Evaluation of an inverse distance weighting method for patching daily and dekadal rainfall over the Free State Province, South Africa

Mokhele Edmond Moeletsi[1,2]*, Zakhele Phumlani Shabalala[1], Gert De Nysschen[1] and Sue Walker[3]

[1]*Agricultural Research Council – Institute for Soil, Climate and Water, Private Bag X79, Pretoria, 0001, South Africa*
[2]*Risks AND Vulnerability Assessment Centre, University of Limpopo, Private Bag X1106, Sovenga, 0727, South Africa*
[3]*Crops for the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia*

ABSTRACT

Climate data recorded by national meteorological agencies is either incomplete or faulty for some periods due to a number of reasons. Multi-functional utilization of climate data in complete form necessitates the filling of these gaps. In this study an inverse distance weighting (IDW) method was used to estimate rainfall utilizing neighbouring station data in the Free State Province of South Africa. Six weather stations evenly distributed across the province, and with data for 1950 to 2008, were used to evaluate this patching IDW approach at daily and dekadal time steps. Coefficient of determination (r^2), mean absolute error (MAE) and mean bias error (MBE) were the statistics used in the assessment. Firstly, the study conducted a sensitivity analysis of the IDW exponent (p) which showed that the best results are obtained when p is either 2 or 2.5. The estimated values at all six stations were highly correlated with the measured rainfall data with an overall r^2 value exceeding 0.70 for both daily and dekadal estimates. MAE showed low miscalculations with values with an average of 1 mm per day and 4.4 mm per dekad. MBE was very low for both daily and dekadal evaluations but the disaggregated data showed underestimation of the IDW mostly for daily rainfall exceeding 10 mm. Thus, IDW methodology proved to be an acceptable approach for estimating both daily and dekadal rainfall in the Free State Province.

Keywords: estimation, missing data, neighbouring stations

INTRODUCTION

Measuring and archiving of different weather elements like rainfall, temperature or humidity is an important exercise. Long-term meteorological data can be used to influence decisions in different sectors including agriculture, aviation, hydrology and engineering. Accurate and complete climatological data is important for the successful design and operation of natural resource management systems (Gyau-Boakye and Schultz, 1994; Jeffrey et al., 2001). Weather data in South Africa dates back to the 1800s with rainfall being the most common element measured. There are a few stations around the country which recorded temperature, humidity and wind in the early 1900s but the number increased significantly from 1950 onwards. With the improvement in technology, automatic weather stations were introduced in South Africa in the 1990s and this helped to address shorter time-scale measurements like minute or hourly values. Increased frequency of measurements means better climate monitoring and understanding of weather occurrences with a life-cycle shorter than 1 day.

Regardless of the technology used to take the recordings, whether manual measurements or the use of electronic sensors, some of the data stored will be either faulty or missing. The gaps in the meteorological archives are caused mainly by absence of observers, vandalism, loss of records, data contamination, data-processing errors, effects of natural disasters like tornadoes or human-induced factors like wars, lack of funds for replacing broken instruments as well as instrument malfunctioning (Tang et al., 1996; Elshorbagy et al., 2000; Smithers and Schulze, 2000; Kim and Ahn, 2009; Villazón and Willems, 2010). Faulty data is mainly caused by observer's negligence, uncalibrated sensors and faultiness of the electronic sensors.

Missing or faulty climate data have to be estimated in order to provide a complete dataset, especially for modelling purposes. The accuracy of the estimations is dependent on a number of factors including the closeness of the stations used and the location of the patching stations in relation to barriers like mountains. There are three main techniques for estimating missing meteorological data, namely, empirical methods, statistical methods and function-fitting methods (Xia et al., 1999). The application of patching methods is dependent on the length of the gap, the season, climatic region, density of stations, and the characteristics of the data archived (Gyau-Boakye and Schultz, 1994).

There are a number of methods used to estimate missing rainfall values. The widely-used patching methods include: closest station, simple arithmetic averaging, inverse distance weighting, multiple regression and normal ratio (Tang et al., 1996; Makhuvha et al., 1997; Xia et al., 1999; Xia et al., 2001). In utilizing the closest station method, the nearest weather station with data corresponding to the period of concern is identified and missing values are either replaced directly by the value at the neighbour station or adjusted by a factor from the ratio of long-term means between the two stations (Xia et al., 2001). In simple arithmetic averaging, the missing data are obtained by arithmetically averaging data of the 2 to 5 closest weather stations around a station (Tang et al., 1996; Xia et al., 1999; De Silva et al., 2007). Inverse distance weighting utilizes the distances from the target station of 2 to 5 neighbour stations, giving more weight to data from the nearest weather station (Tang et al., 1996; Xia et al., 1999; De Silva et al., 2007; Chen and Liu, 2012). The multiple regression model employs step-wise regression to determine the coefficients for all the significant neighbour stations (Makhuvha et al., 1997; Xia et al., 1999), while the normal ratio method utilizes the correlation between the neighbour and target

---

station as well as the number of paired datasets as the weight in estimating values at the target station (Tang et al., 1996; De Silva et al., 2007).

The study aimed to evaluate the use of the inverse distance weighting (IDW) method of estimating rainfall as a possible means of patching missing or faulty rainfall measurements. The method was chosen based on recommendations by Chen and Liu (2012) in a study carried out in Taiwan which yielded a high correlation between measured and estimated values. De Silva et al. (2007) compared the normal ratio, arithmetic averaging and inverse distance methods and concluded that the IDW method performed better than the other two methods in estimating rainfall in all the climate zones of Sri Lanka. Studies by Tang et al. (1996) also revealed that the IDW method has the potential of estimating missing rainfall values with minimum inaccuracy.

METHODOLOGY

The study investigates the use of the IDW method for estimating daily and dekadal rainfall in the Free State Province of South Africa. This is a deterministic estimation method where values at unsampled points are determined by a linear combination of values at known sampled points (Collins and Bolstad, 1996). It employs Tobler's Law by estimating unknown measurements as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points (Collins and Bolstad, 1996; Longley et al., 2001; Oluoch et al., 2013). More specifically, IDW assumes that each measured point has a local influence that diminishes with distance and weights change according to the linear distance of the samples from the unsampled point (Li and Revesz, 2004). The spatial arrangement of the samples does not affect the weights (Collins and Bolstad, 1996). The weighting function is the inverse of the distance to the given exponent, so that the predicted value for a site is given by:

$$w(x,y)=\sum_{i=1}^{N} \lambda_i w_i \quad ; \quad \lambda_i = \frac{\left(\frac{1}{d_i}\right)^p}{\sum_{i=1}^{N} \left(\frac{1}{d_k}\right)^p} \tag{1}$$

where: $w(x,y)$ is the predicted value at location (x,y), N is the number of nearest known points surrounding (x,y), λ_i is the weight assigned to each known point value w_i at location (x_i,y_i), d_i is the Euclidean distance between each (x_i,y_i) and (x,y), and p is the exponent, which influences the weighting of w_i on w (Li and Revesz, 2004).

If Eq. 1 is used to determine w at a location where w has already been measured it will return the measured value, because the weight assigned to a point at zero distance is infinite; for this reason IDW is described as an exact method of interpolation because its interpolated results honour the data points exactly (Longley et al., 2001). The IDW derived value is an average and it always returns a value that is between the limits of the measured values, i.e., no point in the interpolated surface that is more than the largest measured value or less than the smallest value (Longley et al., 2001).

To assess the method of patching rainfall over the Free State Province, 6 weather stations that are evenly distributed over the province were selected. The other criteria for selecting the stations were based on the percentage of missing data, which had to be less than 10%, and the selected stations were also required to have over 20 years of data records. Table 1 shows geographical

TABLE 1
Geographical information of weather stations used in evaluating rainfall and temperature estimation methods

Station Name	Latitude (°)	Longitude (°)	Altitude (m)	Data period
Bethlehem	−28.1626	28.2953	1 631	1951–2008
Bloemfontein-Glen College	−28.9500	26.3333	1 304	1950–2008
Frankfort	−27.2667	28.5000	1 502	1950–2001
Hertzorgville	−28.2000	25.3000	1 326	1978–2000
Oukraal	−29.9333	24.6833	1 143	1978–2002
Welkom	−28.1333	26.6833	1 295	1964–2001

information for the target stations while Fig. 1 shows the spatial distribution of all the weather stations in the Free State Province. Daily rainfall data was obtained from the Agricultural Research Council – Institute for Soil, Climate and Water (ARC-ISCW) Agrometeorology database and the South African Weather Service (SAWS).

Before using the IDW methodology of patching, the accuracy of the method must first be tested. This was done by estimating daily rainfall data for the entire data range for each of the stations in Table 1 using Eq 1. A minimum of 2 and maximum of 5 nearest weather stations with data on that particular day of estimation were used. The methodology also considered only nearby stations within a 50 km radius of the target station (Table 2).

Sensitivity of the exponent (*p*)

To assess the sensitivity of the exponent (*p*) of the IDW equation, rainfall estimates were computed for *p* values ranging from 1 with an increment of 0.5 to 5. Results for each of the *p* values per target station were compared with observed rainfall data using coefficient of determination (r^2), mean absolute error (MAE) and mean bias error (MBE). The values were ranked by assigning the lowest rank on the value that resembles the observed data. For the r^2, the lowest rank was assigned to the exponent that results in the highest correlation. Ranking for the MAE and MBE was done in such a way that the lowest absolute value was assigned the lowest ranking value. An average of all the rankings was then obtained for stations and all the exponents, to determine exponents that result in good estimation of observed rainfall based on r^2, MAE and MBE.

Evaluation of estimation of rainfall at different thresholds

The accuracy of estimates was also determined for low rainfall values (below 5 mm), 5–10 mm, 10–15 mm, 15–20 mm and for high rainfall values exceeding 20 mm. Estimated values in days when observed values were at the stipulated thresholds were then compared with measured values using the MAE and MBE.

Statistical analysis

To analyse the performance of the methodology used at each climate station, both daily and dekadal (10-day basis) values
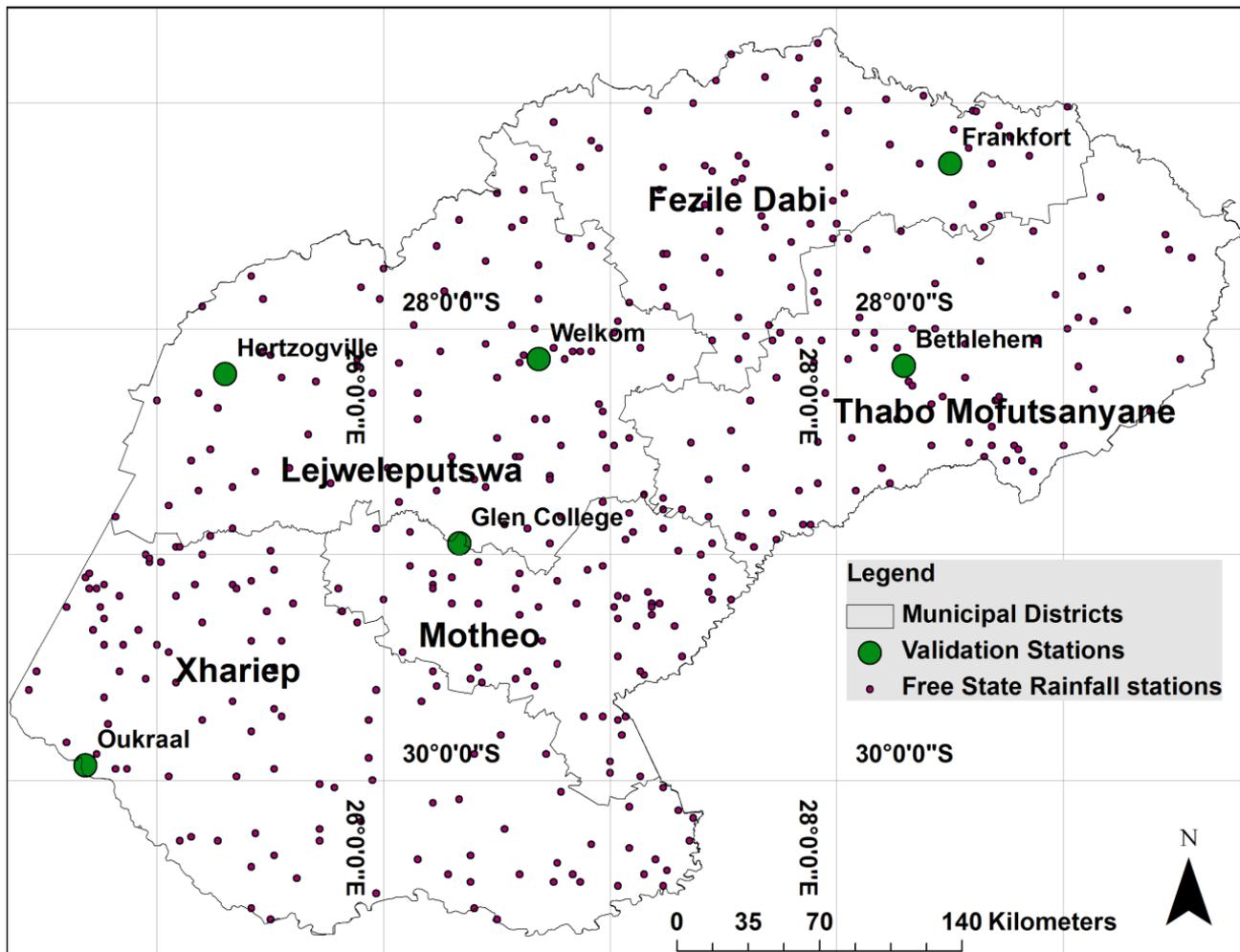
Figure 1
Spatial distribution of weather stations in the Free State Province used to evaluate the rainfall estimation method

Target station											
		Neighbour stations identification number (ID); distance from target station (d) in km									
Bethlehem	ID	14719	14720	14723	14716	14725	14700	14734	15264	14740	14696
	d	0.65	2.35	8.14	9.25	10.41	20.55	22.40	22.59	22.70	26.37
Bloemfontein	ID	19839	20386	14139	13786	20601	13770	20579	13771	13795	14145
	d	1.62	8.09	9.41	12.31	16.99	23.33	23.82	24.96	29.87	30.03
Frankfort	ID	15728	15713	15723	15733	15730	15704	19830	15274	15740	16167
	d	11.1	13.2	16.8	18.1	27.8	27.9	28.0	31.5	34.8	35.3
Hertozgville	ID	14546	14551	20774	21206	14559	15057	15058	21207	19899	14092
	d	14.7	17.0	21.7	21.7	24.6	34.1	34.8	35.3	35.8	37.6
Oukral	ID	13316	13317	13321	13310	13005	13319	12998	13302	13318	13007
	d	7.4	8.8	13	13.7	16.76	22.55	31.66	32.17	34.31	35.8
Welkom	ID	15115	14598	14592	15120	14591	14600	14601	14588	14605	14608
	d	10.0	12.4	13.9	14.9	17.9	19.8	19.8	22.5	22.6	26.9

TABLE 2
Distances from the target weather stations to the neighbour stations

were utilized. In crop water modelling, dekadal time steps can be used to assess rainfall effects on crop production (Araya and Stroosnijder, 2011; Moeletsi and Walker, 2012). Daily values were aggregated to dekadal values. The first 10 days of the month are grouped as the 1st dekad of the month, the second 10 days (11–20) as the 2nd dekad and last 8 to 11 days (depending on the month and year) as the 3rd dekad of the month. The estimated daily and dekadal values were then compared with the measured values in days/dekads where there were no missing values.

To find the correlation between the measured and estimated values, the r^2 statistic was used; for determining the deviations of the estimated values from the measured values, the MAE statistic was used; and to investigate overestimation or underestimation, the MBE statistic was used. The r^2 statistic was first determined by correlating all the daily or dekadal values at each of the weather stations to the estimated values and was determined on a monthly basis (dekadal values grouped by months). The MAE of dekadal estimated values against the measured dekadal values

was also determined on a monthly basis. The following are the formulas for MAE and MBE (Jacovides and Kontoyiannis, 1995):

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |d_i|$$

(2)

where: N is the number of data pairs and d_i is the difference between the ith predicted daily/dekadal value and the ith measured daily/dekadal value.

$$MBE = \frac{1}{N} \sum_{i=1}^{N} d_i$$

(3)

RESULTS AND DISCUSSION

Sensitivity of the inverse distance weighting exponent

Estimation of rainfall using IDW with exponents starting from 1 to 5 resulted in varying rankings across all the weather stations (Table 3). Rankings for r^2 are lower for exponents between 1 and 2.5 while high values are evident for exponents between 3 and 5. This indicates that estimating daily rainfall with the IDW equation assigning extremely high weights to the closest stations results in values that are less correlated to the measured data. The rankings for the MAE are low for exponents between 2 and 3.5 while the other exponents have relatively high rankings. Rankings for the MBE did not show any pattern across all the exponents. The average ranking of the r^2, MAE and MBE of the different exponents of the IDW equation and all the weather stations is lowest for both 2 and 2.5, with a value of 4.22 for both exponents. The results are contrary to the findings of Chen and Liu (2012), who obtained highly variable optimal exponents. Lowest value denotes top ranking and best estimator as compared to the observed values. The other high-ranked exponents are 1.5, 1 and 3, with average rankings of 4.56, 4.72 and 4.72, respectively. Sensitivity analysis of the exponent of the IDW equation showed the optimal values as 2 and 2.5, and thus for the remainder of the paper the exponent of 2 will be utilized to determine the accuracy of the IDW method in estimating measured rainfall values at the selected six weather stations in the Free State Province.

Performance of daily estimations against measured values

Coefficient of determination (r^2) for the comparison of daily measured rainfall and inverse distance weighting estimated values varied from one location to another (Fig. 2). The best correlated estimates were obtained for Bethlehem in the eastern Free State with a r^2 value of 0.77, followed by Bloemfontein with 0.71 over the central parts. A moderate r^2 of 0.569 was obtained in Welkom corresponding to a correlation coefficient of 0.754. The lowest r^2 values were obtained in Frankfort, Hertzogville and Oukraal with 0.406, 0.466 and 0.468 corresponding to correlation coefficients (r) of 0.637, 0.683 and 0.684, respectively. High values of r^2 are an indication that the IDW estimates of daily rainfall in the Free State Province resemble the measured daily data. Correlation coefficients (0.63–0.877) obtained in this study are lower than those obtained using IDW in the study by Chen and Liu (2012) in Taiwan where values exceeding 0.9 were attained. The contributing factors to high correlation are

mostly the distances of the stations from the target station with stations with lo correlation values mostly having relatively distant neighbour stations (Table 2). In contrast, the r^2 values obtained from this study are higher than those obtaine by Villazón and Willems (2010) in Bolivia.

Mean absolute error (MAE) for estimating all the rainfall values was lower than 1 mm over Bloemfontein, Oukraal and Bethlehem, at 0.70 mm, 0.72 mm and 0.81 mm, respectively (Table 4). Low MAE is an indication that, on average, the magnitude of the estimated values is close to the measured values. MAE values for Welkom and Hertzogville were between 1 and 1.5 mm while the highest MAE of 1.55 mm was attained for Frankfort. In other studies, MAE values of daily rainfall estimates exceeding 2.5 mm were achieved by Villazón and Willems (2010) while Xia et al. (1999) obtained a relatively lower average MAE of 0.75 mm in Germany. Furthermore, Teegavarapu et al. (2009) obtained good performance by the IDW method in estimating daily rainfall with MAE of 0.07 mm in Kentucky, USA.

Highly variable results were obtained when comparing daily rainfall estimates for measured values of less than or equal to 5 mm, between 5 and 10 mm, between 10 and 15 mm, between 15 and 20 mm and exceeding 20 mm (Table 4). MAE for measured rainfall of less than or equal to 5 mm ranged from 0.32 mm in Bloemfontein to 0.81 mm in Frankfort. This is an indication that the IDW method for all the stations estimated low rainfall values well. Measured rainfall in the range of 5 to 10 mm, as compared to estimated values, had MAE of between 2.38 and 5.37 mm with relatively low values at Bethlehem, Bloemfonte and Oukraal. For measured daily values between 10 and 15 mm, MAE ranged from 3.45 to 7.95 mm with an average of 5.69 mm. Measured values of between 15 and 20 mm had the lowest MAE for Bloemfontein and Bethlehem with values not exceeding 5mm, while all the other stations had a relatively higher MAE exceeding 7.5 mm. MAE for measured rainfall exceeding 20 mm was also low for Bloemfontein and Bethlehem with values not exceeding 10 mm, while the highest values of 19.1 mm and 20.1 mm were obtained for Oukraal and Hertzogville, respectively. The results clearly show that the IDW method has a large error in extremely high daily rainfall. This might be caused by the fact that most excessively high rainfall events are obtained from convective clouds which are distinctly localized in most southern African regions (Yakir and Morin, 201 Blamey and Reason, 2013).

Mean bias error (MBE) for estimated daily rainfall as compared with measured data has a slight negative bias with an average of −0.08 mm (Table 5). All stations recorded a negative bias of magnitude less than 0.1 mm with the excep tion of Hertzogville with −0.17 mm. The results differ from findings in a study by Bennett et al. (2007), whereby the IDW method resulted in both negative an positive MBE with magnitude ranging from −0.05 to 0.25. Low bias error is an indication of good performance by the IDW method in estimating daily rainfal Disaggregated data shows a slight positive bias for measured rainfall of 5 mm or less with an average of 0.32 mm, implying that IDW tends to overestimate daily rainfall for low rainfall values. All the other measured rainfall categories show negative bias indicating underestimation of the IDW method in estimatin rainfall exceeding 5 mm. Average biasness is 1.27 mm, 3.68 mm, 4.68 mm and 12.72 mm for 5–10 mm, 10–15 mm, 15–20 mm and greater than 20 mm, respec tively. In the upper end of all of the categories, performance of the IDW method at Frankfort, Hertzogville, Oukraal and Welkom is relatively poor as compared Bethlehem and Bloemfontein.

Performance of dekadal estimations against measured values

The results of the coefficient of determination (r^2) for the comparison of the measured dekadal rainfall and the estimated dekadal rainfall for the entire data are high at all the stations (Fig. 3). The r^2 values range from 0.776 in Frankfort to a high of 0.90 in Bloemfontein (Glen College), with all stations having an r^2 exceeding 0.75. This shows that the estimated dekadal rainfall using the IDW method over the Free State results in more than 75% variation in measured dekadal values. The values obtained are higher than that of daily values implyin that cumulative rainfall estimates tend to eliminate daily fluctuations resulting higher correlation. Hence, the IDW method using values from the 2 to 5 neares stations performs well in estimating actual dekadal rainfall amounts.

TABLE 3
Rankings for r^2, MAE and MBE per station and for each IDW equation exponent

r^2									
Stations	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
Bethlehem	1	2	3	4	5	6	7	8	9
Bloemfontein	4	1	2	3	5	6	7	8	9
Frankfort	1	2	3	4	5	6	7	8	9
Hertzogville	3	4	1	1	5	6	7	8	9
Oukraal	9	7	8	6	5	4	3	1	2
Welkom	1	2	3	4	5	6	8	8	7
MAE									
Stations	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
Bethlehem	9	5	2	1	3	4	6	7	8
Bloemfontein	9	8	6	3	1	2	4	5	7
Frankfort	1	2	3	4	5	6	7	8	9
Hertzogville	9	8	7	6	5	4	3	2	1
Oukraal	9	8	5	6	7	4	3	2	1
Welkom	1	2	3	4	5	6	7	8	9
MBE									
Stations	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
Bethlehem	1	2	3	4	5	6	7	8	9
Bloemfontein	8	9	6	7	5	4	3	2	1
Frankfort	1	4	3	2	5	8	7	6	9
Hertzogville	2	1	4	3	6	7	5	8	9
Oukraal	9	5	8	7	6	2	4	3	1
Welkom	6	9	7	8	2	5	3	4	1
Average of r^2, MAE & MBE	4.72	4.56	4.22	4.22	4.72	5.11	5.50	5.78	6.11

TABLE 4
Mean absolute error (MAE) values (in mm) of the comparison between measured daily rainfall and inverse distance weighting (IDW) estimated daily values

Stations	All values	≤ 5 mm	5–10 mm	10–15 mm	15–20 mm	>20 mm
Bethlehem	0.81	0.36	2.38	3.45	4.81	9.31
Bloemfontein	0.70	0.32	3.23	4.05	4.50	8.75
Frankfort	1.55	0.81	4.58	7.02	8.30	16.82
Hertzogville	1.21	0.56	5.37	7.95	9.76	20.14
Oukraal	0.72	0.35	3.34	5.47	7.78	19.08
Welkom	1.13	0.54	4.31	6.18	8.29	14.27
Average	**1.02**	**0.49**	**3.87**	**5.69**	**7.24**	**14.73**

TABLE 5
Mean bias error (MBE) values (in mm) of the comparison between measured daily rainfall and Inverse Distance Weighting (IDW) estimated daily values

Stations	All values	≤ 5 mm	5–10 mm	10–15 mm	15–20 mm	>20 mm
Bethlehem	−0.08	0.24	−0.62	−2.04	−3.22	−6.83
Bloemfontein	−0.01	0.19	−0.30	−1.62	−2.30	−5.83
Frankfort	−0.04	0.59	−1.56	−3.63	−6.16	−15.64
Hertzogville	−0.17	0.33	−1.99	−6.08	−5.93	−17.36
Oukraal	−0.09	0.23	−1.73	−4.19	−4.72	−18.74
Welkom	−0.08	0.37	−1.40	−4.49	−5.67	−11.89
Average	**−0.08**	**0.32**	**−1.27**	**−3.68**	**−4.67**	**−12.72**

Looking specifically at the correlation of this method within different months, one can see a lot of variation in r^2 for dekadal estimates from month to month and per station (Table 5). The values are still mostly above 0.70 in all the months for all the stations, indicating a good performance with few exceptions. On average the r^2 statistic ranges from 0.71 to 0.88 and in 10 out of 12 months the average exceeded 0.75. Bloemfontein shows consistently high values with all the monthly correlations exceeding 0.80. The monthly r^2 values in Bethlehem also exceed 0.75, showing good correlation except in December when the value is below 0.70. In contrast, Frankfort monthly r^2 values show a lot of variation between 0.55 and 0.91. The results clearly show that there is a good relationship between estimated and actual rainfall values.
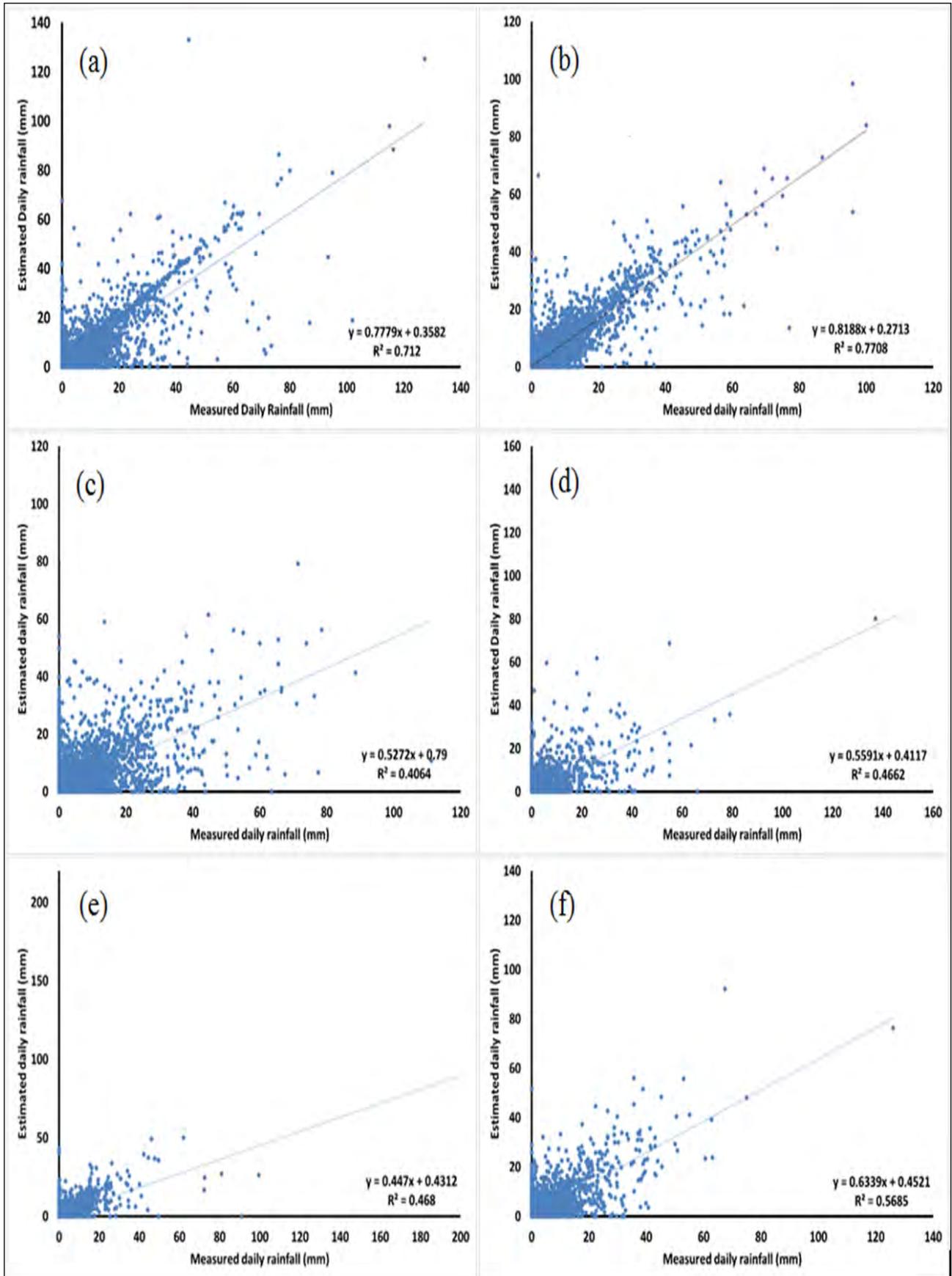
Figure 2

Comparison between measured daily rainfall AND *inverse distance weighting (IDW) estimated daily rainfall for (a) Bethlehem, (b) Bloemfontein (Glen College), (c) Frankfort, (d) Hertzogville, (e) Oukraal* AND *(f) Welkom*
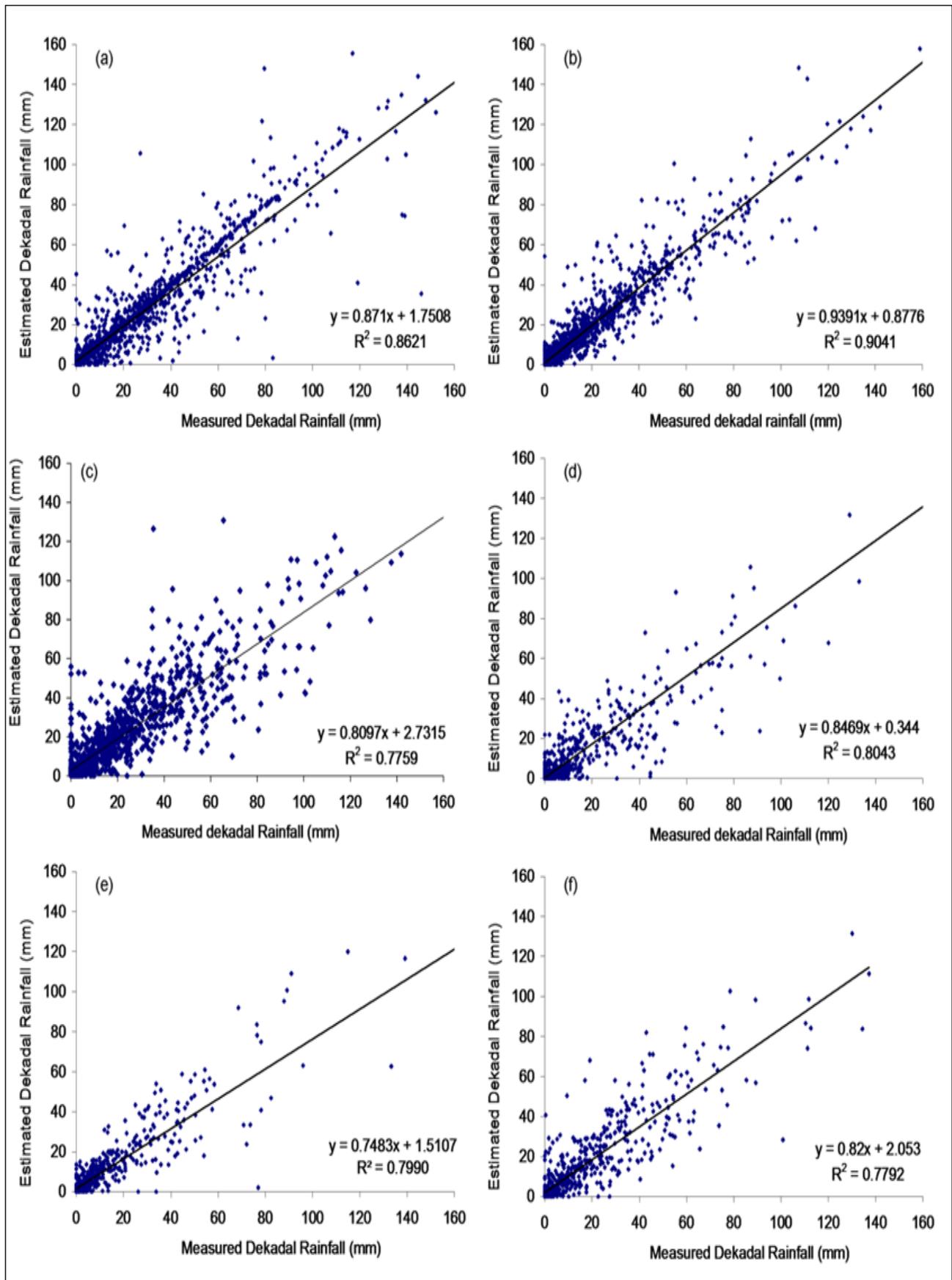
Figure 3

Comparison between measured dekadal rainfall AND inverse distance weighting (IDW) estimated dekadal rainfall for (a) Bethlehem, (b) Bloemfontein (Glen College), (c) Frankfort, (d) Hertzogville, (e) Oukraal AND (f) Welkom

	All values	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
r²													
Bethlehem	0.86	0.79	0.88	0.87	0.81	0.96	0.94	0.90	0.88	0.96	0.91	0.78	0.61
Bloemfontein	0.90	0.84	0.88	0.88	0.92	0.94	0.94	0.96	0.97	0.95	0.94	0.91	0.86
Frankfort	0.78	0.66	0.55	0.75	0.77	0.87	0.77	0.67	0.64	0.90	0.75	0.70	0.72
Hertzogville	0.80	0.70	0.85	0.87	0.88	0.53	0.73	0.63	0.95	0.71	0.79	0.73	0.83
Oukraal	0.80	0.58	0.89	0.81	0.83	0.82	0.92	0.95	0.98	0.86	0.95	0.71	0.80
Welkom	0.78	0.70	0.78	0.58	0.77	0.93	0.69	0.93	0.88	0.91	0.88	0.81	0.57
Average	**0.81**	**0.70**	**0.79**	**0.78**	**0.84**	**0.82**	**0.81**	**0.83**	**0.89**	**0.87**	**0.86**	**0.77**	**0.76**
MAE													
Bethlehem	3.80	7.4	5.4	5.1	3.3	1.2	0.7	0.8	1.4	2.0	4.3	5.9	7.5
Bloemfontein	3.17	7.8	7.3	5.9	3.7	1.3	0.8	0.6	0.8	1.4	3.5	5.3	5.4
Frankfort	4.92	14.5	11.9	8.1	5.3	1.9	1.0	1.0	2.6	2.8	7.0	9.5	11.2
Hertzogville	5.76	13.7	10.5	9.0	5.1	2.3	1.0	1.0	0.6	3.4	6.7	7.0	7.4
Oukraal	2.65	8.3	5.7	5.0	3.6	1.1	0.9	0.6	0.8	1.4	2.9	4.7	7.3
Welkom	5.60	11.9	10.8	10.6	5.2	1.3	0.9	0.4	0.8	2.4	6.3	7.3	8.7
Average	**4.42**	**11.2**	**9.24**	**7.71**	**4.59**	**1.55**	**0.91**	**0.74**	**1.11**	**2.26**	**5.28**	**6.76**	**8.01**
MBE													
Bethlehem	−0.83	−1.70	0.01	−0.86	−0.40	−0.35	−0.07	−0.01	−0.41	−0.73	−0.92	−2.77	−1.08
Bloemfontein	−0.06	0.14	2.56	0.14	−0.58	−0.33	0.04	−0.05	0.06	−0.37	−1.21	−1.30	0.14
Frankfort	−0.55	−4.55	0.52	−1.62	0.89	0.14	−0.33	0.40	−0.03	−0.41	−0.62	−2.74	−2.19
Hertzogville	−1.79	−7.47	1.58	−6.06	−3.68	−0.53	−0.51	−0.44	−0.01	−0.02	−0.38	−1.39	−2.31
Oukraal	−0.74	−3.93	−0.80	−1.14	−1.04	−0.12	−0.30	−0.10	0.20	0.46	−0.17	−0.48	−4.35
Welkom	−0.60	1.83	−1.72	−0.65	−0.36	0.11	0.20	0.28	−0.40	−0.89	−0.81	−1.33	−3.40
Average	**−0.75**	**−2.80**	**0.43**	**−1.87**	**−0.95**	**−0.15**	**−0.18**	**0.02**	**−0.04**	**−0.25**	**−0.64**	**−1.45**	**−2.42**

Dekadal MAE values show a lot of spatial and temporal variability over the Free State. Small MAE values were obtained in the winter months due to the low rainfall received, while in the other months the MAE values were relatively large (Table 6). Bethlehem, Bloemfontein and Oukraal show low error with MAE of less than 10 mm/dekad for all of the months. In Frankfort, Hertzogville and Welkom the errors obtained exceed 10 mm/dekad mostly in December to March while in other months there is a relatively low error of less than 10 mm/dekad. On average the MAE is below 5 mm/dekad.

MBE for dekadal measured rainfall versus dekadal IDW estimated values is mostly negative over all the months and stations denoting a tendency of underestimation (Table 6). The magnitude of underestimation is mostly less than 3 mm per dekad with winter months recording MBE of less than 0.2 mm in absolute terms whereas summer months show values exceeding 1 mm/dekad in magnitude.

The results obtained in evaluating the IDW method's capability for correctly estimating measured rainfall are in line with the findings by Xia et al. (1999), Bennett et al. (2007) and Teegavarapu et al. (2009), which show that IDW estimates rainfall well and is a recommended method of filling in the missing rainfall data. According to these researchers, the method performs best if the density of the stations is high. Similarly, the results also showed relatively lower accuracy at stations like Hertzogville, Oukraal and Frankfort which had distant neighbour stations, as compared to stations like Bethlehem and Bloemfontein where IDW performed very well. Overall, in the Free State there are over 500 rainfall stations with data within the period of 1950 to 2008, which makes it an ideal area for using the IDW method. The only constraint is the varying climate data length and period of recording. Even though IDW is a good estimator of daily rainfall, the number of rainy days increased by 5% to 29% with an average of 17% over the Free State. This is mainly attributed to the localized nature of the rainfall in the region: there are few days of rainfall from stratified clouds which mostly cover a large area, with the rain coming mainly from convective clouds which are generally enhanced by local conditions.

CONCLUSION

The accuracy of the inverse distance weighting (IDW) method of estimating daily and dekadal rainfall in the Free State study area was evaluated using 6 selected stations. The coefficient of determination (r²) for estimated rainfall versus observed rainfall for all the stations shows a good correlation, while mean absolute error (MAE) and mean bias error (MBE) values were low. This is an indication that the IDW method performs well and is thus recommended to fill in the gaps in rainfall data in the Free State Province of South Africa.

ACKNOWLEDGEMENTS

REFERENCES

ARAYA A and STROOSNIJDER L (2011) Assessing drought risk and irrigation needs in northern Ethiopia. *Agric. For. Meteorol.* **151** 425–436. http://dx.doi.org/10.1016/j.agrformet.2010.11.014

BENNETT ND, NEWHAM LTH, CROKE BFW and JAKEMAN AJ (2007) Patching and disaccumulation of rainfall data for hydrological modelling. In: Oxley L and Kulasiri D (eds) *Proceedings of the International Congress on Modelling* AND *Simulation MODSIM2007.* Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Christchurch. 1526–1532.

BLAMEY RC and REASON CJC (2013) The role of mesoscale convective complexes in Southern Africa summer rainfall. *J. Clim.* **26** 1654–1668. http://dx.doi.org/10.1175/JCLI-D-12-00239.1

CHEN F and LIU C (2012) Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy Water Environ.* **10** 209–222. http://dx.doi.org/10.1007/s10333-012-0319-1

COLLINS FC and BOLSTAD PV (1996) Comparison of spatial interpolation techniques in temperature estimation. In: *Proceedings of the Third International Conference/Workshop on Integrating GIS* AND *Environmental Modeling*, 21–25 January 1996, Santa Fe, New Mexico.

DE SILVA R, DAYAWANSA N and RATNASIRI M (2007) A comparison of methods used in estimating missing rainfall data. *J. Agric. Sci.* **3** 101–108.

ELSHORBAGY AA, PANU US and SIMONOVIC SP (2000) Group-based estimation of missing hydrological data: I. Approach and general methodology. *Hydrol. Sci. J.* **45** (6) 849–866. http://dx.doi.org/10.1080/02626660009492388

GYAU-BOAKYE P and SCHULTZ GA (1994) Filling gaps in runoff time series in West Africa. *Hydrol. Sci. J.* **39** (6) 621–636. http://dx.doi.org/10.1080/02626669409492784

JACOVIDES CP and KONTOYIANNIS H (1995) Statistical procedures for evaluation of evapotranspiration computing models. *Agric. Water Manage.* **27** 365–371. http://dx.doi.org/10.1016/0378-3774(95)01152-9

JEFFREY SJ, CARTER JO, MOODIE KB and BESWICK AR (2001) Using spatial interpolation to construct a comprehensive archive of Australia climate data. *Environ. Model. Softw.* **16** 309–330. http://dx.doi.org/10.1016/S1364-8152(01)00008-1

KIM T and AHN H (2009) Spatial rainfall model using a pattern classifier for estimating missing daily rainfall data. *Stoch. Environ. Res. Risk Assess.* **23** 367–376. http://dx.doi.org/10.1007/s00477-008-0223-9

LI L and REVESZ P (2004) Interpolation methods for spatio-temporal geographic data. *J. Comput. Environ. Urban Syst.* **28** (3) 201–227. http://dx.doi.org/10.1016/S0198-9715(03)00018-8

LONGLEY PA, GOODCHILD MF, MAGUIRE DJ and RHIND DW (2001) *Geographic Information Systems and Science.* Wiley, New York.

MAKHUVHA T, PEGRAM G, SPARKS R and ZUCCHINI W (1997) Patching rainfall data using regression methods: 1. Best subset selection, EM and pseudo-EM methods: Theory. *J. Hydrol.* **198** (1–4) 289–307. http://dx.doi.org/10.1016/s0022-1694(96)03285-4

MOELETSI ME and WALKER S (2012) Assessment of agricultural drought using a simple water balance model in the Free State Province of South Africa. *Theor. Appl. Climatol.* **108** 425–450. http://dx.doi.org/10.1007/s00704-011-0540-7

OLUOCH K., MARWAN N, TRAUTH M and KURTH J (2013) Evolving complex networks analysis of space-time multi-scale wave-like fields: application to African rainfall dynamics. *Geophys. Res. Abstr.* **15** EGU2013-13237.

SMITHERS JC and SCHULZE RE (2000) Long duration design rainfall estimates for South Africa. WRC Report No. 811/1/00, Water Research Commission, Pretoria.

TANG WY, KASSIM AHM and ABUBAKAR SH (1996) Comparative studies of various missing data treatment methods – Malaysian experience. *Atmos. Res.* **42** (1–4) 247–262. http://dx.doi.org/10.1016/0169-8095(95)00067-4

TEEGAVARAPU RSV, TUFAIL M and ORMSBEE L (2009) Optimal functional forms for estimation of missing precipitation data. *J. Hydrol.* **374** 106–115. http://dx.doi.org/10.1016/j.jhydrol.2009.06.014

VILLAZÓN MF and WILLEMS P (2010) Filling gaps and daily disaccumulation of precipitation data for rainfall-runoff model. *BALWOIS 2010*, 25–29 May 2010, Ohrid, Republic of Macedonia.

XIA Y, FABIAN P, WINTERHALTER M and STOHL A (1999) Forest climatology: estimation of missing values for Bavaria, Germany. *Agric. For. Meteorol.* **96** 131–144. http://dx.doi.org/10.1016/s0168-1923(99)00056-8

XIA Y, FABIAN P, WINTERHALTER M and ZHAO M (2001) Forest climatology: estimation and use of daily climatological data for Bavaria, Germany. *Agric. For. Meteorol.* **106** (2) 87–103. http://dx.doi.org/10.1016/S0168-1923(00)00210-0

YAKIR H and MORIN E (2011) Hydrologic response of a semi-arid watershed to spatial and temporal characteristics of convective rain cells. *Hydrol. Earth Syst. Sci.* **15** 393–404. http://dx.doi.org/10.5194/hess-15-393-2011