

A modified distance-weighted approach for filling annual precipitation gaps: application to different climates of Iran

Gholamreza Khosravi · Ali Reza Nafarzadegan ·
Ahmad Nohegar · Hasan Fathizadeh · Arash Malekian

Received: 11 August 2013 / Accepted: 3 January 2014 / Published online: 19 January 2014
© Springer-Verlag Wien 2014

Abstract The continuity of precipitation data is critical in climate, hydrological and environmental studies. The main objective of this study is to improve the geographical coordinate (GC) method for filling gaps in annual precipitation observations. The altitude ratio (A_i) and the mean annual precipitation (MAP) were employed through a modification procedure to enhance the efficiency of the GC technique. In this respect, 24 precipitation gauge stations located in six different regions of Iran were taken into consideration in order to investigate the performance of the modified approach. Furthermore, estimates of methods such as the normal ratio (NR), the linear regression (LR) and also the conventional GC were compared with the outcomes obtained from the improved methodology to evaluate whether it has adequate suitability in a broad range of geo-climatic conditions. Average values of model accuracy measures and also results of the priority ranking process indicated that the best method was modified GC and the second priority technique was NR. It was also revealed that the conventional GC occupied no place in the first priority class for all stations, and more interestingly, the modified GC has not been in the lowest priority at all. Hence, the proposed methodology was proven to be a reliable procedure for estimating missing annual

precipitation records. In addition to improved performance, the modified approach utilizes input data that are usually available in the most of data-limited areas such as developing countries.

1 Introduction

Having complete and reliable time series of meteorological inputs is one of the most basic issues in efficient environmental monitoring and modelling. However, weather stations are frequently subject to some malfunctioning in the monitoring period (Tardivo and Berti 2012). This happens due to many reasons such as improper data registration, absence of the observer, destruction of gauging devices, power failure and elimination of the incorrect data. The uninterrupted precipitation data is critical for climate and hydrological studies, especially in cases such as the frequency analysis of extreme events (e.g., floods and droughts) and the design of water systems. Breaks in records of data acquisition systems are more serious in developing countries; in these countries, for most parts, hydrological data series are full of gaps (e.g., Dastorani et al. 2009). Thus, in order to be able to employ these data more efficiently, the missing observations need to be reconstructed.

Reliable estimation of missing climatological data has been a topic of interest for meteorologists, hydrologists and environment managers all over the world (e.g., Schneider 2001; Jeffrey et al. 2001; Geerts 2003; Kashani and Dinpashoh 2012; Li et al. 2013). The imputation of missing values is typically based on measurements for the same location (within-station) when data is available or on observations from nearby stations (between-station) for the same day (Kim and Pachepsky 2010). Short breaks in weather records can be reconstructed by simple within-station approaches, such as interpolation between available observations or

G. Khosravi · A. R. Nafarzadegan (✉) · A. Nohegar
Department of Watershed Management Engineering, Faculty of
Agriculture and Natural Resources, Hormozgan University,
P.O. Box 3995, Bandar Abbas, Iran
e-mail: a.r.nafarzadegan@gmail.com

A. Nohegar
Faculty of Environment, University of Tehran, Tehran, Iran

H. Fathizadeh
Faculty of Agriculture, Ilam University, Ilam, Iran

A. Malekian
Faculty of Natural Resources, University of Tehran, Tehran, Iran

moving averages (Kemp et al. 1983), or by using records before and after the date of missing observations in a nonlinear regression (Acock and Pachepsky 2000). However, utilizing the measurements of neighboring stations for filling data gaps has been done much more often, and also it seems that it has been much more accurate. In this regard, Tardivo and Berti (2012) have pointed out that “when the length of the missing period increases, between-station approaches, considering the specific variability of climate in the period to be reconstructed, tend to give better results.” On application of between-station approaches, Kim and Pachepsky (2010) have referred to some studies that take advantage of the geometrical distances between stations to use the values from the nearest station (Xia et al. 1999), the arithmetic averaging of observations from several adjacent stations (Willmott et al. 1994), linear regressions (LRs) to estimate the missing value as a median of the distribution of regression outputs (Ramos-Calzado et al. 2008) and distance-weighted techniques (Teegavarapu and Chandramouli 2005).

Many researchers assessed between-station approaches for reconstructing missing meteorological and hydrological measurements in different climates of Iran. Sadatinejad (1997) investigated the reference station, normal ratio (NR), geographical coordinate (GC), simple and multiple LR and autoregressive methods for the completion of annual precipitation gaps in Esfahan province. He suggested the NR method for arid and Mediterranean climates and the LR method for semi-dried climates as the best methods. Lookzadeh (2004) compared the NR, inverse squared distance, simple and multiple LR and geostatistical techniques for filling gaps in monthly, seasonal and yearly precipitation data in the central Alborz region. The results showed that the NR method is the best method in 69.2 % of cases; meanwhile, this method gives overestimates and underestimates in wet and dry years, respectively. Sadatinejad et al. (2009) examined the NR, inverse squared distance, GC, simple and multiple LR, autoregressive, artificial neural network (ANN) and fuzzy regression methods for reconstructing missing monthly discharge measurements of great Karoon Basin and suggested ANN as the most accurate method. In order to evaluate the performance of ANN and ANFIS models in filling breaks in flow data, Dastorani et al. (2009) employed two traditionally used methods (the NR and the correlation methods) and concluded that all four techniques presented acceptable predictions, however, the ANFIS technique indicated a superior ability to predict missing flow values, especially in arid land stations with variable and heterogeneous data. Khorsandi et al. (2011) compared the ANN, NR, inverse distance weighting (IDW), and GC methods to compensate missing monthly precipitation records at three stations around the city of Esfahan. They suggested ANN as the most efficient approach in comparison with the other three methods. Kashani and Dinpashoh (2012) investigated the capabilities of 11 different traditional and data-

driven approaches to identify the best gap-filling techniques for hydro-meteorological observations in three distinct climates of Iran. Their results demonstrated the high importance of choosing the appropriate method for completing climatological datasets in Iran as well as other arid and semi-arid regions.

Finally, it is noteworthy that meteorological data in developing countries are full of gaps, and there is a pressing need to propose gap-filling techniques with no detailed requirements (i.e., appropriate for most parts of developing countries) and, of course, with satisfactory outcomes. The main objective of this work is to modify the GC method (Mahdavi 1998; Khorsandi et al. 2011) for reconstructing missing annual precipitation measurements based on a procedure that is applicable for regions having no detailed data. Furthermore, the improved methodology is carried out for different sites of Iran, and results are compared with estimates of the conventional GC, the NR, and the LR approaches to evaluate the efficiency of the modified GC technique and to study its suitability in various climatic and geographical conditions of Iran.

2 Materials and methods

2.1 Data description

In this study, 24 precipitation gauge stations located in different parts of Iran (Fig. 1) were selected and organized into six groups according to geographical proximity and climatic similarities. These gauges are under the supervision of Iran Water Resources Management Company. As can be seen in the web portal of the company, it acts as an agency of the Ministry of Energy to enforce the Law of Fair Distribution of Water and other rules and regulations related to water, including the management and control of water resources operations, and also to collect, prepare, provide and analyze basic information required for studying Iran water resources quality and quantity (<http://www.wrm.ir/english/tabid/198/Default.aspx>). Considering its consistency and relevance, the data provided by the company has been used in many of water resources management projects and research studies in Iran. However, the homogeneity of precipitation data series at selected stations was examined by the runs test, which was accepted at the 95 % confidence level.

In each group, every station was experienced to be considered as a target station and also as a neighboring station. Then, by omitting a portion (about 15 %) of annual precipitation measurements for the target station, the attempt was made to estimate them using different “between-station” methods.

The general properties of stations and the monitoring period (i.e., common record-length) for each group are presented in Table 1. The De Martonne aridity index was applied to determine climate types. As shown in Table 1, six groups of

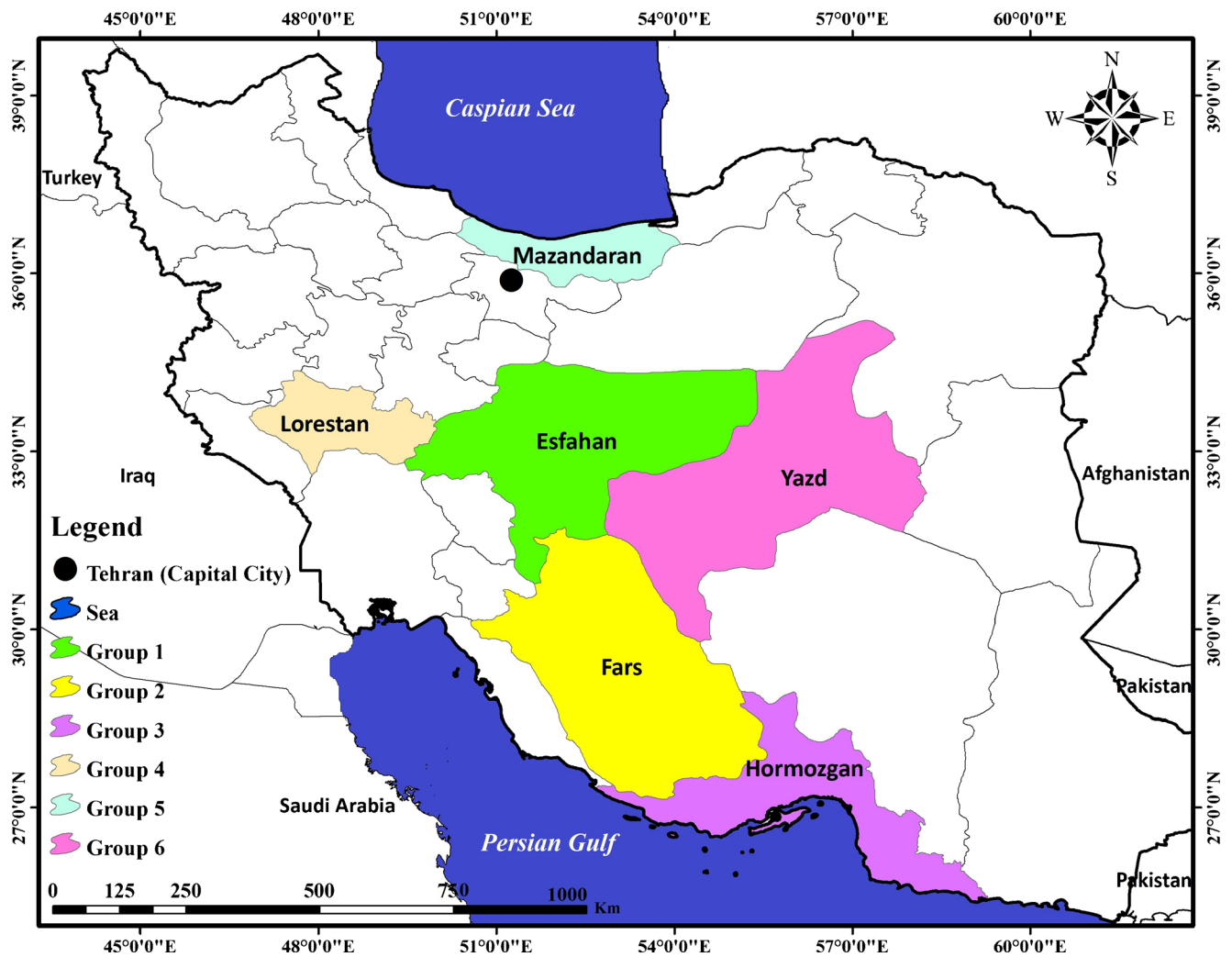


Fig. 1 General location of studied regions in the map of Iran

stations cover a wide variety of climatic conditions ranging from arid to humid.

2.2 Estimation of missing precipitation data

2.2.1 Normal ratio method

This traditional statistical pattern recognition method utilizes the normal precipitation (Chow 1964; Linsley et al. 1988) of the target station and nearby stations (usually three) to reconstruct missing values.

The data recorded at index (neighboring) stations are weighted by the ratios of the normal annual precipitation values. Thus, the formula is given as (Abebe et al. 2000):

$$P_X = \left(\frac{1}{3}\right) \left(\frac{N_X}{N_A} P_A + \frac{N_X}{N_B} P_B + \frac{N_X}{N_C} P_C\right) \quad (1)$$

where X is the target station, A , B and C are index stations, P_X is the missing precipitation record at the target station to be estimated, and P and N represent precipitation and normal precipitation (mean of index period) values, respectively.

2.2.2 Linear regression

In this traditional common approach, the correlation between observations at the target station and each of neighboring stations is assessed using the following equation (Eq. 2):

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} \quad (2)$$

Table 1 General characteristics of stations arranged in six groups

Group	Province	Station	Latitude (N)	Longitude (E)	Elevation (m)	Index period (years)	Climate
1	Esfahan	Esfahan	32°38'10"	51°41'19"	1,568	1968–2010	Arid
		Pol kale	32°22'23"	51°15'50"	1,771	1968–2010	
		Ziar braan	32°30'22"	51°56'24"	1,540	1968–2010	
		Varzaneh	32°25'10"	52°38'49"	1,475	1968–2010	
2	Fars	Kaftar	30°31'15"	52°24'18"	2,350	1982–2007	Semi-arid
		Sedeh	30°30'35"	52°12'27"	2,160	1982–2007	
		Emamzadeh Ismael	30°19'12"	52°35'32"	1,680	1982–2007	
		Mansur abad	30°12'33"	52°27'11"	1,724	1982–2007	
3	Hormozgan	Faryab	27°32'24"	56°14'02"	480	1988–2011	Arid
		Sargalm khorgu	27°24'24"	56°20'28"	260	1988–2011	
		Kuh-e Geno	27°25'60"	56°10'08"	1,600	1988–2011	
		Sarkhoon	27°23'60"	56°20'01"	160	1988–2011	
4	Lorestan	Ghasem abad	33°25'32"	49°29'24"	1,830	1980–2010	Sub-humid
		Cham zaman	33°24'19"	49°24'18"	1,840	1980–2010	
		Dareh takht	33°21'24"	49°22'13"	1,940	1980–2010	
		Kamandan	33°18'13"	49°26'32"	2,080	1980–2010	
5	Mazandaran	Nozar abad	36°48'38"	53°14'31"	10	1972–2010	Humid
		Sefid chah	36°40'06"	53°53'05"	1,040	1972–2010	
		Abelou	36°12'60"	53°41'60"	1,740	1972–2010	
		Bar kola	36°38'36"	53°45'37"	1,430	1972–2010	
6	Yazd	Deh bala	31°36'12"	54°07'39"	2,606	1990–2009	Semi-arid
		Feiz abad	31°42'27"	53°59'17"	1,935	1990–2009	
		Tezerjan	31°39'16"	54°09'23"	2,220	1990–2009	
		Soltaneb	31°35'22"	54°01'13"	2,540	1990–2009	

where y is the available data series at the target station, x is the data series of the index station, and n is the number of pairs of observations.

The amount of correlation coefficient (r) between the measurements of the target site and the index site must be whether in an acceptable level according to the Fisher table (Dastorani et al. 2009) or significant at a confidence level according to the Student t -test (Eq. 3).

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (3)$$

If the value of the test statistic (t) falls in the critical region which is defined based on the degrees of freedom (i.e., $n-2$) and the desired significance level (e.g. $\alpha=0.05$), there is a statistically significant linear relationship between y and x (see Chatterjee and Hadi 2006).

As shown in this approach, only one index station is applied to fill gaps of the target station; Thus, in this study, for each group of stations, the correlation coefficient (r) between the records of the target station and the other sites in the same group was separately

computed, and the station having higher significant correlation with the target site was selected as index station. The missing observations at the target station were then predicted using the following equation:

$$Y = a + bX \quad (4)$$

where Y is the missing record to be estimated and X is the record at the index site corresponding to the missing observation at the target site (Y). The values of a and b can be estimated using Eqs. 5 and 6, respectively:

$$a = \bar{y} - b\bar{x} \quad (5)$$

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (6)$$

where \bar{y} and \bar{x} are mean values of the data series, respectively, at the target and the index stations.

It is also feasible to estimate the precipitation at the target station (Y) using the precipitation at more than one index station (i.e., X_1, X_2, X_3, \dots) through multiple linear regression (MLR). However, since the MLR relies on more than one neighboring station, it intrinsically assumed that the distance–correlation relationship is stationary over a possibly large area that covers all index stations. This is not satisfactory, especially for precipitation because of its higher spatial variation (Tardivo and Berti 2013) compared with other meteorological variables such as temperature.

2.2.3 Geographical coordinate method

This approach utilizes geographic coordinates (GCs) of stations to determine weight coefficients for them. So, the target station is considered as the center point of a Cartesian coordinate system and location points for surrounding stations is specified in the coordinate plane using their GCs (the longitude and the latitude). Then, the distance of each surrounding station is computed according to the center point of the coordinate system.

This technique, as well as other IDW methods, considers a more significant role for closer surrounding stations in order to fill gaps in measurement series at the target station. Thus, the higher weight coefficients will be assigned to nearer stations as follows:

$$W_i = \frac{1}{X_i^2 + Y_i^2} \tag{7}$$

where W_i is the weight coefficient for station i , and X_i and Y_i are the longitude and latitude of the station, respectively.

By calculating the weight coefficients for neighboring stations, precipitation values at the target station can be reconstructed using the following equation (Mahdavi 1998; Khorsandi et al. 2011):

$$P_x = \frac{\sum_{i=1}^N (W_i \times P_i)}{\sum_{i=1}^N W_i} = \frac{W_A \cdot P_A + W_B \cdot P_B + \dots}{W_A + W_B + \dots} \tag{8}$$

where P_i is the precipitation record at surrounding station i (A, B, \dots), P_x is the missing precipitation record at the target station to be predicted and N is the number of surrounding stations.

2.3 Performance criteria

In this study, two performance evaluation criteria are employed. The Nash–Sutcliffe model efficiency coefficient (E) and the mean absolute error (MAE) are calculated by Eqs. 9 and 10, respectively.

$$E = 1 - \frac{\sum_{i=1}^N (O_i - R_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \tag{9}$$

$$MAE = n^{-1} \sum_{i=1}^N |(O_i - R_i)| \tag{10}$$

where O_i is the observed value, R_i is the reconstructed value, \bar{O}_i is the average of the observed values, and n is the number of data points.

The MAE is chosen because of the results obtained by Willmott and Matsuura (2005). They recognized MAE as a more natural and a less ambiguous measure of average error compared with the RMSE (root mean square error). Based on their findings, inter-comparisons of average model-performance error should carry out based on MAE.

3 Results and discussion

3.1 Shortcomings of the GC method

According to the GC-based estimates for annual precipitation amounts in different sites of Iran, the GC method exhibits a low accuracy. This could be due to the following shortcomings in this approach.

- (a) When the mean annual precipitation (MAP; for the index period) at control stations is more or less than its amount at the target station, the GC method could not estimate the missing data with appropriate accuracy. For instance, if the MAP at all or the most of adjacent stations is over 500 mm and at the target station is less than 500 mm, then the estimated amount obtained by the GC method would be more than 500 mm that is an overestimation. Also, if the MAP amount at control stations is less than its amount at the target station, this would lead to underestimated values.

(b) According to the basic assumption of the IDW approaches stated by Nafarzadegan et al. 2012, “nearby points ought to be more closely related than distant points to the value at the interpolate location. In other words, the IDW method applies the idea that the rate of influence in specific location decreases with increasing the distance from particular points”, the GC method as well as other IDW techniques allocate a more contributing role (weight) to the closer control stations to estimate the missing values of the target station. Hence, it is possible that an adjacent station having more altitude and climate similarity with the target station receives a lower weight (or even the lowest weight) than the other neighboring stations.

1.If there is a significant correlation at the 95 % confidence level between the altitude and the MAP of stations in each group, the following equation is presented.

$$P_x = \frac{\sum_{i=1}^N (W_i \times P_i) \cdot C_i \cdot A_i}{\sum_{i=1}^N W_i} = \frac{W_A \cdot P_A \cdot \frac{MAP_T}{MAP_A} \cdot \frac{\log E_T}{\log E_A} + W_B \cdot P_B \cdot \frac{MAP_T}{MAP_B} \cdot \frac{\log E_T}{\log E_B} + \dots}{W_A + W_B + \dots} \tag{11}$$

$$C_i = \frac{MAP_T}{MAP_i} \tag{12}$$

$$A_i = \frac{\log E_T}{\log E_i} \tag{13}$$

where P_x is the annual precipitation at the target station to be reconstructed for a specific year, P_i is the annual precipitation at the control station i (A, B, \dots), W_i is the weighting coefficient of

3.2 Modification of the GC method

The authors have taken advantage of the altitude ratio (A_i) and the MAP in order to modify the GC method. Eventually, two formulae are suggested. It is worth noting that choosing each of them to estimate annual precipitation records depends on physiographic (e.g., altitudes of stations) and climatic (e.g., MAP) conditions of the target region.

Table 2 Selection of the suitable equation for each group of stations based on the proposed methodology for reconstructing omitted observations

Group	Station	Altitude (m)	MAP (mm)	Correlation (<i>r</i>)	P value	Preferred equation	Omitted data
G1	Esfahan	1,568	132.6	0.99	0.006	Eq. 11	6 years: 1969, 1989, 1992, 1999, 2007, 2010
	Pol kale	1,771	190.3				
	Ziar braan	1,540	113.5				
	Varzaneh	1,475	99.7				
G2	Kaftar	2,350	525.1	0.79	0.203	Eq. 14	5 years: 1988, 1994, 1997, 2002, 2005
	Sedeh	2,160	521.2				
	Emamzadeh Ismael	1,680	506.2				
	Mansur abad	1,724	519.1				
G3	Faryab	480	170.1	0.43	0.569	Eq. 12	5 years: 1991, 1997, 1999, 2001, 2009
	Sargalm khorgu	260	275.3				
	Kooh-e Geno	1,600	294.1				
	Sarkhoon	160	247.3				
G4	Ghasem abad	1,830	533.5	0.97	0.028	Eq. 9	6 years: 1981, 1986, 1997, 2003, 2005, 2008
	Cham zaman	1,840	567.6				
	Dareh takht	1,940	609				
	Kamandan	2,080	663.3				
G5	Nozar abad	10	676.5	0.07	0.934	Eq. 12	6 years: 1979, 1988, 1990, 1993, 1998, 2003
	Sefid chah	1,040	424.9				
	Abelou	1,740	734.4				
	Bar kola	1,430	502.5				
G6	Deh bala	2,606	347.5	0.98	0.024	Eq. 9	5 years: 1990, 1993, 1995, 1998, 1999
	Feiz abad	1,935	120.8				
	Tezerjan	2,220	255.7				
	Soltaneb	2,540	308.4				

control stations, MAP_T is the MAP at the target station during the index period, MAP_i is the MAP at control stations $i(A, B, \dots)$ during the index period, E_T is the elevation at the target station, E_i is the elevation at the control station $i(A, B, \dots)$, C_i is the MAP ratio that is calculated by Eq. 12, and A_i is the logarithmic ratio of altitude that is calculated by Eq. 13.

As is obvious, when the MAP_i value is less than MAP_T , the value of C_i (the MAP ratio) becomes more than 1, thus increasing the estimated amounts for reconstructing gaps, and vice versa. The logarithmic ratio of altitude (A_i) also can play a significant role in reaching more accurate gap-filling. As mentioned above, this could occur when a high mutual connection exists between MAPs of stations under consideration and their elevation values (for instance, the higher the altitude values, the higher the MAP amounts become). In such circumstances, this coefficient (A_i) will alter (e.g., increase or decrease) the values estimated by the GC method for filling the precipitation gaps according to the nature of altitude–MAP relationship in each region.

2. If there is no acceptable correlation between the MAP and the elevation of stations in the target area, the altitude ratio will not be used in the modified equation of the GC method; thus, it suggested that:

$$P_x = \frac{\sum_{i=1}^N (W_i \times P_i) \cdot C_i}{\sum_{i=1}^N W_i}$$

$$= \frac{W_A \cdot P_A \cdot \frac{MAP_T}{MAP_A} + W_B \cdot P_B \cdot \frac{MAP_T}{MAP_B} + \dots}{W_A + W_B + \dots} \tag{14}$$

3.3 Comparison with other methods

In the following, outcomes obtained from the modified GC method for different parts of Iran are compared with the

Table 3 Values of the Nash–Sutcliffe efficiency coefficient (E) and MAE criterion used for showing the accuracy of estimates obtained from applied methods

Group	Station	Gap-filling technique							
		Geographical coordinates				Normal ratio		Linear regression	
		Conventional		Modified		E	MAE	E	MAE
		E	MAE	E	MAE				
G1	Esfahan	0.91	8.30	0.91	7.86	0.94	6.40	0.86	11.50
	Pol kale	-1.73	66.56	0.52	21.53	0.45	24.17	0.49	24.27
	Ziar braan	-0.59	31.89	0.84	10.08	0.89	8.45	0.85	9.07
	Varzaneh	-0.77	33.23	0.91	6.50	0.83	8.72	0.88	6.65
G2	Kaftar	0.96	25.32	0.97	21.73	0.97	24.13	0.97	18.54
	Sedeh	0.98	20.17	0.98	21.2	0.98	21.36	0.99	15.08
	Emamzadeh Ismael	0.88	43.79	0.94	28.47	0.95	23.16	0.96	25.20
	Mansur abad	0.99	10.60	0.99	10.47	0.99	12.28	0.99	14.58
G3	Faryab	-3.79	87.37	0.45	30.09	0.50	28.47	0.54	30.84
	Sargalm khorgu	0.66	57.00	0.91	31.29	0.92	29.66	0.95	20.50
	Kuh-e Geno	0.79	48.51	1.0	5.70	0.99	6.48	0.94	23.08
	Sarkhoon	0.97	15.98	0.99	7.26	0.99	8.01	0.98	12.60
G4	Ghasem abad	-0.22	91.54	0.94	15.98	0.92	18.22	0.83	28.65
	Cham zaman	0.79	33.45	0.92	20.82	0.92	21.90	0.53	48.50
	Dareh takht	0.99	7.09	0.99	7.36	0.99	6.05	0.98	12.97
	Kamandan	0.26	85.03	0.99	6.71	0.98	10.82	0.94	21.10
G5	Nozar abad	0.82	59.85	0.91	50.76	0.87	54.28	0.75	77.73
	Sefid chah	-5.06	176.08	0.75	27.92	0.71	27.08	0.71	35.93
	Abelou	-0.98	205.43	0.67	66.14	0.60	79.21	0.73	52.75
	Bar kola	0.69	57.07	0.78	52.30	0.70	55.05	0.62	62.79
G6	Deh bala	-0.74	82.97	0.84	22.96	0.86	21.09	0.87	19.82
	Feiz abad	-53.19	168.25	0.55	14.77	0.51	14.30	0.38	15.32
	Tezerjan	0.91	14.08	0.93	11.39	0.89	16.30	0.87	16.06
	Soltaneb	0.80	24.43	0.90	18.36	0.88	18.47	0.96	12.22

estimates of the conventional GC, the NR, and the LR techniques to evaluate the efficiency of the improved methodology and also to study whether the suitability of each approach varies with factors like topography and climatic zone.

As mentioned earlier, 24 stations located in different parts of Iran were selected and classified in six groups due to geographical proximity and climatic similarities. For performing the modified GC approach, the appropriate equation (Eq. 11 or Eq. 14) should be chosen based on the proposed methodology. Table 2 contains values of the altitude and the MAP for all stations. This table reveals the preferred equation for each group of stations based on the recommended procedure for modifying the GC approach. As shown in Table 2, outputs of correlation analysis for six groups demonstrate a broad range of values that are between 0.07 and 0.99. This could be interpreted as existence of various geo-climatic (Eccel et al. 2012) conditions in six selected territories.

Table 3 presents the performance of considered techniques in gap-filling at each station (i.e., a target station) based on the data of the rest of the stations in the group (i.e., neighboring stations). Of the four methods, the modified GC approach provided the best results, with an average MAE of 21.57 mm and an average E of 0.86. The NR technique had an average MAE of 22.67 mm and an average E of 0.84, and the LR method had an MAE of 25.66 mm and an average E of 0.82. The conventional GC method exhibits the worst outcomes with an average MAE of 60.58 mm and an average E of -2.28 .

It is worth to emphasize that the accuracy of a gap-filling procedure using nearby stations varies among climate variables and has been the least successful for precipitation. This is due to the highly stochastic nature of precipitation and its higher space gradient compared with other weather variables such as temperature (Kim and Pachepsky 2010; Tardivo and Berti 2013), and consequently, difficulties in formulating it into precipitation estimation algorithms (Daly et al. 1994; Thornton et al. 1997; Xia et al. 1999). Despite the above limitations, it still seems feasible to recognize one or possibly even two satisfying techniques (among examined methods) for the significant number of stations under consideration.

Utilizing MAE results (Table 3), a priority setting process was performed at each station. In this way, the accuracy order for reconstructing methods was determined based on the fact that the method with the least MAE value at a given station would reasonably be the best technique (i.e., the first priority) for gap-filling. Ranking results for methods at each station according to priority scores are showed in Table 4. This table indicates that none of the examined approaches for completing precipitation gaps has dominant preference for all studied regions.

For a more rigorous analysis to address the issue of reliability of results, the bootstrap technique (Good 1999) could be employed to evaluate the uncertainty in outcomes of

Table 4 Sorting of methods according to their priority rank (class) for each station

Group	Station	Priority class			
		P1	P2	P3	P4
G1	Esfahan	NR	MGC	LR	GC
	Pol kale	MGC	NR	LR	GC
	Ziar braan	NR	LR	MGC	GC
	Varzaneh	MGC	LR	NR	GC
G2	Kaftar	LR	MGC	NR	GC
	Sedeh	LR	GC	MGC	NR
	Emamzadeh Ismael	NR	LR	MGC	GC
G3	Mansur abad	MGC	GC	NR	LR
	Faryab	NR	MGC	LR	GC
	Sargalm khorgu	LR	NR	MGC	GC
	Kuh-e Geno	MGC	NR	LR	GC
G4	Sarkhoon	MGC	NR	LR	GC
	Ghasem abad	MGC	NR	LR	GC
	Cham zaman	MGC	NR	GC	LR
	Dareh takht	NR	GC	MGC	LR
G5	Kamandan	MGC	NR	LR	GC
	Nozar abad	MGC	NR	GC	LR
	Sefid chah	NR	MGC	LR	GC
	Abelou	LR	MGC	NR	GC
G6	Bar kola	MGC	NR	GC	LR
	Deh bala	LR	NR	MGC	GC
	Feiz abad	NR	MGC	LR	GC
	Tezerjan	MGC	GC	LR	NR
	Soltaneb	LR	MGC	NR	GC

MGC modified geographical coordinate, *GC* geographical coordinate, *NR* normal ratio, *LR* linear regression

performance criteria. It should be noted that the required analysis is relatively cumbersome and quite time-consuming, because the procedure needs to repeat many times (e.g., 1,000 times) in order to provide an adequate number of bootstrapped data series to estimate probability distribution functions of applied accuracy measures (MAE and E). However, if the iteration number of same priority score for each method receives special attention, one can obtain some valuable information. Thus, repetition numbers of gap-filling approaches in each priority class are counted and then presented in

Table 5 Iteration numbers of methods for each priority class

Method	Number in each priority class			
	P1	P2	P3	P4
MGC	11	7	6	0
NR	7	10	5	2
LR	6	3	10	5
GC	0	4	3	17

Table 5. Based on Table 5, the maximum number of iterations in the first priority class belongs to the proposed methodology (modified GC approach), and the NR technique places in the second rank. Meanwhile, it is worth noting that the priority order of methods for the lowest priority class (i.e., P4) is as follows: GC, LR, NR and MGC.

As shown in Tables 4 and 5, the conventional GC method has occupied no place in the first priority class for all stations, and more interestingly, the modified GC approach has not been in the lowest priority at all. This reveals that the authors, by taking the advantage of two variables (A_i and MAP), have succeeded to dramatically heighten the performance of an inverse-distance technique (i.e., conventional GC) for a wide variety of climatic and topographical conditions.

4 Conclusion

The main goal of the study was to modify the GC method for completing gaps in annual precipitation datasets. For this purpose, authors utilized the altitude ratio (A_i) and the MAP to enhance the accuracy of the GC technique. In this respect, 24 precipitation gauge stations situated in six different territories of Iran were taken into account in order to investigate the performance of the proposed methodology. Furthermore, estimates of other techniques such as the NR, the LR and the conventional GC were assessed to examine whether the improved approach has adequate suitability for various geoclimatic conditions.

Average values of model error (MAE) and model efficiency (E) measures and also outcomes of the priority setting process for examined methods demonstrated that in the general point of view, the first priority method for reconstructing missing precipitation measurements in surveyed climatic regions was the modified geographical coordinate (MGC) technique, and the second priority approach was the NR method. Meanwhile, the interesting point was revealed for the conventional GC technique, so that it occupied no place in the first priority class for all stations under consideration, and more interestingly, the modified GC method had not been in the lowest priority at all. Hence, the proposed methodology was proven to be a reliable procedure for filling gaps in precipitation records. In other words, authors have succeeded to dramatically improve the efficiency of an inverse-distance approach (i.e., GC method) that shows poor performance in reconstructing missing precipitation observations and to give it remarkable preference in comparison with traditional techniques such as the LR and the NR. Additionally, the applicability of the proposed modification procedure is inevitable due to its requirements (e.g., A_i and MAP), because these data are usually readily available in a large array of areas including developing countries.

Acknowledgments The authors greatly appreciate Dr. Reza Modarres and Raoof Mostafazadeh for their helpful feedback.

References

- Abebe AJ, Solomatine DP, Venneker RGW (2000) Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. *Hydrol Sci J* 45(3):425–436
- Acock MC, Pachepsky YA (2000) Estimating missing weather data for agricultural simulations using group method of data handling. *J Appl Meteorol* 39(7):1176–1184
- Chatterjee S, Hadi AS (2006) Regression analysis by example. John Wiley & Sons, Hoboken, NJ. ISBN-I3 978-0-471-74696-6
- Chow VT (1964) Handbook of applied hydrology. McGraw-Hill, New York
- Daly C, Neilson RP, Phillips D (1994) A statistical topographic model for mapping climatological precipitation over mountainous terrain. *J Appl Meteorol* 33(2):140–158
- Dastorani MT, Moghadamia A, Piri J, Rico-Ramirez M (2009) Application of ANN and ANFIS models for reconstructing missing flow data. *Environ Monit Assess* 166(1–4):421–434. doi:10.1007/s10661-009-1012-8
- Eccel E, Cau P, Ranzi R (2012) Data reconstruction and homogenization for reducing uncertainties in high-resolution climate analysis in Alpine regions. *Theor Appl Climatol*. doi:10.1007/s00704-012-0624-z
- Geerts B (2003) Empirical estimation of the monthly-mean daily temperature range. *Theor Appl Climatol* 74(3–4):145–165
- Good PI (1999) Resampling Methods: A Practical Guide to Data Analysis. Birkhauser Boston, c/o Springer-Verlag, New York, Inc., New York
- Jeffrey SJ, Carter JO, Moodie KB, Beswick AR (2001) Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ Model Softw* 16:309–330
- Kashani MH, Dinpashoh Y (2012) Evaluation of efficiency of different estimation methods for missing climatological data. *Stoch Environ Res Risk Assess* 26:59–71. doi:10.1007/s00477-011-0536-y
- Kemp WP, Burnell DG, Everson DO, Thomson AJ (1983) Estimating missing daily maximum and minimum temperatures. *J Clim Appl Meteorol* 22:1587–1593
- Khorsandi Z, Mahdavi M, Salajeghe A, Eslamian S (2011) Neural network application for monthly precipitation data reconstruction. *J Environ Hydrol* 19(5):1–12
- Kim J-W, Pachepsky YA (2010) Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *J Hydrol* 394:305–314
- Li X, Li L, Wang X, Jiang F (2013) Reconstruction of hydrometeorological time series and its uncertainties for the Kaidu River Basin using multiple data sources. *Theor Appl Climatol* 113(1–2):45–62
- Linsley RK, Kohler MA, Paulhus JLH (1988) Hydrology for engineers. McGraw-Hill Co., Singapore
- Lookzadeh S (2004) Evaluation of several different reconstruction methods for precipitation data in central alborz region. MSc thesis, Tehran University, Iran. (In Persian)
- Mahdavi M (1998) Applied hydrology, vol 1. Tehran University Publications, Tehran (In Persian)
- Nafarzadegan AR, Rezaeian-Zadeh M, Kherad M, Ahani H, Gharekhani A, Karampoor MA, Kousari MR (2012) Drought area monitoring during the past three decades in Fars province, Iran. *Quatern Int* 250:27–36. doi:10.1016/j.quaint.2010.12.009
- Ramos-Calzado P, Gomez-Camacho J, Perez-Bernal F, Pita-Lopez MF (2008) A novel approach to precipitation series completion in climatological datasets: application to Andalusia. *Int J Climatol* 28(11):1525–1534

- Sadatinejad SJ (1997) Statistical comparison of the reconstruction methods of precipitation data in Isfahan province. MSc thesis, Tarbiat Modares University, Iran. (In Persian)
- Sadatinejad SJ, Shayannejad M, Honarbakhsh A (2009) Investigation of the efficiency of the fuzzy regression method in reconstructing monthly discharge data of hydrometric stations in Great Karoon River Basin. *J Agric Sci Technol* 11:111–119
- Schneider T (2001) Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J Clim* 14(5):853–871
- Tardivo G, Berti A (2012) A dynamic method for gap filling in daily temperature datasets. *J Appl Meteorol Climatol* 51:1079–1086
- Tardivo G, Berti A (2013) Comparison of four methods to fill the gaps in daily precipitation data collected by a dense weather network. *Sci J Environ Eng Res* 2013, 9 pp. doi:10.7237/sjeer/265
- Teegavarapu RSV, Chandramouli V (2005) Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *J Hydrol* 312:191–206
- Thornton PE, Running SW, White MA (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *J Hydrol* 190(3–4):214–251
- Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 30:79–82
- Willmott CJ, Robeson SM, Feddema JJ (1994) Estimating continental and terrestrial precipitation averages from rain-gauge networks. *Int J Climatol* 14(4):403–414
- Xia Y, Fabian P, Stohl A, Winterhalter M (1999) Forest climatology: estimation of missing values for Bavaria, Germany. *Agric For Meteorol* 96:131–144