Indian Journal of Ecology (2024) 51(4): 828-836
DOI: https://doi.org/10.55362/IJE/2024/4317

Manuscript Number: 4317
NAAS Rating: 5.38

Estimation of Monthly Evaporation in Hilly Regions of Uttarakhand using Machine Learning Techniques

Suman Markuna and Basant Ballabh Dumka[1*]

Department of Civil Engineering, Govind Ballabh Pant Institute of Engineering and Technology
Pauri Garhwa- 246 194, India
[1]Department of Civil Engineering, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun-248 007, India
**E-mail: dumkabasant@gmail.com*

Abstract: Evaporation estimation is one of the complicated and important measures of the hydrological cycle due to its complex behavior for planning, management, and development of water resources. In this study, the Multivariate Adaptive Regression Splines (MARS) and Random Forest (RF) models were utilized to estimate mean monthly evaporation at Ranichauri. The mean monthly meteorological variables as, maximum and minimum temperature, morning and afternoon relative humidity, rainfall, morning and afternoon vapor pressure, wind speed, morning and afternoon wind direction, solar radiation, and evaporation were used for the development of the models. The results obtained by MARS and RF models were compared based on statistical indices root mean square error (RMSE), correlation coefficient (r), percent bias (PBIAS), and Nash-Sutcliffe efficiency (NSE). The results indicated that the performance of the RF model is better than the MARS model during the training period and the MARS model is better than the RF model during the testing period with nine input variables in estimating mean monthly evaporation at Ranichauri. Based on the present results when compared with the previous study's results done by many researchers, reveal that the methods can easily be implemented in hilly regions with greater accuracy.

Keywords: Gamma Test, Evaporation, Multivariate adaptive regression splines, Random forest

Evaporation is a nonlinear process that converts a liquid state into a gaseous state under different kinds of atmospheric processes. The meteorological variables (Temperature, wind speed, solar radiation, vapor pressure, humidity etc.) have a great association with the formation of evaporation. Due to the conversion of liquid into a gaseous form, the clouds are formed and then precipitation may occur. Evaporation is a major kind of water loss from the water bodies and water channels. Evaporation is considered a time-dependent function and a cumulative process, further, the estimation of isotopic evaporation is reliable during the dry season while estimations of isotopic loss fraction in the wet season behave as an isotopic mixing model of evaporated water-precipitation-runoff (Perez et al 2020). Evaporation estimation is one of the complicated measures of the hydrological cycle due to its complex behavior. For improving the development and management plans of the water resources the evaporation should be well estimated. Therefore, proper estimation and prediction of evaporation especially in arid and semiarid areas is of great importance to integrated water resources management and modeling studies (Ribot et al 2005, Nourani and Fard 2012, Malik and Kumar 2015). Modern hydrology is concerned with the distribution of water on the surface of the earth and its movement over and beneath the surface, and through the atmosphere (Davie 2008). The rate of evaporation depends upon the temperature variation, wind speed, atmospheric pressure, relative humidity, size of the water body, etc. Estimation of the water loss by evaporation is important for modeling, surveying and management of many projects related to hydrology and water resources systems (Shrisath and Singh 2010).

The direct and indirect methods can be used for estimation and evaluation of evaporation loss of any region. One of the direct methods for evaporation measurement is pan evaporation (Ep) (Eslamian et al 2008, Goyal and Ojha 2011). By using different types of techniques, the statistical parameters can be defined easily with a high degree of accuracy. Several studies have emphasized the need for accurate estimates of evaporation in hydrologic modeling studies (Sudheer et al 2002, Szilagyi and Jozsa 2009). Machine learning algorithms build a mathematical model based on sample data. For this, machine learning techniques are widely used in the estimation of time series, rainfall-runoff modeling, hydro informatics, and trend analysis. Machine learning is a computational statistics, that focuses on predictions and estimations using computers. It is a sub-area of artificial intelligence. Evaporation is one of the common examples, as it exhibits unique characteristics of complexity in nature and chaotic patterns so the estimation of evaporation is complex. Using machine learning techniques such as linear regression, decision tree, wavelet artificial neural network (WANN), random forest (RF), clustering algorithms, support vector regression (SVR), multivariate adaptive regression splines (MARS), artificial neural network (ANN) are well adapted for evaporation estimation (Cramer et al 2017, Dodla et al 2017, Khan et al 2018, Pham et al 2020). Recently for the monthly pan evaporation prediction the MARS, M5 model tree (MT) was coupled with a maximum overlap discrete wavelet transform (MODWT) for Turkey's Siirt and Diyarbakir stations (Ghaemi et al 2019), results revealed that the hybrid MARS model improves the accuracy for both the stations compared with the standalone models. The random forest (RF), which is tree-based simple, and robust to the outliers, has been widely used for pan evaporation and evapotranspiration estimation. It can easily deal with small and large datasets.

The CAT Boost was found to be the best and the RF model was the second best-performing model to map weekly pan evaporation with higher accuracy and effectiveness, followed by MLR, MARS, MNLSR, and M5Tree (Vishwakarma et al 2024).The daily pan evaporation modeling of Poyang lake, China was performed by four different models (M5 model tree, random forest, and gradient boosting decision tree) using daily meteorological variables out of which the gradient boosting decision tree is superior to the random forest model (Lu et al 2018). The machine learning techniques indicated that MARS model is superior to the LSSVM and M5 Tree for the modeling of daily evaporation of coastal cities of Turkey with a lack of input variables (Kisi 2015). The weekly evaporation prediction of different regions of India as Raipur, Karnal, Pattambi, and Anantapur using discriminant analysis, logistic regression, and random forest demonstrates that the random forest model predicts weekly evaporation accurately than the logistic regression, discriminant analysis (Rakhee et al 2020). The performance of the hybrid artificial neural network (ANN) models may be affected by importing meteorological variables such as solar radiation and deficit vapor pressure as inputs for daily evaporation estimation (Seifi *et al*., 2020 and Dumka *et al*., 2018). Among the four models (MLR, SVR, MARS, and RF) the RF model outperformed in both training and testing phases for weekly rainfall prediction (Markuna et al 2023).

Based on the above reviews, this study was conducted to analyze the potential of machine learning techniques (MARS, and RF) for the estimation of mean monthly evaporation of Ranichauri located at the foothills of the Himalayas in the Uttarakhand state of India. The combination of input variables was selected using the Gamma test to improve the performance of the models. A comparison between the models was described using statistical indices such as root mean square error (RMSE), correlation coefficient (r), Percent bias (PBIAS), and Nash Sutcliffe efficiency (NSE).

MATERIAL AND METHODS

Study area and data: Ranichauri is a hill station located in the Northern Indian state of Uttarakhand, in the Tehri Garhwal district. The altitudinal expansion of the study area is 1827 meters above sea level, the latitude is 30°18'40" N and the longitude is 78°24'35" E. The historically available meteorological data obtained for the analysis from 2008-2017 from the College of Forestry, Agromet Field Unit (AMFU), V.C.S.G. Uttarakhand University of Horticulture and Forestry, Ranichauri, Uttarakhand. The total data points were 120, out of the total data 75% (90 data points) were used in training, and the remaining 25% (30 data points) were used in testing periods. The mean monthly meteorological data of maximum and minimum temperatures (T_{max} and T_{min}), morning and afternoon relative humidity (RH_m, RH_a), evaporation (E), wind speed (W_s), morning & afternoon wind direction (Wd_m and Wd_a), solar radiation (SR), morning and afternoon vapor pressure (Vp_m and Vp_a) and rainfall (Rf) were collected from the study area. The warmest month of the study area was June with the highest mean monthly temperature of 29.6 °C, while the coolest month of the study area was January with the lowest temperature -0.551°C.

Gamma test (GT): The gamma test is a non-parametric approach and is based on the minimization of mean square error in the modeling (Agalbjorn et al 1997) and selects the combination of the most appropriate input variables from the original dataset to model the output with greater accuracy. The limitation of the GT is the input and output variables should be continuous to form a smooth system. The test can be achieved using win Gamma software and is efficient in evaluating the error present in the dataset. The combination of the most appropriate input variables is selected based on the obtained minimum value of Gamma static and V_{ratio}. The value of V_{ratio} is between 0 and 1, when its value is closer to 0 means the forecast ability of input combinations is high (achieved smooth system). Mathematically it can be elaborated as:

The estimate for Gamma static (Γ) is as follows:

$$y = A\delta + \Gamma \qquad \ldots(2.1)$$

Where, y= output variable, A=Gradient, and Γ = Intercept (when $\delta = 0$)

$$V_{ratio} = \frac{\Gamma}{\sigma^2} \qquad \ldots(2.2)$$

Where, σ^2 is the variance of output.

Multivariate adaptive regression splines (MARS): The Multivariate adaptive regression splines (MARS) model is the statistical technique and is a nonlinear and nonparametric approach that considers the separation of the datasets into splines. The MARS model is used for high dimensional analysis and non-linear multiple variable modeling. It can identify the nonlinear pattern with greater variables in the datasets. The MARS model was used as a regression method for high dimensional data and produced the appropriate results. It is based on the forward pass and backward pass approach. Using forward pass in the model results in potential knots to enhance the accuracy and better performance of the model. The backward pass is used to prune the modeling by eliminating the basis functions that do not significantly contribute to the best fit of the model.

The two parameters are the maximum number of basis functions and the maximum interaction level considered in the algorithm. Some other kinds of parameters are observed which can affect the generalized cross-validation (GCV), type of cross-validation, and degree of optimization in the modeling. The flexibility of the model was enhanced while selecting the adaptive knot in the data-fitting procedure. The appropriate knot selection method is the minimization of the least square method. The model generally constructs a pair of matrices for model fit representation in the analysis. The advantage of the MARS model to co-operate with the huge variety of predictors naturally includes continuous and categorical variables. The criteria for the selection of the model is that it evaluates the corresponding future lack-of-fit on the dataset.

The general form of the MARS model is as:

$$y = \beta_o + \sum_{i=1}^{m} \beta_i Y_i(x) \ldots (2.3)$$

Where, x = function of explanatory variables and their interactions, m= number of spline functions, Y_i (x) = basis function, β_i = set of coefficients and β_0= constant

The GCV method was developed in 1979 by spline pioneer Grace Wahba. In the MARS technique, the optimum model is selected based on the lowest value of GCV as well in the analysis and it does not correspond to the cross-validation. The GCV is a less expansive approach used to compare the subsets of the model. The GCV for the v^{th} model can be evaluated using the following formula:

$$GCV = \frac{1}{n} \frac{\sum_{i=1}^{n} [y_i - f_v(x_i)]^2}{(1 - \frac{c_v}{n})^2} \ldots (2.4)$$

Where, $f_v(x_i)$= evaluated model in v^{th} step of the backward stage, C_v= cost-complexity = (1+ϕ) b+1, Φ= smoothing parameters value varies between 2-4,b= number of basis functions

Random forest (RF): Random Forest (RF) is a nonparametric approach in which explanatory variables may be either continuous or categorical (Leo Breiman in 2001). It is used for both regression and classification. The basic difference between the regression and classification in the model is that the correlation enhances with the number of features. In the regression approach of the RF model, the output is numerically identified and the training set is derived by the selection of random distribution of input and output features. It is an unsupervised learning model and helps to minimize the generalization error and detect the outliers in the analysis and is a tree-based approach and each tree depends on the selection of variables randomly from the dataset and the trees are selected in the model based on the binary recursive partitioning.

The classification and regression tree (CART) based approach uses the binary tree data algorithm for modeling and further the decision tree approach is improved by increasing the number of trees for better generalization of model (Breiman 2001. The selection of the variables is based on the minimum residual sum of the square error of the explanatory variables for the section of the split. The RF model contained a collection of tree predictors (Yu et al 2017). The out-of-bag (OOB) estimates are defined as the rate of error of the out-of-bag classifier used in the training set. The bagging concept is used in the model to increase the accuracy when features are randomly selected and to evaluate the generalization error, strength, and correlation in the model. For complex problems, this model is used to identify the analysis by making trees. The number of trees (n tree) and the number of variables on each node (mtry) are important terms that can easily affect the accuracy and performance of the model. There are some parameters like the total number of explanatory variables that are selected randomly, the total number of trees in the model, and the size of trees that are required to be tuned to improve the accuracy of the model for its better performance. The error in regression is defined as follows:

$$MSE_{oob} = \frac{1}{N} \sum_{i=1}^{N} y_i - f_{oob}\,(x_i)^2 \ldots (2.5)$$

Where, $f_{oob}\,(x_i)$= OOB forecasting for i^{th} observations.

Performance evaluation of models: The performance evaluation of the models is based on the graphical comparison between the observed and estimated evaporation. The quantitative performance evaluation is based on statistical parameters such as root mean square error (RMSE), correlation coefficient (r), percent bias (PBIAS), and Nash-Sutcliffe efficiency (NSE).

Root Mean Square Error (RMSE): It is defined as the square root of the mean of the squared difference between the observed and estimated values. The lower value elaborates on the evaluation accuracy of the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (E_{oi} - E_{pi})^2}{n}} \ldots (2.6)$$

Correlation coefficient (r): The correlation coefficient evaluates the variance and proximity between the observed and estimated values. It varies between -1 and 1 and zero value represents that there is no relationship between observed and estimated values.

$$r = \frac{\sum_{i=1}^{n}(E_{oi} - \overline{E_{oi}})(E_{pi} - \overline{E_{pi}})}{\sqrt{\sum_{i=1}^{n}(E_{oi} - \overline{E_{oi}})^2}\sqrt{\sum_{i=1}^{n}(E_{pi} - \overline{E_{pi}})^2}} \quad ...(2.7)$$

Percent Bias (PBIAS): The PBIAS evaluates the average tendency of estimated values to be larger or smaller than the observed values. The desirable value of PBIAS is zero. The model with low-magnitude values indicates the modeling accuracy. The positive and negative values of PBIAS indicate the over-evaluation bias and under-evaluation bias.

$$PBIAS = \frac{\sum_{i=1}^{n}(E_{pi} - E_{oi})}{\sum_{i=1}^{n}(E_{oi})} \times 100 \quad ...(2.8)$$

Nash Sutcliffe efficiency (NSE): It was introduced by Nash and Sutcliffe (1970) and value varies between –∞ to 1. The NSE was used to define the accuracy and precision of the model. Its value near 1 represents the perfectness of the model and sensitivity to the outliers present in the dataset. It results in the variance of the observation for models and is used widely in the field of hydrology.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(E_{oi} - E_{pi})}{\sum_{i=1}^{n}(E_{oi} - \overline{E_{oi}})^2} \quad ...(2.9)$$

Where E_{oi}= i^{th} observed evaporation, E_{pi} = i^{th} predicted evaporation, E_{oi} = mean of i^{th} observed evaporation, E_{pi} = mean of i^{th} predicted evaporation and n= total observations.

RESULTS AND DISCUSSION

Input variables selection using GT: The selection of a combination of appropriate input variables for evaporation estimation is a complex process when the data is non-linear. In the present study, combinations of appropriate input variables were selected using the Gamma test to map mean monthly evaporation at Ranichauri for MARS and RF models. The combination of best input variables was selected with the minimum value of gamma (0.00095), gradient (0.10541), standard error (SE) (0.00700), and V_{ratio}(0.00382). Therefore, the combination of appropriate input of nine variables as maximum temperature (T_{max}), minimum temperature (T_{min}), rainfall (Rf), morning relative humidity (RH_m), morning vapor pressure (Vp_m), wind speed (W_s), morning wind direction (Wd_m), afternoon wind direction (Wd_a), and solar radiation (SR) was used for the MARS and RF models to estimate mean monthly evaporation for Ranichauri. The maximum mean monthly evaporation was found 5.65 mm and the minimum mean monthly evaporation was 0.93 mm (Table 1 and Fig. 1). The GT performed on a different set of input variables and parameters estimated for Ranichauri (Table 2).

Estimation of mean monthly evaporation using MARS and RF Models at Ranichauri: According to the GT results, the best combination of input variables was selected as T_{max}, T_{min}, Rf, RH_m, Vp_m, W_s, Wd_m, Wd_a, and SR with a minimum value of Gamma and V_{ratio} which indicates that these variables have a significant effect on evaporation. The mean monthly evaporation for the study area was estimated using MARS and RF models based on historically available data. The data is divided into training (75%) and testing periods (25%) for the validation of models. Twelve meteorological input variables were used to estimate daily pan evaporation of summer maize using MARS model with empirical Shuttleworth-Wallace, Priestley-Taylor, back-propagation neural networks and, Two-Patch models (Shan et al 2020). The values of statistical indices for the MARS model during training and testing periods are RMSE=0.32 and 0.17, r=0.97 and 0.98, PBIAS=0.00 and 0.00 (the value of PBIAS is desirable to 0, it means the estimation of mean monthly evaporation using MARS model is accurate), and NSE=0.93 and 0.96. Similarly, the values of statistical indices for the RF model during training and testing periods are RMSE=0.20 and 0.23, r=0.99 and 0.98, PBIAS=0.10 and 0.80, and NSE=0.97 and 0.90, respectively (Table 3). Thus, the performance of the MARS and RF models was different according to the statistical indices for the estimation of evaporation.

The scatter plot between observed and estimated values of mean monthly evaporation for the MARS and RF models are plotted during the training and testing period (Figs. 2, 4, 6

Table 1. Basic statistics of the selected input variables

Statistical parameter	Input variables (2008-2017)									
	T_{max} (℃)	T_{min} (℃)	Rf (mm)	RH_m (%)	Vp_m (mm Hg)	W_s (Km/h)	Wd_m (◦)	Wd_a (◦)	SR (h)	E (mm)
Maximum	29.06	17.8	559.5	97.74	16.34	12.56	198.4	220.6	10.15	5.65
Minimum	9.3	-0.55	0	45.1	4.22	1.28	2.58	4.90	2.38	0.93
Mean	19.90	10.02	83.87	78.7	9.71	4.01	62.4	102.6	6.612	2.53
Median	21.19	10.00	28.1	78.26	8.78	4.14	29.37	80.55	6.89	2.25
Variance	20.51	28.18	1680	158.0	16.97	1.98	3532	7269	3.35	1.34
S. D.	4.52	5.31	129.6	12.57	4.12	1.41	59.43	85.26	1.83	1.16

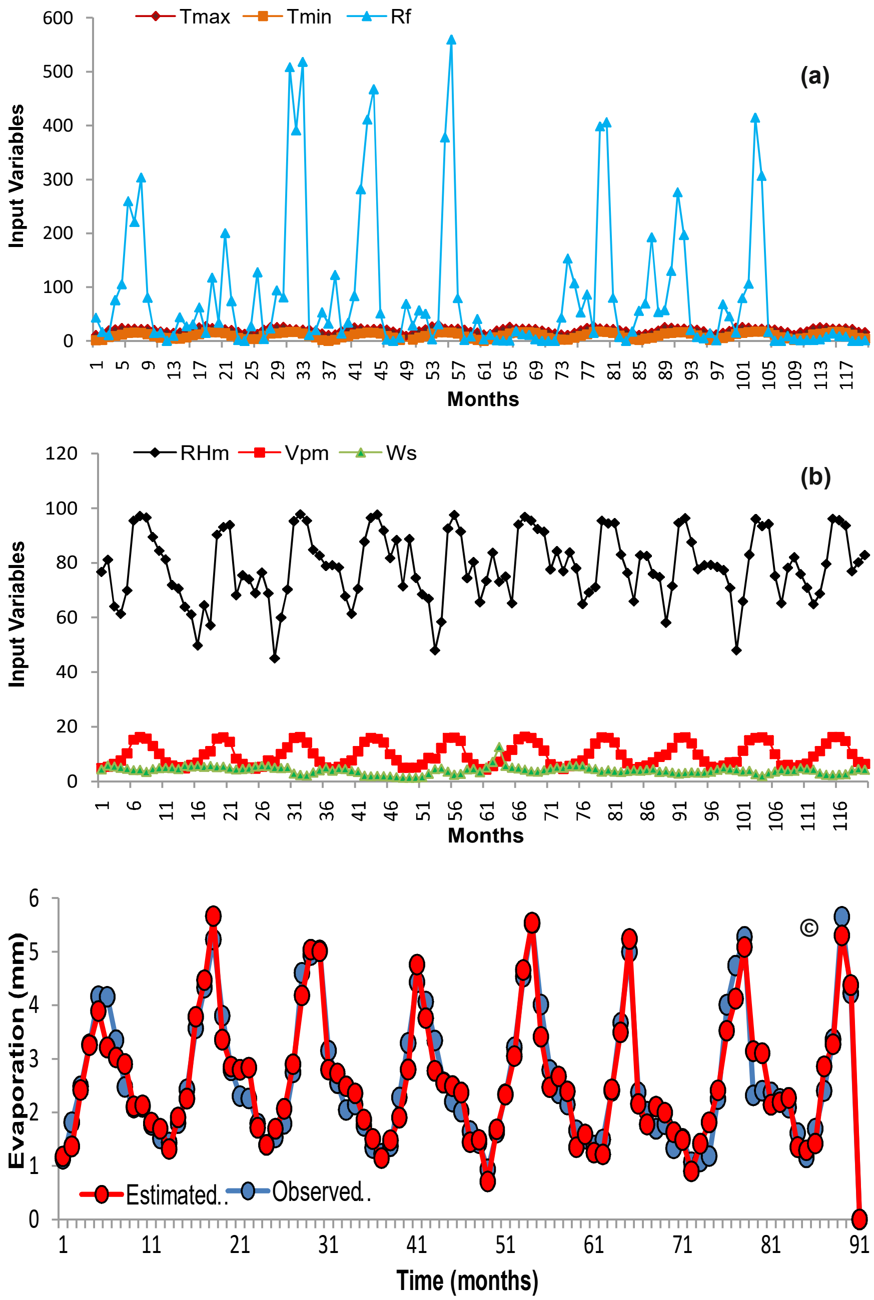

Fig. 1. (a), (b), (c). Time series plot of monthly input variables at Ranichauri

Table 2. Gamma test for input variable selection of Ranichauri

Output	Input combination	Mask	Gamma	Gradient	SE	V_{ratio}
E	T_{max}, T_{min}, Rf, RH_m, RH_a, Vp_m,Vp_a, W_s, Wd_m,Wd_a, SR	11111111111	0.001	0.091	0.008	0.006
E	T_{max}, T_{min}, Rf, RH_m,RH_a, Wd_m,Wd_a, SR	11111000111	0.011	0.011	0.009	0.045
E	T_{max}, RH_m, RH_a, Vp_m	10011100000	0.032	0.021	0.004	0.131
E	T_{max}, Rf, RH_a, W_s	10101001000	0.036	0.068	0.004	0.144
E	T_{min}, Rf, Vp_a,Wd_m,Wd_a, SR	01100010111	0.010	0.249	0.009	0.040
E	Vp_m, Wd_a	00000100010	0.145	-0.267	0.016	0.583
E	T_{max}, T_{min}, Vp_m,Vp_a,W_s, Wd_m,Wd_a	11000111110	0.006	0.227	0.007	0.025
E	T_{max}, Vp_m,Wd_m	10000100100	0.027	0.583	0.005	0.108
E	T_{max}, T_{min}, Rf, RH_m, Vp_m,W_s, Wd_m,Wd_a, SR	11110101111	0.001	0.105	0.007	0.003
E	T_{max}, T_{min}, Rf, RH_m, Vp_m,W_s, Wd_m	11110101100	0.014	0.093	0.004	0.057
E	T_{max}, T_{min}, Rf, Vp_m,Vp_a,W_s, Wd_m,Wd_a, SR	11100111110	0.008	0.119	0.006	0.035

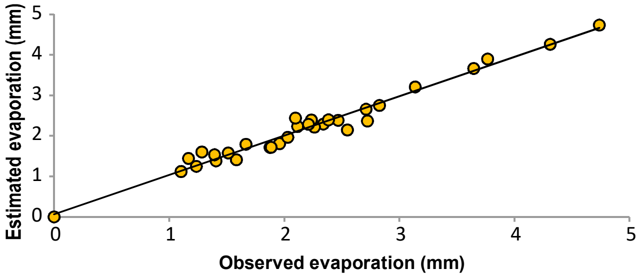

Fig. 2. Scatter plot between observed and estimated mean monthly evaporation using the MARS model for the training period

Table 3. Statistical indices for MARS and RF Models

Statistical indices	Training period		Testing period	
	MARS	RF	MARS	RF
RMSE	0.32	0.20	0.17	0.23
r	0.97	0.99	0.98	0.98
PBIAS	0.00	0.10	0.00	0.80
NSE	0.93	0.97	0.96	0.90

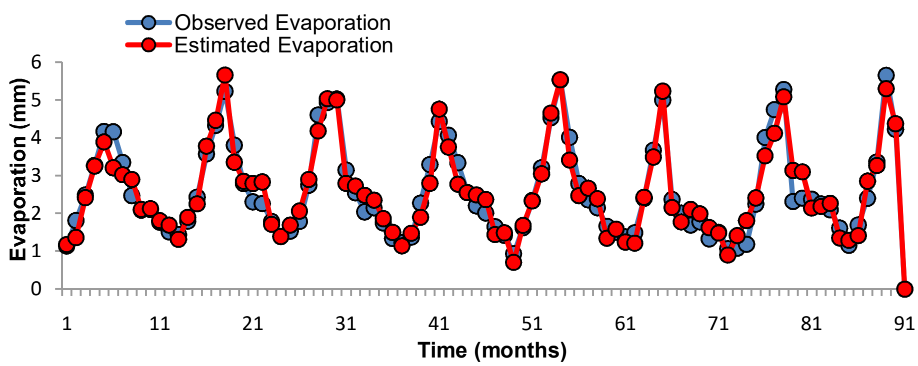

Fig. 3. Temporal variation between observed and estimated mean monthly evaporation by MARS model for the training period

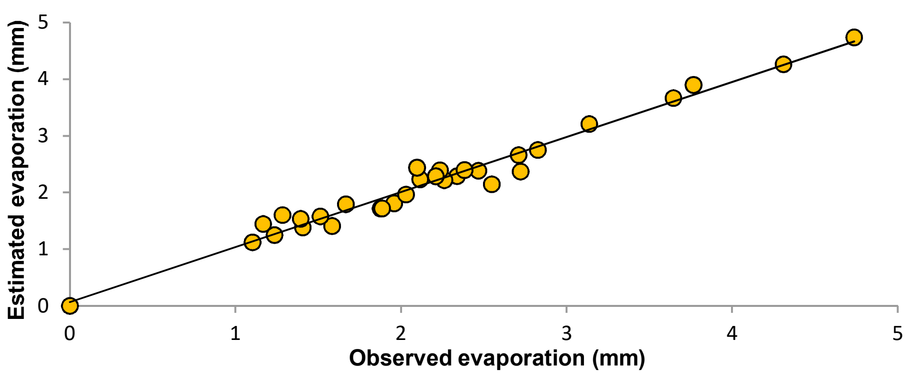

Fig. 4. Scatter plot between observed and estimated mean monthly evaporation using MARS model for the testing period

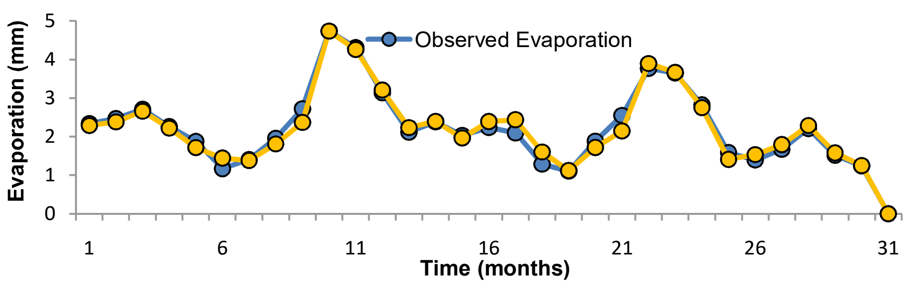

Fig. 5. Temporal variation between observed and estimated mean monthly evaporation by MARS model for the testing period

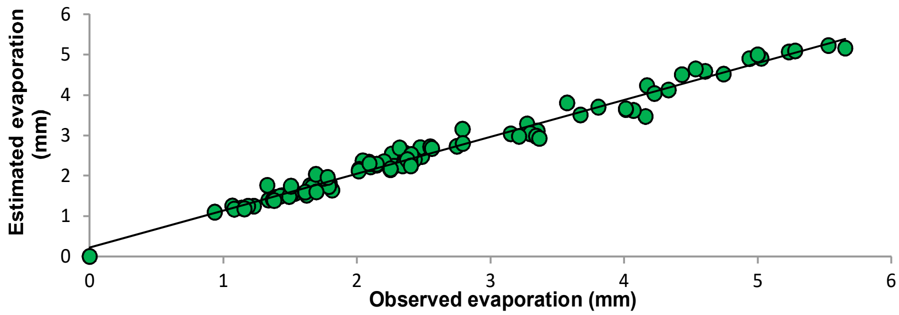

Fig. 6. Scatter plot between observed and estimated mean monthly evaporation using RF model for the training period

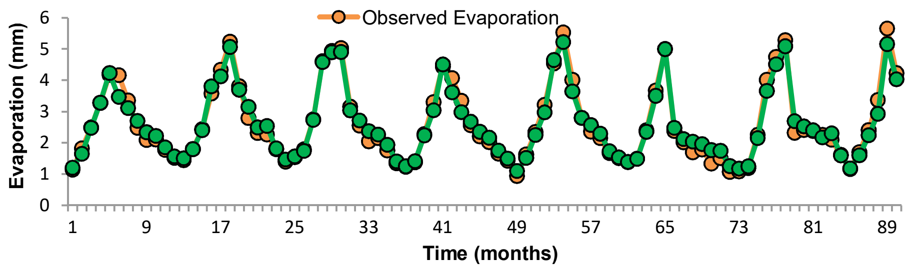

Fig. 7. Temporal variation between observed and estimated mean monthly evaporation by RF model for the training period

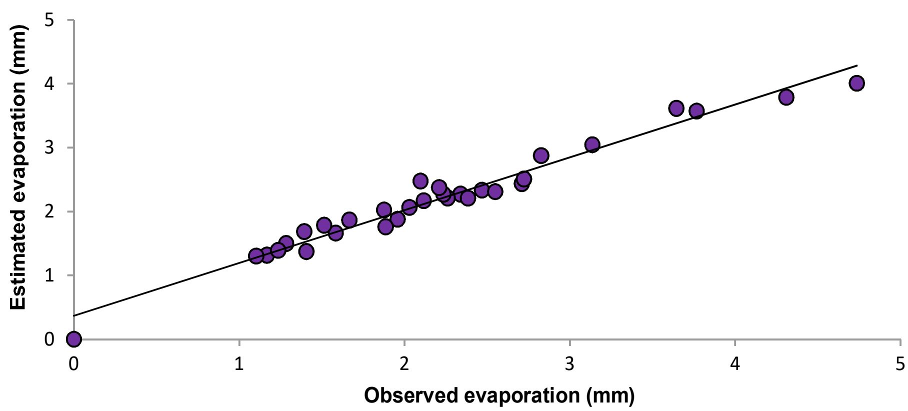

Fig. 8. Scatter plot between observed and estimated mean monthly evaporation using RF model for the testing period

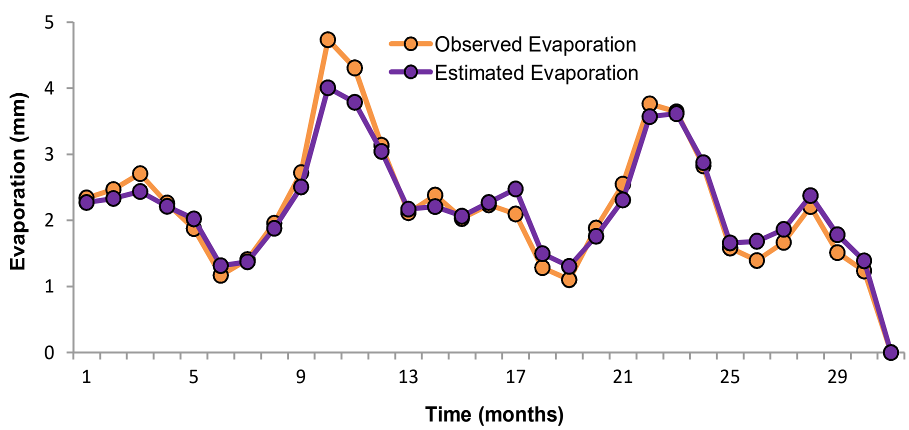

Fig. 9. Temporal variation between observed and estimated mean monthly evaporation by RF model for the testing period

and 8, respectively). The temporal variation between observed and estimated mean monthly evaporation by MARS and RF model for the training and testing period are plotted (Figs. 4, 5, 8, 7 and 9, respective). The observed and estimated values of mean monthly evaporation match closely for the RF model in the training period and the MARS model in the testing period. The value of R^2 and RMSE of approximately (0.950 and 0.910) for training phase and for testing phase (0.62 and 0.910) respectively using RF and MARS models which shows model simulation accuracy for evaporation (Wang et al 2023). The performance of the models with respect to the statistical parameters can be easily identified from Figure 11 for the training set and Figure 12 for the testing set.

Comparison of statistical indices (RMSE, r, and NSE) indicates that the MARS model performs better during the testing period and the RF model performs better during the training period. The models perform with better accuracy concerning RMSE, r, and NSE. According to the given values of PBIAS, the MARS model performs better for both training and testing periods as compared to the RF model. Furthermore, the comparison between our models and the MLR, MNLSR, MARS RF, CAT Boost, M5Tree proposed by Vishwakarma et al (2024) for estimation of evaporation revealed significant findings.

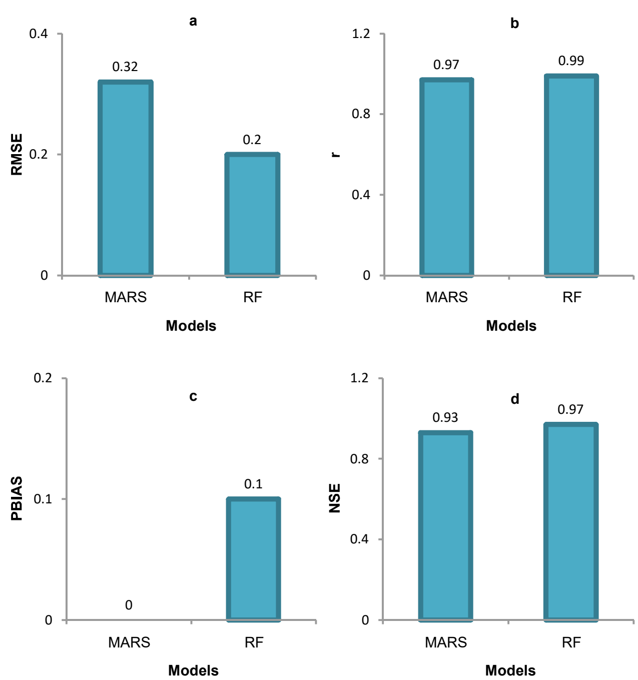

Fig. 10. Comparison of different performance measures between MARS and RF models for training set: (a) RMSEvs. models, (b) rvs. models, (c) PBIASvs. models, (d) NSEvs. models

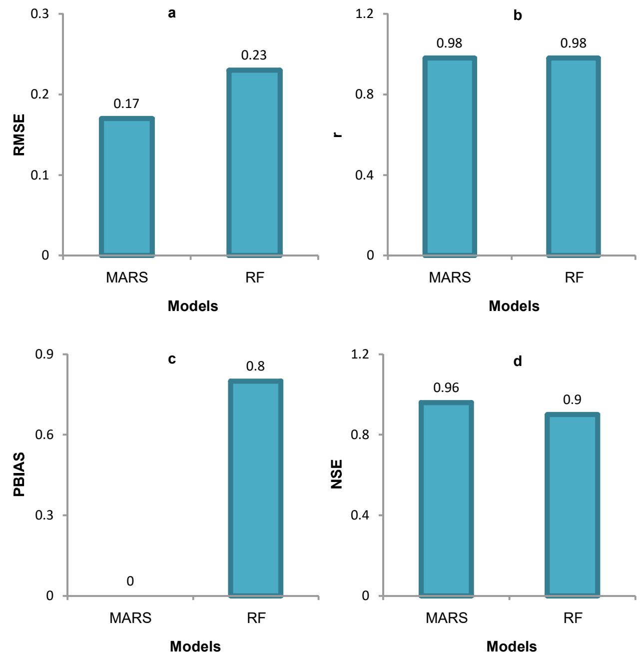

Fig. 11. Comparison of different performance measures between MARS and RF models for testing set: (a) RMSEvs. models, (b) rvs. models, (c) PBIASvs. models, (d) NSEvs. models

CONCLUSIONS

This study evaluates the performance of the machine learning models (MARS, and RF) for the estimation of mean monthly evaporation at Ranichauri. The combination of appropriate input variables was selected using the Gamma test to map evaporation for MARS and RF models. The combination of nine input variables as T_{max}, T_{min}, Rf, RH_m, Vp_m, W_s, Wd_m, Wd_a, and SR based on GT was identified as the most significant for mean monthly evaporation estimation of the study area. The performance of the RF model was found to be better than the MARS model in the training period and the MARS model was found to be better than the RF model in the testing period for the estimation of mean monthly evaporation at Ranichauri. So the performance of both the models was found to be significant for mean monthly evaporation estimation during the study period.

ACKNOWLEDGEMENT

The authors thank AMFU, V.C.S.G. Uttarakhand University of Horticulture and Forestry, Ranichauri, Tehri Garhwal district, Uttarakhand for providing support to collect the dataset to make the present study possible.

REFERENCES

Agalbjörn S, Koncar N and Jones AJ 1997. A note on the gamma test. *Neural Computing and Applications* **5**(3): 131-133.

Cramer S, Kampouridis M, Freitas A and Alexandridis AK 2017. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications* **85**: 169-181.

Davie T 2008. *Fundamentals of hydrology, Routledge – Taylor & Francis Group.*

Dodla VB, Satyanarayana GC and Desamsetti S 2017. Analysis and prediction of a catastrophic Indian coastal heatwave of 2015. *Natural Hazards* **87**: 395-414.

Dumka B, Kashyap PS and Saran B 2018. Modelling reference evapotranspiration of Pantnagar using various training functions in artificial neural network, *Indian Journal of Eco*logy **45**(1): 19-24.

Dumka BB and Kumar P 2021. Modelling rainfall-runoff using Artificial neural networks (ANNs) and wavelet based ANNs (WANNs) for Haripura dam, Uttarakhand, *Indian Journal of Ecology* **48**(1): 271-274

Eslamian S, Gohari SA, Biabanaki M and Malekian R 2008. Estimation of monthly pan evaporation using artificial neural networks and support vector machines. *Journal of Applied Sciences* **8**(19): 3497-3502.

Friedman JH 1991. Multivariate adaptive regression splines. *Annals of Statistics* **19**(1): 1-67.

Ghaemi A, Balf Adamowski J, Kisi O and Quilty J 2019. On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. *Agricultural and Forest Meteorology*. **278**: 107647.

Goyal MK and Ojha CSP 2011. Estimation of scour downstream of a ski-jump bucket using support vector and M5 model tree. *Water Resources Management* **25**(9): 2177-2195.

Khan N, Shahid S, Ahmed K, Ismail T, Nawaz N and Son M 2018. Performance assessment of the general circulation model in simulating daily precipitation and temperature using multiple gridded datasets. *Water* **10**: 1793.

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.

Lu X, Ju Y, Wu L, Fan J, Zhang F and Li Z 2018. Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models. *Journal of Hydrology* **566**: 668-684.

Malik A and Kumar A 2015**.** Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. *Water Resources Management* **29**(6): 1859-1872.

Markuna S, Kumar P, Ali R, Vishwakarma ADK, Kushwaha KS, Kumar R, Singh VK, Chaudhary S and Kuriqi A 2023**.** Application of innovative machine learning techniques for long-term rainfall prediction. *Pure and Applied Geophysics* **180**: 335-363.

Received 11 March, 2024; Accepted 30 June, 2024

Nourani V and Fard MS 2012**.** Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Advances in Engineering Software* **47**: 127-146.

Perez EH, Levresse G, Hernández JC and Martínezc RG 2020**.** Short term evaporation estimation in a natural semiarid environment: Newperspective of the Craig-Gordon isotopic model. *Journal of Hydrology* **587**(2020) 124926.

Pham BT, Le LM, Le T, Bui KT, Le VM, Ly HB and Prakash I 2020**.** Development of advanced artificial intelligence models for daily rainfall prediction. *Atmospheric Research* **237**: 104845.

Rakhee Singh A, Mittal M and Kumar A 2020. Qualitative analysis of random forests for evaporation prediction in Indian Regions. *Indian Journal of Agricultural Sciences* **90**(6): 1140-1144.

Ribot JC, Magalhães AR and Panagides S 2005**.** *Climate Variability, Climate Change and Social Vulnerability in the Semi-arid Tropics.* Cambridge University Press.

Seifi A and Soroush F 2020**.** Pan evaporation estimation and derivation of explicit optimized equations by novel hybrid meta-heuristic ANN based methods in different climates of Iran. Computers and Electronics in Agriculture**.**

Shan X, Cui N, Cai H, Hu X and Zhao L 2020**.** Estimation of summer maize evapotranspiration using MARS model in the semi-arid region of northwest China. *Computers and Electronics in Agriculture* **174**: 105495.

Shirsath PB and Singh AK 2010**.** A comparative study of daily pan evaporation estimation using ANN, regression and climate-based models. *Water Resources Management* **24**(8): 1571-1581.

Sudheer KP, Gosain AK, Rangan DM and Saheb SM 2002. Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes* **16**: 3189-3202.

Szilagyi J and Jozsa J 2009. Analytical solution of the coupled 2-D turbulent heat and vapor transport equations and the complementary relationship of evaporation. *Journal of Hydrology* **37**(1-4): 61-67.

Vishwakarma DK, Kumar P, Yadav K, Ali R, Markuna S, Chauhan S, Heddam S, Kuriqi A, Srivastava A, Alam M and Vinayak V 2024. Evaluation of CatBoost Method for Predicting Weekly Pan Evaporation in Subtropical and Sub-Humid Regions. *Pure and Applied Geophysics*.

Wang H, Sun F, Liu F, Wang T, Liu W and Feng Y 2023. Reconstruction of the pan evaporation based on meteorological factors with machine learning method over China. *Agricultural Water Management* **287**: 108416.

Yu PS, Yang TC, Chen SY, Kuo CM and Tseng HW 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology*, **552**: 92-104.