COMPARATIVE PERFORMANCE ASSESSMENT OF EIGHT UNIVARIATE DATA FILLING TECHNIQUES IN MISSING PRECIPITATION RECORDS

P. Kabbilawsh[1], D. Sathish Kumar[2] and N.R. Chithra[3]
[1]Research Scholar, Department of Civil Engineering, National Institute of Technology, Calicut,
Email: kabbilawsh_p160006ce@nitc.ac.in
[2]Assistant Professor, Department of Civil Engineering, National Institute of Technology, Calicut,
Email: sathish@nitc.ac.in
[3]Assistant Professor, Department of Civil Engineering, National Institute of Technology, Calicut,
Email: chithranr@nitc.ac.in

ABSTRACT: Weighting function which uses squared inverse of the distance as a parameter in the estimation of missing precipitation records are extensively applied for the past five decades. Despite its wide acceptability, it suffers from several drawbacks like the inability to account for the existence of negative spatial autocorrelation, thereby many conceptual enhancements which are distance-based are brought in to overcome the drawbacks. Similarly, regressive models are another set of techniques which are traditionally applied, which also suffers from the conceptual limitation of defining the function behaviour prior to applying it. To overcome this, the non-parametric ranking system is induced in regressive models.

A detailed study is carried out to compare eight univariate techniques, three belonging to Inverse distance weighting methods (IDWM) and rest five to regressive schemes. Precipitation record from 10 gauging stations from the districts of Palakkad, Thrissur, and Malappuram of Kerala, India are selected and used to test the efficiency of the improvements carried out in the existing techniques.

Results suggest that no distinct regressive methods can be adjudged as the best overall. Orthogonal regression performed well for all stations having a least Inter-quartile range (IQR) despite having a non-zero median and appreciable amount of standard deviation. Among distance weighted methods, CCWM (coefficient of correlation weighting method) outperformed traditional IDWM and IEWM (Inverse exponential weighting method) by a large margin having an inter-quartile range close to zero. A generic conclusion cannot be arrived because no specific method is found to be suitable for all situations. It depends on the nature of the data that is being used.

KEYWORDS: Distance weighting schemes; Parametric and non-parametric regression; Missing completely at random; Orthogonal and geometrical regression; Five-number summary

1.INTRODUCTION

Statistical studies done to quantify and predict climate change using meteorological data should be free from missing observations that are measured over a long period of time. Missing data records can bring a considerable amount of bias, induce stressfulness during data analysis and reduce the efficiency of analysis. Traditional approaches applied on time series data, containing missing values use case deletion and imputation with mean value approaches. By default, statistical packages apply list-wise deletion, that is to completely remove the case which contains a missing data (Dow and Eff, 2009). This process decreases the representativeness of the analysis.

Presence of gaps in precipitation series is a common problem while dealing with long-term data (Allison, 2002). There are several data filling techniques available, ranging from simple mean to mathematically complex multiple imputation methods (Singh, 1986). Numerous studies in past three decades have been carried out for comparing several univariate, bivariate, stochastic, deterministic and data-driven approaches (Tung and Asce, 1984).

The purpose of current study is to compare five regressive schemes and three distance weighted methods, followed by performance assessment by five-number summary with mean and standard deviation instead of using popular error measures such as root mean squared error (RMSE), mean absolute error (MAE) and goodness of fit (R^2).

2. MISSING COMPLETELY AT RANDOM (MCAR) IN RAINFALL DATA SERIES

The literature is flooded with a generic explanation of missing data mechanism. However, it is necessary to understand and contemplate the issue in relation to precipitation data. Therefore, a brief discussion on MCAR mechanism is presented here.If the probability with which the missing data occurring in a specific variable is completely non-dependent on any other observed variable and is also non-dependent to the variable with the missing value is called MCAR. In other words, missingness in the variable does not follow any pattern and its unsystematic (Rubin, 1976). For example, when data are missing for the station which their recordings were lost in reporting. This hypothesis can be tested by separating the missing and complete cases and examine the characteristics of the two groups. If the characteristics are approximately equal for both groups, then MCAR hypothesis holds good. Else it needs to be rejected. It is the only missing data mechanism that can be quantitatively calculated using Little's chi-square test, also known as Little's MCAR test.

In Little's chi-square test, the null hypothesis states the data to be missing randomly with no specific pattern if the significance level denoted by p-value is at a level of 0.05. To put in simple terms, if the value is less than 0.05, data are not missing at random. Little's MCAR was applied on the entire dataset considering it as 21X12 matrix and value obtained was about 0.998. A bigger *p*-value indicates weak proof against the null hypothesis, thereby failing to reject it. In the current study, it can be concluded that no pattern exists.

3.METHODS

Inverse distance weighting method (IDWM) is one of the oldest methods for estimation of missing data in the field of hydrological sciences (Teegavarapu and Chandramouli, 2005). With improvements in computational ability, several variants of IDWM are proposed and adopted with emphasis on choice of weighting function in order to capture the arbitrariness in missing data.

3. 1 Inverse distance weighting method (IDWM)

The reciprocal distance method is given by

$$X_m = \frac{\sum_{i=1}^{n} X_i d_{mi}^{-2}}{\sum_{i=1}^{n} d_{mi}^{-2}} \qquad (1)$$

where X_m is the value of precipitation at the base station, *n* is the number of stations, X_i is the value at station *i*, d_{mi} is the distance from the location of station *i* to station *m*.

3.2 COEFFICIENT OF CORRELATION WEIGHTING METHOD (CCWM)

The weighting function in IDWM, d_{mi}^{-2} is replaced by R_{mi} in (2), where R_{mi} is the coefficient of correlation (Teegavarapu and Chandramouli, 2005).

$$X_m = \frac{\sum_{i=1}^{n} X_i R_{mi}}{\sum_{i=1}^{n} R_{mi}} \qquad (2)$$

3.3 INVERSE EXPONENTIAL WEIGHTING METHOD (IEWM)

The weighting function in IDWM, d_{mi}^{-2} is replaced by e_{mi}^{-2d} for IEWM (Teegavarapu and Chandramouli, 2005).

$$X_m = \frac{\sum_{i=1}^{n} X_i e_{mi}^{-2d}}{\sum_{i=1}^{n} e_{mi}^{-2d}} \qquad (3)$$

REGRESSIVE METHODS

In recent times, enormous interest has arisen in the estimation of missing data using single imputative regressive techniques. The following techniques were applied in this study for the assessment of their ability in filling the missing data.

a) Parametric ordinary least-squares regression (POLSR)
b) Non-parametric ranked regression (NPRR)
c) Non-parametric simplified Theil's method (NPSTM)
d) Orthogonal regression (OR)
e) Geometric mean functional regression (GMFR)

3.4 PARAMETRIC ORDINARY LEAST-SQUARES REGRESSION (POLSR)

Parametric least-square regression having a straight line functional form is the most widely used modelling method. It has been applied in wide range of studies, which are beyond its direct scope. The definition, derivation, the criterion is explained with clarity in several literatures (Rubin, 1972). A brief note on its principle, advantage, and limitations are discussed here.

The principle lies in minimizing the sum of squared deviations between the observed response and the functional response produced by the model. The process of minimization decreases the initial large system of equations formed by observed data (which are overdetermined by default) to a balanced system consisting of n equations with n unknowns. Then the new set of equations are simultaneously solved to obtain the numerical value of the parameter. The base station (the one with missing values) is given by Y and X denotes the station used for filling. The equation is given by

$$Y = b * X + a \tag{4}$$

where *b* is the regression coefficient and *a* is the intercept value.

The major disadvantage of least square is the linear shape that it assumes over long ranges leading to weak extrapolation ability where the difference between the observed responses and predicted response is appreciably large. It is also highly insensitive to outliers. Few outliers can sometimes skew the results of the analysis in a specific direction, which would make the model validation incapable of obtaining a correct output.

3.5 NON -PARAMETRIC RANKED REGRESSION (NPRR)

Two-time series belonging to rainfall stations *X* (the predictor stations which is used as the dependent variable) and *Y* (response station which contains the missing data) stations are considered. *X* and *Y* are ordered and ranked in ascending order. Sequential ranks were given to unique values. Ranked *X* $R_O(X)$ as a predictor and ranked *Y* $R_O(Y)$ as a response is modelled by least square linear regression method.

$$R_O(Y) = b * R_O(X) + a \tag{5}$$

where *b* is the regression coefficient and *a* is the intercept value.
Estimated rank $R_E(Y_I)$ is back transformed for finding the functional response by implementing the following criterion. Let $R_O(X_i)$ gives an estimated value of $R_E(Y_I)$ by the equation Eq.2. From the new rank, the value of *Y* is obtained by the following criterion.

$$\text{If } R_E(Y_i) = R_O(Y_a) \quad \text{then } Y_i = Y_a \tag{6}$$

$$\text{If } R_O(Y_a) < R_E(Y_i) < R_O(Y_b) \quad \text{then } Y_I = Y_A + \left[\frac{R_E(Y_i) - R_{OE}(Y_a)}{R_O(Y_a) - R_O(Y_b)}\right] * (Y_b - Y_a) \tag{7}$$

$$\text{If } R_E(Y_i) > Max(R_O(Y)) \quad \text{then } R_E(Y_i) = Max(R_O(Y)) \tag{8}$$

$$\text{If } R_E(Y_i) < Min\, R_O(Y)) \quad \text{then } R_E(Y_i) = Min\, R_O(Y)) \tag{9}$$

where Y_a and Y_b are observed values and $Y_a < Y_b$. By ranking the original data, it gets converted into an ordinal form(Iman and Conover, 1979). The main advantage of inducing ordinality is to ease the collation and categorization of rainfall values. By ranking the data, its inherent behaviour is removed by making it distribution free, which helps in inferring more information from the dataset. The basic assumption of all regression models is that the residuals are normally distributed. It's highly likely to have residuals to be normal if both the dependent variable and response variables are normally distributed. Generally, all the datasets are not normally distributed. In such cases, non-parametric methods are more reliable than the parametric counterpart(Iman and Conover, 1979).

3.6 NON-PARAMERIC SIMPLIFIED THEIL'S METHOD (NPSTM)

Theil (1950) developed a method which does not need the assumption of normality of residuals for the validity of the significant test and at the same time will not be highly affected by the presence of outliers in comparison to parametric least square regression(Theil, 1950). The estimate of the slope is robust in nature and is computed by comparing each pair of data in a pairwise style. A total of n pairs *(X, Y)* will result in $n*(n-1)/2$ pairwise comparisons. For each of these comparisons, a slope $\Delta Y/\Delta X$ is computed. Non-parametric slope estimate is equal to the median of all possible pairwise slope. Since this method is computationally intensive, a simplified approach also exists.

In the simplified method, both predictor and response variables are sorted in ascending order of the predictor variable. Then data is split into two halves. If *N* (number of observations) is odd, one observation, the median value of *X* is left out. Then a new variable having *N/2* differences are calculated for both the variables, *X* and *Y* according to following relationships:

$$\bar{X}_I = X_{I+(N/2)} - X_I \tag{10}$$

$$\bar{Y}_I = Y_{I+(N/2)} - Y_I \tag{11}$$

The regression line is given by $Y = \bar{b}*X + \bar{a}$
Where $\bar{b}$ is called the angular coefficient and $\bar{a}$ is the intercept

$$\bar{b} = \frac{Y_{median}}{X_{median}} \tag{12}$$

$$\bar{a} = \bar{Y}_{median} - \bar{b} * \bar{X}_{median} \tag{13}$$

This way, a regression line is obtained passing through the crossing point of the median (instead of the mean), which is considered as the nonparametric centre of the cloud of points (Theil, 1950). The fitted line goes through the median point (X_{median}, Y_{median}) which is similar to the mean point (X_{mean}, Y_{mean}) in the least square regression. This way Theil's regressive line passes through the crossing point of medians which is then taken to be the centre of cloud formed by the non-parametric points.

3.7 ORTHOGONAL REGRESSION (OR)

One of six main assumptions of least square fit are that each value of the predictor variable is known exactly. All the uncertainty and errors is in response variables. However, this is not true in the case of the variables obtained by measurements, where they are inherently exposed to several errors.

It is the type of regression which is used in case of non-negligible uncertainty in both the response and predictor variable. Since *X* and *Y* being the point rainfall data, will have some appreciable amount of uncertainty in them(Poller *et al.*, 1998).

This model minimizes the squared perpendicular distances between the observed response and functional response. It takes the normal distance instead of vertical distance that is used in the least square method.
The regression line is given by $Y = \bar{b} * X + \bar{a}$. The slope is found by

$$\bar{b} = -L + \frac{\sqrt{L^2 + R^2}}{R} \quad (14)$$

$$\text{where } L = 0.5 * \left[\frac{s_X}{s_Y} - \frac{s_Y}{s_X}\right] \quad (15)$$

$$\text{where } \bar{a} = \bar{Y}_{mean} - \bar{b} * \bar{X}_{mean} \quad (16)$$

3.8 GEOMETRIC MEAN FUNCTIONAL REGRESSION (GMFR*)*

It is used in situations where measurement error is present in both response and predictor. The benefit is that it forms an exclusive regressive model where predictor and the response variable can exchange places and still the model is valid.

$$X_I = X_{mean} + r\frac{s_Y}{s_X}[Y_I - Y_{mean}] \quad (17)$$

$$Y_I = Y_{mean} + r\frac{s_X}{s_Y}[X_I - X_{mean}] \quad (18)$$

The slope of the first line is m $\left(r * \left[\frac{s_Y}{s_X}\right]\right)$ and the slope of the second line is $\bar{m}$ $\left(r * \left[\frac{s_X}{s_Y}\right]\right)$. The slope of the geometric mean functional line $\bar{M}$ equals the geometric mean of the Y on X and X on Y linear least square fit slopes.

$$\bar{M} = (m * \bar{m})^{0.5} \quad (19)$$

$$Y_I = Y_{mean} + sign(r) * \frac{s_X}{s_Y}[X_I - X_{mean}] \quad (20)$$

The principle lies in minimizing the area of the right-angle triangle formed by horizontal and vertical distances between the observed data points and functional data points obtained from the model (Halfon, 1985).

STRENGTH OF ASSOCIATION

To find stations with similar statistics around the base (station with missing data) stations, three correlation statistic was chosen (Myers 2003). uniqueness, strengths, and limitations of these indices are discussed.

3.9 Pearson product-moment correlation coefficient (*r*)

The most repetitively used statistic in literature to calculate the strength of linear relationship between two variables. It is a bivariate correlation statistic to measure the degree of the relationship between linearly related variables. It quantifies the amount of distance by which points lie away from a best linear fit. In case of Pearson, both predicted and response variables are considered to be normally distributed, linear and homoscedastic.

3.10 Spearman's rho (rank correlation coefficient, r_s)

Spearman’s rho determines whether there exists any monotonic relation between the variable. It penalizes the dislocation by the square of the distance. It has better capabilities to detect monotonic non-linear relationships when compared to Pearson product moment correlation. Similar to non-parametric ranked regression, ordinality is induced to remove the inherent distribution (Myers 2003).

3.11 Kendall's tau-b coefficient (τ_B)

Kendall's tau-b penalizes dislocations by the distance of the dislocation. Therefore, Kendall's tau-b penalizes two independent swaps the same as two sequential swaps, but Spearman's rho gives a stronger penalty to the latter than to the former.

4. STUDY AREA AND DATA

Ten rain-gauge stations from the districts of Palakkad, Malappuram, and Thrissur of Kerala state in India, were chosen as shown in Figure 1. for the study. The type of data is monthly precipitation data and length of data ranges from 1991 to 2011. The number of available and missing data values is listed in Table 1.

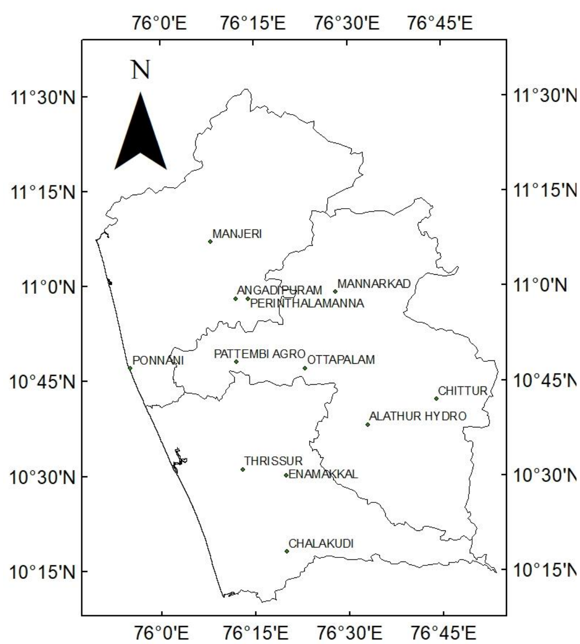

Figure 1. Position of rain gauge statins in the study area comprising of three districts of Kerala

5. METHODOLOGY

The process is comprising of two consecutive stages. First one is locating the station with missing data and determination of station with similar meteorological characteristics from its neighbourhood. The selection is based on strength of association. Using the three indices described earlier, for a station minimum of three station needs to be present for filling. The second stage, eight univariate methods are applied to estimate the missing data. The methodology is explained using Perinthalamanna as a base station and Manjeri, Mannarkad, Angadipuram, Pattamabi Agro and Ottapalam (Table 2). In an earlier study, the maximum distance between stations to be selected a probable candidate for 'twin station' is specified as10 km (Hubbard, 1994). However, due to unavailability of stations within 10 km, the neighbourhood distance is increased to 25 km and used as the threshold.

Table 1. Amount of missing data in chosen 10 stations falling within study area

Number	Station name	Available values	Missing values
1	Manjeri	246	6
2	Mannarkkad	250	2
3	Angadipuram	245	7
4	Perinthalamanna	250	2
5	Pattambi Agro	248	4
6	Ponnani	248	3
7	Ottapalam	250	2
8	Chittur	250	2
9	Thrissur	250	2
10	Chalakudi	249	3

6. RESULT

Selection of best fit regressive model is done using mean, standard deviation, median and Inter-quartile range as criterion parameters. Five- number summary is adopted to calculate the amount of spread since each parameter describes a specific portion of the time series. Median gives the centre of data series; the upper and lower quartiles span the middle half of a data set, and the maximum and minimum values act as a supplementary information describing the dispersion of the data. Boxplot diagrams were used for visual inspection for choosing the appropriate technique. CCWM gave the least error in comparison with IDWM and IEWM, which proves that the conceptual improvements were rightly chosen (Table 3).

Table 2. Amount of association measured by three similarity indices

Gauging station	Person product-moment coefficient (r)	Spearman's rho (r_s)	Kendall tau-b (τ_B)
Manjeri	0.959	0.961	0.846
Mannarkkad	0.903	0.934	0.784
Angadipuram	0.988	0.986	0.92
Pattambi Agro	0.957	0.96	0.841
Ottapalam	0.964	0.957	0.834

Table 3. Comparison of descriptive statistics of distribution of three different weighting models

Variable	IDWM	IEWM	CCWM
Mean	9.66	8.3	-1.369
Standard deviation	38.29	41.24	7.123
Median	0.81	0.35	0
IQR	26.17	26.37	5.472
Mode	0	0	0
Maximum	295.43	338.2	42.768
Minimum	-99.46	-110.6	-32.222
Skewness	1.9	2.16	0.09
Kurtosis	12.95	16.85	9.46

Table 4. Descriptive statistics of error distribution of different regressive models

Station name	**Statistic**	**POSLR**	**NPRR**	**NPSTM**	**OR**	**GMFR**
Angadipuram	Mean	-0.01	6.84	16.27	-8.04	-8.028
	Standard deviation	40.92	49.42	41.13	2.154	2.189
	Minimum	-111.78	-118.85	-92.36	-10.01	-10.03
	Median	-9.18	-0.01	5.46	-8.589	-8.586
	Maximum	327.01	327.19	340.84	0.822	0.978
	IQR	29.13	28.57	31.86	3.065	3.115
	Skewness	2.36	2.39	2.39	1.34	1.34
	Kurtosis	16.88	11.69	15.89	1.86	1.86
Mannarkkad	Mean	-0.01	17.56	19.93	-0.03	-0.01
	Standard deviation	115.26	118.09	115.53	118.16	118.15
	Minimum	-438.69	-309.4	-440.25	-510.02	-509.82
	Median	-25.52	-2.54	-0.03	-3.65	-3.69
	Maximum	543.26	555.59	562.37	540.48	540.51
	IQR	76.83	73.91	71.24	69.71	69.58
	Skewness	1.16	1.63	0.95	0.41	0.42
	Kurtosis	6.29	5.36	6.37	6.36	6.36
Ottapalam	Mean	0	42.88	25.72	-35.77	-35.85
	Standard deviation	70.95	82.45	71.5	28.71	28.66
	Minimum	-247.93	0	-213.93	-154.47	-154.32
	Median	-14.73	22.54	10.35	-28.24	-28.33
	Maximum	315.53	635.04	339.8	-9.59	-9.72
	IQR	60.85	44.4	62.63	40.76	40.69
	Skewness	0.35	4.2	0.58	-1.34	-1.34
	Kurtosis	3.38	20.75	2.88	1.87	1.87
Pattambi Agro	Mean	0	66.7	20.34	-20.587	-20.599
	Standard deviation	78.06	179.4	78.75	1.345	1.399
	Minimum	-219.34	-204.9	-217.75	-26.13	-26.364
	Median	-22.1	-2.1	1.5	-20.231	-20.229
	Maximum	477.09	1002.8	487.89	-19.36	-19.323
	IQR	70.92	119.5	67.23	1.934	2.011
	Skewness	1.68	2.4	1.43	-1.33	-1.33
	Kurtosis	7.07	6.92	6.52	1.85	1.85
Manjeri	Mean	0	-5.55	2.67	-0.06	0.04
	Standard deviation	76.03	69.64	80.54	76.83	76.83
	Minimum	-307.98	-396.84	-365.98	-333.22	-333.06
	Median	-19.08	4.86	-0.01	-13.96	-13.89
	Maximum	400.54	294.02	408.98	402.88	402.97
	IQR	66.85	53.37	64.67	63.85	63.85
	Skewness	0.81	-0.62	-0.17	0.41	0.42
	Kurtosis	5.02	6.01	5.55	5.28	5.28

Few methods though had a zero-median value were not rejected since they have very large inter-quartile range, which was recognized from boxplots (Figure.2b to Figure. 2f). Orthogonal and geometric mean regression techniques produced similar results for all the five stations considered. For Mannarkkad and Ottapalam, non-parametric Theil's method gives exceptional low median compared to other methods. Ranked regression outperforms other methods in all four descriptive parameters in Manjeri station For Angadipuram and Pattamabi Agro station, filling by both orthogonal and geometrical models is most appropriate as these have the least IQR and standard deviation (Table 4).

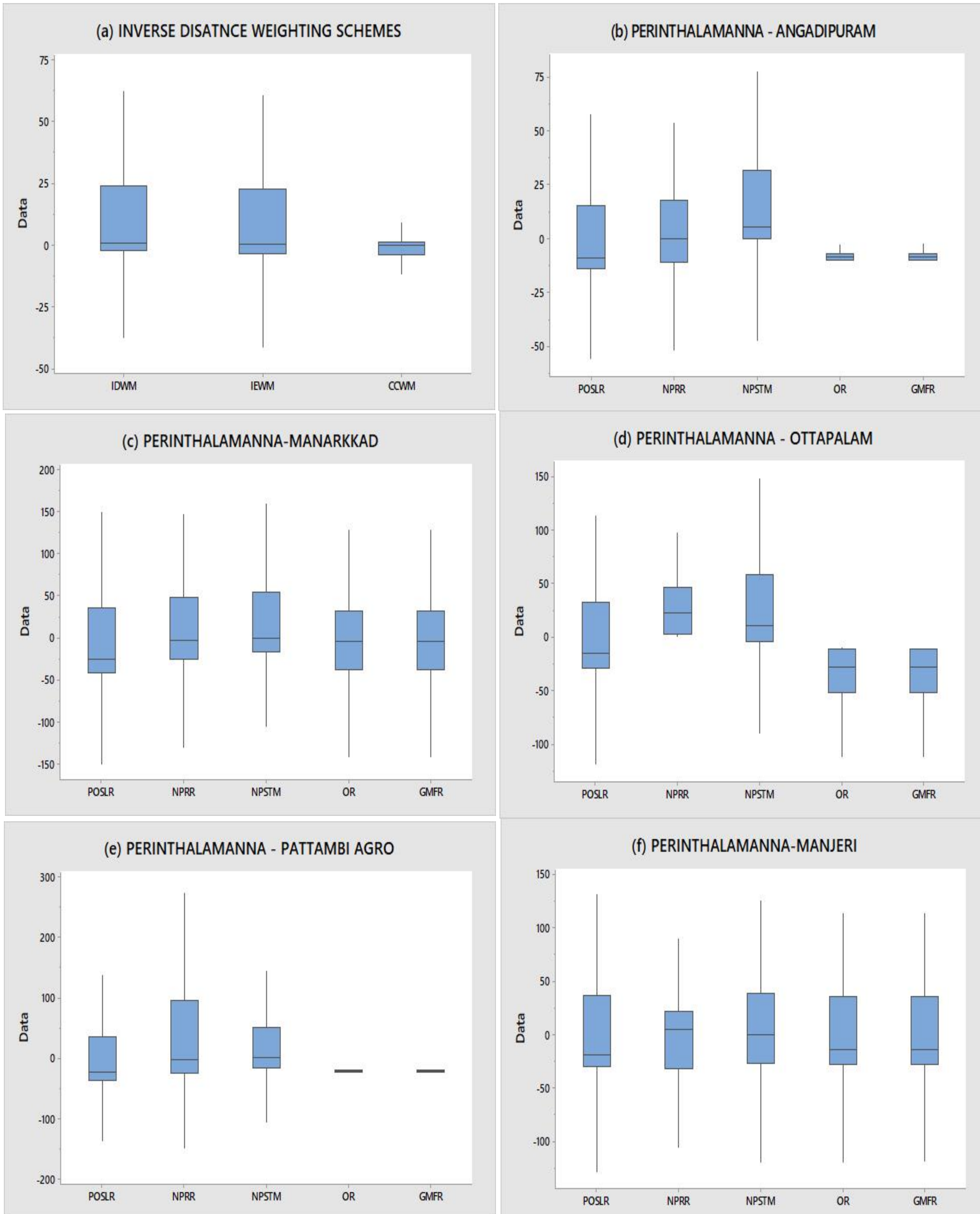

Figure. 2 (a). Box plot representing the relative distribution of absolute errors by different weighting methods. (**b)** Box plot representing the relative distribution of errors using Angadipuram station. (**c)** Box plot representing the relative distribution of absolute errors using Mannarkard station. (**d)** Box plot representing the relative distribution of absolute errors using Ottapalam station**. (e)** Box plot representing the relative distribution of absolute errors using Pattambi Agro station. **(f)** Box plot representing the relative distribution of absolute errors using Manjeri station

7. CONCLUSION

This study provides a unique example of comparing techniques for filling missing precipitation records belonging to two different schemes. The reason for the similarity in the results of orthogonal and geometrical mean functional regression needs to be further investigated. It's noteworthy to conclude that no method is found to be inferior or superior to other. Therefore, it is not advisable to rely on one single technique to fill the missing data. The distribution of the data plays a major role in the selection of specific technique to get better results. Therefore, before application of any of these techniques discussed, a thorough investigation on the distribution of the data is necessary.

8. References

Allison, P. D. (2002) ‘Missing Data’, *Quantitative Applications in the Social Sciences*, p. 104. doi: 10.1136/bmj.38977.682025.2C.

Dow, M. M. and Eff, E. A. (2009) ‘Multiple Imputation of Missing Data in Cross-Cultural Samples’, *Cross-Cultural Research*, 43(3), pp. 206–229. doi: 10.1177/1069397109333362.

Halfon, E. (1985) ‘Regression Method in Ecotoxicology: A Better Formulation Using the Geometric Mean Functional Regression’, *Environmental Science and Technology*, 19(8), pp. 747–749. doi: 10.1021/es00138a016.

Hubbard, K. G. (1994) ‘Spatial variability of daily weather variables in the high plains of the USA’, *Agricultural and Forest Meteorology*, 68(1–2), pp. 29–41. doi: 10.1016/0168-1923(94)90067-1.

Iman, R. L. and Conover, W. J. (1979) ‘The use of the rank transform in regression’, *Technometrics*, 21(4), pp. 499–509. doi: 10.1080/00401706.1979.10489820.

Myers, Jerome L.; Well, Arnold D. (2003) 'Research Design and Statistical Analysis (2nd ed.)’, *Lawrence Erlbaum*. pp. 508. ISBN 0-8058-4037-0**.**

Poller, L. *et al.* (1998) ‘A comparison of linear and orthogonal regression analysis for local INR determination in ECAA coagulometer studies’, *British Journal of Haematology*, 102(4), pp. 910–917. doi: 10.1046/j.1365-2141.1998.00866.x.

Rubin, D. B. (1972) ‘A Non-Iterative Algorithm for Least Squares Estimation of Missing Values in Any Analysis of Variance Design’, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 21(2), pp. 136–141. doi: 10.2307/2346485.

Rubin, D. B. (1976) ‘Inference and Missing Data’, *Biometrika*, p. 581. doi: 10.2307/2335739.

Singh, V. P. (1986) ‘Comparing some methods of estimating mean areal rainfall’’, 22(2), pp. 275–282.

Teegavarapu, R. S. V and Chandramouli, V. (2005) ‘Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records’, *Journal of Hydrology*, 312(1–4), pp. 191–206. doi: 10.1016/j.jhydrol.2005.02.015.

Theil, H. (1950) ‘A rank-invariant method of linear and polynomial regres- sion analysis, I’, *Proceedings of the Koninklijke Nederlandse Academie van Wetenschappen*, 53(1950), pp. 386–392. doi: 10.1007/978-94-011-2546-8_20.

Tung, B. Y. and Asce, A. M. (1984) ‘Point rainfall estimation for a mountainous region’, 109(10), pp. 1386–1393.