Review Article
Published: 17 Jun, 2020

SF Journal of Environmental and Earth Science

Techniques of Filling Missing Values of Daily and Monthly Rain Fall Data: A Review

*Muluken LE**

[1]Department of Bio-systems Engineering, Hawassa University, Hawassa, PO.Box 05, Ethiopia

[2]Sirinka Agriculture Research Center, Woldia, PO.Box 74, Ethiopia

Abstract

Precipitation or rainfall is a pertinent climatic parameter and the research on rainfall is mostly troubled due to lack of continuous data. Representation of climatological characteristics demands a good quality of rainfall data for an efficient environmental analysis. The outcome of data analysis depends on the quality and completeness of data. Lack of good quality rainfall data will have bad implication on the process of analyses and subsequently, leads to biased results of the analysis. The accurate planning and management of water resources depends on the presence of consistent and exact precipitation data in meteorology stations. Spatial interpolation techniques are widely used methods for filling the gaps in daily rainfall series through estimating the unknown rainfall amount for a point from the known data from adjacent stations. In this review paper different methods of filling missing data belonging to many environmental phenomenon have been extracted from a lot of literatures. Finally it is advised to use the missing data filling methods for a certain purpose through assuring their validity before application.

Keywords: Rainfall data; Missing data; Data filling techniques; Daily rainfall

Introduction

Rainfall data is important for hrtydrological modeling, agricultural and water budget estimation. Environmental models typically require a complete time series of meteorological inputs, thus reconstructing missing data is a key issue in the functionality of such physical models. Weather stations in the watershed sometimes malfunction during the monitoring period. Missing data has been a problem and therefore have to be estimated for models to function. Estimation of missing data, also known as infilling [1], reconstruction [2], meteorological data series completion [3], or imputation of meteorological data [4], has been a topic of interest in many hydrologic and other environmental modeling studies.

A good quality of rainfall data is very important in representing the climatological characteristics truly for an efficient environmental analysis. Moreover, the results accuracy of climatological analyses is extremely dependent on the quality of rainfall data used. Lack of good quality rainfall data will have bad implication on the process of analyses and subsequently, leads to biased results of the analysis. However, rainfall time series often carries an uncertainty (such as outliers) due to temporal and spatial variability of rainfall measurement, which will affect the quality of data. Furthermore, factors such as faulty instruments, relocation of stations, and human negligence during the measurement of rainfall may affect the continuity of the rainfall records [5]. These problems will affect the estimation output and subsequently, produce inaccurate results. Concerning this situation, this study introduced a practical and reliable approach for developing estimation methods to impute the incomplete (in other words, missing) rainfall data in effort of providing a good quality dataset for public domain.

The accurate planning and management of water resources depends on the presence of consistent and exact precipitation data in meteorology stations. In countries where it has not been possible to accurately and consistently record precipitation data in a particular time section, it is necessary to use methods to estimate the missing precipitation data and apply it in hydrological models. To overcome the problem, a number of interpolation techniques have been developed over the decades, aimed at estimating missing observations in climatic time series, mainly on a monthly, seasonally and daily time scales, are based on spatial interpolation that is, imputed values at a target station are calculated by using synchronous observations from surrounding stations. As it is discussed above many precipitation stations have short breaks in their records because of either absence of the

OPEN ACCESS

***Correspondence:**
Muluken Lebay Egigu, Department of Bio-systems Engineering, Hawassa University, Hawassa, PO.Box 05, Ethiopia.
E-mail: *mulerbruk2016@gmail.com*

Received Date: *25 May 2020*
Accepted Date: *12 Jun 2020*
Published Date: *17 Jun 2020*

Citation: *Egigu ML. Techniques of Filling Missing Values of Daily and Monthly Rain Fall Data: A Review. SF J Environ Earth Sci. 2020; 3(1): 1036.*

ISSN 2643-8070

observers or instrumental failure so, it is mandatory to estimate this missing record with techniques that are described as follows.

Literature Review

Techniques for estimating missing daily and monthly rainfall data

Spatial interpolation techniques are widely used methods for filling the gaps in daily rainfall series through estimating the unknown rainfall amount for a point from the known data from adjacent stations [6]. Paulhus and Kohler (1952) [7] explored two methods of interpolation, the normal-ratio and 3-station-average, to fill the missing values in monthly rainfall data. The Inverse Distance Weighting (IDW) methods estimate the rainfall amount of a location as a weighted average of the rainfall amount of adjacent stations and the weights are considered as a function of the distances [8]. Various IDW methods have been developed on the basis of the functional form of the distances. For instance, inverse of squares, higher powers or exponential of distances [9] are used as weights. The correlation coefficients between data series are also explored to estimate the weights [10]. Spatial interpolation methods yield non-zero rainfall amounts for a station if even just one of the neighboring stations has rainfall, and hence, due to the poor sampling by rainfall gauge networks, tend to significantly overestimate the number of rainy days. Moreover, the errors in estimating the missing records due to the faulty measurement process of rainfall at neighboring stations can't be ignored [11].

Regression based methods are also used for estimating missing precipitation values [12]. Regression models consider climate data, elevation, topography, proximity to coastal area etc. as explanatory variables to estimate missing rainfall series of a station [13]. Some regressive techniques explored in filling missing daily rainfall series include: simple substitution, parametric regression, ranked regression, and the Theil method [12]. In addition to spatial interpolation and regression methods, neural network algorithms are also explored for imputation of missing precipitation values [14]. The neural network algorithms adapt the weighted interpolation technique from neighboring stations. Regression based methods also underestimate the number of days with no rainfall [15].

Simple Arithmetic Mean Method/ Local Mean Method / AA: This is the simplest method commonly used to fill in missing meteorological data in meteorology and climatology. If the normal annual precipitations at surrounding gauges are within the range of 10% of the normal annual precipitation at station X, then the Arithmetic procedure could be adopted to estimate the missing observation of station X [16]. This assumes equal weights from all nearby rain gauge stations and uses the arithmetic mean of precipitation records of them as estimate [17].

$$Px=1/m[P1+P2+\ldots+Pm] \quad \text{(Eq-1)}$$

Where Px is the estimated value of the missing data, P is the value of same parameter at mth nearest weather station, and m is the number of the nearest stations. The AA method is satisfactory if the gauges are uniformly distributed over the area and the individual gauge measurements do not vary greatly about the mean.

Normal ratio method: This method is used if any surrounding gauges have the normal annual precipitation exceeding 10% of the considered gauge. This weighs the effect of each surrounding station [18]. If the normal precipitations vary considerably then Px is estimated by weighting the precipitation at various stations by the ratios of normal annual precipitation. The normal ration method gives Px as:

$$Px=Nx/m[P1/N1+P2/N2+\ldots+P3/N3] \quad \text{(Eq-2)}$$

Where, Px=Estimate for the ungagged station Pi=Rainfall values of rain gauges used for estimation Nx=Normal annual precipitation of X station Ni=Normal annual precipitation of surrounding stations m=No. of surrounding stations.

Modified normal ratio method: Normal ratio method is modified to incorporate the effect of distance in the estimation of missing rainfall. It is Modified by Young (1992) [19] is a common method for estimation of rainfall missing data. This method is used if any surrounding gauges have normal annual precipitation exceeding 10% of the considered gauge. This weighs the effect of each surrounding station [18]. The estimated data is considered as a combination of parameters with different weights, as shown below

$$V_o=\frac{\sum_{i=1}^{n} W_i V_i}{\sum_{i=1}^{n} W_i} \quad \text{(Eq-3)}$$

Where, 'V_O' is the estimated value of the missing data, W_i is the weight of ith nearest weather station expressed as:

$$W_i=\left[R_i^2\left(\frac{N_i-2}{1-R_i^2}\right)\right] \quad \text{(Eq-4)}$$

Where, R_i is the correlation coefficient between the target station and the i^{th} surrounding station, and N_i is the number of points used to derive correlation coefficient.

Inverse distance method: In this method, weights for each sample are inversely proportionate to its distance from the point being estimated [20].The inverse distance method has been advocated to be the most accurate method as compare to other methods discussed above as it is a function of;

A. Rainfall measured at the surrounding index stations

B. Distance to each index station from the ungagged location

The Inverse Distance (reciprocal-distance) Weighting Method (IDWM) [21] is the method most commonly used for estimating missing data. This weighting distance method for estimating the missing value of an observation, which uses the observed values at other stations, is determined by;

$$X_m=\frac{\sum_{i=1}^{n} X_i d_{mi}^{-2}}{\sum_{i=1}^{n} d_{mi}^{-2}} \quad \text{(Eq-5)}$$

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.

Coefficient of Correlation Weighting Method (CCWM): The weighting function in IDWM, d_{mi}^{-2} is replaced by R_{mi}^{-2} in (2), where R_{mi} is the coefficient of correlation [8].

$$X_m=\frac{\sum_{i=1}^{n} X_i R_{mi}}{\sum_{i=1}^{n} R_{mi}} \quad \text{(Eq-6)}$$

Inverse Exponential Weighting Method (IEWM): The weighting functions in IDWM, d_{mi}^{-2} is replaced by e_{mi}^{-2d} for IEWM [8].

$$X_m=\frac{\sum_{i=1}^{n} X_i e_{mi}^{-2d}}{\sum_{i=1}^{n} e_{mi}^{-2d}} \quad \text{(Eq-7)}$$

Aerial Precipitation Ratio (APR) method: This method was developed based on spatial distribution of daily rainfall without accounting for the historical recurrence. The method leads the extension of point rainfall records to Thiessen Polygon areas. The APR method assumes the contribution of rainfall from surrounding stations is proportionate to the aerial contribution of each sub catchment (Thiessen polygon area claimed by each station without considering the missing gauge), when the station of missing values is excluded. The formula of the method can be given as follows;

$$P_X = \frac{\sum_{i=1}^{N}[(A_j - A_i)P_i]}{\sum_{i=1}^{N}(A_j - A_i)} \text{ .... (Eq-8)}$$

Where,= Thiessen Polygon area for the station with missing values, A_j Thiessen Polygon area when station with missing values is excluded, A_i Thiessen Polygon area when station with missing values is included, P_i annual precipitation of surrounding stations, P_x estimate for monthly rainfall for the station with missing observations.

Kriging Method (KM): The commonly used kriging method among the kriging families is ordinary kriging. In this estimation method, optimal and unbiased estimation of regionalized variables at unsampled location is carried out and the nature of the data at the sampled locations should be free of trend [22].

$$\check{z}(X_O) = \sum_{i=1}^{n} \mu_i * z(x_i) \text{ .... (Eq-9)}$$

Where: $\check{z}(Xo)$=estimated value of an attribute or variable at point of interest Xo, z=observed value at the sampled point (xi); μ=weight assigned to the sampled point; n=number of sample points used for estimation; xi=represents geographic coordinate (x,y) at location i. The considerations taken in assigning weights in kriging are two:

1. The estimate, $\check{z}(X_O)$, of the true value, $Z(X_O)$, is unbiased:

$E[\check{z}(X_O)-Z(X_O)]=0$......... (Eq-9.1)

2. The prediction variance is minimum:

$\partial^2(X_O)=Var[\check{z}(X_O)-Z(X_O)]$=minimum....... (Eq-9.2)

Thin-Plate-Spline (TPS): To estimate missing values of any meteorological variable at base station, Thin Plate-Spline method is one option among the available methods. As it is explained in [23], the basic equation for Partial Spline method for n number of measurement stations for the variable of interest to be estimated at base station, say X0, represented by Zo is given as follows:

$Z_O=F(x_i)+D^TY_i+\beta_i (i=1....,n)$......... (Eq-10)

Where: Zo: predicted value of a missing meteorological variable at base station; xi: is a d-dimensional vector of independent variable; F: is an unknown smooth function of the xi; Yi: is a p-dimensional vector of independent co-variates; D: is an unknown p-dimensional vector of coefficient of the Yi; βi: is zero mean random error term. The above model is reduced to ordinary Thin-Plate-Spline (or Thin-Plate-Spline method) when the value of p is zero (or, p=0), which means there is no covariance. Therefore, for this particular study Thin-Plate-Spline method is selected and this method is relatively robust and developed for climatic data [24].

Multiple linear regression: Multiple Linear Regressions (MLR) is a statistical method for estimating the relationship between a dependent variable and two or more independent, or predictor, variables. MLR identifies the best-weighted combination of independent variables to predict the dependent, or criterion, variable. Eischeid et al. (1995) [25] highlighted many advantages of this method in data interpolation and estimation of missing data. The missing data (V_o) is estimated as:

$$V_O = a_o + \sum_{i=1}^{n}(a_i V_i) \text{ .... (Eq-11)}$$

Where, Vi is the value of same parameter at i[th] nearest weather station, and a_i, a_1,...,a_n are the regression coefficients.

SIB: In the SIB method, the closest neighbor station is used as an estimate for a target station. The target station rainfall is estimated using the same data from the neighbor station that has the highest positive correlation with the target station [26].

EM method: The Expectation Maximization method (EM) consists of a conditional expectation of missing data and a maximization step that finds the estimates of the model parameter to maximize complete data log likelihood function from the expectation step in an iteration process [27].

In recent times, enormous interest has arisen in the estimation of missing data using single imputative regressive techniques. The following techniques are also been used for filling the missing rain fall data.

Parametric Ordinary Least-Squares Regression (POLSR): Parametric least-square regression having a straight line functional form is the most widely used modeling 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. 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 over determined by default) to a balanced system consisting of n equations with n unknowns. Then the new set of equations is 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:

$Y=b*X+a$......... (Eq-12)

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.

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 are modeled by least square linear regression method.

$R_O(Y)=b*R_O(X)+a$......... (Eq-13)

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 $R_O(Xi)$ gives an estimated value of $R_E(Y_I)$ by the equation Equation 6. From the new rank, the value of Y is obtained by the following criterion.

If, $R_E(Y_i)=R_O(Ya)$

Then, $Y_i=Y_a$......... (Eq-14)

If $R_O(Y_a)<R_E(Y_i)<R_O(Y_b)$

then,

$Y_I = Y_A + [R_E(Y_i) - R_{OE}(Y_a)/R_O(Y_a) - R_O(Y_b)]*(Y_b - Y_a)$ .... (Eq-15)

If, $R_E(Y_i) > Max(R_O(Y))$

then, $R_E(Y_i) = Max(R_O(Y))$ .... (Eq-16)

If, $R_E(Y_i) < Min(R_O(Y))$

then, $R_E(Y_i) = Min(R_O(Y))$ .... (Eq-17)

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 [28]. The main advantage of inducing ordinality is to ease the collation and categorization of rainfall values. By ranking the data, its inherent behavior 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 [28].

Non-Parametric Simplified Thiel's Method (NPSTM): Thiel (1950) [29] 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 [29]. 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 ΔY/ΔX 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:

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

$\bar{Y}_I = Y_{I+(N/2)} - Y_I$ .... (Eq-18)

The regression line is given by Y=b̄*X+ā, Where, b̄ is called the angular coefficient and ā is the intercept

$\bar{b} = \bar{Y}_{median} / \bar{X}_{median}$ ...... (Eq-19)

$\bar{a} = \bar{Y}_{median} - \bar{b} * \bar{X}_{median}$ .... (Eq-20)

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 center of the cloud of points [29]. 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.

Orthogonal Regression (OR): 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 [30]. 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=b̄*X+ā. The slope is found by:

$$\bar{b} = -L + \frac{\sqrt{L^2 + R^2}}{R} \text{ .... (Eq-21)}$$

Where, $L = 0.5*[S_X/S_Y - S_Y/S_X]$ .... (Eq-22)

$\bar{a} = \bar{Y}_{mean} - \bar{b} * \bar{X}_{mean}$ .... (Eq-23)

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.

$X_I = X_{mean} + r(S_Y/S_X)[Y_I - Y_{mean}]$......... (Eq-24)

$Y_I = Y_{mean} + r(S_X/S_Y)[X_I - X_{mean}]$......... (Eq-25)

The slope of the first line is m $(r*[S_Y/S_X])$ and the slope of the second line is $\bar{m}$ $(r*[S_X/S_Y])$. 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.

$\bar{M} = (m * \bar{m})^{0.5}$......... (Eq-26)

$Y_I = Y_{mean} + sign(r) * S_X/S_Y[X_I - X_{mean}]$......... (Eq-27)

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 [31].

Conclusive Remarks

This review report provides a number of techniques for filling missing precipitation records belonging to many different environments. The outcome of data analysis depends on the quality and completeness of data. The degree of suitability of rainfall missing values estimation methods for different climatic zones needs to be determined and validated before applying on certain purposes. 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.

References

1. Coulibaly P, Evora ND. Comparison of neural network methods for infilling missing daily weather records. Journal of Hydrology. 2007; 341: 27-41.
2. Abebe AJ, Solomatine DP, Venneker RGW. Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. Hydrological Sciences Journal. 2000; 45: 425-436.
3. Ramos-Calzado P, Gomez-Camacho J, Perez-Bernal F, Pita-Lopez MF. A novel approach to precipitation series completion in climatological datasets: application to Andalusia. International Journal of Climatology. 2008; 28: 1525-1534.
4. Schneider T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. Journal of Climate. 2001; 14: 853-871.
5. Suhaila J, Deni SM, Jemain AA. Revised spatial weighting methods for

estimation of missing rainfall data. Asia-Pacific J Atmos Sci. 2008; 44: 93-104.

6. Burrough PA, McDonnell RA. Principles of geographical information systems. Oxford University Press. 1998.
7. Paulhus JLH, Kohler MA. Interpolation of Missing Precipitation Records. Monthly Weather Review. 1952; 80: 129-133.
8. Teegavarapu RSV, Chandramouli V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Journal of Hydrology. 2005; 312: 191-206.
9. Chen FW, Liu CW. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy & Water Environment. 2012; 10: 209-222.
10. Ahrens B. Distance in spatial interpolation of daily rain gauge data. Hydrology and Earth System Sciences. 2006; 10: 197-208.
11. Teegavarapu RSV. Estimation of missing precipitation records integrating surface interpolation techniques and spatio-temporal association rules. Journal of Hydroinformatics. 2009; 11: 133-146.
12. Presti RL, Barca E, Passarella G. A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin (Italy). Environmental Monitoring and Assessment. 2010; 160: 1-22.
13. Daly C, Neilson RP, Phillips DL. A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. Journal of Applied Meteorology. 1994; 33: 140-158.
14. Malek MA, Harun S, Shamsuddin SM, Mohamad I. Reconstruction of Missing Daily Rainfall Data Using Unsupervised Artificial Neural Network. International Journal of Electrical and Computer Engineering. 2009; 4: 340-345.
15. Simolo C, Brunetti M, Maugeri M, Nanni T. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. International Journal of Climatology. 2010; 30: 1564-1576.
16. Chow VT, Maidment DR, Mays LW. Applied Hydrology, Mc Graw Hill Book Company. 1988.
17. Tabios GQ, Salas JD. A comparative analysis of techniques for spatial interpolation of precipitation. Water Resource Bull. 1985; 21: 365-380.
18. Singh VP. Elementary Hydrology. 1994.
19. Young KC. A three-way model for interpolating monthly precipitation values. Monthly Weather Review. 1992; 120: 2561-2569.
20. Lam NS. Spatial Interpolation Methods review. The American Cartographer. 1983; 10: 129-149.
21. Wei TC, McGuinness JL. Reciprocal Distance SquaredMethod: A Computer Technique for Estimating Area Precipitation. US Agricultural Research Service. 1973.
22. Mardikis MG, Kalivas DP, Kollias VJ. Comparison of Interpolation Methods for Prediction of Reference Evapotranspiration-An Application in Greece. Water Resources Management. 2005; 19: 251-278.
23. Eklundh L, Pilesjo P. Regionalization and Spatial Estimation of Ethiopian Mean Annual Rainfall. International Journal of Climatology. 1990; 10: 473-494.
24. Seleshi Y, Zanke U. Recent Changes in Rainfall and Rainy Days in Ethiopia. International Journal of Climatology. 2004; 24: 973-983.
25. Eischeid JK, Baker CB, Karl TR, Diaz HF. The quality control of long-term climatological data using objective data analysis. Journal of Applied Meteorology and Climatology. 1995; 34: 2787-2795.
26. Kashani HM, Dinpashoh Y. Evaluation of efficiency of different estimation methods for missing climatological data. Journal of Stochastic Environment Research Risk Assessment. 2012; 26: 59-71.
27. Schneider T. Analysis of incomplete climate data: Estimation of Mean Values and covariance matrices and imputation of Missing values. Journal of Climate. 2001; 14: 853-871.
28. Iman RL, Conover WJ. The use of the rank transform in regression. Technometrics. 1979; 21: 499-509.
29. Theil H. A rank-invariant method of linear and polynomial regres- sion analysis. Henri Theil's Contributions to Economics and Econometrics. 1950; 53: 386–392.
30. Poller L, van den Besselaar AM, Jespersen J, Tripodi A, Houghton D. A comparison of linear and orthogonal regression analysis for local INR determination in ECAA coagulometer studies. British Journal of Haematology. 1998; 102: 910-917.
31. Halfon E. Regression Method in Ecotoxicology: A Better Formulation Using the Geometric Mean Functional Regression. Environmental Science and Technology. 1985; 19: 747-749.