Modeling Earth Systems and Environment (2021) 7:495–502
https://doi.org/10.1007/s40808-020-00921-8

ORIGINAL ARTICLE

Modeling of reference evapotranspiration for temperate Kashmir Valley using linear regression

Syed Mohsin[1] · M. A. Lone[1]

Received: 16 June 2020 / Accepted: 29 July 2020 / Published online: 7 August 2020

Abstract

Reference evapotranspiration (ET_O) estimation is a prerequisite for the estimation of evapotranspiration rates of different crops. There is global consensus on the consistency of the FAO-56 Penman–Monteith (FAO-56 PM) method, and it has lately become the first-choice for the estimation of ET_O, but despite its consistent performance, other methods continue to be popular for their simple structure and data requirement, more so in field practice. In this study, regression models were developed for predicting monthly ET_O using monthly averages of minimum temperature (T_{min}), relative humidity (RH), wind speed (WS), and sunshine hours (SSH) for three stations in the temperate Kashmir Valley. The models are based on the ET_O data generated using the FAO-56 PM method. The weather data required for the analysis were for a period of 20 years. The models were evaluated using the coefficient of determination (R^2) and residual analysis. R^2 values for the developed models were greater than 0.96. In addition to the ET_O estimation, the developed models may also be used for trend evaluation in the ET_O data in climatologically similar regions. In this study, for the cases of very limited data availability where the FAO-56 PM equation cannot be used, the Hargreaves and Samani (HAR) equation, which only uses temperature as an input, was calibrated for each station using linear regression between FAO-56 PM ET_O and HAR ET_O.

Keywords FAO-56 penman–monteith equation · Hargreaves and samani · Linear regression · Reference evapotranspiration · Temperate

Introduction

In the very simplest of terms, evapotranspiration (ET) could be defined as the amount of water lost to the surrounding environment of a crop through transpiration done by leaves and evaporation from the soil surface. From an agricultural and hydrological point of view, ET estimation is at the heart of many operations. From irrigation water allocation and subsequent design of conveyance systems to regional and global water balance studies, ET finds its relevance.

Different crops have different ET characteristics, which also vary spatially and temporally. For instance, a paddy crop of a particular variety, growing at a certain location, requires varying amounts of water at different growth stages. The same variety of paddy may demand a different irrigation regime at a different location. Therefore, a single ET model may not be appropriate for the very same crop at two geographical locations. To address the complexity arising from such situations, the idea of a reference surface was introduced. The ET rate of a crop is related to the reference surface evapotranspiration or reference evapotranspiration (ET_O) by a multiplicative factor known as the crop coefficient. FAO-56 (Allen et al. 1998) puts forth the definition of a reference surface as "a hypothetical reference crop with an assumed crop height of 0.12 m, and a fixed surface resistance of 70 sm^{-1} and an albedo of 0.23.

Prior to the consultation of experts in 1990, the Food and Agriculture Organization (FAO) endorsed the use of Blaney–Criddle, radiation, modified Penman, and pan evaporation methods. However, a performance evaluation study of the above mentioned and some more estimation procedures carried out by the committee on the Irrigation Water Requirement of the American Society of Civil Engineers (ASCE) led to the understanding that the methods showed variable performance in varying climates. This led to the recommendation for the adoption of the Penman–Monteith combination method as the new standard method for ET_O

✉ Syed Mohsin
mohsin_29phd17@nitsri.net

[1] Department of Civil Engineering, National Institute of Technology, Srinagar, India

estimation, and it was called the FAO-56 Penman–Monteith (FAO-56 PM) equation. It was seconded by multiple subsequent studies evaluating the performance of the FAO-56 PM equation (Kashyap and Panda 2001; Irmak et al. 2003a, b; Garcia et al. 2004; Temesgen et al. 2005; Allen et al. 2005, 2006; Jabloun and Sahli 2008). The equation is given by:

$$ET_o = \frac{0.408\Delta\left(R_n - G\right) + \gamma\frac{900}{T+273}U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)}, \quad (1)$$

where R_n is the net radiation available at the crop surface [MJ m^{-2} day^{-1}], G is the soil heat flux density [MJ m^{-2} day^{-1}], T [°C]and U_2 [m s^{-1}] are the mean air temperature and wind speed measured at 2 m from the ground surface, respectively, e_s and e_a are the saturation and actual vapour pressure, respectively, in kPa, e_s-e_a is the saturation vapour pressure deficit [kPa], Δ the slope of vapour pressure curve [kPa °C^{-1}] and γ is the psychrometric constant [kPa °C^{-1}].

It has been observed very commonly how planners tend to rely on simpler temperature-based equations for calculation of ET (Xu and Singh 2001). In the process of doing so, the resulting ET estimates can be fairly biased because of ET being sensitive to a number of variables. In such circumstances, linear regression equations could be used as alternatives owing to their structural simplicity. Linear regression based models have been consistently used to estimate ET_O (Irmak et al. 2003b; Kovoor and Nandagiri 2007; Cristea et al. 2013; Yirga 2019). In this paper, we analyse the relationship between meteorological variables and ET_O on a monthly time-scale in the temperate Kashmir Valley.

Materials and methods

Study area and meteorological data

The study area for this analysis is the Valley of Kashmir. With a mean altitude of 1545 m, the Valley is situated between 32° 22′ and 34° 43′ N latitude and 73° 52′ and 75° 42′ E longitude (Fig. 1). The Valley, bound by the Pir-Panjal range and the greater Himalayas, is famous world over for its snow-clad mountains, pristine lakes, and gushing streams. There are four marked seasons in the Valley; Spring, Summer, Autumn, and Winter. The climate of the Valley puts it in the *dfb* category in the Köppen climate classification.

The meteorological data required for the study corresponding to the Srinagar, Qazigund, and Kupwara stations were acquired from the National Data Centre (NDC), Pune. The dataset comprised of monthly values of maximum temperature (T_{max}), minimum temperature (T_{min}), Wind Speed (WS), Relative humidity (RH), and sunshine hours (SSH) for

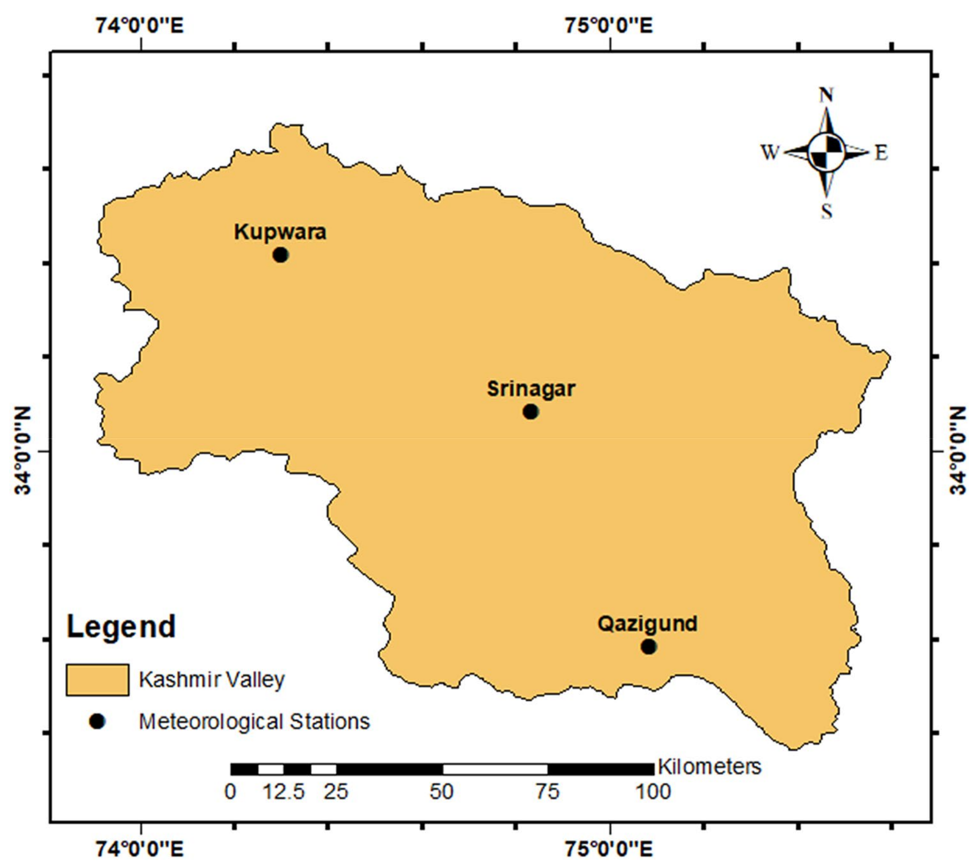

Fig. 1 Meteorological stations considered for the study

a period of 20 years (2000–2019). Missing data, if any, were estimated by following the procedure laid out in FAO-56 (Allen et al. 1998). Table 1 presents the location and primary information about the average values of the meteorological data.

Reference evapotranspiration estimation

The FAO-56 PM equation (Allen et al. 1998) was used for calculating ET_O for the Srinagar, Qazigund, and Kupwara stations for a period of 20 years using DSS_ET, which is a decision support system for ET estimation (Bandyopadhyay et al. 2012). DSS_ET is a very handy tool for ET calculations using a variety of models for making the calculations. It also has features for estimating missing data, visualization of the results, and primary statistical analysis for performance evaluation of ET_O models, which makes it a very useful tool for researchers.

Multiple linear regression

A generic equation for multiple linear regression (MLR) can be expressed as

$$Y = \propto_0 + \mu_1 X_1 + \mu_2 X_2 + \ldots + \mu_n X_n + \in, \quad (2)$$

where Y is the dependent variable, $\propto_0$ is the intercept, μ_1, $\mu_2 \ldots \mu_n$ are the coefficients of the multiple linear regression equation, X_1, X_2, X_n are the independent variables and $\in$ is the error term associated with the multiple linear regression equation. Multiple linear regression equation is essentially a minimization problem wherein the coefficients are estimated for the minimum sum of squared error terms.

Using the ET_O results based on the calculations made by the FAO-56 PM equation, multiple regression analysis was performed to develop linear models for the estimation of monthly ET_O values from the meteorological variables (T_{max}, T_{min}, WS, RH, and SSH) for each of the stations. The basic assumptions of multiple linear regression were checked, and the variables showing multicollinearity were removed. Moreover, HAR ET_O values were also calculated using DSS_ET, and the same were used for carrying out a linear regression with FAO-56 PM $ET_{O,}$ and linear models were developed for FAO-56 PM ET_O estimation in terms of HAR ET_O for each station.

Results and discussion

Spatiotemporal variation in the ET_O values in the study area

Figure 2 presents the overall variation of the average daily ET_O values for the stations under consideration calculated by the FAO-56 PM equation. The ET_O values peak between May and August with the mean maximum values for Srinagar, Kupwara, and Qazigund as 4.187, 4.112, and 3.920 mm/day. The overall annual mean ET_O values in the same order are 2.549, 2.409, and 2.424 mm/day.

Modeling ET_O using multiple linear regression

After elimination of highly correlated explanatory variables which induced the problem of singularity in the regression and subsequent recognition of linear relationships between the independent variables T_{min}, RH, WS, SSH, and the dependent variable ET_O (Fig. 3a–l), based on the linear regression analysis between explanatory variables and the independent variable, linear models for reference evapotranspiration, which passed the significance test at a p value of 0.05 were developed for the Srinagar, Kupwara and Qazigund stations. Brief results of the multiple linear regression are presented in Table 2. The units of measurement of all the variables are presented in Table 3. Figure 4a–c illustrates the

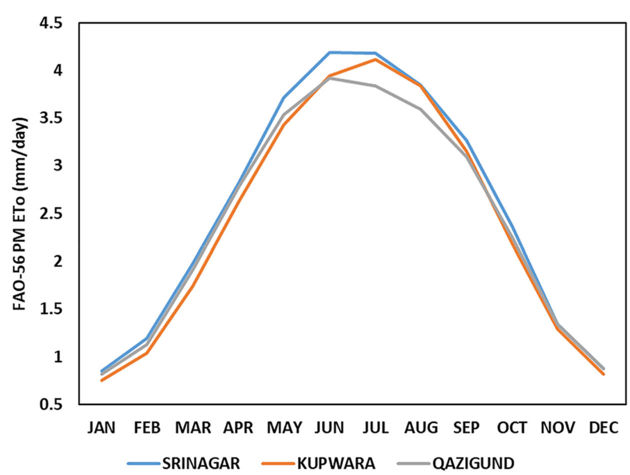

Fig. 2 Temporal variation of ET_O values for the stations

Table 1 Information regarding the location and the annual average values of the primary meteorological data

Station	Latitude (°N)	Longitude (°E)	Elevation (m)	T_{max}(°C)	T_{min}(°C)	*RH* (%)	*W* (Km/h)
Srinagar	34° 05′	74° 50′	1588	20.65	7.91	66.20	1.91
Kupwara	34° 25′	74° 18′	1609	20.87	6.64	75.69	0.84
Qazigund	33° 35′	75° 05′	1690	19.62	6.50	68.60	1.19

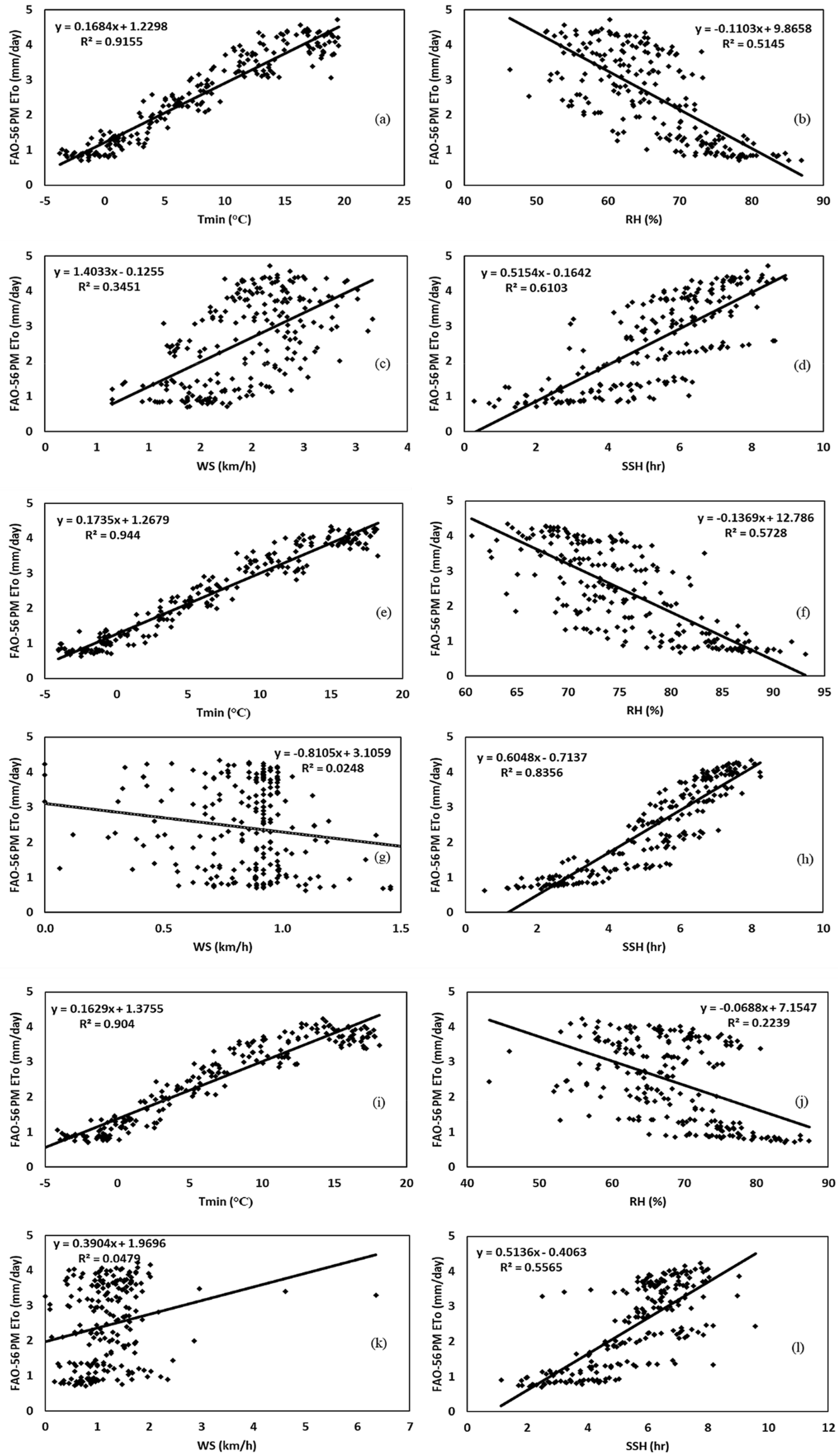

Fig. 3 Correlations of the FAO-56 PM ET_O and all other predictor variables for Srinagar (**a**–**d**), Kupwara (**e**–**h**) and Qazigund (**i**–**l**)

Table 2 Multiple linear regression models developed for the stations

Station	Variables	Coefficients	SEE	R^2
Srinagar	Intercept	1.660	0.215	0.971
	T_{min}	0.127		
	RH	– 0.017		
	WS	0.190		
	SSH	0.130		
Kupwara	Intercept	2.430	0.202	0.973
	T_{min}	0.134		
	RH	– 0.021		
	WS	0.152		
	SSH	0.113		
Qazigund	Intercept	6.194	0.229	0.962
	T_{min}	0.169		
	RH	– 0.057		
	WS	– 0.170		
	SSH	-0.135		

Table 3 Units of measurement of the variables in the developed regression models

Variable	Unit of measurement
Reference evapotranspiration (ET_O)	mm/day
Minimum temperature (T_{min})	°C
Relative humidity (RH)	%
Wind speed (WS)	km/h
Sunshine hours (SSH)	H

R^2 values obtained for the plots between predicted ET_O and actual ET_O. The normal probability plots (Fig. 5), which are approximately straight lines, imply an insignificant departure from the normal distribution. Moreover, the residual plots (Fig. 6a–l) did not show any signs of heteroscedasticity. As the variables are meteorological variables, some level of multicollinearity will always be there. Some studies suggest the upper bound of the variance inflation factor (VIF) as 5 to indicate significant collinearity while others fix a limit of 10 to indicate significant collinearity. Some even argue that a VIF of the order of 40 or even higher may not essentially be harmful to the model performance (O'Brien 2007). For this study, the VIFs were lower than 5, the highest being 3.73.

Station-wise calibration of the Hargreaves and Samani equation

The Hargreaves and Samani (HAR) equation (Hargreaves and Samani 1985) is a very simple equation for the determination of ET_O from records of minimum and maximum temperature. In data-scarce regions, where the temperature is often the only variable recorded and in places where meteorological data acquisition is wearying and uneconomical, this method is frequently put to use. In certain places, despite the availability of data, field practitioners are often tempted to rely on this method owing to the structural simplicity of the model. ET_O values calculated using the HAR equation when compared to the ET_O values calculated using the FAO-56 PM equation showed that the HAR equation significantly over-estimated the ET_O values for Srinagar (199 mm/year), Kupwara (266 mm/year) and Qazigund (209 mm/year). Keeping the same in mind, a linear regression between the two models was set up with HAR ET_O as the independent variable and the FAO-56 PM ET_o as the dependent variable. The results of the regression analysis are presented in Table 4.

Conclusions

Reference evapotranspiration (ET_O) values calculated using the FAO-56 PM method were used to carry out multiple linear regression to develop linear models for the estimation of ET_O for three stations in the Kashmir Valley. For all the stations, the strongest correlation of ET_O was found with minimum temperature (T_{min}) followed by sunshine hours (SSH), relative humidity (RH), and wind speed (WS). The analysis of variance of the residuals showed no signs of heteroscedasticity. Moreover, the normal probability plots were also satisfactory. All the models performed exceedingly well with $R^2 > 0.96$ for each model. The developed models are simple in their structure and do not mandate any complex intermediate calculations and can be therefore used effectively in field practice. For cases wherein the data measured are only limited to maximum and minimum temperature, the Hargreaves and Samani equation was calibrated for all the stations considering its overestimation of ET_O in comparison to the FAO-56 Penman–Monteith equation.

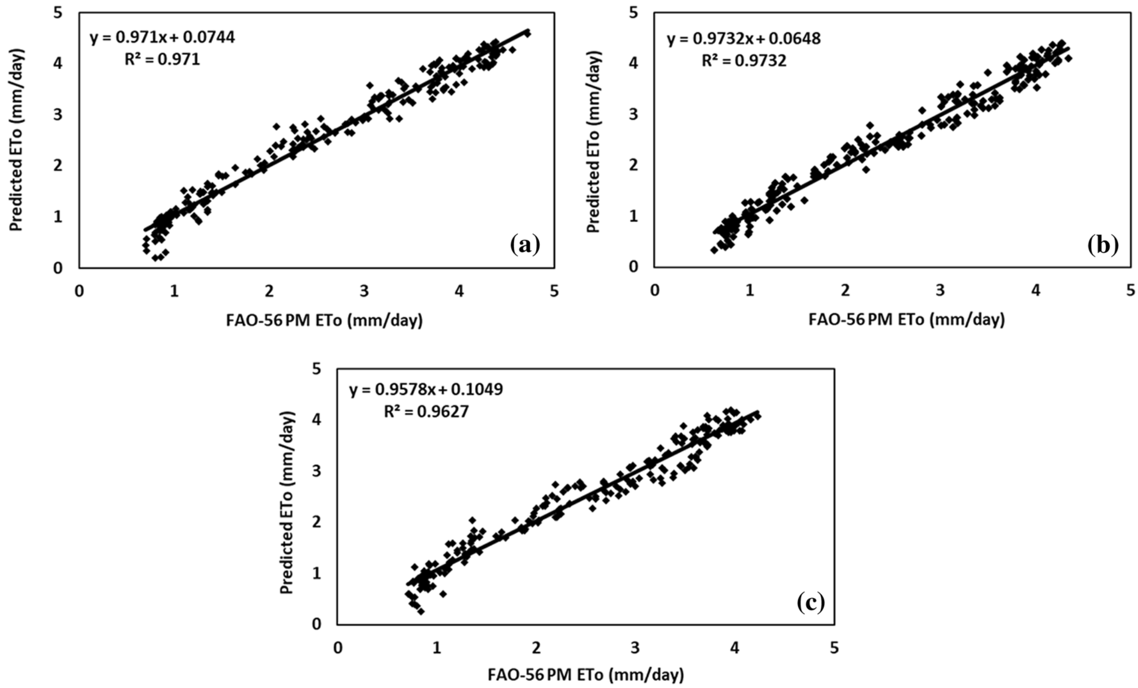

Fig. 4 Plots of FAO-56 PM ET_O and predicted ET_O using the developed linear models for the Srinagar (**a**), Kupwara (**b**) and Qazigund (**c**) stations, respectively

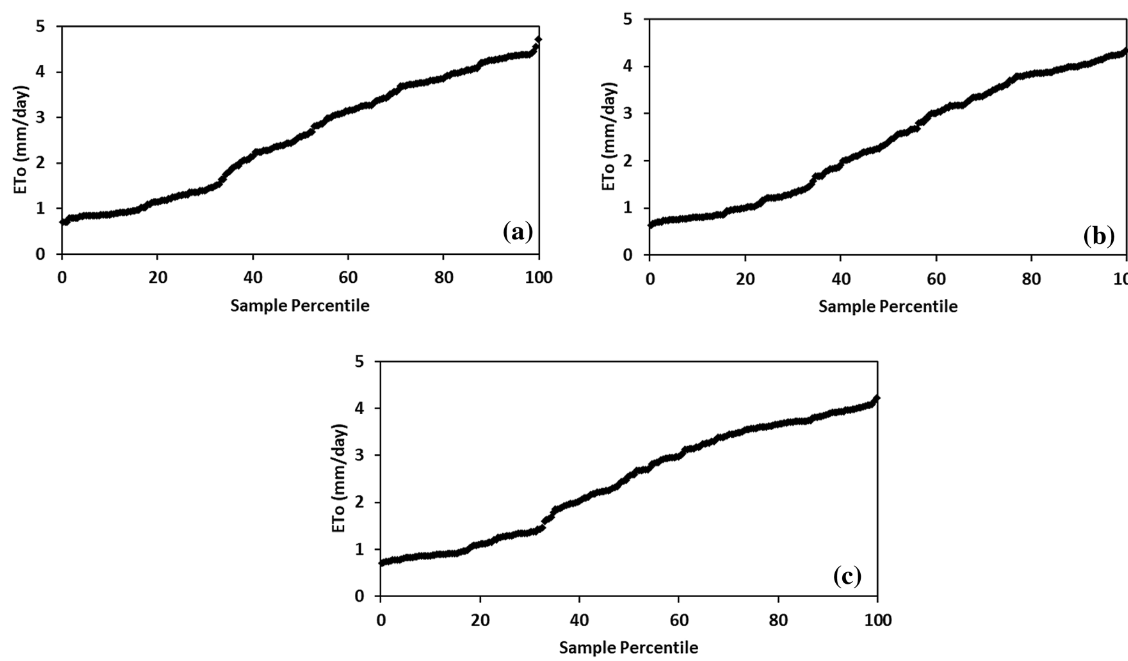

Fig. 5 Normal probability plots of the developed linear models for the Srinagar (**a**), Kupwara (**b**) and Qazigund (**c**) stations, respectively

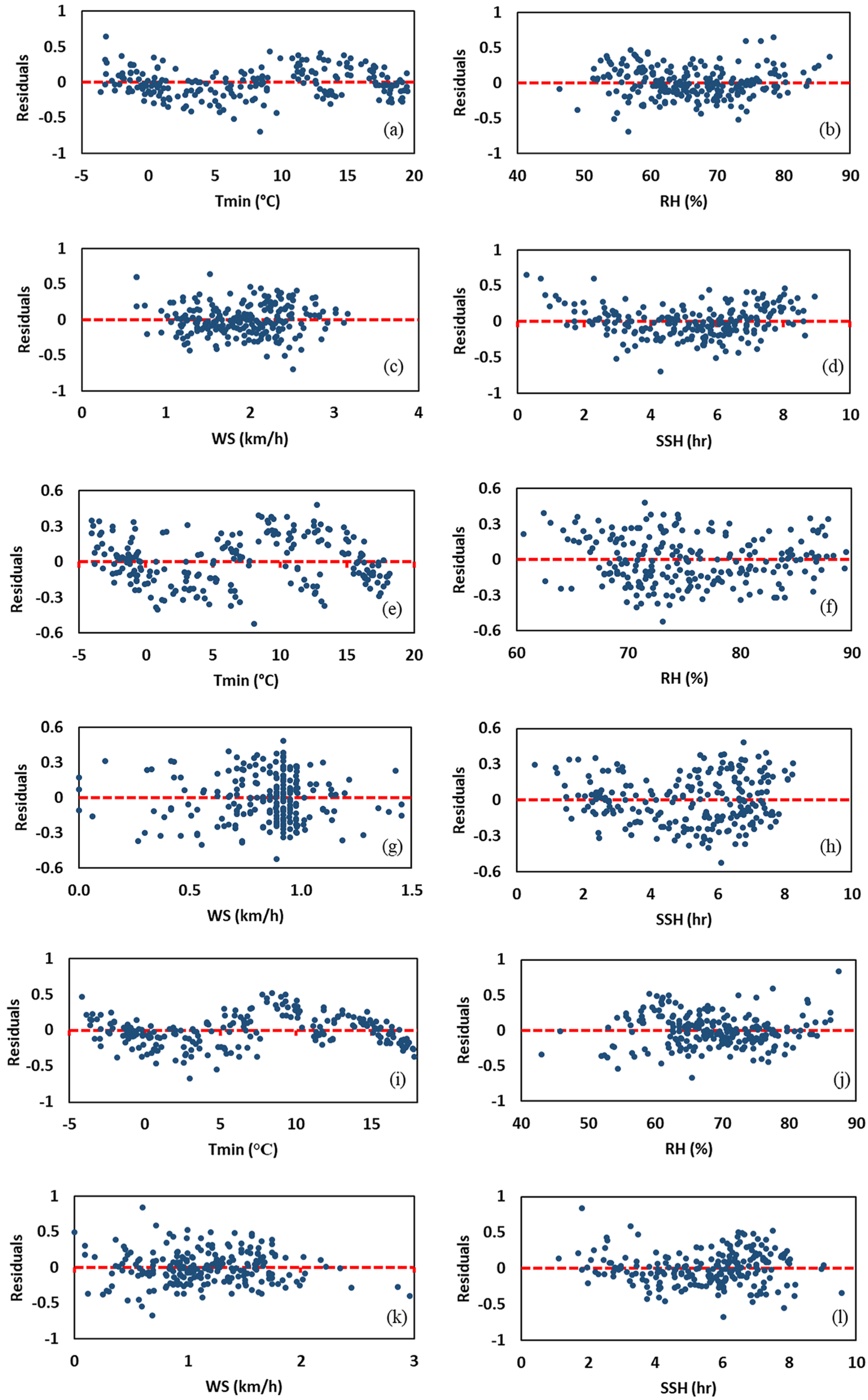

Fig. 6 Residual plots for the linear models developed for the Srinagar (**a**–**d**), Kupwara (**e**–**h**), and Qazigund (**i**–**l**) stations, respectively

Table 4 Regression equation for FAO-56 ET_O in terms of HAR ET_O

Station	Regression equation	SEE	R^2
Srinagar	FAO-56 $ET_O = 0.759 \times$ HAR $ET_O - 0.049$	0.134	0.988
Kupwara	FAO-56 $ET_O = 0.679 \times$ HAR $ET_O - 0.023$	0.146	0.985
Qazigund	FAO-56 $ET_O = 0.733 \times$ HAR $ET_O - 0.028$	0.104	0.991

ET_O in mm/day

Acknowledgements Acknowledgments are due to the Ministry of Human Resources Development (MHRD), Government of India, for providing the Doctoral Fellowship to the author, IMD Pune, for providing the meteorological data and to NIT Srinagar for providing space for computational analysis.

Funding The research is funded by the Ministry of Human Resources Development (MHRD), Government of India.

Code availability Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Availability of data and material The data that support the findings of this study are available from IMD PUNE. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors with the permission of IMD PUNE.

References

Allen RG, Clemmens AJ, Burt CM, Solomon K, O'Halloran T (2005) Prediction accuracy for project wide evapotranspiration using crop coefficients and reference evapotranspiration. J Irrigat Drain Eng 131:24–36

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration. Irrigation and Drainage Paper No. 56. FAO and United Nations: Rome, Italy; 300 pp

Allen RG, Pruitt WO, Wright JL, Howell TA, Ventura F, Snyder R, Itenfisu D, Steduto P, Berengena J, Beselga J, Smith M, Pereira LS, Raes D, Perrier A, Alves I, Walter I, Elliott R (2006) A recommendation on standardized surface resistance for hourly calculation of reference ET_O by the FAO56 Penman-Monteith method. Agric Water Manag 81:1–22

Bandyopadhyay A, Bhadra A, Swarnakar RK, Raghuwanshi NS, Singh R (2012) Estimation of reference evapotranspiration using a user-friendly decision support system: DSS_ET. Agric For Meteorol 154:19–29

Cristea NC, Kampf SK, Burges SJ (2013) Linear models for estimating annual and growing season reference evapotranspiration using averages of weather variables. Int J Climatol 33:376–387

Garcia M, Raes D, Allen R, Herbas C (2004) Dynamics of reference evapotranspiration in the Bolivian highlands (Altiplano). Agric For Meteorol 125:67–82

Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Appl Eng Agric 1:96–99

Irmak S, Allen RG, Whitty EB (2003a) Daily Grass and alfalfa-reference evapotranspiration estimates and alfalfa to grass evapotranspiration ratios in Florida. J Irrigat Drain Eng 129:360–370

Irmak S, Irmak A, Allen RG, Jones JW (2003b) Solar and net radiation- based equations to estimate reference evapotranspiration in humid climates. J Irrigat Drain Eng 129:336–347

Jabloun M, Sahli A (2008) Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited climatic data, application to Tunisia. Agric Water Manag 95:707–715

Kashyap PS, Panda RK (2001) Evaluation of evapotranspiration estimation methods and development of crop-coefficients for potato crop in a sub- humid region. Agric Water Manag 50:9–25

Kovoor G, Nandagiri L (2007) Development of regression models for predicting pan evaporation from climatic data – a comparison of multiple least-squares, principal components and partial least squares approaches. J Irrigat Drain Eng 133:444–454

O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690

Temesgen B, Eching S, Davidoff B, Frame K (2005) Comparison of some reference evapotranspiration equations for California. J Irrigat Drain Eng 131:73–84

Xu CY, Singh VP (2001) Evaluation and generalization of radiation-based methods for calculating evaporation. Hydrol Process 15:305–319

Yirga SA (2019) Modelling reference evapotranspiration for Megecha catchment by multiple linear regression. Model Earth Syst Environ 5:471–477