A Comparison between *MLR*, *MARS*, *SVR* and *RF* Techniques: Hydrological Time-series Modeling

Pankaj Kumar [1], Abhinav Kumar Singh [1*]

[1] *Department of Soil and Water Conservation Engineering, GBPU A&T Pantnagar,263145, India*

Abstract

Pan evaporation modeling is an essential part of water resources management and water budget governance. The study's objective was to examine the suitability of regression and tree-based techniques for estimating pan evaporation from climatic variables. Multiple linear regression (MLR), multivariate adaptive regression splines (MARS), support vector machine (SVM) and random forest (RF) techniques are employed for weekly pan evaporation modeling for the Ranichauri station situated in the Mid-Himalayan region of Uttarakhand, India. The determination of the most appropriate inputs among climatic variables to map evaporation was done by regression approaches. The data was divided into two parts: the first three years of data used for calibration and the remainder of the one-year data used for model validation. Statistical indices such as root mean square error (RMSE), Nash Sutcliffe coefficient of efficiency (NSE), and coefficient of determination (R^2) were used to assess the performance of weekly pan evaporation estimating models. Based on scatter plots, the results are under-predicted and over-predicted for the weekly pan evaporation values. The results showed that the values of the RMSE values ranged from 0.542 to 0.689, the NSE values ranged from 0.953 to 0.974, and the highest R^2 value was found for the SVR model for the testing period. Therefore, the SVR model was found to be superior and can be applied to predict weekly pan evaporation values for the Ranichauri site.

Keywords: Pan Evaporation; MLR; MARS; SVM; Random Forest; Hydrology.

1. Introduction

Techniques for quantifying evaporation loss need to be strengthened to meet essential needs such as water requirements for domestic use, industrial use, agricultural uses, and commercial and cultural purposes. The main factor that should be precisely determined for the water budget is evaporation. Evaporation is usually a form of water depletion, and the entire world faces this problem. Evaporation is responsible for climatic variables such as relative humidity, temperature, wind speed, surface area, etc. According to the report, an accurate and precise assessment of evaporation is essential for managing, planning, and growing water supplies [1, 2]. Evaporation is a process that is nonlinear, unstable, and very complex [3-5]. The impact of climate change has been significant on the water system [6]. In water resource modeling, Multivariate Adaptive Regression Splines (MARS) and Random Forest are successfully used [7]. Various learning algorithms have been used to create models by using meteorological variables to show their ability to perform pan evaporation [8–19].

A number of research studies have been conducted in recent years on the implementation of machine learning (ML) models for evaporation estimation across various regions [20]. Yaseen (2020) [19] employed classification and regression tree (CART), gene expression programming (GEP). The cascade correlation neural network (CCNNs), and the support vector machine (SVM) for prediction of pan evaporation in two stations in Iraq were tested and it was found that SVM yielded the best performance. Along with appropriate model selection, the input-output selection is of prime importance as it is very complex and involves mathematical procedural interventions. Using different techniques and then choosing the best model or various methods to pay attention to hydrology is also a part of the planning, management, and development of water resources [19].

The paper explores and compares the estimation of pan evaporation in Ranichauri, Uttarakhand, India using four approaches: (1) Multiple linear regression (2) Multivariate adaptive regression splines (MARS) approach (MARS), (3) Random Forest (RF), and (4) Support Vector Machine (SVM). The study also examines the association of climatic variables with pan evaporation and selects the most suitable model for pan evaporation estimation. The adequacy of developed models was examined by statistical indices: root mean square error (RMSE), coefficient of determination (R^2), and Nash-Sutcliffe coefficient of efficiency (NSE) between observed and estimated pan evaporation values.

2. Material and Methods

2.1. Description of the Study Area and Data

Ranichauri is a hill station and a small village located in Uttarakhand's northern Indian state, in the Tehri Garhwal district. The Ranichauri is known for the College of Forestry of Uttarakhand University of Horticulture and Forestry, situated in the Mid-Himalayan region at 1827 m altitude (30°18' N latitude and 78°24' E longitude). The location map of the study area is shown in Figure 1. The weekly meteorological variables such as maximum temperature (T_{max}), minimum temperature (T_{min}), wind speed (W), maximum relative humidity (Rh_1), minimum relative humidity (Rh_{min}), sunshine hours (Sn), and evaporation (E) were obtained from the College of Forestry of Uttarakhand University of Horticulture and Forestry from January 2014 to December 2017. Out of the 4-year total data, 3-year data were used for calibration (2014 to 2016), and 1-year remaining data were used for validation (2017).
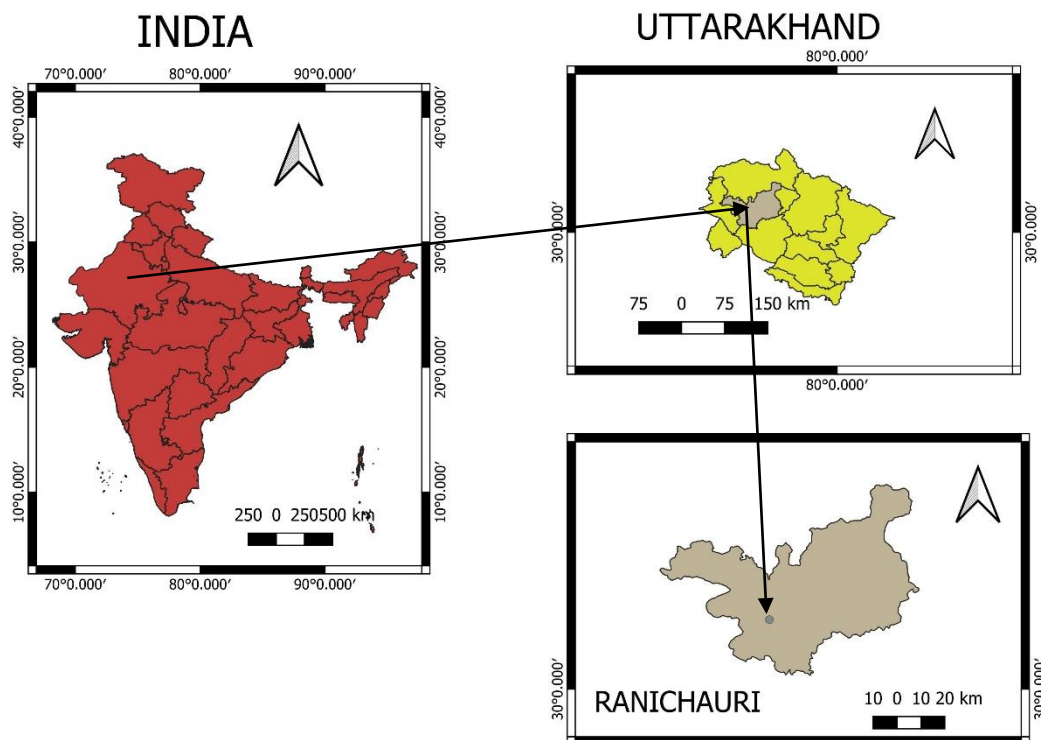

Figure 1. Location map of watershed: Ranichauri, Uttarakhand's Northern Indian State

2.2. Multiple Linear Regression (MLR)

It is the most commonly used statistical approach to predict the output of several input variables and develops the linear relationship between multiple variables. MLR is a type of model in which the number of independent and dependent variables forms a quantitative relationship. The values of independent variables are affiliated with the value of dependent variables. MLR can be used for estimating impacts, predicts sequence & future values and to elaborate on the effect of independent variables on the dependent variables. In this study, MLR is used to predict pan evaporation, by taking the dataset into training and testing periods. The expression for MLR is defined as follows:

$$Y = \alpha_o + \sum_{i=1}^{n} \alpha_i X_i \tag{1}$$

where, Y = dependent variable's estimated value; α_o = value of the dependent variable when all independent variables become zero; α_i = regression coefficients and; X_i = independent variables.

2.3. Multivariate Adaptive Regression Splines (MARS)

MARS proposes a set of extensible regression models where the solution space of each model is divided into various intervals of predictor variables, and the splines are fit to such interval space [21]. For the comparison of subsets of the model, a less expansive technique is used called generalized cross-validation [22] it and expressed as follows:

$$GCV = \frac{MSE}{\left[1 - \frac{(f+1) + pf}{n}\right]^2} \tag{2}$$

where, MSE = Mean squared error; f = number of basis functions; p = Basis function penalty and; n= number of observations.

MARS is a model which performs under two type of functions, forward and backward [23]. In the case of forward pass, the over fitted model developed through a huge quantity of basis functions introduced to the MARS model. The generalized form of the MARS model is given [24]:

$$Y = \beta_o + \sum_{i=1}^{m} \beta_i H_{ei}(X_v(e, i)) \tag{3}$$

where, Y = Output parameter; β_o = Constant value; i = number of basis functions; $H_{ei}(X_v(e, i))$= ith basis function and; β_i= Corresponding coefficient of $H_{ei}(X_v(e, i))$

The MARS is an extended form of a linear model that describes the relationship between dependent and independent variables which is not presumed. It is an algorithm which constructs a model in two stages, the first one is a collection of basis function and in the second stage, it estimates the least square model. MARS model defined as follows [25]:

$$Y = \partial_o + \sum_{i=1}^{n} \partial_i h_i(X) \tag{4}$$

Where; $h_i(X)$= Spline functions; ∂_o= Coefficients from spline functions can be determined by the minimum sum of the square errors; i= total function number in model.

2.4. Support Vector Regression

The Support Vector Machine (SVM) is a nonlinear generalization algorithm used for classification and regression problems, introduced by Vapnik (1992) [26]. It is based on the concept of minimizing structural risk and the algorithm converts the patterns that are not linearly separable to the higher dimensional feature space using kernel functions. In the algorithm, the kernel function was built to map the actual data into higher-dimensional space data without increasing the computational cost. Instead of fixing the empirical error, the SVR attempts to minimize the upper limit of the generalization error. The hard margin solution is considered the first formulation of the SVR, which contributes to overfitting. The definition of soft-margin, which appears to generalize well in the presence of noise, outliers, and prevents overfitting, which is widely used for forecasting research, was then included in further approaches. The overall form of the SVM model regression equation is as follows:

$$y = w^T \phi(x) + b \tag{5}$$

where, $w^T = i = 1nw_ix_i$; w = weight vector; $\phi(x)$= nonlinear mapping function of inputs; and b= bias, w and b can be estimated by minimizing the loss function.

2.5. Random Forest (RF)

Random forest is the model used for both classification and regression problems and also comes under the umbrella of classification and regression tree (CART) tools. It is the algorithm that operates the model by constructing several decision trees and can say that it deals with many numbers of features. Random Forest is an aggregation of tree predictors. Each tree is estimated according to the values of a random vector computed from the same distribution for all the trees in the forest [27]. In random forest, each tree in the model is grown with random subsets of variables [28]. Random forest ensures that the bagged trees' bias happens in the same way as that of the single tree reduces the correlation and variance between the trees of the model [29]. Random Forest, which breaks variables at each node, is also chosen from a random subset of the available features, which reduces the association between trees. The mean squared error of the model can be evaluated as follows [22]:

$$MSE = N^{-1} \sum_{i=1}^{m} (Z_i - i)^2 \tag{6}$$

where, Z_i= Measured variable value and; i= Mean of all out-of-bag (OOB) predictions.

Several trees in random Forest modeling are also key parameters, and the performance of the model can be evaluated through out-of-bag (OOB). Using random Forest, the over-fitting of the model for the training data can be evaded by selecting input variables during the training cycle to establish variability in weak learners [30]. The model builds multiple decision trees, and the model's output can be calculated by taking the average or mean output of each tree [31]. The predicted values can be estimated as follows:

$$Y = \frac{1}{n}\sum_{i=1}^{n} R(x) \tag{7}$$

where; Y = Predicted output from Random forest; n = Total tree (ntree) utilized in the model; $R(x)$ = Result of each Random Tree.

2.6. Input Variable Selection

The acquisition of appropriate inputs for the model is an essential exercise. To avoid the models' overfitting and minimize the complications in the prediction, the Gamma test (GT) used to eliminate the unnecessary input parameters that have an irrelevant role in predicting the output. The GT avoids complexity with the selection of best inputs, which is genuinely influenced by the results. The combinations of input variables were designated using Gamma value (Γ), standard error (SE), and V-ratio for the pan evaporation estimation model. The Gamma Test is a nonparametric approach based on the minimization of mean square error in the modeling [32-34]. The best input combination is determined based on the obtained minimum value of Gamma static (Γ).

3. Results and Discussion

For the study area, the input combination maximum temperature (T_{max}), Maximum Relative Humidity (RH_{max}), Minimum Relative Humidity (RH_{min}), Wind Speed (W), and sunshine hours (S) have minimum Gamma value (0.213), standard error (0.057), and V-ratio (0.134) therefore this combination was selected for modeling approach. There are no specific guidelines for separating training and testing data in predictive modeling for learning. However, Zounemat-Kermani et al. (2019) [35] suggested 80% of total data to develop the models. Thus, we divided the data into training (80%) and testing to set up the studied models for evaporation estimation (20%). Therefore, data had been used for training in the period 2014-2016 and the remaining data from 2017 was used for evaluating the applied models [36-39]. The statistical investigation viz. minimum, maximum, mean, standard deviation, first quartile, and third quartile of the training, testing, and the complete dataset is given in Table 1. It can be seen in Table 1 that the maximum value of evaporation for the training set is 7.15 mm, and for the test set, it is 4.38. Similarly, the minimum values are 0.77 and 0.48 mm, respectively. The standard deviation for the training set is 1.31, and for the test set, it is 1.05. There seems a slight variation between the first and third quartile values of the evaporation for training and testing sets.

Table 1. **Statistical parameters of training, testing, and complete weekly climatic dataset of maximum temperature (T_{max}), maximum relative humidity (RH_{max}), maximum relative humidity (RH_{min}), Wind Speed (W), sunshine hours (S), and pan evaporation (E) of the study area. x_{min}, x_{max}, x_{mean}, SD, $1^{st}Q$, $3^{rd}Q$, are minimum, maximum, mean, standard deviation first quartile, and third quartile.**

Statistical parameters	Climatic Variables					
	T_{max} (°C)	RH_{max} (%)	RH_{min} (%)	W (km/h)	S (h)	E (mm)
Training (2014-2016)						
X_{min}	8.586	33.86	18.71	1.47	0.57	0.77
X_{max}	29.29	100	91.71	6.42	11.21	7.15
x_{mean}	19.69	78.91	62.51	3.84	6.74	2.59
Median	21.44	79.64	62.17	3.87	7.05	2.15
$1^{st}Q$	16.46	70.39	50.11	3.26	5.55	1.70
$3^{nd}Q$	23.31	91.46	77.86	4.35	8.14	3.16
SD	4.84	13.62	15.91	0.91	2.05	1.31
Testing (2017)						
X_{min}	7.12	54.43	34.86	1.54	1.68	0.48
X_{max}	27.029	98.71	91.71	5.271	10.2	4.38
x_{mean}	20.389	80.64	62.15	3.533	6.675	2.11
Median	22.679	80.07	59.64	3.757	7.164	1.72
$1^{st}Q$	16.807	72.46	51.25	2.611	5.189	1.41
$3^{nd}Q$	23.586	93.50	77.36	4.379	8.193	2.66
SD	4.64	12.12	15.21	1.005	2.30	1.05

| Complete Data (2014-2017)						
X_{min}	7.12	33.86	18.71	1.47	0.57	0.48
X_{max}	29.22	100	91.71	6.42	11.21	7.15
x_{mean}	19.87	79.34	62.42	3.76	6.72	2.47
Median	21.74	79.79	60.79	3.86	7.11	2.1
1st Q	16.59	70.68	50.50	3.18	5.43	1.59
3nd Q	23.38	91.61	77.86	4.35	8.15	3.12
SD	4.79	13.26	15.70	0.94	2.11	1.26

## 3.1. Modeling Daily Pan Evaporation

The *MLR, MARS, SVR, and RF* models were trained with the training data set. In the case of the MARS model, for comparison of subsets of the model, the value of generalized cross-validation (GCV) for training and testing was found to be 0.273 & 0.636 [40]. In the SVR model, the radial basis kernel function was implemented with kernel parameter gamma of 1.0. Moreover, convergence epsilon was 0.001. For RF model, the number of variables randomly sampled as candidates at each split was chosen 3 with 100 numbers of trees. The results of training and testing periods of the *MLR, MARS, RF, and SVM* models in estimating pan evaporation for the Ranichauri station are shown in Table 2. According to Table 2, it is evident for the training period, the MLR model values of RMSE, NSE, and R^2 are 0.53, 0.80 and 0.84. SVR model has values of RMSE, NSE, and R^2 as 0.33, 0.94 and 0.94. For the MARS model, RMSE, NSE, and R^2 are 0.39, 0.97, and 0.91. Identically, the performance indices, RMSE, NSE, and R^2 values for the RF model were found to be 0.22, 0.97, and 0.98 for the training period. It is visible from the assessment of all models that the RF model has the minimum RMSE value and the highest NSE and R^2 value during the training period. The MLR model is the simplest model, with the highest RMSE value and the lowest NSE and R^2 values for the study area during the training phase.

**Table 2. Comparison of the various models by using statistical indices**

Models	Training			Testing		
	RMSE	**NSE**	**R^2**	**RMSE**	**NSE**	**R^2**
MLR	0.53	0.80	0.84	0.67	0.59	0.82
SVR	0.33	0.94	0.94	0.67	0.64	0.88
MARS	0.39	0.97	0.91	0.68	0.52	0.87
RF	0.22	0.97	0.98	0.73	0.47	0.86

While for the testing period, the MLR model values of RMSE, NSE, and R^2 are 0.67, 0.59 and 0.82. In the case of the SVR model, the indices of RMSE, NSE, and R^2 are 0.67, 0.64, and 0.88. MARS model has 0.68, 0.52, and 0.87 values of RMSE, NSE, and R^2. The performance indices, RMSE, NSE, and R^2 values for the RF model were found to be 0.73, 0.47, and 0.86 for the testing period. It is visible from the assessment of all models during the testing period that the SVR model outperformed.
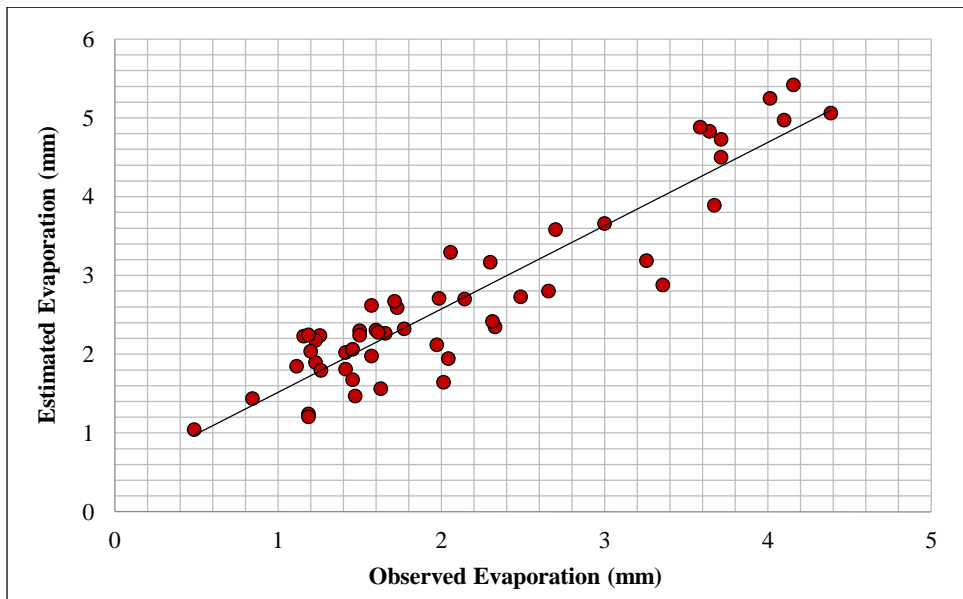

(a)
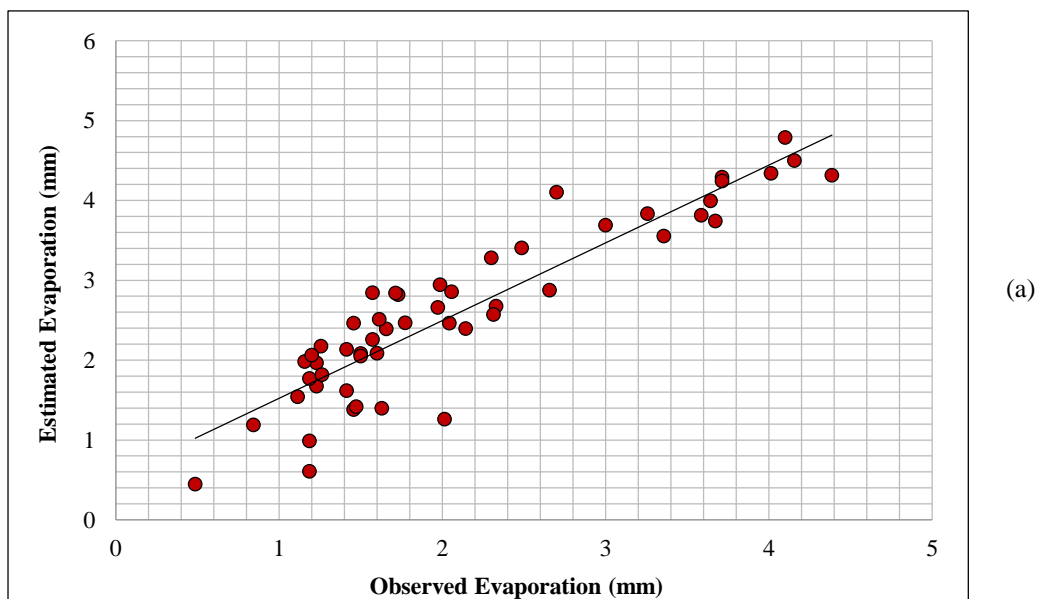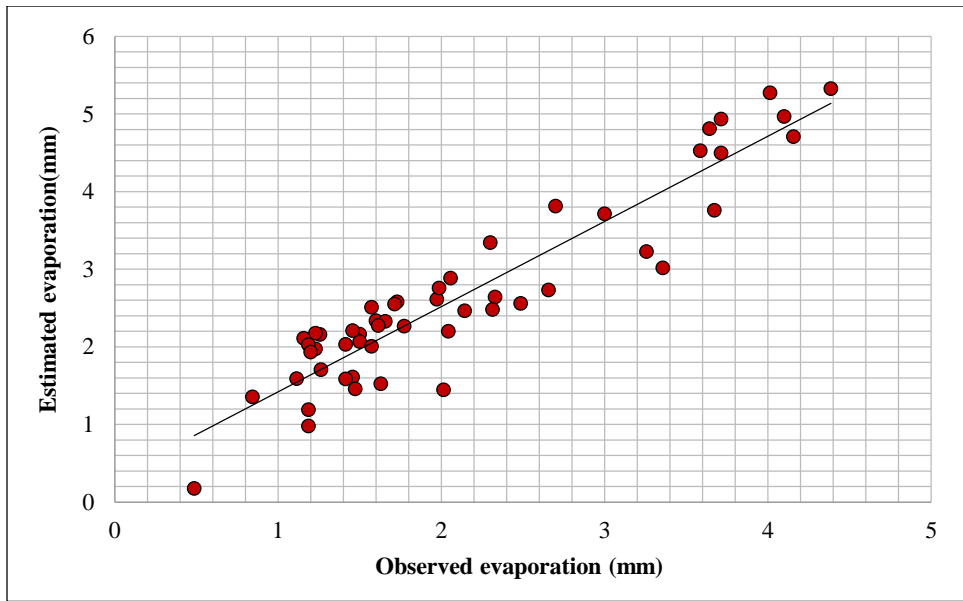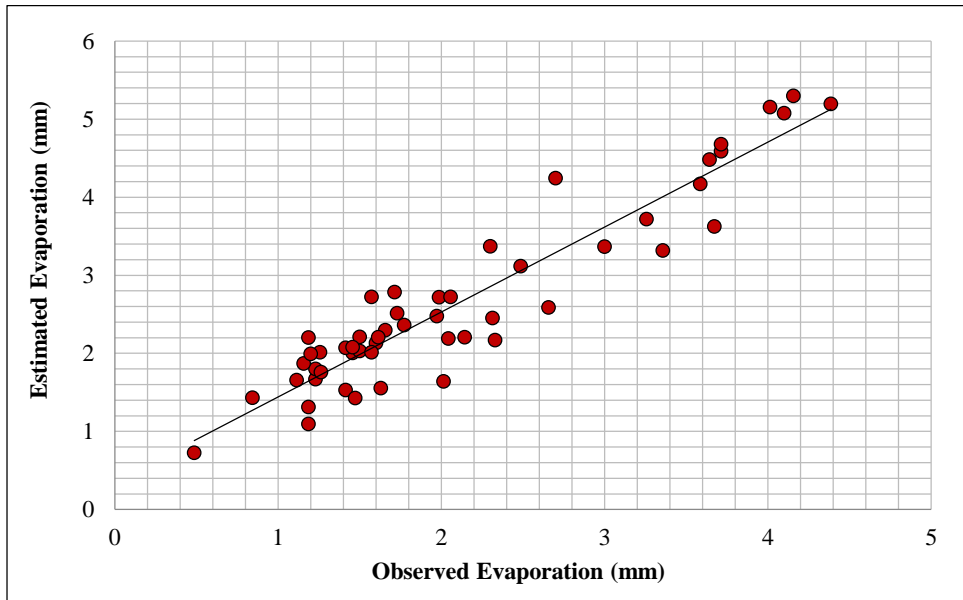
Figure 3. Scatter plot of the observed and estimated weekly pan evaporation values by MLR, MARS, SVR and RF models during the testing period for the study area

The variations of the scatter plot of applied models during the testing period are compared in Figure 3(a-d). It is clear from the illustration that the model's scatterplot follows the 1:1 best-fit line. However, the SVR model has less scatter. It can be seen from Table 2 that all models' performance is not consistent in the testing case. The RF model outperformed in the training period, but it could not generalize results for the testing period. However, the MLR model, which is generally considered the simplest model, has done considerably well for the RMSE and NSE statistics during the testing period. Perhaps these uncertainties in the results obtained were due to data division, input uncertainty, and model parameter optimization. Therefore, the present study results showed that the evaporation values could be calculated in similar climates using a few easily measured meteorological parameters and with sufficient precision. In order to determine the consistency of the considered models, they should be tested in different climates with varying data lengths and training testing split.

As a final comment, the results obtained suggest that the accuracy of MLR, MARS, SVR, and RF techniques was adequate when using maximum temperature (T_{max}), maximum relative humidity (RH_{max}), maximum relative humidity (RH_{min}), wind speed (W), and sunshine hours (S) meteorological parameters. In addition, while for Ranichauri station, different machine learning methods' accuracy varied, and SVR output was better than other models examined.

4. Conclusion

This work employed four approaches (MLR, MARS, SVR, and RF) to determine the pan evaporation, using a short data set for the Ranichauri station. In climatic explanatory variables, in particular, the input combination of the maximum temperature (T_{max}), minimum relative humidity (RH_{min}), maximum relative humidity (RH_{max}), wind speed (W), and sunshine hours (S) has a significant influence on pan evaporation, thus they are chosen as inputs for models. All four approaches can estimate pan evaporation accurately ($R^2 > 0.82$). The MLR approach is the simplest one and showed considerable generalization capability for the Ranichauri station for the given data set. The RF model performed very well in the training period and could not generalize for the testing period. The performance wise ranking of models was done as SVR, MARS, MLR, and RF. Among the four models, the SVR model outperformed in testing periods and also performed well in the training period. Consequently, it can be surmised that the SVR model delivers the strongest potential for estimating pan evaporation in Ranichauri.

5. Declarations

5.1. Author Contributions

Conceptualization, P.K. and A.K.S.; methodology, P.K.; software, A.K.S.; validation, P.K. and A.K.S.; formal analysis, P.K.; investigation, P.K.; resources, P.K.; data curation, A.K.S.; writing—original draft preparation, A.K.S.; writing—review and editing, P.K.; visualization, A.K.S.; supervision, P.K.; project administration, P.K. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

6. References

[1] Penman, H. L. (1948). Natural evaporation from open water, hare soil and grass. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 193(1032), 120–145. doi:10.1098/rspa.1948.0037.

[2] Stephens, J. C., & Stewart, E. H. (1963). A comparison of procedures for computing evaporation and evapotranspiration. Publication, 62, 123-133.

[3] Kim, S. J., Ko, S. H., Kwak, R., Posner, J. D., Kang, K. H., & Han, J. (2012). Multi-vortical flow inducing electrokinetic instability in ion concentration polarization layer. Nanoscale, 4(23), 7406. doi:10.1039/c2nr32467a.

[4] Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. Journal of Hydrology, 528, 312–320. doi:10.1016/j.jhydrol.2015.06.052.

[5] Wang, L., Niu, Z., Kisi, O., Li, C., & Yu, D. (2017). Pan evaporation modeling using four different heuristic approaches. Computers and Electronics in Agriculture, 140, 203–213. doi:10.1016/j.compag.2017.05.036.

[6] Resnick, K., Nelson, M. (2011). Reflections from the 2009 NEWWA/EPA Water Resources Symposium: Water Resiliency - Adapting Water Supply to Changing Climate, Land Use, and Regulation. Journal of the New England Water Works Association, 125(3), 244.

[7] Tezel, G., & Buyukyildiz, M. (2016). Monthly evaporation forecasting using artificial neural networks and support vector machines. Theoretical and Applied Climatology, 124(1-2), 69–80. doi:10.1007/s00704-015-1392-3.

[8] Bruton, J. M., McClendon, R. W., & Hoogenboom, G. (2000). Estimating daily pan evaporation with artificial neural networks. Transactions of the American Society of Agricultural Engineers, 43(2), 491–496. doi:10.13031/2013.2730.

[9] Sudheer, K. P., Gosain, A. K., & Ramasastri, K. S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrological Processes, 16(6), 1325–1330. doi:10.1002/hyp.554.

[10] Kişi, Ö. (2006). Daily pan evaporation modelling using a neuro-fuzzy computing technique. Journal of Hydrology, 329(3–4), 636–646. doi:10.1016/j.jhydrol.2006.03.015.

[11] Doğan, E., Işik, S., & Sandalci, M. (2007). Estimation of daily evaporation using artificial neural networks. Teknik Dergi/Technical Journal of Turkish Chamber of Civil Engineers, 18(2), 4119–4131.

[12] Kim, S., & Kim, H. S. (2008). Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. Journal of Hydrology, 351(3–4), 299–317. doi:10.1016/j.jhydrol.2007.12.014.

[13] Piri, J., Amin, S., Moghaddamnia, A., Keshavarz, A., Han, D., & Remesan, R. (2009). Daily Pan Evaporation Modeling in a Hot and Dry Climate. Journal of Hydrologic Engineering, 14(8), 803–811. doi:10.1061/(asce)he.1943-5584.0000056.

[14] Moghaddamnia, A., Gosheh, M. G., Nuraie, M.., Mansuri, M. A., Han, D., & Schmitter, E. (2010). Performance evaluation of LLR, SVM, CGNN and BFGSNN models to evaporation estimation. In WSEAS International Conference. Proceedings. Mechanical Engineering Series (Vol. 5). World Scientific and Engineering Academy and Society.

[15] Kumar, P., D., K., Jaipaul, & Tiwari, A. K. (2012). Evaporation Estimation Using Artificial Neural Networks and Adaptive Neuro-Fuzzy Inference System Techniques. Pakistan Journal of Meteorology, 8(16), 81–88.

[16] Samui, P. (2012). Determination of ultimate capacity of driven piles in cohesionless soil: A Multivariate Adaptive Regression Spline approach. International Journal for Numerical and Analytical Methods in Geomechanics, 36(11), 1434–1439. doi:10.1002/nag.1076.

[17] Kumar, P., Kumar, D., & Panwar, R. (2016). Evaporation estimation from climatic factors. Mausam, 67(4), 897–902. doi:10.54302/mausam.v67i4.1417.

[18] Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agricultural and Forest Meteorology, 263, 225–241. doi:10.1016/j.agrformet.2018.08.019.

[19] Yaseen, Z. M., Al-Juboori, A. M., Beyaztas, U., Al-Ansari, N., Chau, K. W., Qi, C., Ali, M., Salih, S. Q., & Shahid, S. (2020). Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models. Engineering Applications of Computational Fluid Mechanics, 14(1), 70–89. doi:10.1080/19942060.2019.1680576.

[20] Bhardwaj, S., Chandrasekhar, E., Padiyar, P., & Gadre, V. M. (2020). A comparative study of wavelet-based ANN and classical techniques for geophysical time-series forecasting. Computers and Geosciences, 138. doi:10.1016/j.cageo.2020.104461.

[21] Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics, 19(1), 1-67. doi:10.1214/aos/1176347963.

[22] Zhang, Y., Zhao, Z., & Zheng, J. (2020). CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. Journal of Hydrology, 588, 125087. doi:10.1016/j.jhydrol.2020.125087.

[23] Kisi, O., & Parmar, K. S. (2016). Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. In Journal of Hydrology 534, 104–112. doi:10.1016/j.jhydrol.2015.12.014.

[24] Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017). Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. Journal of Hydrology, 553, 356–373. doi:10.1016/j.jhydrol.2017.08.006.

[25] Adnan, R. M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., & Li, B. (2020). Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. Journal of Hydrology, 586. doi:10.1016/j.jhydrol.2019.124371.

[26] Vapnik, V. (1992). Principles of risk minimization for learning theory. Advances in neural information processing systems, NeurIPS Proceedings, 4, 831–838.

[27] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). doi:10.1023/A:1010933404324.

[28] Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. In Ecosystems 9,(2), 181–199. doi:10.1007/s10021-005-0054-1.

[29] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer Series in Statistics, Springer New York, NY, United States. doi:10.1007/978-0-387-84858-7.

[30] Sadler, J. M., Goodall, J. L., Morsy, M. M., & Spencer, K. (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. Journal of Hydrology, 559, 43–55. doi:10.1016/j.jhydrol.2018.01.044.

[31] Srivastava, R., Tiwari, A. N., & Giri, V. K. (2019). Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India. Heliyon, 5(10). doi:10.1016/j.heliyon.2019.e02692.

[32] Stefánsson, A., Končar, N., & Jones, A. J. (1997). A note on the gamma test. Neural Computing and Applications, 5(3), 131–133. doi:10.1007/BF01413858.

[33] Zick, S. E. (2020). Quantifying extreme precipitation forecasting skill in high-resolution models using spatial patterns: A case study of the 2016 and 2018 Ellicott City floods. Atmosphere, 11(2), 136. doi:10.3390/atmos11020136.

[34] Keshtegar, B., Piri, J., & Kisi, O. (2016). A nonlinear mathematical modeling of daily pan evaporation based on conjugate gradient method. Computers and Electronics in Agriculture 127, 120–130. doi:10.1016/j.compag.2016.05.018.

[35] Zounemat-Kermani, M., Seo, Y., Kim, S., Ghorbani, M. A., Samadianfard, S., Naghshara, S., Kim, N. W., & Singh, V. P. (2019). Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida. Applied Sciences, Switzerland, 9(12). doi:10.3390/app9122534.

[36] Kreklow, J., Tetzlaff, B., Burkhard, B., & Kuhnt, G. (2020). Radar-based precipitation climatology in germany-developments, uncertainties and potentials. Atmosphere, 11(2), 217. doi:10.3390/atmos11020217.

[37] Kumar, P., Singh, R., Sharma, A., Singh, G., Kumar, D., & Singh, A. K. (2021). Comparison of Different Interpolation Techniques for Mean Areal Rainfall Estimation of Uttarakhand using Geographical Information System. Journal of the New England Water Works Association, 135(3), 43–50.

[38] Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. Remote Sensing of Environment, 178, 127–141. doi:10.1016/j.rse.2016.03.006.

[39] Kenney, T. (2021). Obituaries: Guy Manning Foss. Journal of the New England Water Works Association, 135(1), 89–89.

[40] Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K.-W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. Journal of Hydrology, 569, 387–408. doi:10.1016/j.jhydrol.2018.11.069.