



## Article

# A Novel Hybrid Spatiotemporal Missing Value Imputation Approach for Rainfall Data: An Application to the Ratnapura Area, Sri Lanka

Shanthi Saubhagya <sup>1,\*</sup> , Chandima Tilakaratne <sup>1,\*</sup> , Pemantha Lakraj <sup>1</sup> and Musa Mammadov <sup>2</sup><sup>1</sup> Department of Statistics, University of Colombo, Colombo P.O. Box 1490, Sri Lanka; pemantha@stat.cmb.ac.lk<sup>2</sup> School of Info Technology, Faculty of Science, Engineering and Built Environment, Geelong Waurm Ponds Campus, Deakin University, Geelong, VIC 3216, Australia; musa.mammadov@deakin.edu.au

\* Correspondence: shanthi@stu.cmb.ac.lk (S.S.); cdt@stat.cmb.ac.lk (C.T.)

**Abstract:** Meteorological time series, such as rainfall data, show spatiotemporal characteristics and are often faced with the problem of containing missing values. Discarding missing values or modeling data with missing values causes negative impacts on the accuracy of the final predictions. Hence, accurately estimating missing values by considering the spatiotemporal variations in data has become a crucial step in eco-hydrological modeling. The multi-layer perceptron (MLP) is a promising tool for modeling temporal variation, while spatial kriging (SK) is a promising tool for capturing spatial variations. Therefore, in this study, we propose a novel hybrid approach combining the multi-layer perceptron method and spatial kriging to impute missing values in rainfall data. The proposed approach was tested using spatiotemporal data collected from a set of nearby rainfall gauging stations in the Ratnapura area, Sri Lanka. Missing values are present in collected rainfall data consecutively for a considerably longer period. This pattern has scattered among stations discontinuously over five years. The proposed hybrid model captures the temporal variability and spatial variability of the rainfall data through MLP and SK, respectively. It integrates predictions obtained through both MLP and SK with a novel optimal weight allocation method. The performance of the model was compared with individual approaches, MLP, SK, and spatiotemporal kriging. The results indicate that the novel hybrid approach outperforms spatiotemporal kriging and the other two pure approaches.

**Keywords:** machine learning; spatial kriging; optimal weight allocation; spatiotemporal data



**Citation:** Saubhagya, S.; Tilakaratne, C.; Lakraj, P.; Mammadov, M. A Novel Hybrid Spatiotemporal Missing Value Imputation Approach for Rainfall Data: An Application to the Ratnapura Area, Sri Lanka. *Appl. Sci.* **2024**, *14*, 999. <https://doi.org/10.3390/app14030999>

Academic Editor: Grzegorz Dudek

Received: 19 November 2023

Revised: 22 December 2023

Accepted: 22 December 2023

Published: 24 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background

Rainfall is a natural phenomenon that shows spatiotemporally varying behavior. These characteristics with chaotic dynamic patterns make it difficult to predict its future behavior. Moreover, a common shortcoming of rainfall data is the inherent inclusion of missing values. This is because the data are collected from different gauging stations located in a widespread geographical area. The malfunctioning of equipment, relocation of gauging stations, network interruptions, natural hazards, and emergencies like pandemics may disturb the continuous measurement of this natural phenomenon [1].

Ignoring the stations with missing data leads to an information loss due to the strong spatial correlation between meteorological stations. On the other hand, modeling rainfall data with missing values negatively impacts the accuracy of rainfall models due to the discontinuity of the time sequence [2]. Hence, identification of a suitable mechanism to address this issue is a crucial step in an effective rainfall forecasting.

When the missing values occur at random, modeling them to estimate their values is not required. However, if the missing values are not random, then the missingness cannot be ignored and must be modeled and predicted [3]. This is known as missing value

imputation and it is carried out as a data preprocessing step. Much research has been conducted using different types of missing value imputation methods. However, studies on imputing the missing values for spatiotemporal data (including meteorological records) are rare [4–7]. This study presents a novel and more efficient spatiotemporal methodology to impute missing values in rainfall data for a few rain gauging stations. The proposed methodology was tested using rainfall data collected from six neighbouring rain gauging stations (which are considered reference stations) of Ratnapura, which was identified as the target rain gauging station in our main study of rainfall forecasting. The reason for our choice is that Ratnapura is a flood-prone area in Sri Lanka due to its frequent exposure to heavy rainfall throughout the year [8,9]. In 2003, flash flood events in Ratnapura accounted for financial damage of LKR 50.6 million [9].

### 1.2. Related Work

Much past research has accompanied traditional statistical and/or machine learning techniques in missing value imputation of meteorological data. Some studies only captured the spatial variations (e.g., [10–13]) using spatial interpolation methods such as the inverse distance method (ID), normal ratio method (NR), geographical co-ordinates method (GC), and spatial kriging, also known as Gaussian process regression, while others only focused on their variation over time (e.g., [14–16]). The most common temporal method used is the autoregressive integrated moving average (ARIMA) models.

Some studies (e.g., [17]) pointed out that the predictions obtained from spatiotemporal methods are more accurate than those of purely spatial predictions. This is mainly because spatiotemporal interpolation can be applied to geo-referencing positions over a space–time grid. Some researchers have used the traditional spatial kriging model after some modifications to model the spatiotemporal behavior of the data [4,5].

There have been several spatiotemporal kriging-based studies conducted with weather variables. The study in [18] developed a spatiotemporal kriging model to predict wind speed and compared the results with the autoregressive moving average (ARMA) and Monte Carlo methods. Spatiotemporal kriging modeling gave a better fit for the data.

Moreover, the authors stated a few advantages of kriging models in comparison with other regression methods. Providing estimates with the mean squared error of the estimation (kriging variance), non-requirement of any distributional assumption related to the data, the ability to use a complete set of spatial and temporal information, and the ability to use a limited number of sampled data points to estimate the value of a variable over a continuous spatial field are some of them.

Due to these prominent advantages of kriging, recent studies have integrated kriging with conventional methods (e.g., regression and time series models) to gain fair and more accurate predictions for the missing observations. Nevertheless, a few research studies that use such hybrid approaches were found. The study [17] applied a spatiotemporal kriging method to model total monthly rainfall data among 269 rain gauge stations, of which nearly 80% of stations had missing data. The deterministic trend component was estimated with multiple linear regression, taking several covariates, including latitude, longitude, and quadratic effect of the corresponding longitude and latitude pairs. Since it only captured 29% of the temporal variability in rainfall, the residuals produced by the model were analyzed using a generalized product-sum spatiotemporal variogram and the interpolation was carried out with spatiotemporal kriging. The final predictor was obtained by combining trend estimate and interpolated value with kriging. In another study [5], spatial association of atmospheric temperature data was modeled with universal kriging and temporal variability was captured with autoregressive (AR) techniques. Then, they were spatiotemporally combined to predict  $k$ -steps (days) ahead temperature in a given spatial domain. A comparison of forecasted values with those obtained from ARIMA model indicated that the novel hybrid model performed better than the ARIMA model.

The recent advancements in machine learning (ML) and deep learning (DL) can be utilized effectively to predict missing cases in meteorological data. Both approaches

have a great ability to handle large-scale problems and provide a flexible architecture in capturing spatial and temporal features [1,2]. In addition, they do not rely on hand-crafted feature engineering and prior assumptions on input data. Moreover, the aptness of deep neural networks on large volumes of spatiotemporal data compared to statistical and other classical machine learning techniques has been recognized.

A convolution bidirectional long short-term memory (LSTM) model was applied to capture spatial and temporal patterns in traffic flow data in the study [2]. This model outperformed state-of-the-art missing data imputation models. Some studies, including the study carried out in [19], demonstrated the effectiveness and efficiency of deep learning (DL) methods compared to the ARIMA model and back propagation neural network model. A convolutional neural network (CNN)-based DL model was proposed in [14] for imputing missing values in weather data of an individual weather station on a temporal basis. The performance of the model was evaluated using various stations nearest to the stations without missing values. This study used five optimizers (Rmsprop, Adam, Nadam, Stochastic Gradient Descend, (SGD), and Adagrad) and found that the SGD optimizer provides the most accurate results in predicting missing values.

Past studies reveal that the application of spatial kriging and DL methodology in missing value imputation of weather data has promising results. However, studies that impute missing values in spatiotemporal weather data using hybrid models are extremely rare, especially when there is a lower number of weather gauging stations. The existing research [5,17,20], which applies hybrid models, used more than 50 weather stations when interpolating spatiotemporal missing data. Most importantly, so far, no study has applied a spatiotemporal hybrid model to impute missing values among rainfall data at the Ratnapura gauging station. Therefore, to reduce this gap in the literature and to utilize the potential of spatial kriging, machine learning, and deep learning techniques in dealing with missing values, this study proposes a hybrid model by integrating them. Our target is to verify its appropriateness when the periods with missing values of rainfall gauging stations are different and discontinuous. Our focus is also given to the scenario where data are available only for a few rainfall stations.

The rest of this paper is organized as follows. The next section describes the adopted missing values imputation methods and the development of the new approach. The results and discussion section, followed by the conclusion are lined up last.

## 2. Materials and Methods

### 2.1. Proposed Methodology

This study introduces a hybrid method by integrating multi-layer perceptron (MLP) with spatial kriging as a tool for imputing sequential missing values of spatiotemporal data. The performance of the new model was evaluated against each of the two individual models as well as the spatiotemporal kriging model, since many past studies [17,18,20] have identified the spatiotemporal kriging model as a promising tool for estimating missing values in spatiotemporal data.

The methodology behind this study will be illustrated using the rainfall data set gathered across several neighbouring weather stations for a sufficiently longer period.

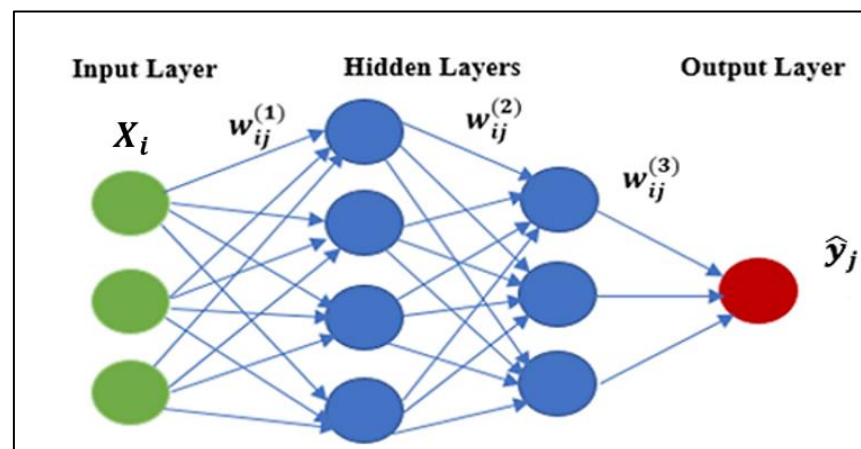
The target variable in this study represents rainfall measurements and each data point has a timestamp (day) and 2D spatial location (longitude and latitude). Our initial experiments performed for missing value imputation using machine learning methods (MLP, long short-term memory (LSTM), convolutional neural network (CNN), CNN-LSTM, ConvLSTM, random forest (RF), and nonlinear autoregressive network with exogenous inputs (NARX)) revealed that MLP can capture the temporal variation of the same set of rainfall data better than the other methods (see Table S1). Therefore, in this study, we chose MLP to capture the temporal variation of rainfall data in our hybrid model. The spatial kriging was used to model spatial variation due its promising behavior in capturing the spatial variations of the weather data and incorporation into a hybrid model [4,17,18,20–22].

### 2.1.1. Multi-Layer Perceptron Neural Network

Multi-layer perceptron (MLP) is a feed-forward neural network consisting of three types of layers: input layer, hidden layer (s), and output layer. In an MLP, the data flow in the forward direction from the input to the output layer, and the MLPs are trained with the backpropagation learning algorithm, which allows nonlinear mapping between the input and the output layer [23]. Its fast operation, ease of implementation, and requirement of a smaller training set (compared to radial basis neural networks) lead it to be popular among researchers [24].

In MLP, each neuron in a layer is connected to all neurons in the succeeding layer and connection strength is called weight. The output of each neuron of hidden layers as well as the output layer is a function of the sum of the inputs to the node modified by applying a linear or nonlinear transfer or an activation function [25,26]. The input layer passes the input vector to the network. An MLP can have one or more hidden layers.

Figure 1 illustrates an example of an MLP model with two hidden layers. Here,  $X_i$  is the input to the  $i^{\text{th}}$  neuron in the input layer.  $w_{ij}^{(1)}$  is the weight associated with the  $i^{\text{th}}$  neuron of the input layer to the  $j^{\text{th}}$  neuron of the first hidden layer,  $w_{ij}^{(2)}$  is the weight of the link connecting the  $i^{\text{th}}$  neuron of the first hidden layer to the  $j^{\text{th}}$  neuron of the second hidden layer, and  $w_{ij}^{(3)}$  is the weight associated with the  $i^{\text{th}}$  neuron of the second hidden layer to the  $j^{\text{th}}$  neuron of the output layer.  $b_i^{(1)}$  and  $b_i^{(2)}$  are the bias values associated with the  $i^{\text{th}}$  neuron of the first hidden layer and the  $i^{\text{th}}$  neuron of the second hidden layer, respectively.  $b_j$  is the bias associated with the  $j^{\text{th}}$  neuron of the output layer.  $\varnothing^{(1)}$ ,  $\varnothing^{(2)}$ , and  $\varnothing^{(3)}$  are the activation functions associated with the first hidden layer, second hidden layer, and the output layer, respectively.



**Figure 1.** A multilayer perceptron model with two hidden layers. The nodes in Input Layer are represented in green colour. The nodes in Hidden Layer are represented in blue colour, and the node in Output Layer is coloured in red.

Then, the net output from the  $i^{\text{th}}$  neuron of the first hidden layer is given by  $h_i^{(1)}$  and the net output from the  $j^{\text{th}}$  neuron of the second hidden layer is  $h_j^{(2)}$  (see Equations (1) and (2)). The output of the  $j^{\text{th}}$  neuron of the output layer will be  $\hat{y}_j$  (in our study,  $j = 1$ ). This is also noted as the imputed rainfall value at Ratnapura by MLP.

Finally, the MLP network computations can be written as follows:

$$h_i^{(1)} = \varnothing^{(1)}(\sum_j w_{ij}^{(1)} X_i + b_i^{(1)}) \text{ where } i = 1, 2, 3 \text{ and } j = 1, 2, 3, 4 \quad (1)$$

$$h_j^{(2)} = \varnothing^{(2)}(\sum_i w_{ij}^{(2)} h_i^{(1)} + b_j^{(2)}) \text{ where } i = 1, 2, 3, 4 \text{ and } j = 1, 2, 3 \quad (2)$$

$$\hat{y}_j = \varnothing^{(3)} \left( \sum_i w_{ij}^{(3)} h_j^{(2)} + b_j \right) \text{ where } i = 1, 2, 3 \text{ and } j = 1 \tag{3}$$

### 2.1.2. Spatial Kriging

Spatial kriging is an optimal interpolation technique that uses the theory of regression to predict the value at an unobserved location using values of observed neighborhood locations. Here, observed values of neighborhood locations are weighted according to the covariance values. As depicted in Figure 2, consider the location of interest,  $X_0$ , that is surrounded by the neighboring observed locations,  $X_i$ .

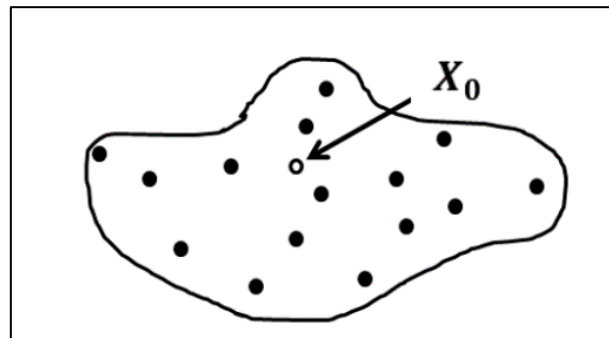


Figure 2. The neighboring observed locations and the location of interest,  $X_0$ .

The value at location  $X_0$ ,  $Z(X_0)$ , is estimated through the values of  $n$  neighboring sample locations taken as a linear combination of observed location values. Let the weight be associated with the observed value at location  $X_i$ , i.e.,  $Z(X_i)$  is  $w_i$ . Then, the spatial kriging estimate of  $Z(X_0)$  can be computed as follows:

$$\hat{Z}(X_0) = k + \sum_{i=1}^n w_i Z(X_i) = k + w^T Z \tag{4}$$

where  $w \equiv (w_1, w_2, \dots, w_n)^T$ ,  $Z \equiv (Z(X_1), Z(X_2), \dots, Z(X_n))$ . The values of  $k$  and  $w$  are calculated by minimizing the mean squared prediction error [18].

The optimal linear predictor of  $Z(X_0)$ ,  $\hat{Z}(X_0)$ , is given by:

$$\hat{Z}(X_0) = \mu(X_0) + c^T \Sigma^{-1} (Z - \mu) \tag{5}$$

where  $\mu$  is the Lagrange multiplier,  $\mu \equiv (\mu(X_1), \mu(X_2), \dots, \mu(X_n))^T$ ,  $(X_i) = E(Z(X_i))$ ,  $i = 0, 1, 2, \dots, n$ ,  $c \equiv (Cov(X_0, X_1), \dots, Cov(X_0, X_n))^T$ , and  $\Sigma$  is the  $n \times n$  matrix whose  $(i, j)$ <sup>th</sup> element is  $Cov(X_i, X_j)$ .

Then, the minimum value of  $MSE(X_0, w, k)$  for a fixed location  $X_0$ ,  $\sigma^2(X_0)$  (spatial kriging variance) is given by [18]:

$$\sigma^2(X_0) = Cov(X_0, X_0) - c^T \Sigma^{-1} c \tag{6}$$

The values of  $c$  and  $\Sigma$  are calculated through the variogram or a covariance function. The different kriging models (simple, ordinary, or universal) can be obtained by changing the way of modeling  $E(Z(X))$ .

### 2.1.3. Hybrid Model for Missing Value Estimation

In the proposed hybrid model, the spatial kriging estimate and the MLP estimate are integrated with optimal weights. The optimal weights must be estimated using a sample of artificially created missing cases. Suppose the spatial kriging estimate of that sample is  $\hat{y}_k$  (which was produced as  $\hat{Z}(X_0)$  in Section 2.1.2) and the MLP estimate is  $\hat{y}_m$  (which was

produced as  $\hat{y}_j$  in Section 2.1.1), weights from each estimate are  $w_k$  and  $w_m$ , respectively, the actual value is  $y_a$ , and the final estimate is  $\hat{y}_f$ . In order to find the final prediction closer to the actual value, the optimal weights from each prediction should be found such that the prediction error is minimized. Therefore, the optimization problem can be defined as mean squared error (MSE) loss function ( $f_{loss}$ ) between predicted and actual values:

$$f_{loss}(w_k, w_m) = \sum_{i=1}^n (w_k \hat{y}_{ki} + w_m \hat{y}_{mi} - y_{ai})^2 \tag{7}$$

with respect to the following constraints that provide a convex combination of two edge points  $\hat{y}_{ki}$  and  $\hat{y}_{mi}$ :

$$w_k + w_m = 1, 0 \leq w_k \leq 1, 0 \leq w_m \leq 1 \tag{8}$$

Following the study [27], it would be logical to consider the case when the weights  $w_k$  and  $w_m$  are greater than 1 or less than 0. Considering this assumption and denoting  $w = w_k$ , Equations (7) and (8) can be formulated as an unconstrained optimization problem with respect to  $w$  in the form:

$$\text{Minimize : } f_{loss}(w) = \sum_{i=1}^n (w \hat{y}_{ki} + (1 - w) \hat{y}_{mi} - y_{ai})^2 \tag{9}$$

The objective function  $f_{loss}(w)$  in this problem is convex and smooth; therefore, optimization Equation (9) can be easily solved. We have:

$$\frac{1}{dw} f_{loss}(w) = \sum_{i=1}^n 2(w \hat{y}_{ki} + (1 - w) \hat{y}_{mi} - y_{ai})(\hat{y}_{ki} - \hat{y}_{mi}) = 0 \tag{10}$$

Solving Equation (10) with respect to  $w$ , we find optimal weight  $w$  and, therefore, weights  $w_k = w$  and  $w_m$  in the form:

$$w_k = (\sum_{i=1}^n (\hat{y}_{mi}^2 + \hat{y}_{ki} y_{ai} - \hat{y}_{mi} y_{ai} - \hat{y}_{ki} \hat{y}_{mi})) / (\sum_{i=1}^n (\hat{y}_{ki}^2 + \hat{y}_{mi}^2 - 2\hat{y}_{ki} \hat{y}_{mi})) \tag{11}$$

$$\text{and } w_m = 1 - w_k \tag{12}$$

The found optimal weights were used to obtain the final prediction  $\hat{y}_f$  at the point  $i$  where the particular value is missing:

$$\hat{y}_f = w_m * \hat{y}_m + w_k * \hat{y}_k \tag{13}$$

The imputed missing values using the proposed method were compared with those obtained from spatial kriging, MLP as well as spatiotemporal kriging. The reason for considering spatiotemporal kriging is that it is a widely used spatiotemporal missing value imputation method in past studies.

#### 2.1.4. Spatiotemporal Kriging

The spatial kriging can be extended to obtain spatiotemporal predictions by incorporating an additional temporal dimension,  $t$ , to the spatial kriging [4,17,18,20].

The spatiotemporal set of observed data are given by  $\{Z(X_1, t_1), Z(X_2, t_2), \dots, Z(X_n, t_n)\}$ , and let  $Z_i = Z(X_i, t_i)$  be the value of the variable  $Z$  at location  $X_i$  and time  $t_i$ . The spatiotemporal kriging prediction at an unobserved location  $X_0$  at time  $t$  is defined by:

$$\hat{y}_f = \hat{Z}(X_0, t) = \sum_{i=1}^n w_i(X, t)(Z(X_i, t_i)) \text{ subject to } \sum_{i=1}^n w_i(X, t) = 1 \tag{14}$$

where  $w_i$  are weights for  $Z(X_i, t_i)$ ,  $i = 1, 2, \dots, n$ . Following the study in [18], the optimal prediction is taken by minimizing the mean squared prediction error subject to the constraint:

$$E[\hat{Z}(X_0, t) - Z(X_0, t)] = 0 \tag{15}$$

The kriging weights can be computed through the covariance function or by means of a semivariogram. Once the covariance values are calculated, the optimal weights of the kriging estimator can be computed through Lagrange multiplier application. Finally, the optimal value of  $\hat{Z}(X_0, t)$  is calculated for spatiotemporal location  $(X_0, t)$  [18].

The spatiotemporal covariance function has different models such as linear, exponential, spherical, Gaussian, or logistic. Capturing the spatial and temporal variability is difficult for the variogram due to their dissimilarities. Therefore, the two-dimensional kriging is modified to incorporate the time component. The type of variogram (metric, separable, SumMetric, and product sum models) varies according to the way it combines and models the temporal and spatial dependencies. The best-fitting model can be identified by comparing the mean squared errors (MSE) between the sample variogram and fitted variogram.

### 2.1.5. Model Evaluation

In order to evaluate the performance of the fitted MLP, spatial kriging and spatiotemporal kriging models, mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination ( $R^2$ ) metrics were used.

The MAE, RMSE, and  $R^2$  are calculated as below.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{ai} - \hat{y}_{fi}| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ai} - \hat{y}_{fi})^2} \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{ai} - \hat{y}_{fi})^2}{\sum_{i=1}^n (y_{ai} - \bar{y}_{ai})^2} \quad (18)$$

where  $\hat{y}_{fi}$  is the predicted value,  $y_{ai}$  is the actual value, and  $\bar{y}_{ai}$  is the mean value of  $y_{ai}$ . The smaller the values of MAE and RMSE and larger the  $R^2$  value, the better the forecasting is.

The hybrid model, which integrates MLP and spatial kriging, was implemented in python and the spatiotemporal kriging model was fitted using R software (version 4.3) and the related packages. In the spatiotemporal model, the data were represented as Space Time Full Data Frame (STFDF) object before fitting the variogram since data considered in this study were with complete space time grid with  $s$  substations and  $t$  time intervals of their observations.

### 2.2. Description of Data

For this study, daily rainfall readings gathered by nine rainfall neighbouring gauging stations, including the target station (Ratnapura), were considered. The other eight stations are Balangoda (St1\_Bal), Detanagalla (St2\_Det), Elston (St3\_Els), llumbuluwa (St4\_Ill), Keragala (St5\_Ker), Moraliyoa (St6\_Mor), Nivithigala (St7\_Niv), and Ulinduwawa (St8\_Uli). These eight stations are considered as reference stations.

The total number of days accounted for in this study is 1826. As mentioned earlier, these reference stations consist of various discontinuous missing patterns within the study period (1 January 2015–31 December 2019). The following heatmap (Figure 3) depicts the missing patterns in the data of eight reference stations and Table 1 summarizes percentages.

Figure 3 depicts the data availability with blue color and inconsistent missingness with white blocks along the time and space. With respect to the figure, it can be noted that, in most cases, the missingness has occurred as a block (i.e., a sequence of missing values). The percentage of missing values of the last two sub-stations, as indicated in Table 1, are 68.4% and 73.55%, respectively. Since more than half of the data values are missing, these two sub-stations were excluded from the analysis. The stations St1\_Bal, St2\_Det and St5\_Ker had random missing cases (noncontinuous missing values). Those were interpolated using the average values of the neighbouring rainfall values of the

corresponding station. Finally, there were six reference stations considered for assessing the suitability of the proposed methodology.

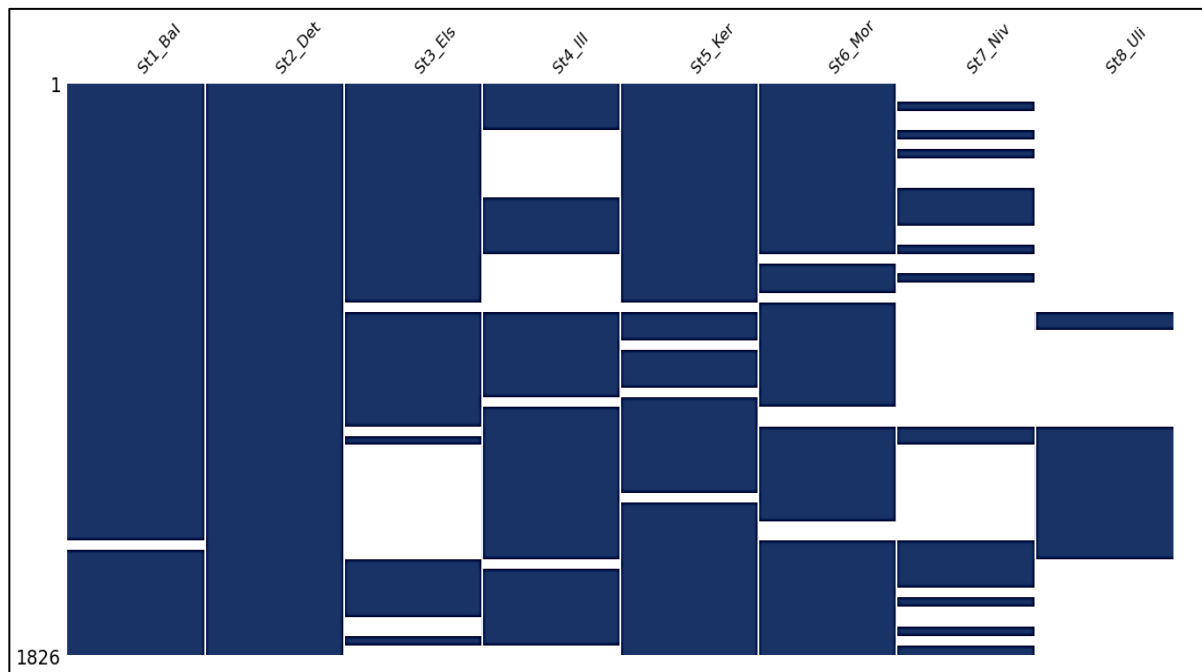


Figure 3. Missing patterns of sub-stations' rainfall data.

Table 1. Missing percentages of sub-stations' rainfall data.

Sub Station	Missing Count	Missing Percentage
St1_Bal	32	1.75
St2_Det	1	0.05
St3_Els	519	28.42
St4_Ill	491	26.89
St5_Ker	123	6.74
St6_Mor	183	10.02
St7_Niv	1249	68.4
St8_Uli	1343	73.55

### 3. Results and Discussion

#### 3.1. Preliminary Analysis

The correlation between the rainfall values of different stations may be either linear or nonlinear. Therefore, to capture any type of correlations, the Spearman correlation coefficients were calculated (refer to Figure 4).

The correlation between all the considered sub-stations and the target station Ratnapura and among sub-stations are significant at 5% significance level (i.e.,  $p$  value  $< 0.05$ ). A strong correlation ( $>0.6$ ) can be seen between Ratnapura and sub-stations Keragala, Moraliyoa, and Elston. The results indicate that all six sub-stations are important in modeling rainfall at Ratnapura station. Thus, it confirms the need for imputing missing values of the reference stations.

#### 3.2. The Hybrid Model Formation

The structure of the performed hybrid model with MLP and spatial kriging can be depicted as in Figure 5. In the notation,  $\hat{y}_k$  and  $\hat{y}_m$  refers to the spatial kriging and the MLP estimates corresponding to the missing rainfall value of the  $i^{th}$  station, for instance, St2 (there are situations where the rainfall values of more than one station are missing on the same day).



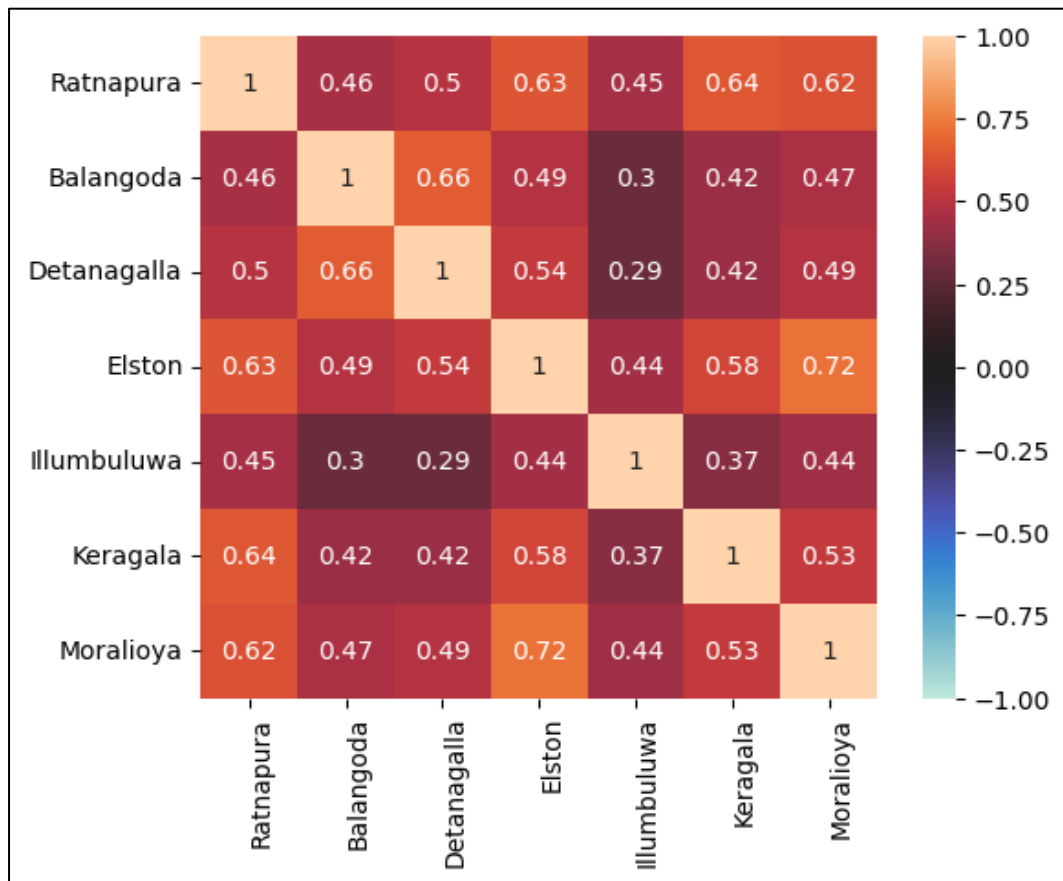


Figure 4. The heatmap of correlation between rainfall gauging stations.

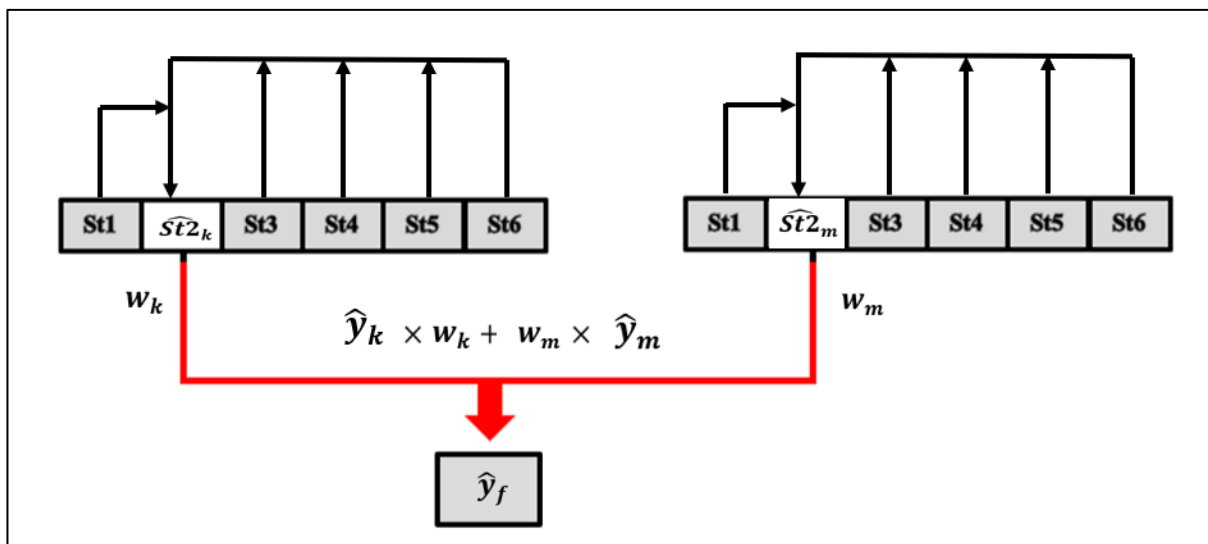


Figure 5. The hybrid model structure.

Note that MLP (for multivariate time series data) was trained using the parameter values tuned with grid search approach (selected from a range of values) tabularized in Table 2.

**Table 2.** Common parameter settings in MLP models.

Parameters	Value
Activation function	ReLU
Epochs	80
Number of hidden layers	1
Number of neurons in hidden layer	20
Batch size	72
Learning rate	0.01
Optimizer	Adam
Loss Function	MAE

The regularization parameters and number of training/testing samples were modified to obtain a convergent model in each MLP model. As depicted in Figure 5, in order to integrate the two estimates obtained from spatial kriging and MLP, the optimal weights from each estimate should be found.

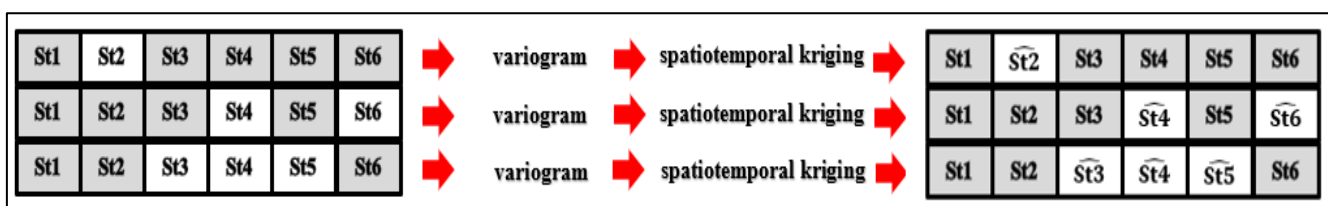
For this, a data set without missing values needs to be used. Therefore, the following procedure was applied:

- (a) From the original data set, three complete subsets (with no missing cases for all the stations) were selected. Consequently, subsets from the years 2015, 2016, and 2019 were selected.
- (b) Consecutive missing periods were artificially created randomly within the three subsets to capture the missing pattern of the original data set.
- (c) The spatial kriging and MLP were applied to estimate the missing cases.
- (d) Those estimated rainfall values were combined to form a single data set of 120 estimated rainfall values.
- (e) Then, from the final data set mentioned in (d), 100 simple random samples (with replacement) with a moderate sample size of 60 were drawn.
- (f) For each sample, the optimal values of two weights,  $w_k$  and  $w_m$ , were calculated by applying Equations (11) and (12) as described in Section 2.1.3. The procedure resulted in 100 optimal pairs of weights.
- (g) Then, the first optimal weight pair was used to calculate the weighted estimates ( $\hat{y}_f$ ) of the rest of the samples (99). For each sample, the RMSE values (including the first sample considered) were calculated. Using these RMSE values, the average RMSE was computed.
- (h) Likewise, steps (a) to (g) were repeated using the second optimal weight pair, then third optimal weight pair, and so on.

Finally, there were 100 average RMSE values. Out of them, the pair of weights that produced the lowest average RMSE value was selected as the best optimal pair of weights for spatial kriging and MLP (Table S2). The results indicated that the 11th random sample produced the final optimal weights,  $w_k = 0.50013$  and  $w_m = 0.49987$ . Those weights were taken to obtain optimal weighted estimates for the missing values.

### 3.3. The Spatiotemporal Kriging Model (SPTK Model)

Figure 6 summarizes the steps followed in applying the spatiotemporal kriging model.



**Figure 6.** The spatiotemporal kriging method.

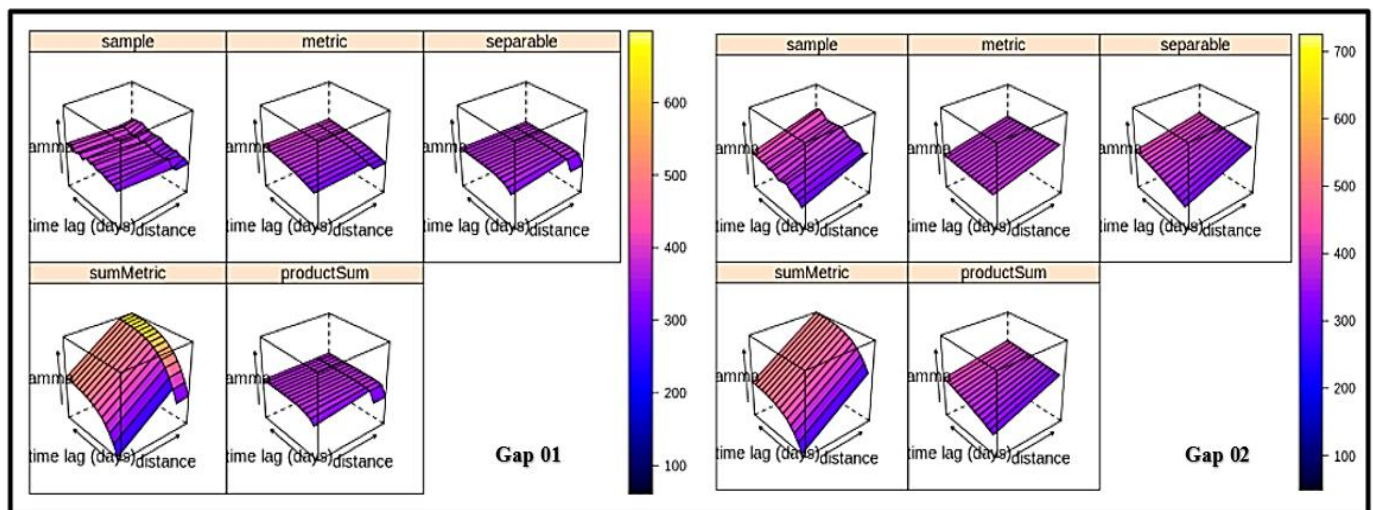
In this approach, a variogram for each missing sequence was fitted using the available rainfall values of stations. For example, according to Figure 6, the first gap of the above data set can be seen in station St2. Then, the data from the remaining stations were employed to fit a variogram. Similarly, the missing blocks of St4 and St6 were estimated using the available rainfall values of the other stations and so on.

Four variogram types (metric, separable, sum metric, and product sum) were compared in terms of RMSE values. Using the variogram with the lowest RMSE, the spatiotemporal kriging was applied to estimate the missing values in a substation(s) that was/were in the created grid. Table 3 provides the outputs resulting from 11 blocks of missing values.

**Table 3.** Comparison of prediction errors of variogram types.

Gap no	Metric	Separable	SumMetric	Product Sum
1	<b>13.79</b>	23.16	64.75	23.33
2	38.56	22.871	52.43	<b>22.868</b>
3	<b>15.46</b>	16.19	31.14	16.25
4	<b>30.31</b>	30.86	71.54	30.82
5	136.15	132.79	<b>128.77</b>	132.79
6	17.53	10.03	31.8	<b>9.99</b>
7	11.02	9.67	<b>9.13</b>	9.66
8	<b>88.75</b>	89.13	110.29	89.28
9	<b>21.58</b>	21.95	41.33	22.13
10	<b>81.02</b>	81.03	118.85	81.04
11	34.95	<b>13.54</b>	58.8	13.63

According to Table 3, it can be noticed that, in most of the cases, the metric variogram has become the best fit and Figure 7 depicts a sample of figures (out of 11 figures), which illustrates the comparison between variograms. In each case, the sample variogram and fitted four types of variograms are given.



**Figure 7.** Differences between the sample and the best fitting spatiotemporal variogram of each family of the first two gaps. The half of the variogram function ( $\gamma$ —gamma) which gives measures of spatiotemporal correlation of the random field is plotted against each space–time co-ordinate.

In fitting the spatiotemporal model for the rainfall data, the empirical variogram surface (or sample variogram) based on space and time is computed and used as inputs for fitting different models: metric, separable, sum metric, and product sum models. Out of them, the best fitting model can be selected, comparing with the surface of the empirical variogram. When considering Missing Gap 01 in Figure 7, the metric-type variogram shows the most similar surface shape with respect to distance and time lags as that of the original

sample. Nevertheless, in Missing Gap 02, the product sum variogram is more similar to the pattern of the actual sample. Likewise, in 11 gaps, the optimal variogram type deferred with the missing pattern of the data.

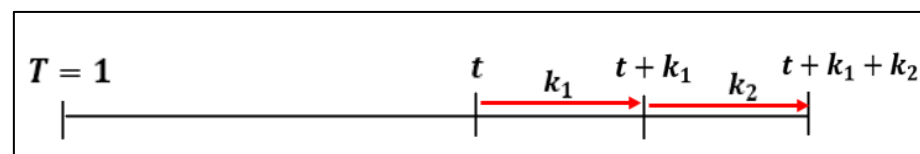
Through optimal variograms the covariances of the data were calculated and, thereby, the optimal weights to predict the missing value at the targeted location (see Section 2.1.4).

### 3.4. Model Evaluation

To validate the models, a two-way evaluation mechanism was carried out.

#### 3.4.1. Model Evaluation along with Actual Missing Data

Assume that the rainfall values of a station are missing from  $t$  to  $t + k_1$  (see Figure 8). The hybrid model was fitted, and these missing values were estimated using the available values from 1 to  $t$ . To check the validity of the fitted model, rainfall values from  $(t + k_1)$  to  $(t + k_1 + k_2)$  were estimated based on the same fitted model (refer to Section 3.2) and compared with the actual values. This validation data set was with size 55 (here, the total number of observations in all  $k_2$  periods). The value  $k_2$  was chosen according to the continuous availability of data within the considered time period.



**Figure 8.** The model evaluation criterion.

Spatiotemporal kriging (SPTK) was applied to the same validation set. Finally, a comparison of RMSE, MAE, and  $R^2$  (Table 4) was conducted using the results produced by individual approaches and the hybrid approach.

**Table 4.** Comparison of prediction error of all models.

Method	Spatial Kriging	MLP	Hybrid	SPTK
RMSE	12.76	11.68	<b>10.72</b>	15.97
MAE	7.12	<b>4.01</b>	5.12	10.25
$R^2$	<b>13%</b>	3%	<b>13%</b>	0%

Among the considered four models, the minimum RMSE is produced by the hybrid weighted method. It also has an MAE close to its lowest value, which is produced by MLP, and the highest  $R^2$  value as of the spatial kriging (although it still needs to be improved). Thus, in the first phase of model comparison, it was proved that the hybrid model outperformed the other individual models. Moreover, the hybrid model outperformed the spatiotemporal kriging model with respect to all validation measurements.

#### 3.4.2. Model Evaluation on Artificially Generated Missing Data

Since the previous evaluation could not capture all the actual missing gaps (due to the inconsistency of the data set), we thought to generalize the validation using another validation set that could capture all the missing patterns (which are and are not in the original data set) that could be found in an original data set of six rainfall gauging stations.

In year 2016, for 6 months (from January to June), there were no missing observations. Therefore, we took the data of that period to create 56 subsets with different sequential missing patterns, as shown in Table 5.

**Table 5.** Types of missing gaps within six sub-stations.

Number Stations with Missing Observations (M)	Number of Combinations (of Stations)
01	${}^6C_1 = 6$
02	${}^6C_2 = 15$
03	${}^6C_3 = 20$
04	${}^6C_4 = 15$
Total number of data sets	<b>56</b>

When two stations have missing observations in the same period (consider one incident out of 15 combinations in case 02), each station was modeled separately. Thus, it resulted in two trials. Finally, all combinations of two stations with missing observations (i.e., 15) generated 30 trials ( $15 \times 2$ ). This was the same for the rest of the cases where there were at least two stations with missing data. The case where the number of stations with missing observations,  $M = 5$ , was not considered because there should be at least two stations without missing observations in the spatiotemporal grid to model those gauging stations with missing values using spatial kriging (a model requirement). In each subset, missingness was created with the last 20 observations (without breaking the time sequence and allocating enough data points for training) of the relevant sub-station. The rest of the data were used to train and test the models and, hence, identify the best-fit model for a selected sample of data.

The estimated values from each model were compared with the actual values and tabularized in Table S3. In terms of the lowest values of RMSE and MAE and the highest values of  $R^2$ , the hybrid model has outperformed the other approaches. Moreover, it is important to point out that, when kriging performs better than other individual methods in terms of three validation measurements, the hybrid model has performed approximately equally in most of the cases.

The outperformance with respect to the total (where lowest RMSE and MAE values and highest  $R^2$  value occur) shows the hybrid method could gain the score of 41% (see Table 6).

**Table 6.** Final comparison of model performance.

Method	Kriging	MLP	Hybrid	SPTK
No. of times that each approach produced best results (as a percentage with respect to the sum of trials)	<b>30</b>	<b>20</b>	<b>41</b>	<b>9</b>

In addition to the above outcome, all the cases (in Table S3) were taken to obtain the graphs shown in Figure 9.

Figure 9 illustrates that kriging produced most cases within the lowest value range of MAE and within the highest value range of  $R^2$ , whereas the MLP gained the majority of cases within the lowest value range of RMSE. In all these occurrences, the hybrid model has shown approximately the same performance as the optimal model.

This study demonstrated that the spatial and temporal variations should not be neglected in missing value imputation in spatiotemporal data. Further, considering the two-way evaluation, we could confirm that the hybrid model had the ability to outperform other pure approaches or almost equally performed compared to the best approach whenever it was selected as the second-best approach.

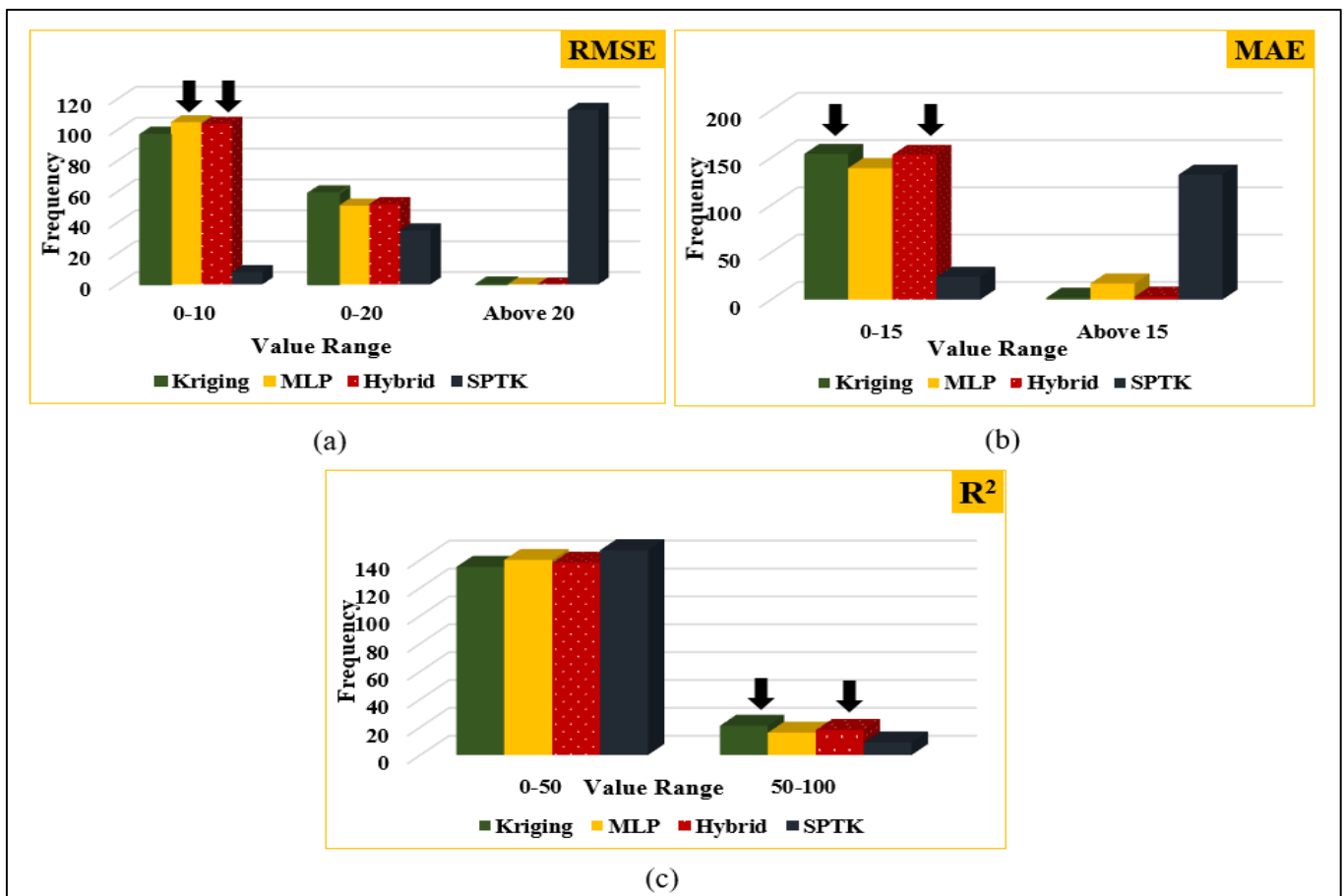


Figure 9. The outperformance of models with respect to RMSE (a), MAE (b), and  $R^2$  (c) values.

#### 4. Conclusions

In this paper, we proposed a novel approach for spatiotemporal missing value imputation. Furthermore, this approach was illustrated by applying it to impute rainfall data in six reference stations to compile a complete data set that could be used in building a rainfall forecasting model for a target station. The proposed approach is a hybrid approach combining MLP that captures the temporal variability and spatial interpolation (spatial kriging) that captures the spatial variations. This hybrid model was designed for imputing missing values when the missingness of reference stations was spread discontinuously and not uniformly over the given period and a few rainfall gauging stations are spatially correlated. Moreover, the results obtained by the hybrid model were compared with those resulting from spatiotemporal kriging. The study supported the claim that the proposed hybrid approach provided the lowest prediction error when estimating a sequence of missing values.

Only a single published research study that imputes the missing rainfall values of Ratnapura gauging stations is available. The accuracy measurements of the proposed method are far better than those of said study [28].

The availability of a few rain gauging stations around Ratnapura station and unavailability of data in some stations for a considerable period due to relocation or other reasons were the challenging limitations confronted during this study.

In order to verify the optimality of weight estimates and the outperformance of the hybrid method, the data for an extended period need to be gathered and tested.

While considering the aforesaid limitations, the future direction of this research will be the development of more hybrid models combining other missing value imputation methods and applying this approach to estimate missing values of spatiotemporal data obtained from other fields.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app14030999/s1>, Table S1: Missing value imputation of sub-station rainfall data using DL methods; Table S2: Final optimal weights; Table S3: Comparison of prediction error of all models in validation under Section 3.4.2.

**Author Contributions:** Conceptualization, C.T. and S.S.; methodology, S.S., C.T., P.L. and M.M.; software, S.S.; validation, S.S.; formal analysis, S.S.; investigation, C.T., M.M. and P.L.; resources, S.S.; data curation, S.S.; writing—original draft preparation, S.S.; writing—review and editing, C.T. and M.M.; visualization, S.S.; supervision, C.T., M.M. and P.L.; project administration, C.T.; funding acquisition, C.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Collaborative Research Grant of the University of Colombo, Sri Lanka, grant number AP/3/2019/CG/30.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the authors upon reasonable request.

**Acknowledgments:** The authors wish to acknowledge the University of Colombo, Sri Lanka, for the monetary support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Mital, U.; Dwivedi, D.; Brown, J.B.; Faybishenko, B.; Painter, S.L.; Steefel, C.I. Sequential Imputation of Missing Spatio-Temporal Precipitation Data Using Random Forests. *Front. Water* **2020**, *2*, 20. [\[CrossRef\]](#)
- Asadi, R.; Regan, A. A convolution recurrent autoencoder for spatio-temporal missing data imputation. *arXiv* **2019**, arXiv:1904.12413.
- Soley-Bori, M. *Dealing with Missing Data: Key Assumptions and Methods for Applied Analysis*; Boston University: Boston, MA, USA, 2013; Volume 4, pp. 1–19.
- Yang, H.; Yang, J.; Han, L.; Liu, X.; Pu, L.; Chin, S.; Hwang, H. A Kriging based spatiotemporal approach for traffic volume data imputation. *PLoS ONE*. **2018**, *13*, e0195957. [\[CrossRef\]](#) [\[PubMed\]](#)
- Agarwal, A. A New Approach to Spatio-Temporal Kriging and Its Applications. Master's Thesis, Ohio State University, OhioLINK Electronic Theses and Dissertations Center, Columbus, OH, USA, 2011. Available online: [http://rave.ohiolink.edu/etdc/view?acc\\_num=osu1306871646](http://rave.ohiolink.edu/etdc/view?acc_num=osu1306871646) (accessed on 18 November 2023).
- Aguilera, H.; Guardiola-Albert, C.; Hidalgo, C.S. Estimating extremely large amounts of missing precipitation data. *J. Hydroinformatics* **2020**, *22*, 578–592. [\[CrossRef\]](#)
- Feng, L.; Nowak, G.; O'Neill, T.J.; Welsh, A. CUTOFF: A spatio-temporal imputation method. *J. Hydrol.* **2014**, *519*, 3591–3605. [\[CrossRef\]](#)
- Hettiarachchi, P. *Inundation Maps of the Kalu Ganga Basin during the Flood in May 2003*; Department of Irrigation: Colombo, Sri Lanka, 2013.
- Nandalal, K.D.W. Use of a hydrodynamic model to forecast floods of Kalu River in Sri Lanka. *J. Flood Risk Manag.* **2009**, *2*, 151–158. [\[CrossRef\]](#)
- Rafii, F.; Kechadi, T. Collection of historical weather data: Issues with missing values. In Proceedings of the 4th International Conference on Smart City Applications (SCA'19), Casablanca, Morocco, 2–4 October 2019.
- Burhanuddin, A.; Zahrah, S.N.; Deni, S.; Ramli, N.M. Imputation of Missing Rainfall Data Using Revised Normal Ratio Method. *Adv. Sci. Lett.* **2017**, *23*, 10981–10985. [\[CrossRef\]](#)
- Jahan, F.; Sinha, N.; Rahman, M.; Rahman, M.M.; Mondal, M.S.; Islam, M. Comparison of Missing Value Estimation Techniques in Rainfall Data of Bangladesh. *Theor. Appl. Climatol.* **2019**, *136*, 1–17. [\[CrossRef\]](#)
- Ahmad, R.; Noor, F.; Zakaria, R.; Azman, M.A. Estimation of missing rainfall data using spatial interpolation and imputation methods. *AIP Conf. Proc.* **2015**, *1643*, 42–48. [\[CrossRef\]](#)
- Gad, I.; Doreswamy, H.; Manjunatha, B. A robust deep learning model for missing value imputation in big NCDC dataset. *Iran J. Comput. Sci.* **2021**, *4*, 67–84. [\[CrossRef\]](#)
- Afrifa-Yamoah, E.; Mueller, U.; Taylor, S.; Fisher, A. Missing data imputation of high-resolution temporal climate time series data. *Meteorol. Appl.* **2020**, *27*, e1873. [\[CrossRef\]](#)
- Nassir, S.; Badr, A.; Mousa, W. Estimation the Missing Data of Meteorological Variables in Different Iraqi Cities By using ARIMA Model. *Iraqi J. Sci.* **2018**, *59*, 792–801.
- Medeiros, E.; De Lima, R.; Olinda, R.; Santos, C. Modeling Spatiotemporal Rainfall Variability in Paraíba, Brazil. *Water* **2019**, *11*, 1843. [\[CrossRef\]](#)

18. Cuenca, J.; Correa-Flórez, C.; Patino, D.; Vuelvas, J. Spatio-Temporal Kriging Based Economic Dispatch Problem Including Wind Uncertainty. *Energies*. **2020**, *13*, 6419. [[CrossRef](#)]
19. Duan, Y.; Lv, Y.; Liu, Y. An efficient realization of deep learning for traffic data imputation. *Transp. Res. Part C Emerg. Technol.* **2016**, *72*, 168–181. [[CrossRef](#)]
20. Zoest, V.V.; Osei, F.; Hoek, G.; Stein, A. Spatio-temporal regression Kriging for modelling urban NO<sub>2</sub> concentrations Spatio-temporal regression Kriging for modelling urban NO<sub>2</sub> concentrations. *Int. J. Geogr. Inf. Sci.* **2019**, *34*, 851–865. [[CrossRef](#)]
21. Bae, B.; Kim, H.; Lim, H.; Liu, Y.; Han, L.; Freeze, P. Missing data imputation for traffic flow speed using spatio-temporal cokriging. *Transp. Res. Part C Emerg. Technol.* **2018**, *88*, 124–139. [[CrossRef](#)]
22. Srinivasan, B.; Duraiswami, R.; Murtugudde, R. Efficient Kriging for real-time spatio-temporal interpolation. In Proceedings of the 20th Conference on Probability and Statistics in the Atmospheric Sciences, Atlanta, Georgia, 18–21 January 2010; p. 228.
23. Abirami, S.P.; Chitra, P. Chapter Fourteen-Energy-efficient edge based real-time healthcare support system. *Adv. Comput.* **2020**, *117*, 339–368.
24. Orhan, U.; Hekim, M.; Ozer, M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.* **2011**, *38*, 13475–13481. [[CrossRef](#)]
25. Gardner, M.W.; Dorling, S.R. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [[CrossRef](#)]
26. Grosse, R. Lecture 5: Multilayer Perceptrons. 2018. Available online: [https://www.cs.toronto.edu/~rgrosse/courses/csc321\\_2018/readings/L05%20Multilayer%20Perceptrons.pdf](https://www.cs.toronto.edu/~rgrosse/courses/csc321_2018/readings/L05%20Multilayer%20Perceptrons.pdf) (accessed on 16 January 2022).
27. Zhou, Y.; Guan, Z.; Ji, P.; Wang, X. Drug Abuse Prediction Model Based on Relevance Analysis. In *Data Processing Techniques and Applications for Cyber-Physical Systems (DPTA 2019)—Advances in Intelligent Systems and Computing*; Huang, C., Chan, Y.W., Yen, N., Eds.; Springer: Singapore, 2020; Volume 1088. [[CrossRef](#)]
28. Jeewanthi, P.W.; Wijesuriya, W.; Liyanarachchi, L.A.T.S.; Gunatathne, L.H.P. Appropriate conventional methods forestimating missing precipitation values in Sri Lanka. *J. Agric. Value Addit.* **2023**, *5*, 45–53. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.