

Supplementary Material

Geostatistical and deterministic methods for rainfall interpolation in the Zayandeh Rud basin, Iran

Farshad Jalili Pirani, Reza Modarres

Semi-variogram models

The semi-variogram models used in the study are given below.

Spherical model

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(\frac{3h}{2\alpha} - \frac{h^3}{2\alpha^3} \right) & 0 < h \leq \alpha \\ c_0 + c & h > \alpha \\ 0 & h = 0 \end{cases} \quad (S1)$$

Gaussian model

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(1 - \exp\left(-\frac{h^2}{\alpha^2}\right) \right) & h > 0 \\ 0 & h = 0 \end{cases} \quad (S2)$$

Exponential model

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(1 - \exp\left(-\frac{h}{\alpha}\right) \right) & h > 0 \\ 0 & h = 0 \end{cases} \quad (S3)$$

Circular model

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(1 - \frac{2}{\pi} \cos^{-1}\left(\frac{h}{\alpha}\right) + \sqrt{1 - \frac{h^2}{\alpha^2}} \right) & 0 < h \leq \alpha \\ c_0 + c & h > \alpha \\ 0 & h = 0 \end{cases} \quad (S4)$$

Tetraspherical model

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(\frac{2}{\pi} \arcsin\left(\frac{h}{\alpha}\right) + \frac{2h}{\pi\alpha} \sqrt{1 - \left(\frac{h}{\alpha}\right)^2} + \frac{4h}{3\pi\alpha} \right) & h \leq \alpha \\ c_0 + c & h > \alpha \\ 0 & h = 0 \end{cases} \quad (S5)$$

Pentaspherical model

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(\frac{15h}{8\alpha} - \frac{5}{4} \left(\frac{h}{\alpha}\right)^3 + \frac{8}{3} \left(\frac{h}{\alpha}\right)^5 \right) & h \leq \alpha \\ c_0 + c & h > \alpha \\ 0 & h = 0 \end{cases} \quad (S6)$$

Rational quadratic

$$\gamma(h, c_0, c, \alpha) = \begin{cases} c_0 + c \left(\frac{19 \left(\frac{h}{\alpha} \right)^2}{1 + 19 \left(\frac{h}{\alpha} \right)^2} \right) & h > 0 \\ 0 & h = 0 \end{cases} \quad (S7)$$

Kriging methods

The kriging methods used in the study are explained below in detail.

Ordinary kriging (ORK)

The basic theory behind the kriging method is that weights are produced to ensure that the estimations are unbiased and have the least variance. To ensure the use of unbiased weights, the sum of all weights must be equal to 1. Thus, we are confident that the difference between measured Z_{s_o} and generated data is low or almost equal to zero:

$$(Z_{s_i} - \sum_{i=1}^{i=n} \lambda_1 Z_{s_i})^2 = 0 \quad (S8)$$

The ORK approach suggests that the weight of each measured point used in estimating unmeasured points is a linear function of the covariance between each active point and other points, including the passive point. Therefore, a system of multiple equations and multiple variables (here weights) is characteristic of this method, which can be summarized as follows:

$$\begin{cases} \sum_{i=1}^{i=n} (\lambda_1 \cdot \gamma_{ij}) - \mu & \text{for } j = 1, 2, \dots, n \\ \sum_{i=1}^{i=n} \lambda_1 = 1 \end{cases} \quad (S9)$$

$$\mathbf{\Gamma} \times \boldsymbol{\lambda} = \mathbf{G} \Rightarrow \boldsymbol{\lambda} = \mathbf{\Gamma}^{-1} \times \mathbf{G} \quad (S10)$$



$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ m \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix} \quad (S11)$$

where, $\mathbf{\Gamma}$ is a matrix with each row having a couple of elements, such as the variance between each measured point and other measured points excluding the passive point; $\boldsymbol{\lambda}$ is the weight vector (each row having just one element, which is the weight of each point corresponding to the same row of \mathbf{G}); and \mathbf{G} is the variance vector between unmeasured points and surrounding measured points (again with each row containing only one element corresponding to the same row of $\mathbf{\Gamma}$ and $\boldsymbol{\lambda}$), which is obtained through the semi-variogram. Consequently, the weight vectors, λ_1 to λ_n are obtained by multiplying the inverse of $\mathbf{\Gamma}$ by \mathbf{G} . The weights obtained by this process are then inserted into Eq. (1) to produce the prediction for a passive point. For each unmeasured point, the variance matrix of $\mathbf{\Gamma}$ and variance vector of \mathbf{G} are formed, and related weights are calculated.

Universal kriging (UNK)

Ordinary kriging assumes that the desired parameter (e.g. precipitation) is solely a function of the spatial correlation. However, universal kriging considers a structural trend for a variable, as well as a spatial correlation between known points. The weights are determined such that the estimation is unbiased (Eq. 14) and variance is minimized (Eq. 15). Using the Lagrange multiplier (μ) one minimizes Eq. (15) in a manner that is subject to Eq. (14), resulting in the following UNK system:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma_{ij} + \sum_{i=1}^L \mu_i f_i^l = \gamma_{i0} & \text{for } j = 1, 2, \dots, n \\ \sum_{i=1}^{ns} \lambda_i f_i^l = f_0^l \end{cases} \quad (\text{S12})$$

where γ_{ij} is the semivariogram between two points i and j , and μ_i is the Lagrange parameter associated with l^{th} unbiased condition, and f is the mean and is a function of spatial coordinates. This study considered a linear trend, thus $L=2$, $f_i^0=1$, $f_i^1=x_i$, $f_i^2=y_i$, where x and y are the abscissa and ordinate of the points when $L=0$, $\sum_{i=1}^{ns} \lambda_i = 1$.

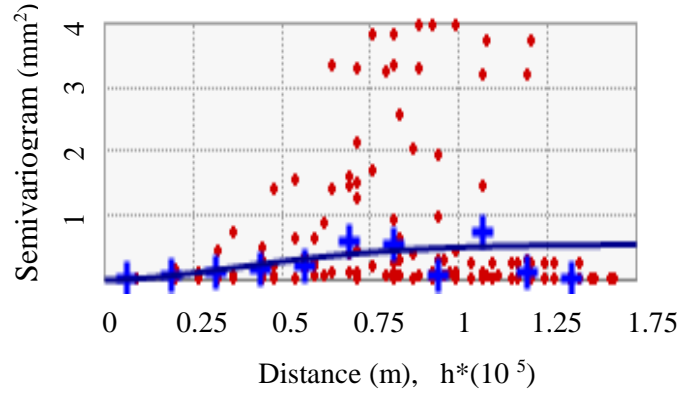


Figure S1. The fitted Gaussian model (blue line) to the experimental semi-variogram in interpolation for Chelgerd station using 27 gauges. Red dots are the semi-variogram values *versus* different distances; blue crosses are the average semi-variogram values for each distance lag.

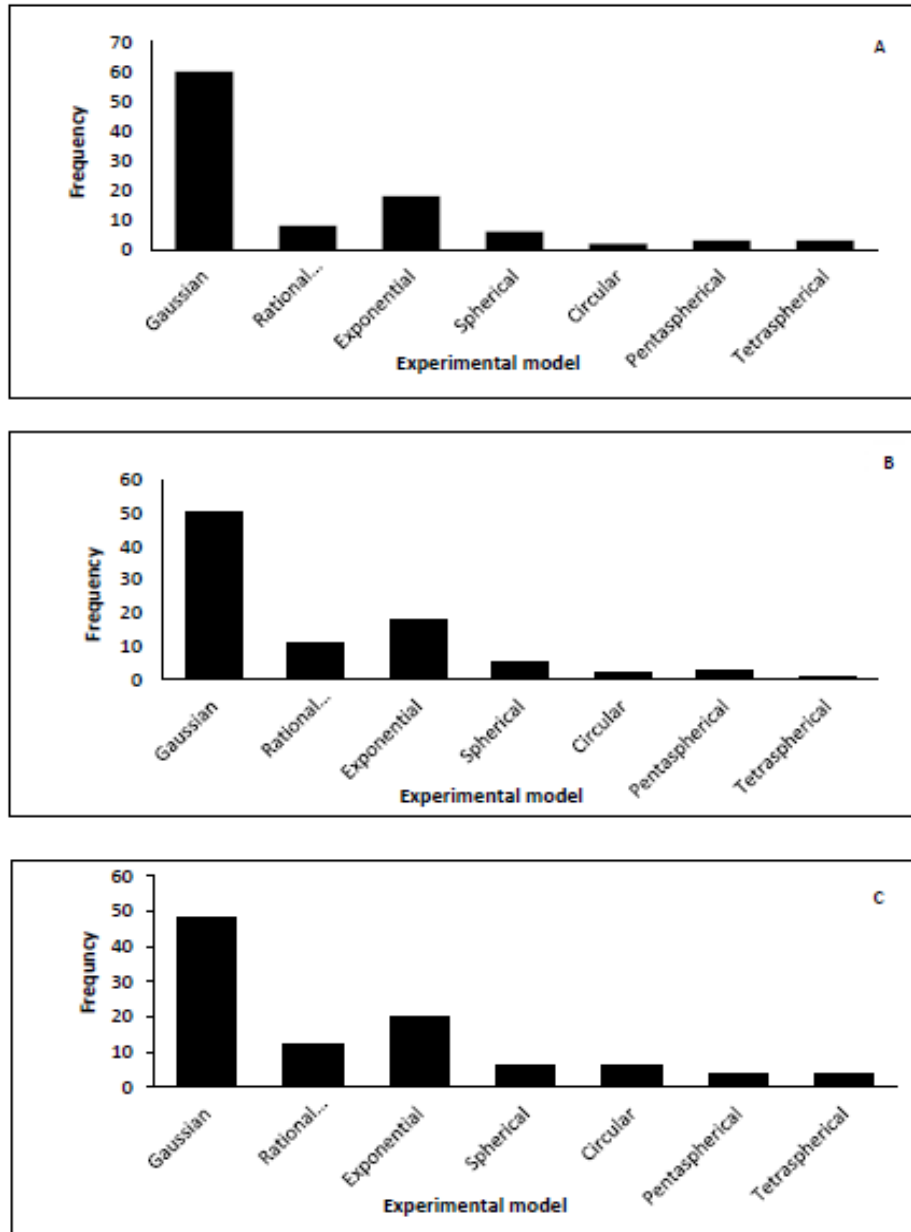


Figure S2. Number of each fitted experimental model (seven models) to semi-variogram of seasonal average for all 20 years: A: Chelgerd, B: Chadegan and C: Mirabad.

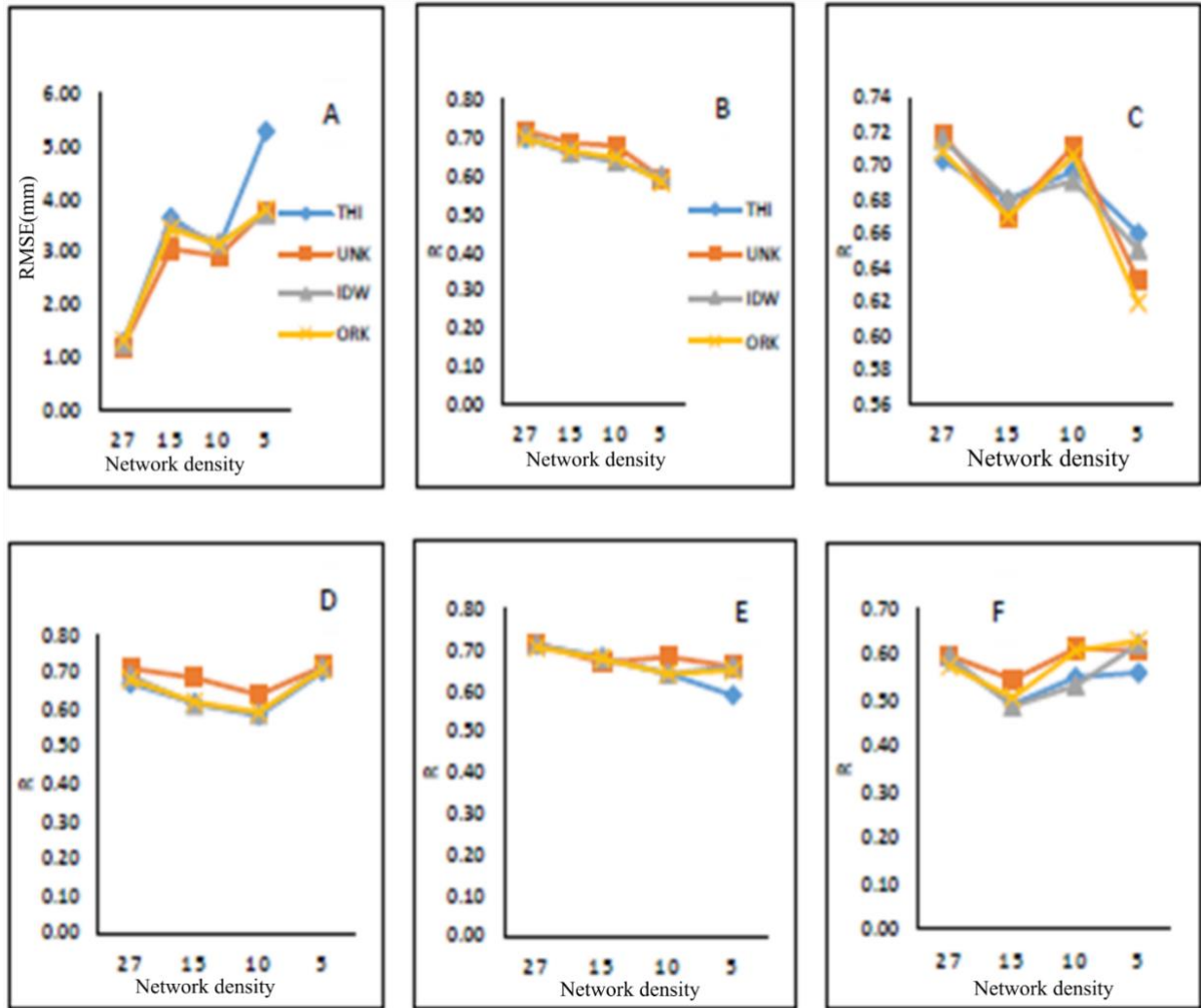


Figure S3. RMSE and R values obtained using different methods for the Mirabad station. A and B represent the RMSE and R , respectively, for the whole 20-year period; C, D, E and F represent R values of spring, autumn, winter and summer, respectively.

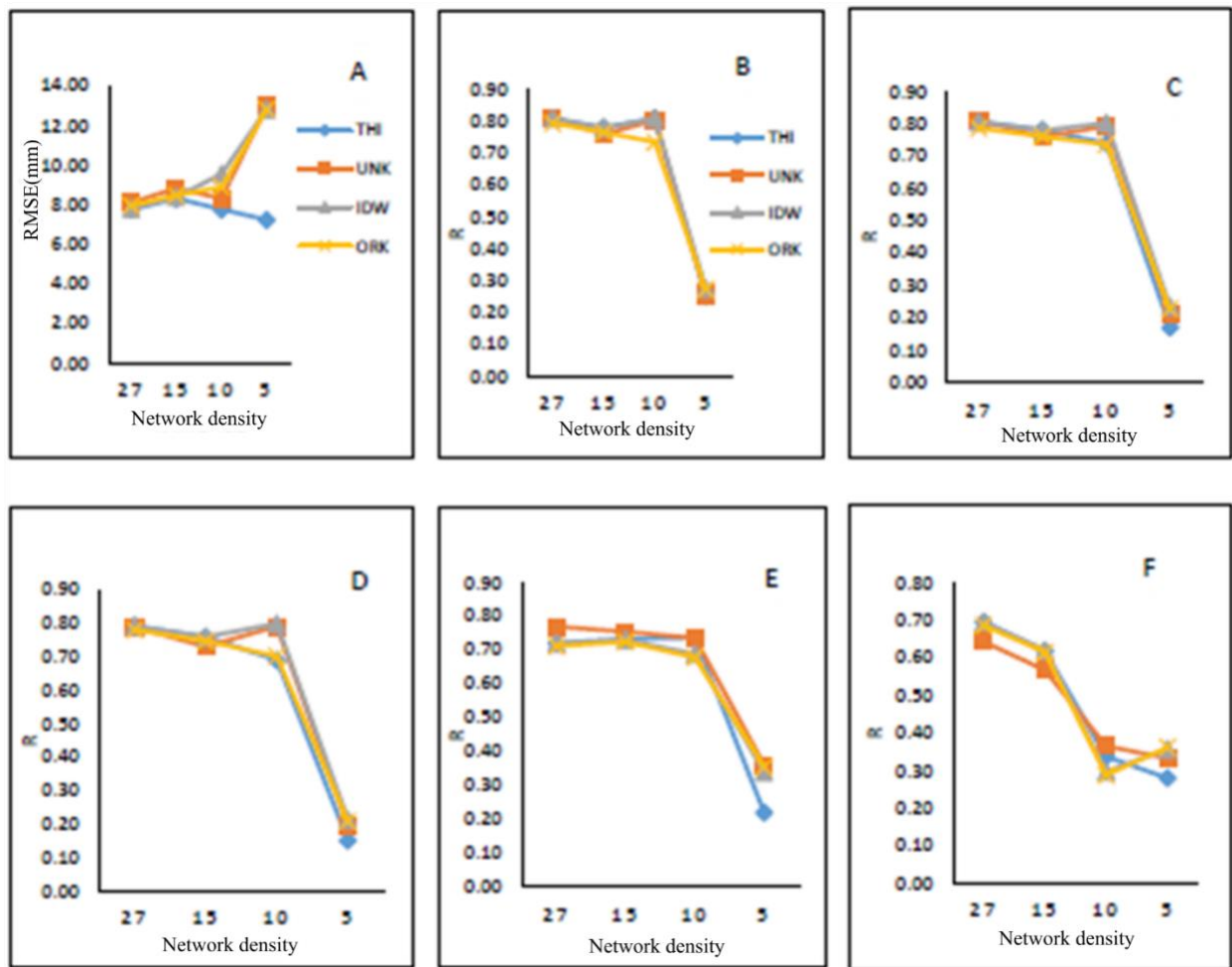


Figure S4. RMSE and R values obtained using different methods for the Chelgerd station. A and B represent the RMSE and R , respectively, for the whole 20-year period; C, D, E and F represent R values of spring, autumn, winter and summer, respectively.

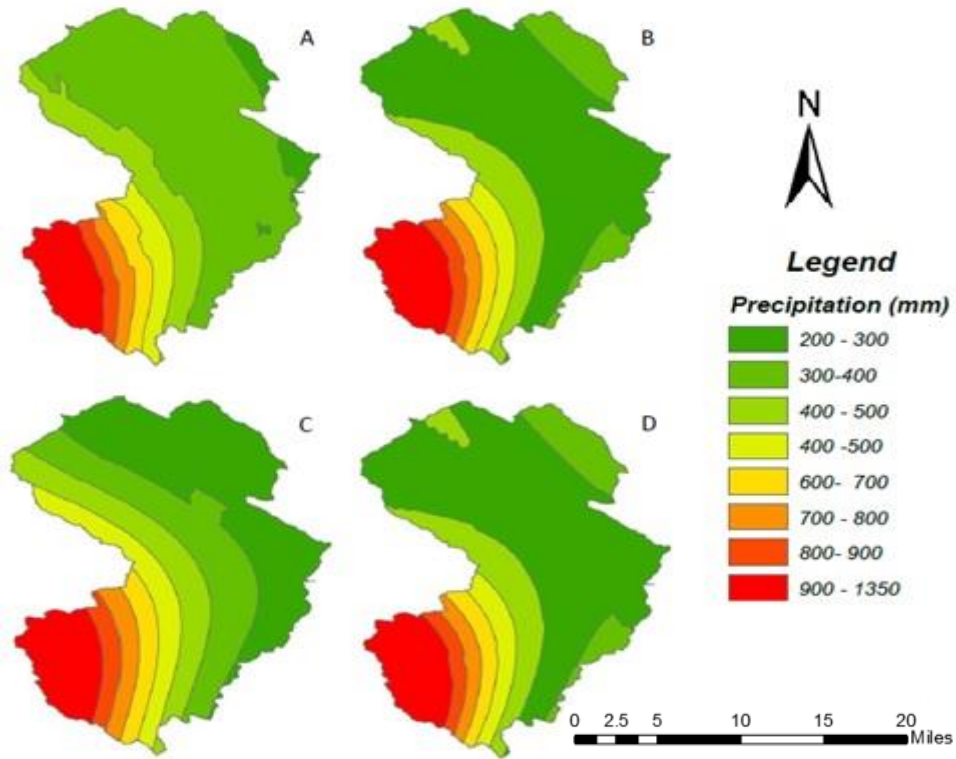


Figure S5. Annual mean rainfall map generated through the best interpolation algorithm (UNK) and network sizes of 27 (A), 15 (B), 10 (C) and 5 (D). The interpolation was carried out without the Mirabad station.