a mean difference of residuals of tendency to zero and a variance tendency to 1. There are several ways to implement cross validation (CV). Data splitting is the CV method commonly used, which according to Sluiter (2008), although easy to implement, may have limited success if the spatial distribution of rainfall is too heterogeneous. Another simpler method of CV is the leave-one-out method where each sampling point is left out from the dataset one after another and at its point the value is interpolated and the residue is obtained before moving to the next point (Burrough and McDonnell, 1998; Bivand *et al.*, 2008).

The CV method adopted in this study is the leave-one-out which was applied to the IDW. The prediction errors from cross validation can help to find the most accurate prediction model. For an accurate model, the mean error should be close to 0 and the root-mean-square error should be as small as possible (Johnston *et al.*, 2001). The ArcGIS spatial interpolation tool produces the global mean error and RMSE values for each interpolation, taking into account all the points which ware used to interpolate the rainfall and this results in higher RMSE caused by the edge effect. To overcome this problem, a second leave-one-out CV analysis was done where the measured and predicted values from the rain gauges that fall within the study area boundary were extracted from the resulting rainfall interpolation map for further analysis. The R^2, mean absolute error and root mean square error were calculated for the resulting residuals and compared.
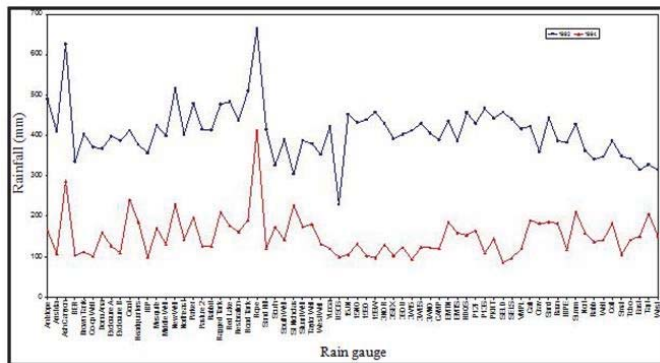

Figure 3. Spatial and temporal differences in hydrological year rainfall for 1992 and 1994 at the study site.

5. Results and Discussion

5.1 Rainfall analysis

Figure 3 shows the spatial and temporal variability of the hydrological year rainfall for 1992 and 1994 in the study area. The highest rainfalls received at both years were recorded at the Rope Springs station, which is located in the San Andreas mountain area in the eastern part of the study area. The mean values of hydrological year rainfall for the wet (1992) and dry (1994) years were 409 mm and 151 mm respectively (Fig. 4 and 5*)*. In 1992 the most frequent mean hydrological year rainfall occurred in the range from 400-450 mm, occurring mainly in the bajada slope areas of Jornada basin. In contrast, in 1994, the most frequent mean hydrological year rainfall ranged from 120 mm to 150 mm and primarily occurred around the plain in the centre, towards the western part of the Jornada basin.

Overall results showed that, in both years, the rainfall stations located in the northeast and southwest parts of the Jornada basin show high spatial variation of rainfall, with the 1994 data displaying greater spatial variation of the two. However low variations in rainfall were found throughout the central, northwest and southeast stations of the Jornada basin in both years (Fig. 6 and 7).
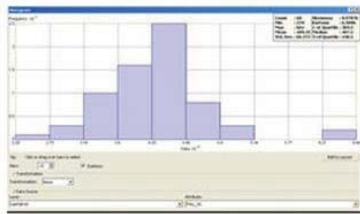

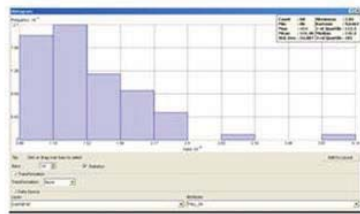**Figure 4** Frequency distribution of hydrological year rainfall for 1992.

Figure 5 Frequency distribution of hydrological year rainfall for 1994.
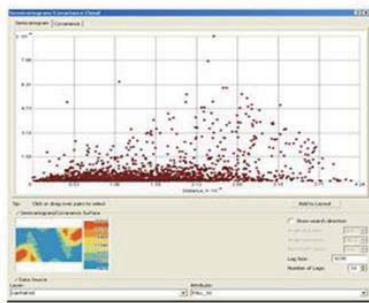

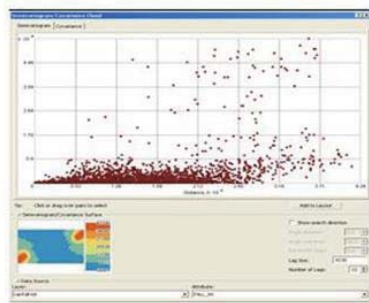**Figure 6** Semivariogram cloud of hydrological year rainfall for 1992.

Figure 7 Semivariogram cloud of hydrological year rainfall for 1994.

The QQ plot was used to compare the distribution of rainfall data to a standard normal distribution by plotting data values versus the value of a standard Normal distribution where their cumulative distributions are equal (Johnston *et al.*, 2001). A plot can be assumed to be close to a Normal distribution if the data plotted were close to a straight line. In Figure 8a, the plotted point data fall on a straight line except for the data points from the northeast and southwest of the region. In Figure 8b, logarithmic transformation of the data was applied and the resulting plot was closer to a straight line, exhibiting similarity to a Normal distribution. The 1994 data also shows the same trend and needs transformation to Normalize the dataset (Figure 9a and 9b). Thus, both datasets can be assumed to be log-normally distributed.

The results in Figure 10 and Figure 11 shows a non-flat curve in blue on the projected plane indicating presence of global trends in the dataset. The Jornada rainfall displayed a strong trend from the southwest to the northeast, as also demonstrated in general by Wainwright (2005).
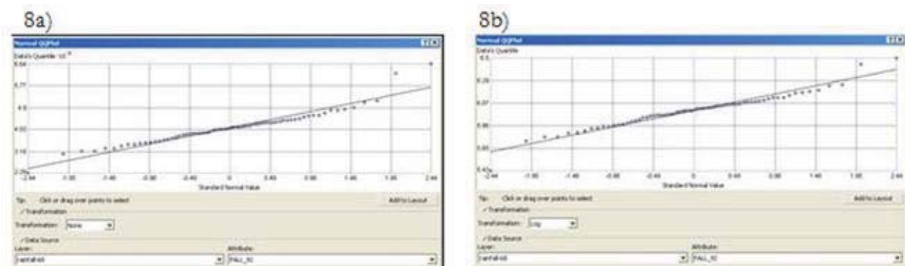

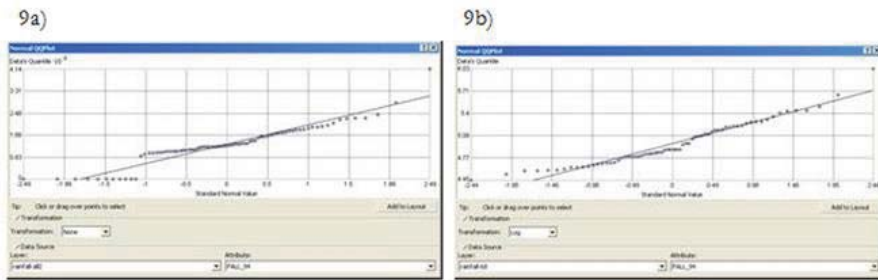**Figure 8:** a) Normal QQ plot, and b) Normal QQ plot with Log transformation of hydrological year for 1992.

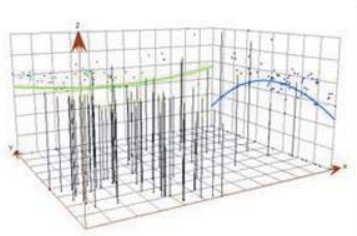**Figure 9:** a) Normal QQ plot, and b) Normal QQ plot with Log transformation of hydrological year for 1994.

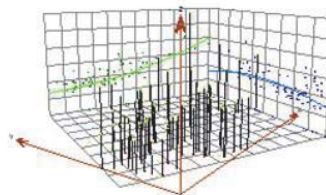**Figure 10** Trend analysis for hydrological year rainfall for 1992.

Figure 11 Trend analysis for hydrological year rainfall for 1994.

5.2 *Inverse Distance Weighted (IDW) interpolation result*

The results for both 1992 and 1994 from the IDW interpolation model are summarised in Table 1. Mean error for 1992 is -7.8 mm and the RMSE is 55.69 mm, while the mean error for 1994 is -3.77 mm and the RMSE is 32.36 mm. Figure 12 and 13 show the spatial rainfall interpolation results for 1992 and 1994 rainfall respectively. The CV result for the points within the boundary of the study area is presented in Table 2. The R^2 is 0.1141, MAE is 31.84 mm and RMSE is 43.65 mm for 1992. In 1994 the R^2 is 0.6373, MAE is 17.67 mm and RMSE is 21.36 mm. In both years, the CV using the data points within the study boundary produced lower RMSE compared to the first method. Figure 14a and 14b shows the scatter plot of the measured and estimated values from the data within the study area for both 1992 and 1994.

Table 1. Results from the IDW spatial rainfall interpolation of 1992 and 1994 rainfall data for Jornada basin.

	Inverse Distant Weighted	
	1992	1994
Mean error (mm)	-7.811	-3.768
RMS error (mm)	55.69	32.36

Table 2. Regression coefficient, mean absolute error and root mean squared error for the data within the study area boundary.

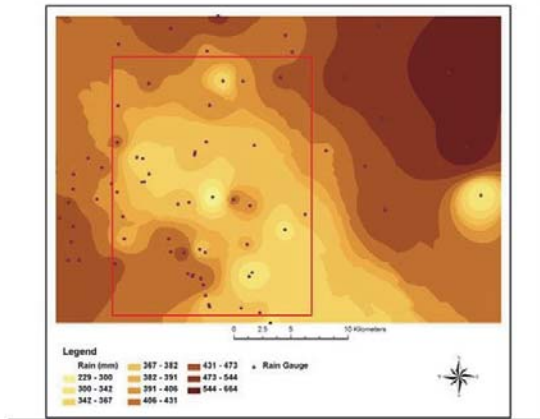|	Inverse Distant Weighted	
	1992	1994
R²	0.1141	0.6373
MAE (mm)	31.84	17.67
RMSE (mm)	43.65	21.36

Figure 12. IDW spatial rainfall interpolation for 1992 rainfall data.
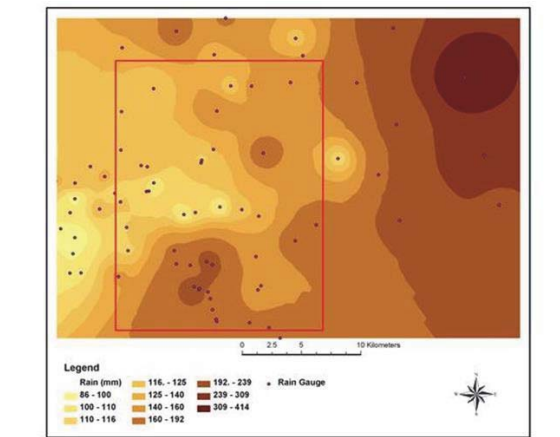

Figure 13. IDW spatial rainfall interpolation for 1994 rainfall data.
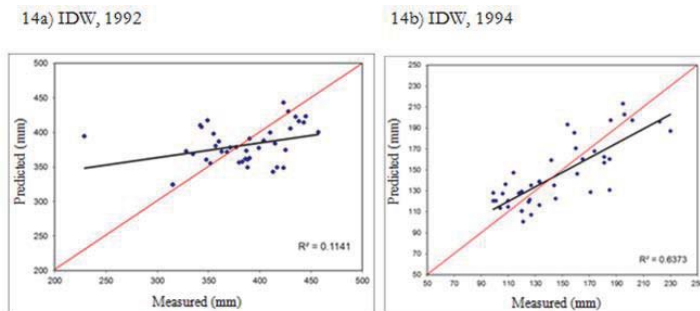

Figure 14. Scatter plots of measured values versus estimated values from IDW rainfall interpolation for 1992 and 1994

5.3 Accuracy of prediction

Table 3 summarized the prediction accuracy of two group of CV where the first group was based on all the available data points for each year and the second group was based on the data points found within the study-area boundary. The results from the IDW interpolation method used showed that the best fit model for both years where the mean error is -7.81 mm (1992), -3.77 mm (1994) and RMSE is 55.69 mm (1992) and 32.36 mm (1994) respectively. In the second CV analysis using the data within the study area boundary, the best fit model where MAE is 31.84 mm (1992), 17.67 mm (1994) and RMSE is 43.65 mm (1992) and 21.36 mm (1994).

Table 3. Prediction accuracy measures for IDW: A) result from using all data points, and B) result from using data points within the study area boundary.

	A) CV analysis on all data points within the Jornada basin				B) CV analysis on data points within the study area.					
	Mean error (mm)		RMSE (mm)		R^2		MAE (mm)		RMSE (mm)	
	1992	1994	1992	1994	1992	1994	1992	1994	1992	1994
IDW	-7.81	-3.77	55.69	32.36	0.11	0.64	31.84	17.67	43.65	21.36

Overall, the CV result improved by using the data points within the study area. This could be due to minimal edge effects as the study area is located in the centre of the distribution of rain-gauge. Moreover, the study area is located at the centre of the Jornada basin where there is little pronounced topography and away from the San Andreas mountain to the west hence is not affected by the higher rainfall recorded by the rain-gauges near the mountain area. The prediction showed that the 1994 data produced lower RMSE values for both set of CV. This result could be due to 1994 being a dry year and the centrally located study area means it received relatively homogenous rainfall throughout the year unlike the bajadas and mountainous area where occasionally it will still receive higher than the annual average rainfall during dry periods.

6. Conclusion

The analysis presented here illustrates the performance of spatial interpolation methods in the Jornada basin which exemplify the semi-arid environment. The Jornada basin rainfall data from a wet (1992) and dry (1994) years were used to test the spatial rainfall-interpolation method in ArcGIS. The IDW interpolation method was tested using both rainfall dataset. The result from the CV analysis of the total data points for both year showed that the IDW interpolation method produced a good fit model for the Jornada basin. A further analysis of CV of the predicted and measured values from data points within the study area boundary produced better fits compared to the CV result from all the data points within the Jornada basin. The result from constraining the rainfall-interpolation within the study-area boundary showed that the interpolation error from the edge effects can be minimised and result predicted values showed a closer fit to the measured values. Based on the outcome of the rainfall interpolation test, the deterministic IDW model provides a simple interpolation method with good results and close fit. The IDW interpolation method with constrained rainfall data points within the study boundary is therefore recommended to be used for the rainfall-interpolation work.

7. Acknowledgement

References

Bivand, R.S., E.J. Pebesma & V. Gomez-Rubio. (2008). Applied Spatial Data Analysis with R: Springer.
Blöschl, G., Grayson, R.B. (2000). Spatial observations and interpolation. In: Grayson, R.B., Blöschl, G. (Eds.), Spatial Patterns in Catchment Hydrology. Observations and Modelling. Cambridge University Press, Cambridge, pp. 17–50.
Burrough, P.A. & R.A. McDonnell. (1998). Principles of Geographical Information Systems. Oxford: Oxford University Press.
Dirks, K. N., J. E. Hay, C.D. Stow, and D. Harris. (1998). High resolution studies of rainfall on Norfolk Island part II: Interpolation of rainfall data, *Journal of Hydrology* 208, 187-193.
Goodale, C. L., J. D. Aber, and S.V. Ollinger. (1998). Mapping monthly precipitation, temperature, and solar radiation for Ireland with

polynomial regression and a digital elevation model, *Climate Research* 10, 35 – 49.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology. Amsterdam, Netherland, 228, 113–129.

Hevesi, J. A., J. D. Istok, and A. L. Flint. (1992). Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: Structural analysis, *Journal of Applied Meteorology* 31, 661 – 676.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei. (2008). Comparison of six methods for the interpolation of daily European climate data, *Journal of Geophysical Research* 113.

Isaaks, E.H, and R.M. Srivastava, (1989). An introduction to applied geostatistics, Oxford University Press: New York, 592 pp.

Johnston, K., Ver Hoef, M. J., Krivoruchko, K. and Lucas, N. (2001). Using ArcGIS Geostatistical Analyst. ESRI.

Justiniano, M. J., and T. S. Fredericksen. (2000). Phenology of tree species in Bolivian dry season, Biotropica, 32, 276–281.

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

Noy-Meir, I. (1973). Desert ecosystems: environments and producers. *Ann. Rev. Ecol. Syst.*, **4**, 398–417.

Peñuelas, J., I. Filella,X.Y. Zhang, L. Llorens, R. Ogaya, F. Lloret, P. Comas, M. Estiarte, and J. Terradas. (2004). Complex spatiotemporal phenological shifts as a response to rainfall changes, New Phytol., 161(3), 837–846.

Price, D. T., D. W. Mckenney, I. A. Nalder, M. F. Hutchinson, and J. L. Kesteven. (2000). A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data, *Agricultural and Forest Meteorology* 101, 81 – 94.

Sluiter, R. (2008). Interpolation methods for climate data: literature review. De Bilt, Royal Netherlands Meteorological Institute (KNMI).

Schwinning, S., Sala, O.E., Loik, M.E. & Ehleringer, J.R. (2004). Thresholds, memory, and seasonality: understanding pulse dynamics in arid/semi-arid ecosystems. *Oecologia*, 141, 191–193.

Wainwright, J. (2005). Climate and climatological variations in the Jornada Experimental Range and neighbouring areas of the US Southwest. Advances in Environmental Monitoring and Modelling, 2, 39–110.

Wainwright, J. (2006). Climate and climatological variations. In: Havstad, K.M., Schlesinger, W.H. and Huenneke, L.F. (eds.) Structure and Function of a Chihuahuan Desert Ecosystem: The Jornada LTER. Oxford University Press, Oxford. 44–80.

Vicente-Serrano, S. M., M. A. Saz, and J. M. Cuadrat. (2003). Comparative analysis of interpolation methods in the middle Ebro valley (Spain): Application to annual precipitation and temperature, *Climate Research* 24, 161–180.