Reference Evapotranspiration Modelling using Support Vector Regression

Arun Raj V.E , P.G Jairaj
Department of Civil Engineering
College of Engineering Trivandrum
Trivandrum, India
rahere2630@gmail.com, jairaj_pg@yahoo.com

Abstract— **Evapotranspiration has a prominent role in the hydrologic cycle and its accurate estimation is of great importance in the fields of hydrology and water resources. Regression techniques can be made use of for developing relationship between Evapotranspiration and associated parameters influencing it. This paper demonstrates the applicability of Support Vector Regression (SVR) in modeling Reference Evapotranspiration in terms of various meteorological parameters such as Temperature, Solar radiation, Relative Humidity and Wind Speed. The SVR model is applied to the data pertaining to the IMD Station Trivandrum, Kerala and Davis region, California, United States. The performance of SVR model was compared with Multiple Linear Regression (MLR) in terms of various performance measures. It is seen that SVR models provide better estimates of Evapotranspiration compared to MLR models, thereby emphasizing the usefulness of SVR technique in estimation of Evapotranspiration.**

Keywords—SVR; Reference Evapotranspiration; MLR

I. INTRODUCTION

Evapotranspiration is one of the major components of the hydrologic cycle and its accurate estimation is important in hydrological practices such as in studies of hydrologic water balance, crop yield estimation, water resources system design and management. The evapotranspiration rate from a reference surface, not short of water is called the reference crop evapotranspiration or reference evapotranspiration. The phenomenon of Evapotranspiration is a complex one and depends on several climatological factors such as Temperature, Solar Radiation, Humidity, Wind speed, Type and Growth stage of crop etc. There are many available direct and indirect methods to estimate Reference Evapotranspiration. The direct method of estimation of Reference Evapotranspiration is by using lysimeter. But the method is time consuming and expensive. Thus, various indirect methods to estimate Evapotranspiration based on meteorological variables have been developed by many researchers. The well-known modified FAO-56 Penman-Monteith method is commonly used for Reference Evapotranspiration estimation. But the data requirement is very large which makes this method less adoptive. Accordingly, Penman equation cannot be applied if one or more of its parameters are not available from meteorological weather station measurements. Thus alternate methods are to be developed to model the complex phenomenon of Evapotranspiration. The methods to be developed should not only be capable of accurate estimation of Reference Evapotranspiration, but also it must have minimum number of input parameters. Here comes the importance of regression techniques, wherein functional relationship between dependent and independent variables are established. The SVR has been made use of in the modelling of Pan Evaporation, wherein relationship between different meteorological variables was established [3].

In the present study, SVR algorithm was used to model Reference Evapotranspiration by establishing relationship with meteorological parameters such as Temperature, Solar Radiation, Relative Humidity and Wind Speed considered as parameters associated with it. The model is applied to the data pertaining to IMD station: Trivandrum, Kerala, India and Davis region, California, United States. The performance of SVR model was compared with a conventional MLR model. A brief overview of the SVR analysis, methodology adopted in the study, analysis and results of the study are discussed in the subsequent sections.

II. SUPPORT VECTOR REGRESSION

Support Vector Machines (SVMs) is an analytical tool that can be used for both classification and regression purposes [4]. Theoretical concepts of SVR and procedures are available in the literature [1]. SVR puts an emphasis on robustness of the results through the use of the ε -insensitive loss function [2]. The SVMs classification methods are based on the principle of optimal separation of classes. The technical nature of SVMs for function estimation is called Support Vector Regression. The main strength of SVR method is that it has got good generalizability. This method provides the simplest function that can describe the relation between the dependent and independent variables. The basic concepts of SVR regarding loss function and function estimation are as follows.

Loss Function: Vapnik proposed Support Vector Regression by introducing an insensitive loss function (ε) [4]. This loss function allows the concept of margin to be used for regression problems. Purpose of Support vector regression is to find a function having at the most ε deviation from the actual target vectors for all given training data and have to be as flat as possible. The loss function employed in SVR is the ε-insensitive loss function that has the form

$$g(r_i) = \left|y_i - f(x_i)\right| = \max\{0, \left|y_i - f(x_i)\right| - \varepsilon\} \quad (1)$$

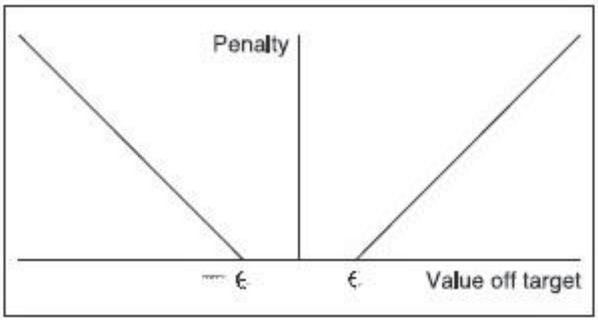

Fig. 1. ε-insensitive loss function

where y_i is the true target value, x_i is a vector of input variables and $f(x_i)$ is the estimated target value for observation i. Fig.1 shows the resulting function for the residual. If the absolute residual is off-target by ε or less, then there is no loss, that is, no penalty should be imposed. But if $\left|y_i - f(x_i)\right| - \varepsilon > 0$, then a certain amount of loss should be associated with the estimate. This loss rises linearly with the absolute difference between y and $f(x)$ above ε.

Function estimation with SVR - Linear case: First of all, consider the simplest case of function estimation with SVR having only one independent variable x_1, one dependent variable y and l data points available for training. The functional relationship between the variables is assumed as linear, of the form

$$y = w_1 x_1 + b \quad (2)$$

where w_1 and b are parameters to be estimated. Here b is the bias term that determines the offset of the hyperplane from origin and w determines the orientation of hyperplane. A smaller value of w indicates the flatness. The linear decision surface $f(x) = w_1 x_1 + b$ should be as flat as possible. Therefore, in order to find the SVR decision surface, SVR minimizes the following optimization problem [4].

$$Minimize \frac{1}{2} w_1^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \quad (3)$$

Subject to

$$y_i - (w_1 x_{i1}) - b \le \varepsilon + \xi_i$$

$$(w_1 x_{i1}) + b - y_i \le \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0 \text{ for } i = 1,2..l$$

The first term, $\frac{1}{2} w_1^2$ captures the degree of complexity. The slack variables ξ_i, ξ_i^* where $i = 1,2..l$, are constrained to be non-negative. The manually adjustable constant C determines the trade-off between function's complexity, $\frac{1}{2} w_1^2$, and the overall loss associated with it, $\sum_{i=1}^{l} (\xi_i + \xi_i^*)$. Thus, the second term of the objective function - the sum $C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$ - stands for the actual amount of loss associated with the estimated function, since loss occurs only if a point lies outside the ε-insensitive region. In practice, a dual representation of the optimization problem is used to solve for the SVR. If there are n independent variables, then the optimal regression function $f(x) = (w' x) + b$ with a vector of independent variables, $x' = (x_1, x_2, x_3, \ldots. x_n)$, weight vector $w' = (w_1, w_2, \ldots w_n)$, and the inner product $(w' x) = (w_1 x_1 + w_2 x_2 + \ldots. + w_n x_n)$ In that case flatness is defined in terms of the Euclidean norm of the weight vector $\|w\| = \sqrt{w_1^2 + w_2^2 + \ldots. + w_n^2}$. The unknown parameters of the linear SVR ie, w, b, ξ_i and ξ_i^*, i = 1, 2,..., l, can be found as the unique solution of the dual of the primal problem of (3).

Function estimation with SVR - Nonlinear case: The fundamental theoretical idea behind constructing nonlinear SVR is to map the available data from the original space into a higher-dimensional space, called feature space, and compute an optimal linear regression function in this feature space. The transformation of the data is carried out via the mapping $X \to \phi(x)$, where $X = (x_1, x_2 \ldots. x_n)$ is a vector of independent variables. The linear regression function constructed in the transformed, feature space corresponds to a nonlinear regression function in the original, non-transformed space. For these transformations, kernel functions are made use of. By definition, any kernel is just a dot product of two vectors (in some space). The advantage of using kernels is that the dot product between two points i and j in the transformed space can be written out as a kernel:

$$\phi(x_i).\phi(x_j) = k(x_i x_j) \quad (4)$$

Examples for such very common kernel functions are the d-order polynomial kernel,

$$k(x_i x_j) = \phi(x_i).\phi(x_j) = (x_i . x_j + 1)^d \quad (5)$$

and the Radial Basis Function (RBF) kernel such as Gaussian kernel:

$$k(x_i x_j) = \phi(x_i).\phi(x_j) = e^{-\gamma \left|x_i - x_j\right|^2} \quad (6)$$

where d and γ are manually adjustable parameters inherent to the polynomial and RBF kernels, respectively. In the nonlinear case the SVR estimates take the form:

$$f(x) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) k(x_i, x_j) + b \quad (7)$$

where the unknowns are the multipliers α_i and α_i^*, i = 1, 2,..., l. They are the weights associated with each data point i. If both α_i and α_i^* for point i are equal to zero, then this point lies inside the ε-insensitive region. It has a (combined $\alpha_i + \alpha_i^*$) weight of zero and plays no role for the final formulation of the SVR function.

III. METHODOLOGY

The present study deals with the development of a model for Reference Evapotranspiration using Support Vector Regression approach. Evapotranspiration calculated from modified FAO-56 Penman-Monteith method was treated as the dependent variable and meteorological parameters like temperature, solar radiation, wind speed and relative humidity were treated as independent variables.

SVR model: Different SVR models were developed considering various input parameter combinations. The loss function adopted for SVR model development is the ε-insensitive loss function. Also two different types of kernel functions, the Polynomial kernel function and Gaussian kernel function were tried in the SVR analysis to identify the sensitivity of these kernel functions. The performance of the SVR model was compared with MLR model to check its suitability as an alternative estimation method for Reference Evapotranspiration.

Multiple linear regression: The linear relationship between a dependent variable and an independent variable can mathematically be expressed by linear regression analysis and the relationship between the variables in simple linear regression analysis is written as

$$y = a + bx \quad (8)$$

where x is the independent variable, y is the dependent variable, and a, b are constants to be obtained from the regression analysis. Similarly, Multiple Linear regression technique was used to model evapotranspiration data with climatological parameters in combination. For a multiple linear regression model, the dependent variable y is assumed to be a function of n independent variables $x_1, x_{2,....}x_n$ and the model is expressed in the form

$$y = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n \quad (9)$$

where $a_0, a_1, ... a_n$ are the regression constants determined using least squares criterion which minimize the sum of squares of error terms.

Model performance: The performance of the model developed from SVR as well as MLR was assessed by calculating Correlation Coefficient (r) and Root Mean Square Error (RMSE). Correlation Coefficient (r) is given by,

$$r = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y} \quad (10)$$

where r = Correlation coefficient
S_y = Standard deviation of dependent variable
S_x = Standard deviation of independent variable

Similarly, Root Mean Square Error is given by,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_{calculated} - Y_{observed})^2} \quad (11)$$

The model having the highest Correlation Coefficient and lowest RMSE value was selected as the best model from each regression procedures.

Model validation: Different models of Reference Evapotranspiration in terms of meteorological parameters were developed using Support Vector Regression approach as well as Multiple Linear Regression analysis. The best models developed from these approaches were used to predict evapotranspiration values on the basis of the respective variables in the model. The model results were compared in terms of correlation coefficient, coefficient of determination (r^2), Root Mean Square Error, Mean Absolute Error (MAE) and Standard Error of Estimate (SEE); where MAE and SEE are defined as

$$MAE = \frac{\sum |Y_{calculated} - Y_{observed}|}{n} \quad (12)$$

$$SEE = \left[\frac{\sum (Yobserved - Ycalculated)^2}{n-1} \right]^{0.5} \quad (13)$$

where n is the number of data points .

IV. ANALYSIS OF THE PROBLEM

The application of proposed methodology was used to model Reference Evapotranspiration considered as dependent variable and meteorological variables considered as independent variables for the Trivandrum meteorological station, Kerala, India as well as for Davis region, California, United States. Different input combinations were tried for the analysis. The details pertaining to the data used and the computations involved are discussed in the subsequent sections.

Data used: Data used in the present study include the monthly meteorological data pertaining to IMD station: Trivandrum from 2001 to 2004 and Davis region, California from 2001 to 2008. The meteorological data under consideration include temperature, solar radiation, wind speed and relative humidity and Reference Evapotranspiration was computed from modified FAO-56 Penman-Monteith method. In the analysis, average daily evapotranspiration in a month was taken and a total of 36 points were used in training and 12 points were used in testing of the models for IMD Station: Trivandrum whereas for Davis region, a total of 72 points were used for training and 24 points were used for testing the models.
.

Input combinations: The parameters used in the modelling include average temperature (T), solar radiation (S), relative humidity (H) and wind speed (W) as independent parameters and Reference Evapotranspiration (ETo) as the dependent parameter. The various input combinations and symbolic representation of models using monthly data are given in Table I.

TABLE I. SYMBOLIC REPRESENTATION OF MLR AND SVR MODELS

Sl. No	Input Combinations	Symbolic Representation of Model			
		IMD Station: Trivandrum		*Davis region*	
		MLR	*SVR*	*MLR*	*SVR*
1	T	LT1	ST1	LD1	SD1
2	S	LT2	ST2	LD2	SD2
3	H	LT3	ST3	LD3	SD3
4	W	LT4	ST4	LD4	SD4
5	S+T	LT5	ST5	LD5	SD5
6	S+H	LT6	ST6	LD6	SD6
7	S+W	LT7	ST7	LD7	SD7
8	T+S+H	LT8	ST8	LD8	SD8
9	T+S+W	LT9	ST9	LD9	SD9
10	S+H+W	LT10	ST10	LD10	SD10
11	T+S+H+W	LT11	ST11	LD11	SD11

The models developed using MLR for IMD Station: Trivandrum were named as LT1 to LT11 whereas for Davis region, the MLR models were labelled as LD1 to LD11 for different input combinations. Similarly, models developed using SVR technique for IMD Station: Trivandrum were labelled as ST1 to ST11 whereas for Davis region, the models were labelled as SD1 to SD11.

Support Vector Regression: Reference Evapotranspiration modelling was done using SVR technique, for single and combination of meteorological parameters. The various SVR parameters used in modelling are given in Table II.

TABLE II. SVR PARAMETERS IN ETO MODELLING

	IMD Station: Trivandrum	*Davis region, California*
Polynomial kernel	C = 10 ε = 0.001 d = 3	C = 10 ε = 0.001 d = 2
Gaussian kernel	C = 10 ε = 0.001 γ = 5	C = 10 ε = 0.001 γ = 3

To study the sensitivity of polynomial and Gaussian Kernels in the model development, different trials were done and the best values of the parameters were selected for the analysis. A MATLAB implementation of SVR was used and different models were developed for various input combinations. The performance of each model was assessed in terms of correlation coefficient and RMSE and the best models from each set of parameter combination were identified.

Multiple Linear Regression: Among the different models developed for IMD Station: Trivandrum by MLR technique using monthly data, the models LT1, LT2, LT3 and LT4 denote the relation of Reference Evapotranspiration with individual parameters of temperature, solar radiation, relative humidity and wind speed respectively. The models LT5 to LT7 represent the relation between ETo and two meteorological parameters. The models LT8 to LT10 are with three parameters and model LT11 incorporate all the input parameters together. Similarly for Davis region, models LD1 to LD4 represent all one variable models; LD5 to LD7 represent all two variable models; LD8 to LD10 represent all three variable models and model LD11 incorporate all the input parameters together. The correlation coefficient and RMSE of different MLR models were calculated to identify the best model from each set of input parameters. Similarly, the analysis was also carried out using daily data corresponding to different input combinations for the models in Table 5.4. The MLR analysis was carried out using IBM SPSS Statistics 20.0.

V. RESULTS AND DISCUSSION

Both SVR and MLR procedures were used to model Reference Evapotranspiration in terms of various meteorological parameters.

Support Vector Regression: The performance of Reference Evapotranspiration models derived using SVR analysis for IMD Station: Trivandrum are tabulated in Table III, for two different kernel functions: Polynomial kernel function and Gaussian kernel function. From Table III, it can be seen that among the single parameter models, the most influencing individual parameter on ETo is the solar radiation (model ST2) with higher correlation coefficient and lower RMSE when compared to other single parameter models.

TABLE III. PERFORMANCE ANALYSIS OF SVR MODELS IN MODEL FITTING FOR IMD STATION: TRIVANDRUM

Model Represe ntation	Parameter Combination	Polynomial kernel		Gaussian kernel	
		r	*RMSE*	*r*	*RMSE*
ST1	T	0.715	0.405	0.716	0.807
ST2	S	0.935	0.206	0.913	1.04
ST3	H	0.665	0.432	0.538	1.71
ST4	W	0.455	0.515	0.416	0.534
ST5	S+T	0.954	0.174	0.928	1.1
ST6	S+H	0.943	0.193	0.908	0.274
ST7	S+W	0.958	0.166	0.923	1.32
ST8	T+S+H	0.967	0.148	0.788	1.68
ST9	T+S+W	0.974	0.132	0.924	1.191
ST10	S+H+W	0.972	0.136	0.95	0.211
ST11	T+S+H+W	0.982	0.109	0.953	0.207

TABLE IV. PERFORMANCE ANALYSIS OF SVR MODELS IN MODEL FITTING FOR DAVIS REGION

Model Representation	Parameter Combination	Polynomial kernel		Gaussian kernel	
		r	*RMSE*	*r*	*RMSE*
SD1	T	0.939	0.774	0.941	0.769
SD2	S	0.988	0.352	0.989	0.657
SD3	H	0.882	1.08	0.913	0.959
SD4	W	0.303	2.15	0.307	2.152
SD5	S+T	0.992	0.28	0.991	0.663
SD6	S+H	0.995	0.216	0.989	0.817
SD7	S+W	0.988	0.349	0.986	0.622
SD8	T+S+H	0.996	0.194	0.989	0.852
SD9	T+S+W	0.997	0.193	0.991	0.665
SD10	S+H+W	0.996	0.211	0.989	0.818
SD11	T+S+H+W	0.999	0.117	0.99	0.853

With two variable models, model ST7 using Polynomial kernel and model ST5 using Gaussian kernel were performing better and were selected for testing. Similarly, when all the three parameters are considered, the model ST9 using Polynomial kernel and model ST10 using Gaussian kernel were having higher correlation coefficient and lower RMSE compared to other models and were selected for testing. Further, model ST11 using Polynomial kernel or Gaussian kernel was also selected for testing when all the four input parameters are considered and shows highest correlation coefficient, r = 0.982 and lowest RMSE among all the SVR models developed. Therefore, models ST2, ST7, ST9 and ST11 with Polynomial kernel and the models ST2, ST5, ST10 and ST11 with Gaussian kernel were selected for testing, with unknown data sets. In general, when the SVR models are compared on the basis of kernel functions, models with Polynomial kernel function shows higher correlation coefficient and lower RMSE value when compared to SVR models with Gaussian kernel function.

Following the same criteria of SVR model selection for Reference Evapotranspiration in the case of IMD Station: Trivandrum discussed above, the better performing SVR models for the Davis region, California, were selected from Table IV as SD2, SD6, SD9 and SD11 with Polynomial kernel and models SD2, SD5, SD9 and SD11 with Gaussian kernel functions.

Multiple Linear Regression: The MLR models developed for the estimation of Reference Evapotranspiration, from meteorological parameters for IMD Station: Trivandrum is shown in Table V. From the table, it can be seen that the second model (LT2) which is based on solar radiation is having highest correlation coefficient and lowest RMSE value when compared to other single variable models. Similarly, when combination of parameters were considered, models LT5, LT9 and LT11 are the best two variable, three variable and four variable models respectively. Therefore the models LT2, LT5, LT9 and LT11 were selected as best models based on MLR analysis and were used in testing. Similarly, the results of MLR analysis of Reference Evapotranspiration for Davis region, California is given in Table VI.

TABLE V. PERFORMANCE ANALYSIS OF MLR MODELS IN MODEL FITTING FOR IMD STATION: TRIVANDRUM

Input	MLR Equation	r	RMSE
LT1	ETo = -11.11+0.545 T	0.695	0.42
LT2	ETo = -0.56+0.245 S	0.933	0.21
LT3	ETo = 9.145-0.066 H	0.599	0.46
LT4	ETo = 3.515+0.077 W	0.283	0.31
LT5	ETo= -4.63+0.212S+0.17T	0.95	0.18
LT6	ETo= -0.54+0.245S+0H	0.933	0.21
LT7	ETo= -0.594+0.241S+0.02W	0.936	0.20
LT8	ETo= -5.388+0.177T+0.218S+0.005H	0.95	0.19
LT9	ETo= -5.45+0.20T+0.198S+0.035W	0.95	0.17
LT10	ETo=0.625+0.22S-0.012H+0.034W	0.938	0.20
LT11	ETo= -4.389+0.198T+0.181S-0.009H+0.046W	0.958	0.166

TABLE VI. PERFORMANCE ANALYSIS OF MLR MODELS IN MODEL FITTING FOR DAVIS REGION

Input	MLR Equation	r	RMSE
LD1	ETo= -1.698+0.365T	0.939	0.78
LD2	ETo= -0.95+0.024S	0.987	0.36
LD3	ETo=12.361-0.138H	0.883	1.06
LD4	ETo= -1.319+2.073W	0.307	2.14
LD5	ETo= -1.327+0.019S+0.096T	0.993	0.28
LD6	ETo= 1.542+0.02S-0.029H	0.992	0.31
LD7	ETo= -1.196+0.024S+0.106W	0.987	0.36
LD8	ETo= 0.435+0.071T+0.018S-0.019H	0.995	0.25
LD9	ETo= -2.791+0.13T+0.016S+0.573W	0.995	0.22
LD10	ETo= 1.234+0.02S-0.029H+0.145W	0.992	0.29
LD11	ETo= -1.186+0.106T+0.016S-0.015H+0.504W	0.996	0.22

From Table VI, it can be seen that in case of Davis region also, the correlation of Reference Evapotranspiration is higher for solar radiation (model LD2) when single parameter was considered and the least with the wind speed (model LD4). Also, models LD5, LD9 and LD11 shows higher correlation coefficient and lower RMSE for Reference Evapotranspiration with two, three and four variable combinations. Therefore, the models LD2, LD5, LD9 and LD11 are the better performing models for Davis region, California, derived using MLR technique and were used in testing.

Model Testing: From the models developed for Reference Evapotranspiration using MLR and SVR techniques considering meteorological variables as independent parameters and ETo as the dependent parameter, best model with combination of meteorological parameters (ie, one, two, three and four), selected on the basis of performance parameters in training, were used in testing against unknown data sets and the performance were evaluated, for IMD Station: Trivandrum and Davis region, California. The performance of the models assessed on the basis of performance parameters r^2, r, RMSE, MAE and SEE are tabulated in Table VII for IMD Station: Trivandrum and in Table VIII for Davis region, California. From the tables, it can be seen that the correlation coefficient values of MLR models goes on increasing with the increase of input variables and the highest value of correlation coefficient was obtained for

TABLE VII. TESTING OF MLR AND SVR MODELS FOR IMD STATION: TRIVANDRUM

Parameter	MLR Estimated ETo (mm/day)				SVR Estimated ETo (mm/day)							
					Polynomial kernel				*Gaussian kernel*			
Model	LT2	LT5	LT9	LT11	ST2	ST5	ST9	ST11	ST2	ST5	ST10	ST11
r²	0.871	0.848	0.899	0.898	0.878	0.889	0.901	0.906	0.85	0.85	0.874	0.874
r	0.933	0.921	0.948	0.948	0.937	0.943	0.949	0.952	0.922	0.922	0.935	0.935
RMSE	0.157	0.178	0.144	0.144	0.156	0.152	0.140	0.137	0.182	0.179	0.163	0.164
MAE	0.126	0.151	0.118	0.112	0.125	0.130	0.116	0.114	0.143	0.146	0.138	0.136
SEE	0.164	0.186	0.151	0.152	0.163	0.159	0.146	0.143	0.189	0.187	0.17	0.171

TABLE VIII. TESTING OF MLR AND SVR MODELS FOR DAVIS REGION

Paramet er	MLR Estimated ETo (mm/day)				SVR Estimated ETo (mm/day)							
					Polynomial kernel				*Gaussian kernel*			
Model	LD2	LD5	LD9	LD11	SD2	SD6	SD9	SD11	SD2	SD5	SD9	SD11
r²	0.975	0.991	0.993	0.995	0.978	0.992	0.994	0.996	0.848	0.846	0.848	0.882
r	0.987	0.995	0.996	0.997	0.989	0.996	0.997	0.999	0.921	0.92	0.921	0.939
RMSE	0.372	0.264	0.182	0.220	0.382	0.199	0.175	0.114	0.981	1.05	1.04	1.02
MAE	0.312	0.221	0.142	0.183	0.321	0.210	0.139	0.132	0.773	0.853	0.859	0.808
SEE	0.381	0.269	0.186	0.225	0.393	0.203	0.179	0.116	1.002	1.07	1.08	1.042

models LT11 and LD11 respectively for IMD Station: Trivandrum and for Davis region, California. The model LT9 of IMD Station: Trivandrum also shows reasonable performance in terms of performance parameters and in the case of Davis region, California the model LD9, can be considered as the best model in terms of various performance measures.

When the performance of SVR models were analysed, it is seen that SVR models with Polynomial kernels gives better performance when compared to SVR models with Gaussian kernel. The best performance measures were obtained for models ST11 and SD11 respectively with Polynomial kernel for IMD Station: Trivandrum and for Davis region, California which make use of all the four input parameters in the model development. Among all the models considered, the models ST2, ST5, ST9 and ST11 with Polynomial kernel function show higher value of correlation coefficient than any other one, two, three and four variable combination models of MLR and SVR using Gaussian kernel function in the case of IMD Station: Trivandrum. Also, the error measures for the above models such as RMSE, MAE and SEE are lower compared to other models. The same results were obtained for Davis region, California also.

VI.CONCLUSION

The present study deals with the development of Multiple Linear Regression (MLR) and Support Vector Regression (SVR) models for modelling Reference Evapotranspiration. The models were developed for different combinations of one, two, three and four input meteorological variables under consideration and the performance of different models were studied so as to identify the best combination of input variables for estimation of Evapotranspiration. The developed models of Reference Evapotranspiration were applied to two study areas; IMD Station: Trivandum and Davis region, California. The specific conclusions from the study are as follows:

- The highest correlation coefficient and lowest Root Mean Square Error values were obtained when all the four input variables; temperature, solar radiation, relative humidity and wind speed were considered in the model development.
- The study reveals that for IMD Station: Trivandrum and for Davis region, California, the most influencing individual parameter on Reference Evapotranspiration is the solar radiation and has got a higher correlation coefficient and lower RMSE when compared with other individual parameters.
- MLR and SVR techniques provide reasonably good estimate of Reference Evapotranspiration for monthly data and among the two regression techniques, SVR shows better performance in terms of performance parameters.
- Based on the sensitivity analysis carried out with two kernel functions, it is seen that the Polynomial kernel based SVR performs better when compared with Gaussian kernel based SVR for estimation of Reference Evapotranspiration for IMD Station: Trivandrum as well as Davis region, California.

Thus, it can be concluded that regression methods such as Multiple Linear Regression (MLR) and Support Vector Regression (SVR) can be used to model Evapotranspiration. It can also be stated that Polynomial based SVR performs better compared to other models for modelling Reference Evapotranspiration for the study area under consideration.

References

[1] A.J. Smola and B. Scholkopf. A Tutorial on Support Vector Regression. NeuroCOLT2 Technical Report, University of London, UK, 1998.

[2] G. Nalbantov, Patrick J.F .Groenen and Jan C. Bioch. Support Vector Regression Basics. Spring, 13(1):16-19, 2005.

[3] S. Deswal and Mahesh Pal. Modelling of Pan Evaporation using Support Vector Machines Algorithm. ISH Journal of Hydraulic Engineering, 14(1):104-116, 2008

[4] Vladimir N. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc.,1995.