International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

Performance Comparison of Interpolation Methods for Precipitation Data Reconstruction: Kriging, Polynomial Interpolation, Cubic Splines, and Neural Networks

Evelin Tayupanta Portilla[1], Adriana Vilela Carrillo[2], Pablo Páliz Larrea[3]
[1]Universidad Nacional de Chimborazo, Author 1, Chimborazo, Ecuador, evelin.tayupanta@unach.edu.ec ORCID https://orcid.org/0009-0007-8222-4698
[2]Universidad Nacional de Chimborazo, Author 2, Chimborazo, Ecuador, adriana.vilela@unach.edu.ec , ORCID https://orcid.org/0009-0001-7087-3064
[3]Universidad de las Fuerzas Armadas ESPE, Author 3, Sangolquí, Ecuador, pipaliz@espe.edu.ec , ORCID https://orcid.org/0009-0000-9860-7465
* Corresponding author: pipaliz@espe.edu.ec

Abstract: *Reliable and complete precipitation data are essential for hydrological modeling and water resource management. However, gaps in records due to sensor failures, human error, or limited station coverage can compromise analysis quality. This study evaluates the performance of four interpolation techniques: artificial neural networks, multiple linear regression, ordinary kriging, and cubic splines for estimating missing daily precipitation values in the Cedar River basin, a mountainous region in Washington State, USA. Prior to interpolation, data quality control was applied using double mass curve analysis and Pettitt's test. Performance was assessed using RMSE, MAE, Pearson correlation, and Nash–Sutcliffe efficiency. Results indicate that artificial neural net-works provided the most accurate estimates (RMSE = 2.64 mm; Correlation coefficient: = 0.98 and Nash–Sutcliffe efficiency= 0.96), followed by cubic splines, kriging, and multiple regression. Neural networks effectively captured nonlinear patterns in precipitation but required specialized knowledge and more computational resources. Kriging offered a robust and simpler alternative when spatial structure was well-defined, while regression worked well when station correlations were high. Cubic splines performed poorly under high temporal variability. The findings suggest that neural networks are best suited for complex conditions, but traditional methods re-main valid in operational settings with limited resources.*
Keywords: Precipitation, Interpolation, Artificial Neural Networks, Kriging, Multiple Regression, Cubic Splines, Data Quality Control

1. INTRODUCTION

Precipitation plays a fundamental role in water resource management, climate modeling, and natural disaster mitigation. Accurate measurement of this variable is essential for forecasting extreme events such as droughts and floods, optimizing water allocation, and improving the calibration and performance of hydrological models) [1], [2]. However, gaps in precipitation time series remain a recurrent issue in climate research, often caused by sensor failures, extreme weather conditions, or human error during data collection [3].

These discontinuities impair the reliability of predictive models, hinder long-term climatic analyses, and introduce uncertainty into water management strategies. Moreover, data acquisition is frequently limited by the uneven spatial distribution of weather stations and by socioeconomic and technological constraints, such as high installation and maintenance costs or the lack of infrastructure in remote areas [3].

To address these limitations, a wide range of interpolation techniques have been developed to estimate missing data and ensure the temporal continuity of precipitation records [4]. These methods have proven critical for enhancing the spatial and temporal resolution of climate datasets in regions with sparse observational networks. Their effectiveness, however, depends on several factors, including the spatial variability of precipitation, the number and distribution of stations, and the temporal granularity of available data [5], [6].

A broad spectrum of techniques has been employed in the literature. Geostatistical methods such as regression kriging have shown good spatial performance in mountainous regions with sparse networks [7]. Statistical regression and machine learning models have also been used to reconstruct incomplete records, demonstrating flexibility in complex or data poor environments [8]. More recently, high-resolution global precipitation models have been developed using non-stationary SPDE structures and latent Gaussian processes, which effectively represent fine-scale spatial patterns [9].

In parallel, deep learning methods have emerged as promising tools for precipitation estimation. For

instance, Temporal Frame Interpolation (TFI) has shown superior performance in recovering extreme events from meteorological imagery [10], while generative models based on radar data have improved short-term rainfall nowcasting [11]. Other architectures, such as U-Net convolutional neural networks, have been applied to enhance precipitation prediction from radar imagery [12].

In the context of gap filling, [4] reported that kriging performed best when auxiliary variables were available, while [13], [14] confirmed its superior accuracy over deterministic approaches in both daily and monthly applications. In addition, Papacharalampous and collaborators in 2023 demonstrated the utility of machine learning algorithms, such as Light GBM and Random Forest, in capturing the spatial complexity of precipitation using satellite data [3].

The Cedar River basin, located in Washington State, USA, is a critical water supply source for the Seattle metropolitan area and plays a key role in maintaining ecological processes such as surface runoff, aquifer recharge, and forest soil moisture retention. Given its hydrological significance and ecological sensitivity, particularly in the context of climate change, accurate characterization of precipitation dynamics in this region is of high practical relevance [15], [16].

This study aims to evaluate and compare the performance of four interpolation methods: ordinary kriging, polynomial regression, cubic splines, and artificial neural networks for reconstructing missing daily precipitation data in the Cedar River basin. The objective is to determine the most accurate and operationally viable method under conditions of low station density and high climatic variability.

2. METHODS

2.1 Study Area and Climatic Characteristics

This study was conducted in the Cedar River Basin, located in the state of Washington, in the northwestern United States, near the city of Seattle. The region combines mountains, rivers, and temperate forests. Its topographic diversity generates broad variability in the local climate, which is relevant for analyzing precipitation pattern [15]. As shown in Figure 1, the analysis focuses on six meteorological stations located within the Cedar River watershed: Cougar Mountain (CM), Olallie Meadows (OM), Meadows Pass (MP), Mount Gardner (MG), Tinkham Creek (TC), and Rex River (RR). The precipitation information was obtained from the databases of the National Oceanic and Atmospheric Administration (NOAA), which is a free global repository of climatological data [17]. For this research, data will be used from January 1, 1996, to December 31, 2023 (a total of 27 years)

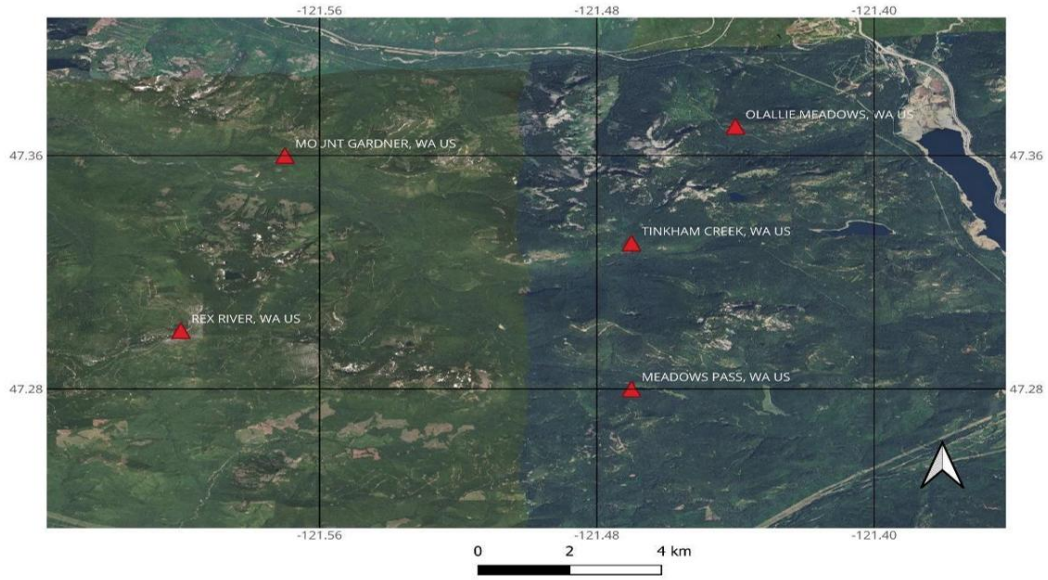

Figure 1. Location of meteorological stations in the study area

This basin belongs to the Cedar River system and is part of the greater Puget Sound watershed, supplying drinking water to a significant portion of the Seattle metropolitan area [15]. The mountainous terrain and dense forests that characterize this area create favorable conditions for the study of hydrological processes. Additionally, the basin exhibits steep altitudinal gradients and microclimatic variations that directly influence the spatial distribution of precipitation, making this region ideal for analyzing estimation and prediction methods [16]

The selection of this region is motivated by its climatic complexity and strategic importance. The Cascade

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

Range acts as a natural barrier, creating diverse microclimates within a single region. For instance, while some areas may receive more than 6000 mm of annual precipitation, others show considerably drier conditions [16]. This variation enables the evaluation of the behavior of different precipitation estimation techniques under contrasting climatic contexts, as illustrated in Figure 2.

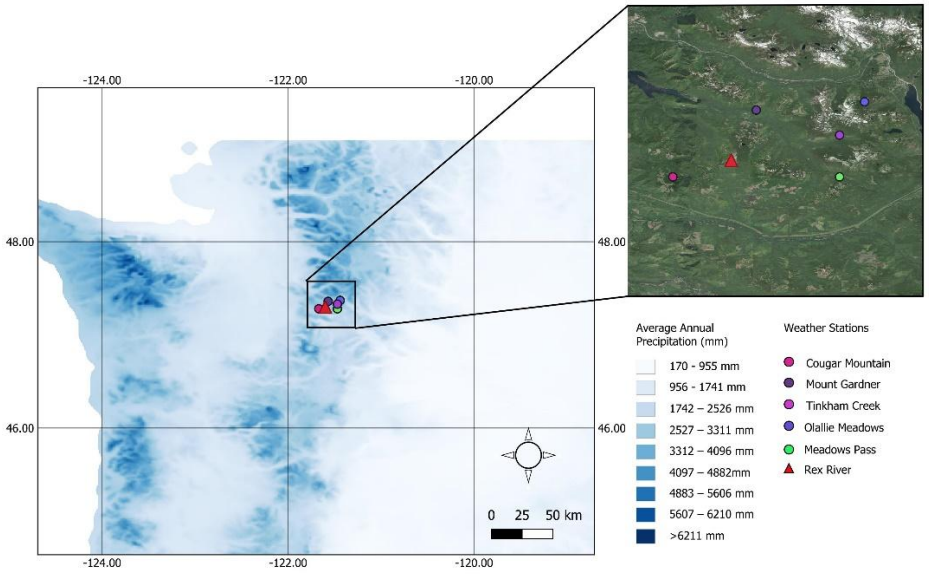

Figure 2. Map of mean annual precipitation in the state of Washington, based on a 30-year reference period

Furthermore, this region has been identified as vulnerable to the effects of climate change, as the state of Washington has experienced sustained warming in recent decades, along with a decrease in snow accumulation during winter and shifts in seasonal precipitation patterns. These changes include reduced summer rainfall and increased winter precipitation, which directly impact the region's hydrological regime and pose challenges for sustainable water management [16]. These changes have direct consequences on the water cycle. Snowmelt now occurs earlier, base flows in rivers decrease during the dry season, and the frequency of extreme events such as droughts or intense rainfall has increased.

2.2 Overview of the Interpolation Models

2.2.1 Artificial Neural Networks

In machine learning, Artificial Neural Networks (ANNs) are used to approximate nonlinear functions through hierarchical structures composed of interconnected layers of neurons [4]. As shown in Figure 3, these layers are divided into three fundamental types: the input layer, where initial data from the problem are received (e.g., precipitation series); the hidden layer, where the data are processed through activation functions that allow the detection of complex patterns; and the output layer, where the final result of the prediction or estimation is obtained [18].

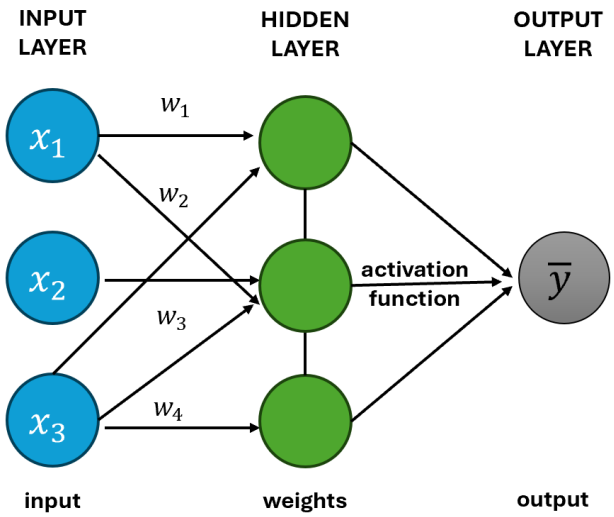

Figure 3. Schematic representation of an artificial neural network with an input layer, a hidden layer, and

an output layer.

Each neuron may take a linear combination of the inputs and pass it through a nonlinearity, enabling the model to learn representations of complex relationships such as precipitation, which may be nonlinear and dependent on multiple terms [12]. This is illustrated in Table 1.

Table 1. Components of a multilayer neural network

Component	Description
Input layer	A set of nodes that receive the input variables (e.g., climate or meteorological data). Each node represents a measured feature or value.
Weights	Parameters that multiply each input before being processed by a neuron; they determine the influence of that value on the output.
Bias	A value added to the weighted sum that allows shifting the activation function, enabling a more accurate output fit.
Activation function	Nonlinear functions applied in hidden and output neurons to generate outputs adapted to complex patterns.
Hidden layer	Intermediate nodes that process signals from the input layer by applying the activation function. These allow modeling nonlinear relationships.
Output layer	Final neurons that deliver the estimated result (e.g., interpolated daily precipitation), after processing the signal through the previous layers.
Prediction error	The difference between the network's estimated value and the actual value. It is used to feed back the model during training.

A basic multilayer perceptron (MLP) is expressed as:

$$y = f\left(\sum_{j=1}^{n_h} w_j \times g\left(\sum_{i=1}^{n_x} w_{ji} \times x_i + b_j\right) + b\right)$$

Where: $x_i$are the inputs, w_{ji} are the weights between the input and the hidden neuron j, $b_j$are the bias terms for the hidden neuron j, g is the activation function, w_j are the weights from hidden neuron j to the output neuron, b is the bias of the output neuron, f is the output layer activation function, n_x is the number of input variables, and n_h is the number of neurons in the hidden layer.

Neural networks are trained using algorithms such as error backpropagation and gradient descent optimization, enabling the model to learn the connections between meteorological stations and to predict available values. They work particularly well under conditions characterized by a large volume of data and when high-degree nonlinearity is present [11]

2.2.2 Multiple Linear Regression

Multiple linear regression is a statistical method that models the linear relation-ship between a dependent variable and two or more independent variables. Its application assumes that the explanatory variables are independent and that the residuals of the model are normally distributed [6]. The prediction is based on a linear combination of the variables considered and follows the general form:

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon$$

Where Y is the predicted value of the dependent variable (e.g., precipitation), β_0 is the intercept, β_i are the coefficients corresponding to each independent variable, X_i represent the values of the independent variables and ϵ is the random error term or residual.

This method is useful when the available meteorological stations are close to each other, and their precipitation series exhibit a strong linear correlation. Its ease of implementation and interpretation makes it an attractive tool for hydrological applications in operational contexts [8].

2.2.3 Cubic Splines

Cubic splines are piecewise polynomial functions used to smoothly interpolate data. They ensure continuity of the function and its first and second derivatives at the knot (joining) points [19]. For each interval $[x_i, x_{i+1}]$, a cubic polynomial is defined as:

$$S_{i(x)} = a_i + b_{i(x - x_i)} + c^2_{i(x - x_i)} + d^3_{i(x - x_i)}$$

With the interpolation conditions that each spline must pass through the points in the table $S_i(x_i) = y_i$ y $S_i(x_{i+1}) = y_{i+1}$, the first derivative must be continuous at interior nodes $S'_i(x_{i+1}) = S'_{i+1}(x_{i+1})$ and the second derivative must also be continuous at interior nodes $S''_i(x_{i+1}) = S''_{i+1}(x_{i+1})$ [19] This method provides a continuous and smooth way to interpolate values in time series where abrupt fluctuations or overshooting are to be avoided. It is particularly useful when there are missing data within well-defined intervals [13].

2.2.4 Kriging

Kriging is a geostatistical interpolation method based on stochastic process theory and semivariance analysis. It assumes that the distance between points affects the similarity of their values, allowing the construction of spatial dependence models[20]. The estimation at a point x_0 is a linear combination of observed values:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad (4)$$

Subject to the constraint:

$$\sum_{i=1}^{n} \lambda_i = 1 \quad (5)$$

Where $\lambda_i$are the weights assigned to each observation $Z(x_i)$. These weights are determined from the variogram, which describes semivariance as a function of the distance hhh, where N(h) is the number of point pairs separated by distance h, and $\gamma(h)$ is the empirical semivariance for that distance [20]):

$$\boldsymbol{\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2} \quad (6)$$

Ordinary kriging adjusts the weights based on the spatial dependence structure described by the semivariogram, which allows minimizing the mean squared error of estimation and calculating the associated uncertainty [20]. This approach is particularly effective in environments with high spatial correlation, such as precipitation series in mountainous regions, where even elevation can introduce variability in values. However, properly modeling the variogram and performing cross-validation entails a considerable computational cost [21].

2.3 Data Quality Control

A homogeneity test was conducted to detect changes in the historical data series caused by station relocations. A double mass curve was used to identify shifts in the reference station relative to the others [22]. When a slope change was detected, the Pettitt test was applied to confirm whether a significant shift existed in the median at a change point for non-parametric data[23] The procedure consists of assigning ranks to the ordered data and calculating the U_k statistic for each possible change point k, where R_i is the rank of the i -th data point in the series and n is the total number of observations:

$$U_k = 2 \sum_{i=1}^{k} R_i - k(n + 1) \quad (7)$$

The test statistic is defined as K, where the point that maximizes $|U_k|$ is considered the most probable moment of change, evaluated with a given level of significance:

$$K = \max_{1 \le k \le n} |U_k| \quad (8)$$

Deterministic methods were used to fill in missing daily precipitation data. Correlation between stations was first established, followed by the fitting of a polynomial trend line to the data. Missing values were imputed based on the highest correlation with the nearest stations. Outliers in daily maximum precipitation time series were identified using the univariate method proposed by the American Water Resources Association (AWRA)[24]. This approach assumes a log-normal distribution of the data and defines thresholds for detecting outliers based on the logarithmic transformation of observations. Given a data set $\{y_1, y_2,,..., y_n,\}$ the following transformation was applied:

$$x_i = \log_{10}(y_i) \quad (9)$$

From the transformed series, the mean and standard deviation were calculated. Subsequently, the detection thresholds were defined as follows:

$$x_H = \underline{x} + k_N \times S_x \quad (10) \qquad x_L = \underline{x} - k_N \times S_x \quad (11)$$

where k_N is an empirical coefficient that depends on the sample size NNN and is computed as:

$$k_N = 3.2201 - 6.6246N^{-0.25} + 2.47832N^{-0.25} - 0.491436N^{-0.75} + 0.037911N^{-1} \quad (12)$$

The original values $y_i$are considered outliers if $\log_{10}(y_i) > x_H$ for a high outlier, or $\log_{10}(y_i) < x_L$ for a low outlier. The decision to evaluate high, low, or both types of outliers was based on the kurtosis of the transformed series. If the kurtosis exceeded +0.4, only high outliers were evaluated; if it was less than –0.4, only low outliers were considered; and if it was within the threshold range, both extremes were assessed.

2.3.1 Model Configuration

The use of cross-validation in the development of machine learning models is common. This methodology consists of dividing the data into three different groups called training, validation, and testing. The first group is used to train the neural network under different configurations. The second group validates the neural network's performance during training. Finally, the test group is used to evaluate the model with new information that the model has not previously processed.

The groups were divided into 80% for training (8115 days), 10% for validation (1114 days), and 10% for testing the models (1114 days as well). Since the goal was to assess the model's ability to fill in missing data, 20% of the data in the test set were randomly removed. A total of 204 values were left for imputation.

In this study, an input–output neural network available in the neural net time series library of ®MATLAB was used[25]. The inputs were normalized within a range of $(-1, 1)$ to prevent certain inputs from having disproportionate weight. A tangential sigmoid transfer function was used to capture the normalization range of the data. The output equation was set as linear to allow the interpretation of high precipitation values that were not included during training. The Levenberg–Marquardt training algorithm was used due to its short training time and satisfactory results [26]. An iterative script was implemented to determine the best neural network architecture, training models with 10 to 50 nodes in a single hidden layer. The best-performing model in the validation stage was selected.

For the multiple linear regression, the method was applied using as independent variables the precipitation records from five nearby stations (Cougar Mountain, Olallie Meadows, Meadows Pass, Mount Gardner, Tinkham Creek), aiming to predict the values at a target station which is Rex River. For the Kriging method, the geographic coordinates of the meteorological stations were used to estimate missing values at the target station based on an exponential covariance function derived from the inter-station distance. Lastly, cubic spline interpolation was applied exclusively to the time series of the test segment of the target station, filling missing values based on their temporal continuity.

2.4 Model Performance Metrics

To evaluate the performance of the implemented models, three main metrics were used: Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Pearson Correlation Coefficient (CC). MAE provides a measure of the average magnitude of the prediction error, regardless of its direction. MSE penalizes large errors more severely, as it squares the differences, making it useful for detecting significant deviations between observed and predicted values. Finally, the Pearson Correlation Coefficient measures the strength and direction of the linear relationship between actual values and the values estimated by the model. Its value ranges from -1 to 1, where values close to 1 indicate a strong positive correlation, and values near 0 indicate little or no correlation. Additionally, the Taylor diagram was used as a multi-objective evaluation tool, as it allows for a graphical representation of the relationship between observed and modeled series using three key statistical metrics: the Pearson correlation coefficient, the Root Mean Square Error (RMSE), and the standard deviation[27].

3. RESULTS AND DISCUSSION

3.1 Results of Data Pre-processing

The precipitation records showed no signs of changes in the slope of the double mass curve, assuming homogeneity of the reference station (Rex River). No outliers were detected at any station based on the AWRA analysis. Nevertheless, the dispersion values of the rain gauge stations indicated high intermonthly and interannual variability in precipitation, with maximum values exceeding 250 mm/day in some cases, as shown in figure 4. Based on these results, it can be inferred that the region exhibits high climatic variability, possibly caused by the area's rugged geography or the vegetation in the study area.

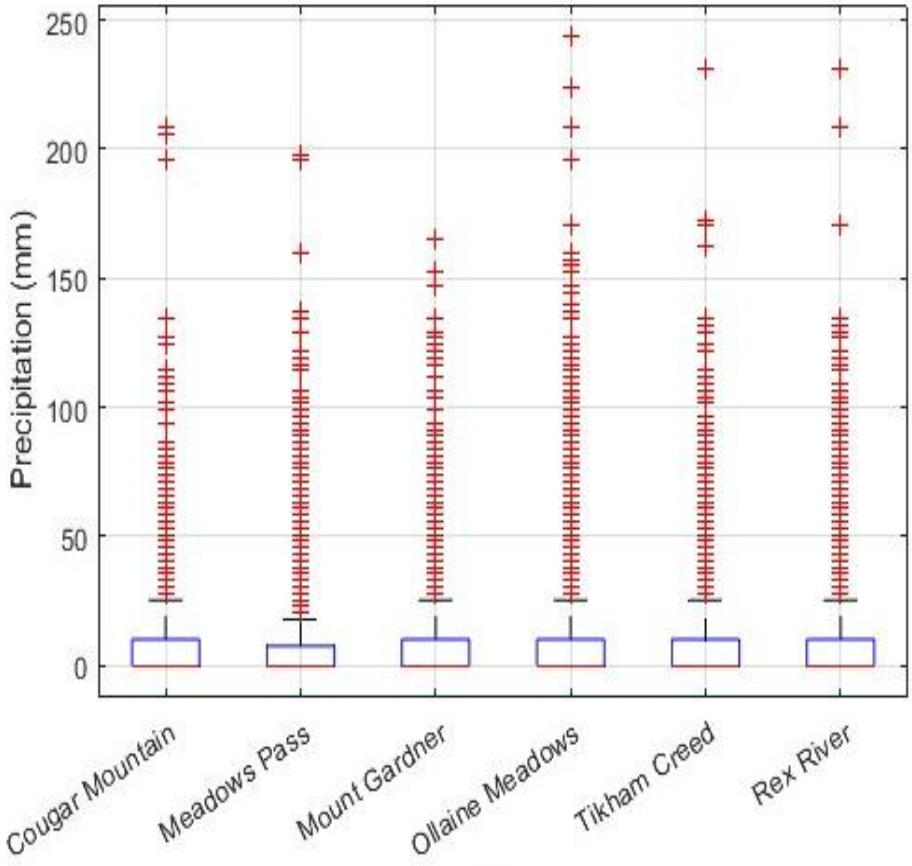

Figure 4. Boxplot of daily precipitation data

A high correlation was observed among most stations, with a maximum value of 0.93 between stations Meadows Pass and Rex River, and a minimum value of 0.72 between stations Ollalie Meadows and Tikham Creed. This pattern of high correlation supports the use of these series as cross-predictor variables in the imputation models, as information from neighboring stations can enhance the estimation of missing data in stations with incomplete records.

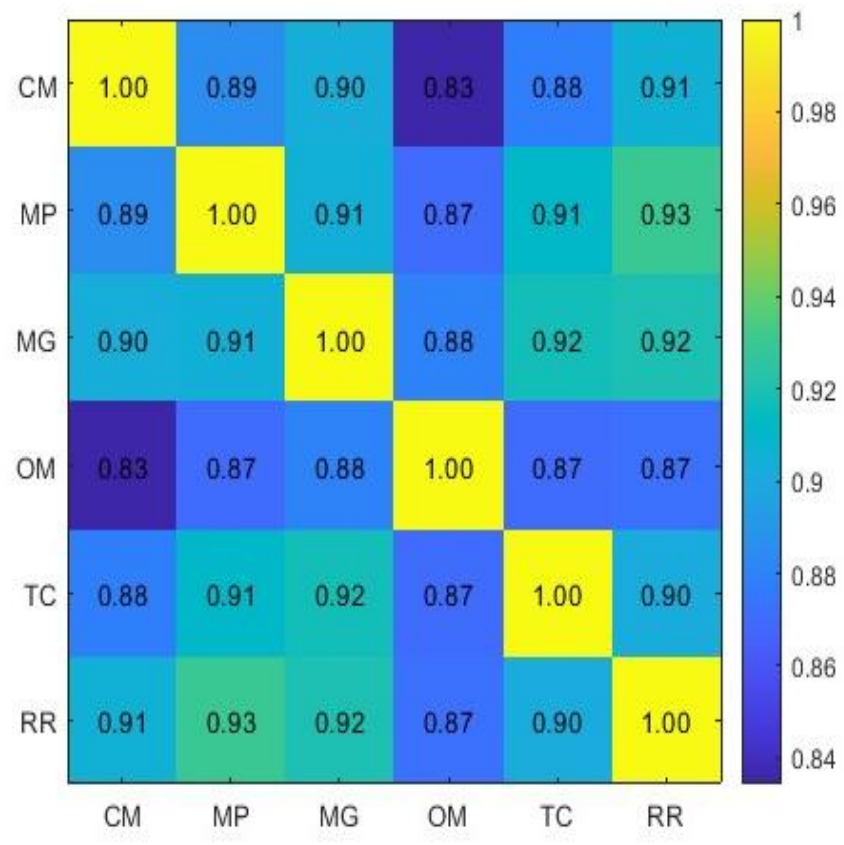

Figure 5. Precipitation heatmap of stations. Symbology: Cougar Mountain (CM), Olallie Meadows (OM), Meadows Pass (MP), Mount Gardner (MG), Tinkham Creek (TC), and Rex River (RR).

3.2 Model Results

Table 2 presents the final results of the neural network, Kriging, multiple linear regression, and cubic spline models along with their performance metrics. Likewise, the final results can be seen in Figure 6

Model	**MAE**	**MSE**	**CC**	**NS**
Neural Networks	1.55	8.46	0.983	0.966
Multiple Linear regression	1.77	11.11	0.978	0.955
Kriging	2.0136	19.8994	0.9665	0.9192
Cubic Splines	9.59	279.98	0.441	-0.137

Simbology: MAE= Mean Average Error, MSE= mean square error, CC= Correlation coeffitient, N=: Nash–Sutcliffe efficiency

In general, the neural network model achieved the best performance among the evaluated methods, presenting a MAE of 1.55, an MSE of 3.32, a correlation coefficient of 0.95, and an NSE of 0.90. These results demonstrate high predictive capability and low error dispersion, indicating that this model fits the actual data well. This behavior is consistent with the findings of [28], who developed a model based on automated neural architecture search (AdaNAS), obtaining a MAE of 0.98 mm/day and an RMSE of 2.04 mm/day in precipitation estimation, significantly outperforming traditional methods. These findings support the use of neural networks as effective tools for hydrometeorological data reconstruction and prediction.

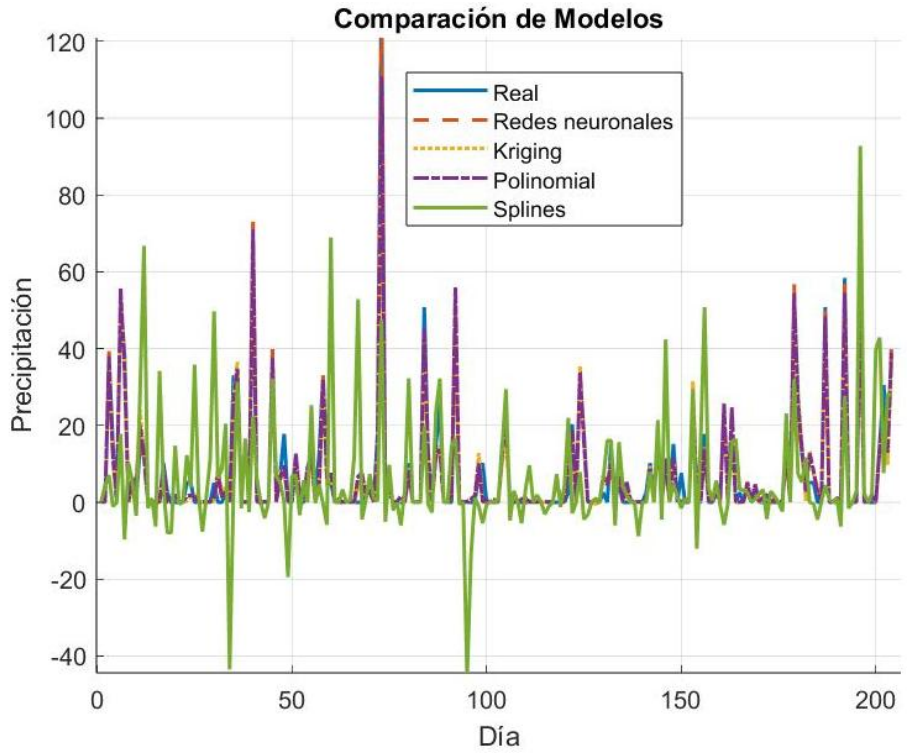

Figure 6. Line plot of data imputation.

The multiple linear regression model achieved a MAE of 1.77, an MSE of 11.11, and a Nash–Sutcliffe Efficiency (NSE) coefficient of 0.955, indicating a good agreement with the observed data. Statistically, all estimated coefficients were significant ($p < 0.001$), highlighting the relevance of the variables included in the prediction. The multiple linear regression equation is shown below, where y represents the precipitation at Rex River, $x_1$is the precipitation at Cougar Mountain, $x_2$at Meadows, x_3 at Mount Gardner, $x_4$at Ollaine Meadows, and x_5 at Tikham Creed.

$$y=0.1322+0.2500x_1+0.4037x_2+0.2579x_3+0.0800x_4+0.0567x_5 \quad (13)$$

The Meadows and Mount Gardner stations showed the highest coefficients (0.4037 and 0.2579, respectively), followed by the other stations. This result suggests a greater contribution from the former to the variation in the estimated value. The Root Mean Square Error (RMSE) was 4.48, reinforcing the model's accuracy within an acceptable margin for daily precipitation series. However, studies such as [8] have shown that while multiple linear regression can be suitable for data imputation, it tends to exhibit greater variability in error compared to machine learning–based methods, especially in contexts of high climatic variability [8]. Despite its good results, the model was outperformed by the neural network in all metrics, suggesting that although the linear model captures relevant relationships, it may not be sufficient for complex or nonlinear patterns present in the time series.

The Kriging method showed solid performance in estimating missing values, with a MAE of 2.01, an MSE of 19.90, and a Nash–Sutcliffe Efficiency (NSE) coefficient of 0.9192. Although its metrics were slightly lower than those of the neural network and polynomial models, Kriging stands out for its interpolation capability based on the spatial and statistical structure of the data, without requiring a supervised training process. This feature gives it a significant advantage in contexts where large volumes of training or cross-validation data are unavailable, or when an immediate solution based on geostatistical principles is required. Teegavarapu (2009) supports this approach, emphasizing that Kriging is particularly suitable for spatial estimation of missing precipitation data, with acceptable error metrics and good correlation even in sparse networks. The strong observed correlation (0.9665) suggests that the model successfully captures inter-station dependencies, leveraging semivariance to produce robust estimates[29].

The results obtained with Kriging are consistent with previous studies that have highlighted the method's effectiveness for imputing environmental and meteorological data, particularly when spatial structure is relevant [30].The findings show that Kriging emerged as the most accurate method among several

interpolation options, validated through cross-validation. Although it may be outperformed by deep learning models in absolute terms of precision, its operational simplicity and statistical foundations make it a reliable and efficient option for imputation tasks in hydrology.

The cubic spline method showed the poorest performance among the evaluated models, with a MAE of 9.59, an MSE of 279.98, and a negative Nash–Sutcliffe Efficiency (NSE) coefficient of –0.137. This negative NSE value indicates that the model's estimates were even less accurate than a simple prediction based on the mean of the observed data, demonstrating an inadequate capacity to represent the actual behavior of the precipitation series. This poor performance may be attributed to the highly oscillatory nature of the cubic spline method when applied to time series with high variability or a lack of continuity. When attempting to fit smooth curves between scattered points or abrupt changes, splines may produce local overfitting or underestimation in key segments of the series, significantly degrading their generalization capability.

The Taylor diagram with the results of all the models is presented, where point A represents the actual data series, B corresponds to the neural network, C to Kriging, D to the Multiple Linear regression, and E to cubic splines, as shown in Figure 6. The results show that the neural network model exhibited the best overall fit, with a standard deviation very close to the observed data, high correlation, and a low RMSE of 2.91. It was followed by the polynomial and Kriging models, which also showed high correlations and moderate errors. In contrast, the cubic spline–based model, despite replicating the series variability with a standard deviation of 15.95, exhibited low correlation and a significantly higher RMSE of 16.73, indicating a poor fit. These results reinforce the suitability of neural network and polynomial regression approaches for estimating missing data in daily precipitation series

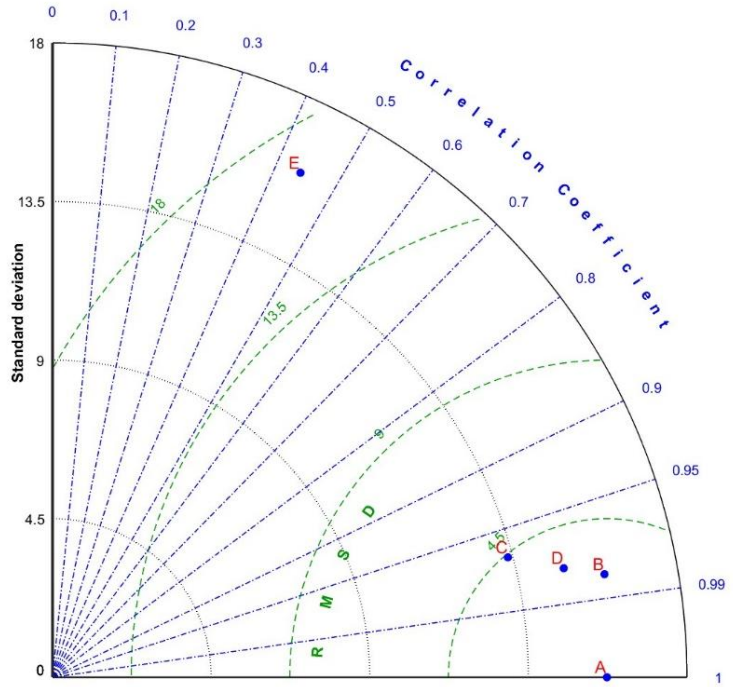

Figure 7. Taylor diagram of the results. Symbology: A= real observations, B= Neural Networks, D= Kriging, C= Multiple Linear regression, E= Cubic Splines

3. CONCLUSIONS

The results obtained allow us to conclude that, although neural networks showed the best quantitative performance in estimating missing data (NSE of 0.966), their implementation in operational environments requires a higher level of technical expertise, as well as a sufficiently large volume of data to enable effective partitioning into training, validation, and testing sets. Therefore, their use is mainly recommended in institutional or academic contexts where computational capabilities and specialized knowledge in machine learning are available.

On the other hand, the multiple linear regression model also produced acceptable results (NSE of 0.955), with the advantage of being easily interpretable and reproducible using standard statistical tools. Its application is suitable when nearby stations with high correlation are available, and technical staff possess

basic statistical knowledge.

The Kriging method, although it yielded a slightly lower NSE (0.9192), represents an excellent alternative for operational contexts due to its ease of implementation, robustness in the presence of well-defined spatial structures, and the advantage of not requiring a complex training process. Its use is particularly useful in hydroclimatic institutions that possess georeferenced data but lack advanced computational infrastructure.

Finally, the cubic spline method demonstrated poor performance (NSE = –0.137), discouraging its direct use in daily precipitation series without proper oscillation control or prior data segmentation. Although easy to implement, it can yield erroneous estimates if the structure of the original data is not properly considered.

REFERENCES

1 H. E. Beck et al., "Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS," Hydrol Earth Syst Sci, vol. 23, no. 1, pp. 207–224, 2019, doi: 10.5194/HESS-23-207-2019,.

2 J. S. Gómez Guerrero, M. I. Aguayo Arias, J. S. Gómez Guerrero, and M. I. Aguayo Arias, "Evaluación de desempeño de métodos de relleno de datos pluviométricos en dos zonas morfoestructurales del Centro Sur de Chile," Investigaciones geográficas, no. 99, 2019, doi: 10.14350/RIG.59837.

3 T. Hırca and G. Eryılmaz Türkkan, "Assessment of Different Methods for Estimation of Missing Rainfall Data," Water Resources Management, vol. 38, no. 15, pp. 5945–5972, 2024, doi: 10.1007/S11269-024-03936-3/FIGURES/2.

4 C. Fagandini, V. Todaro, M. G. Tanda, J. L. Pereira, L. Azevedo, and A. Zanini, "Missing Rainfall Daily Data: A Comparison Among Gap-Filling Approaches," Math Geosci, vol. 56, no. 2, pp. 191–217, Feb. 2024, doi: 10.1007/S11004-023-10078-6/FIGURES/8.

5 J. Ossa-Moreno, G. Keir, N. Mcintyre, M. Cameletti, and D. Rivera, "Comparison of approaches to interpo-lating climate observations in steep terrain with low-density gauging networks," Hydrol Earth Syst Sci, vol. 23, no. 11, pp. 4763–4781, 2019, doi: 10.5194/HESS-23-4763-2019,.

6 G. Papacharalampous, H. Tyralis, N. Doulamis, and A. Doulamis, "Uncertainty estimation of machine learning spatial precipitation predictions from satellite data," Mach Learn Sci Technol, vol. 5, no. 3, 2023, doi: 10.1088/2632-2153/ad63f3.

7 V. D. Agou, E. A. Varouchakis, and D. T. Hristopulos, "Geostatistical analysis of precipitation in the island of Crete (Greece) based on a sparse monitoring network," Environ Monit Assess, vol. 191, no. 6, p. 353, Jun. 2019, doi: 10.1007/s10661-019-7462-

8 M. Portuguez-maurtua, J. L. Arumi, O. Lagos, A. Stehr, and N. M. Arquiñigo, "Filling Gaps in Daily Precip-itation Series Using Regression and Machine Learning in Inter-Andean Watersheds," Water (Switzerland), vol. 14, no. 11, p. 1799, 2022, doi: 10.3390/W14111799/S1.

9 J. Zhang, M. Bonas, D. Bolster, G. A. Fuglstad, and S. Castruccio, "High-resolution global precipitation downscaling with latent Gaussian models and non-stationary stochastic partial differential equation struc-ture," J R Stat Soc Ser C Appl Stat, vol. 73, no. 1, pp. 65–81, 2024, doi: 10.1093/JRSSSC/QLAD084.

10 L. Han, X.-Y. Chen, H.-J. Ye, and D.-C. Zhan, "Learning Robust Precipitation Forecaster by Temporal Frame Interpolation," 2023, [Online]. Available: https://arxiv.org/abs/2311.18341v2

11 S. Ravuri et al., "Skilful precipitation nowcasting using deep generative models of radar," Nature, vol. 597, no. 7878, pp. 672–677, 2021, doi: 10.1038/S41586-021-03854-Z;TECHMETA=129,139,141;SUBJMETA=117,172,639,704,705;KWRD=COMPUTER+SCIENCE,ENVIRONMENTAL+SCIENCES.

12 S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, "Machine Learning for Precipita-tion Nowcasting from Radar Images," 2019, [Online]. Available: https://arxiv.org/pdf/1912.12132

13 K. Konca-Kedzierska, J. ibig, and M. Gruszczyńska, "Comparison and combination of interpolation methods for daily precipitation in Poland: evaluation using the correlation coefficient and correspondence ratio," Meteorology Hydrology and Water Management, vol. 11, no. 2, pp. 1–27, 2023, doi: 10.26491/MHWM/171699.

14 T. Caloiero, G. Pellicone, G. Modica, and I. Guagliardi, "Comparative Analysis of Different Spatial Interpo-lation Methods Applied to Monthly Rainfall as Support for Landscape Management," Applied Sciences 2021, Vol. 11, Page 9566, vol. 11, no. 20, p. 9566, 2021, doi: 10.3390/APP11209566.

15 H. S. Franz, "SAFEGUARDING OUR LANDS, WATERS, AND COMMUNITIES: DNR'S PLAN FOR CLI-MATE RESILIENCE F E B R U A R Y 2 0 2 0 2 LETTER FROM THE COMMISSIONER OF PUBLIC LANDS".

16 M. Chang et al., "Chapter 27 : Northwest. Fifth National Climate Assessment," 2023, doi: 10.7930/NCA5.2023.CH27.

17 M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, "An Overview of the Global Historical Climatology Network-Daily Database," J Atmos Ocean Technol, vol. 29, no. 7, pp. 897–910, Jul. 2012, doi: 10.1175/JTECH-D-11-00103.1.

18 I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," SN Comput Sci, vol. 2, no. 6, pp. 1–20, 2021, doi: 10.1007/S42979-021-00815-1/METRICS.

19 L. Lusa and Č. Ahlin, "Restricted cubic splines for modelling periodic data," PLoS One, vol. 15, no. 10, pp. e0241364–e0241364, 2020, doi: 10.1371/JOURNAL.PONE.0241364.

20 A. Batool, S. Ahmad, A. Waseem, V. Kartal, Z. Ali, and M. Mohsin, "Evaluating variogram models and kriging approaches for analyzing spatial trends in precipitation simulations from global climate models," Acta Geophysica, vol. 73, no. 4, pp. 3677–3697, 2025, doi: 10.1007/S11600-025-01545-1/METRICS.

21 A. Bertoncini and J. W. Pomeroy, "Quantifying Spatiotemporal and Elevational Precipitation Gauge Net-work Uncertainty in the Canadian Rockies," 2024, doi: 10.5194/EGUSPHERE-2024-288.

22 J. K. Searcy, C. H. Hardison, and W. B. Langbein, Double-Mass Curves With a section Fitting Curves to Cyclic Data Manual of Hydrology: Part 1. General Surface-Water Techniques. Washington, 1960.

23 R. Sneyers, On the statistical analysis of series of observations; Technical Note No. 143, WMO No. 415. Secretariat of the World Meteorological Organization, 1990. Accessed: Jun. 20, 2025. [Online]. Available: https://library.wmo.int
24 D. R. Helsel, R. M. Hirsch, K. R. Ryberg, S. A. Archfield, and E. J. Gilroy, "Statistical methods in water re-sources," Techniques and Methods, vol. 2020, no. 4-A3, pp. 1–484, 2020, doi: 10.3133/TM4A3.
25 A. Gilar, "MATLAB: An Introduction with Applications, 6th Edition | Wiley." Accessed: Jun. 15, 2024. [Online]. Available: https://www.wiley.com/en-us/MATLAB%3A+An+Introduction+with+Applications%2C+6th+Edition-p-9781119299257
26 X. Huang, H. Cao, and B. Jia, "Optimization of Levenberg Marquardt Algorithm Applied to Nonlinear Sys-tems," Processes 2023, Vol. 11, Page 1794, vol. 11, no. 6, p. 1794, Jun. 2023, doi: 10.3390/PR11061794.
27 National Center for Atmospheric Research Staff, "The Climate Data Guide: Taylor Diagrams."
28 Y. Wen, W. Yu, F. Zheng, D. Huang, and N. Xiao, "AdaNAS: Adaptively Post-processing with Self-supervised Neural Architecture Search for Ensemble Rainfall Forecasts," Dec. 2023, Accessed: Jun. 21, 2025. [Online]. Available: https://arxiv.org/pdf/2312.16046
29 R. S. V. Teegavarapu, "Estimation of missing precipitation records integrating surface interpolation tech-niques and spatio-temporal association rules," Journal of Hydroinformatics, vol. 11, no. 2, pp. 133–146, 2009, doi: 10.2166/HYDRO.2009.009.
30 M. G. Ebrea, N. Habtamu, and Hundessa, "Simple Kriging for Rainfall Mapping: A Geostatistical Analysis of North East Amhara (Wollo), Ethiopia Using ArcMap Pro.," Apr. 2025, doi: 10.21203/RS.3.RS-6457750/V1.