Multivariate Time Series Evapotranspiration Forecasting using Machine Learning Techniques

Chalachew Muluken Liyew
Dip. di Informatica, Università di Torino

Rosa Meo
Dip. di Informatica, Università di Torino

Elvira Di Nardo
Dip. Matematica "G. Peano", Università di Torino

Stefano Ferraris
DIST, Politecnico di Torino and Università di Torino

ABSTRACT

The actual evapotranspiration (AET) could be forecasted using meteorological variables to manage and plan water resources even though it is challenging to choose the relevant variables for prediction. The Pearson correlation method was applied to select candidate variables and further, tolerance and VIF scores are implemented to avoid multicollinearity problems among variables. As a result, five relevant variables are selected for training the AET prediction models. In this paper, we proposed three methods for forecasting AET: (i) deep learning-based (LSTM, GRU, and CNN), (ii) classical machine learning (SVR and RF), and (iii) a statistical technique (SARIMAX). The performance of each model is measured with statistical indicators (RMSE, MSE, MAE, and R^2). The results showed that relatively high performance is measured in the LSTM model.

CCS CONCEPTS

• **Mathematics of computing** → **Time series analysis**; • **Computing methodologies** → **Machine learning approaches**;

KEYWORDS

evapotranspiration, deep learning, machine learning, multivariate time series analysis

ACM Reference Format:
Chalachew Muluken Liyew, Rosa Meo, Elvira Di Nardo, and Stefano Ferraris. 2023. Multivariate Time Series Evapotranspiration Forecasting using Machine Learning Techniques. In *The 38th ACM/SIGAPP Symposium on Applied Computing (SAC '23), March 27-March 31, 2023, Tallinn, Estonia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3555776.3577838

1 INTRODUCTION

Actual evapotranspiration (AET) is the loss of water that occurs via evaporation from surfaces of soil and water bodies and via the transpiration from the plant that comes into the air as water vapor [6, 15]. When there is low rainfall on average and high evapotranspiration in a region, agricultural development depends on irrigation due to water scarcity in the soil. To handle the water resource challenges, AET forecasting is an essential tool for farmers and water resource managers [16]. AET can be measured directly from devices [1, 3] using either eddy covariance or a lysimeter. It is challenging to obtain direct measure data of AET [1] so the estimation of AET should replace the direct measurement. Again, the estimation made by $FAO-56PM$ equation [17] is challenging due to the difficulty of gathering or measuring about eight variables used by the equation. Hence, designing an AET forecasting model is essential to formulate an effective plan to prevent water resource challenges and manage irrigation water requirements. The works in [11, 13, 15, 18] used the most available meteorological variables as predictors and implement a suitable model to forecast AET.

Different Machine Learning and statistical methods are proposed in the literature. Seasonal Auto-Regressive Integrated Moving Average model with eXogenous factors (SARIMAX) is employed for time series AET prediction [7] to consider the effect of seasonality and exogenous variables in improving the prediction accuracy. According to [3], Temporal Convolutional Neural network (TCN) and Long-Short Memory Neural Network (LSTM) outperformed the Deep Neural Network (DNN), the Support-Vector Machines (SVM) and the Random-forest (RF) in the temperature-based features of AET forecast. In [6] Convolutional Neural Networks (CNN) with different structures were employed to forecast the daily AET that outperformed the SARIMA model and the seasonal naive. In [8] the authors used LSTM, one-dimensional CNN (1D CNN) and a combination of the two previous models (CNN-LSTM) [2, 9], as well as Artificial Neural Network (ANN), Decision Tree (DT) and RF. They showed that the deep learning models slightly outperformed the Machine Learning ones and among the deep learning models, the CNN - LSTM combination outperforms the AET forecasting.

This paper aims to assess an AET forecasting model (SARIMAX, LSTM, GRU, CNN, SVM, and RF) at a specific site: Cogne (Valle d'Aosta, Italy). An eddy covariance station measured AET in an abandoned pasture, which is today an increasingly spread land cover type. The task is challenging since it is well-known that forecasting many days in advance, or using the mean value of AET in a specific time interval (e.g., a month), shows a low accuracy.

The remainder of this article is organized as follows. Section 2 provides a systematic literature review. Section 3 introduced the material and methods. Section 4 presents the results of empirical experiments, and Section 5 draws the conclusion.

2 SYSTEMATIC LITERATURE REVIEW

We perform a systematic literature review of the topic of analysis of evapotranspiration by automatic techniques. Several Machine Learning algorithms have been applied to pursue evapotranspiration prediction research. In this study, a Systematic Literature

SAC '23, March 27-March 31, 2023, Tallinn, Estonia

ACM ISBN 978-1-4503-9517-5/23/03.
https://doi.org/10.1145/3555776.3577838

Review (SLR) is conducted to extract and synthesize the Machine Learning algorithms, the variables used (also referred to as features as synonymous), and the evaluation parameters used in evapotranspiration forecasting studies. Five electronic databases (Google Scholar, Wiley Online Library, ACM Digital Library, Elsevier, and IEEEexplore) have been used to extract data from the literature for this SLR. We retrieved 1854 papers from which we have selected 27 studies (18 of which are in the bibliography). We excluded the remaining 1829 studies using criteria based on relevance, the broadly used language (we preferred English over others), elimination of duplications, full-text unavailability, and excluding surveys.

In Table 1 we show the most used features for the task of evapotranspiration prediction. From the detailed analysis of the selected studies, we found that temperature, relative humidity, solar radiation, and wind speed are the widely used ones for evapotranspiration prediction. This justifies our choice for the features we collected and used in the present work.

Table 1: Most used features in the prediction of ET

Feature name	# of papers
Temperature (minimum and Maximum)	19
Relative humidity (RH)	15
Solar radiation (Rs)	13
Wind speed	10
Evapotranspiration only (in univariate time serie)	7
Sunshine duration	4

In Table 2 we show the most used Machine learning methods and their performance evaluation parameters in the literature for evapotranspiration prediction. We observe that the most widely applied Machine Learning algorithms are LSTM, CNN, SVM, SARIMA, and RF. The most used evaluation parameters are RMSE, MAE, and the coefficient of determination (R^2). They were selected and used for the present work.

These findings justify the choice we made in this work about the Machine Learning methods, the features, and the evaluation measures.

Table 2: Most used methods and evaluation parameters in prediction of ET

ML Method	# of papers	Evaluation measure	# of papers
LSTM	11	RMSE (Root mean squared error)	23
SVM	9	MAE (Mean absolute error)	17
ANN and CNN	7	R^2 (Coefficient of determination)	12
RF	6	MSE (Mean Squared Error)	5
SARIMA	4	-	

3 MATERIALS AND METHODS

3.1 Data collection and Cleaning

The four years (2014 - 2017) growing season (June - September) dataset is collected from the Cogne site, in Italy (1.534m altitude, $45\,°\,36'31.47''$N7 $°21'21.68''E$ of latitude and longitude) in every 30 minutes as shown in Table 3 with some missing values. Those missing values that existed randomly were imputed or predicted using a linear regression algorithm. We applied multiple iterative regression imputation starting from the two independent features that do not have missing values as shown in Table 3 but strongly correlated as shown in Table 4. First, we imputed `Mean temperature` having missing value (0.828%) due to its high correlation value with the variables `Sensible heat flux` and `Net` CO_2 which have no missing values. Then we trained another regression model for `Relative humidity` using as regressors `Sensible heat flux`, `Net` CO_2 and `Mean temperature` and so on. The dataset is published at "$https://github.com/rosimeo/Evapotranspiration-dataset$".

Table 3: missing values in (%) and correlation with AET

Variables - total observations = 23424	Missed observations(%)	Correlation with AET
Evapotranspiration (AET)	0	1
Sensible heat flux	0	0.82
Net CO_2	0	0.84
Mean Temperature	0.828	0.64
Air pressure	10.135	0.01
Wind speed	10.135	0.51
Wind direction	10.135	0.38
Soil surface temperature	4.568	0.41
Net solar radiation	1.878	0.89
Relative humidity	0.845	0.63
Water content	0.726	0.03

Prior to splitting the dataset into training and testing sets, the dataset was normalized into the $[0, 1]$ interval using the min-max normalization due to the application of the Sigmoid function on the deep learning recurrent neural network algorithms. The min-max normalization performs a linear transformation on the original data preserving their relationships. The variables having a correlation greater than 0.5 are taken as relevant candidate features for the model and are studied again the multicollinearity problem among selected variables and justified the irrelevant and redundancy variables presence among the selected variables as shown in Table 4 that degrades the model performance. The tolerance and VIF scores [4] were applied to solve this problem. When the VIF is higher than 10 (or tolerance is lower than 0.1), there is a significant multicollinearity that needs to be corrected [14]. This occurs in this study for the `net radiation` as shown in Table 5.

For the aim of this study, the chosen models were initially fit on the training set (14054), refined on the validation set (3514) by tuning the model parameters, and then evaluated on the test set (5856) for a total of 23424 observations referred to the five meteorological variables in the stamp of time series.

3.2 Machine Learning Methods

The six machine learning techniques used in this study for `AET` prediction are explained in this section.

Table 4: Pearson correlation among the selected features

	Net solar radiation	Net CO_2	Sensible heat flux	Mean temp	Relative humidity	Wind speed
Net solar radiation	1.00					
Net CO_2	-0.83	1.00				
Sensible heat flux	0.92	-0.80	1.00			
Mean temp	0.61	-0.49	0.48	1.00		
Relative humidity	-0.60	0.51	-0.52	-0.67	1.00	
Wind speed	0.64	-0.51	0.62	0.28	-0.48	1.00

Table 5: Tolerance, and VIF scores of independent variables

Variables	VIF	Tolerance	Re-VIF	Re-Tolerance
Net solar radiation	10.959	0.091	-	-
Net CO_2	3.384	0.296	2.911	0.344
Sensible heat flux	7.150	0.140	3.414	0.293
Mean temperature	2.366	0.423	1.956	0.511
Relative humidity	2.167	0.461	2.167	0.461
Wind speed	1.927	0.519	1.751	0.571

Note: Re-VIF and Re-Tolerance means recalculated VIF and tolerance

Long-short term memory Neural Network (LSTM). LSTM has been widely used recently for time series forecasting. For the analysis carried on in this paper, the LSTM model was designed using the five selected features as explained in Section 3.1. Each of these features is forming a time series, divided into windows whose size is 48 representing the observations of one day (one observation every half an hour). Each window is given as input to the first hidden layer of the LSTM network. At any time step, the window advances to the successive one so that the windows given in input at two consecutive times are overlapping for 47 observations of each time series. The target variable in output to the LSTM network is the dependent `AET`. The multivariate training set is used to train the LSTM model of the target `AET` variable using the Adam optimizer. Finally, the model performance is tested using the test set.

Gated Recurrent Unit (GRU). GRU is a special type of optimized LSTM-based recurrent neural network [12]. For this study, the recurrent neural network consists of a one-layer GRU of 32 units, followed by two dense layers of 16 units with ReLU activation function and single units with linear activation. A dropout of 0.2 was applied to the non-recurrent connections. The learning rate was set to 0.001. as recommended by [12]. As regards the setting of the above hyper-parameters of the model, we tried many different combinations of values guided by a grid search over the most common selection. For instance, the number of input units varied among $[8, 16, 32, 48, 64]$. For dropout, we tried with values among $[0.2, 0.5, 0.8]$. The mean squared error loss function was minimized using the Adam optimizer. Finally, the test set was used to evaluate the model.

Convolutional Neural Network (1D-CNN). CNN is a deep learning model already implemented to investigate the prediction of `AET` time series [6]. For the analysis carried on in this paper, one convolutional layer and two fully connected layers are used. In particular, the setup of the convolution layer filter is 32, the ReLU activation kernel size is 2, and the padding is set to “same”[1]. The number of neurons in the first fully connected layer is set equal to 8 and the ReLU activation function is used. In the second fully connected layer instead, it is the output layer that uses one neuron and a linear activation function.

Support Vector Machine (SVM). Xianming Dou and Yongguo Yang in [5] employed the SVM model with three kernel algorithms (Radial Basis Function (RBF), Polynomial (Poly) function, and Sigmoid function) for `AET` forecasting: the experimental result showed that the SVM with the RBF kernel function outperforms the Sigmoid and Poly kernel functions. In this paper, the internal function of the SVM is arranged with the kernel function of RBF and the epsilon parameter[2] with the value of 0.5. This model is tested with the same test data assigned to the other models.

Random Forest (RF). RF is a decision tree-based algorithm. According to RF, various subsets of the training data are fitted with a suitable decision tree [8]. In this paper, since the `AET` data is continuous, the implemented RF is addressed to solve the regression problems of `AET` forecast. Hence, the hyper-parameter `n_estimators` that corresponds to the number of trees used in the ensemble model, is chosen equal to 100 after we applied a grid search technique spanning the values of $[50, 100, 150, 200]$. The `random_state` is set equal to zero and the other hyper-parameters are set to their default values.

SARIMAX. For the analysis carried on this paper, suitable SARIMA models are proposed in [7] aiming to minimize the prediction error by considering seasonality patterns. The SARIMA has seasonal orders (P, Q, D) in addition to the orders (p, d, q). To identify the best SARIMA, the orders are determined by the built-in function of `auto_arima` in the `pmdarima` package. When one further hypothesis is added, those *eXogenous* variables are supposed to affect the time series prediction of the dependent variables.

4 EXPERIMENTAL RESULTS AND DISCUSSION

For the aim of this paper, the forecasting performance of each model was measured and compared to identify the best-performing model. The statistical metrics used to measure the performance of models are the Root Mean Squared Error (RMSE), mean absolute error(MAE), Mean Squared Error (MSE), and coefficient of determination R^2. The higher the value of R^2 the better the model is, while the other measures RMSE, MAE, and MSE are interpreted as measures of the prediction errors. In these cases, the lower is the better.

The three deep learning neural network models (LSTM, GRU, and CNN), the traditional Machine Learning methods (SVR and RF), and SARIMAX for time series are trained and evaluated on the same training, validating, and testing datasets, and the performances are

[1]Using the same padding corresponds to avoiding the use of an aggregation function that reduces the dimension of the output w.r.t the input size during the flow of information towards the successive layers.

[2]Epsilon controls the tolerance where no penalty is given for the prediction errors.

evaluated using the statistical metrics. The results are given in Table 6. With the bold font, we noted the best results.

Table 6: Model performance measures

Models	RMSE	MSE	MAE	R^2
LSTM	**0.0242**	**0.0006**	**0.0155**	**0.8747**
CNN	0.0275	0.0008	0.0169	0.8376
GRU	0.0264	0.0007	0.0161	0.8512
SVR	0.0289	0.0008	0.0221	0.8144
RF	0.0281	0.0008	0.0167	0.8250
SARIMAX	0.0266	0.0007	0.0153	0.8457

The prediction accuracy shows that all the tested models are well-fitted to predict and forecast AET. As shown in Table 6, the deep learning models performed slightly better than others. Among the deep learning models, the LSTM performs better ($R^2 = 0.8747$) compared to the other deep learning models such as CNN ($R^2 = 0.8376$) and GRU ($R^2 = 0.8512$). The GRU performs slightly better than the CNN according to the results of the performance measurements. SVM and RF showed slightly lower performance in AET prediction. Among the six tested models, SVR demonstrated a relatively weak measure of accuracy ($R^2 = 0.8144$) compared to the remaining five models. The RMSE of deep learning ranges from $0.0242 - 0.0275$ and the RMSE of SVM and RF are 0.0289 and 0.0281 respectively. The RMSE of the SARIMAX is 0.0266 which is comparable with the GRU RMSE of 0.0264.

Similar studies [3, 8, 16] are conducted and reported that deep learning outperforms classical machine learning. Ferreira and França da Cunha [8] experimented with the combination of CNN-LSTM to show a slightly better performance.

For this study, the experimental results showed the SARIMAX model ($R^2 = 0.8457$) outperforms the RF ($R^2 = 0.8250$) and the CNN ($R^2 = 0.8376$) models, a result similar to [10] that shows the ARIMA model outperforms the Neural network. Our experimental analysis showed that the LSTM model outperforms the traditional Machine Learning models and the SARIMAX model; a minor result is obtained by the SVM model. Comparable results are obtained in the SARIMAX and GRU models, as shown in Table 6.

5 CONCLUSION

Since obtaining the direct measure of AET is challenging, it is crucial to forecast it using the most readily available meteorological variables as inputs. For this study, we used the most readily available meteorological variables by assessing their relevance to the model development. Six candidate variables were selected for the model using Pearson correlation and further analysis was made using tolerance and VIF scores to select the most relevant variables by avoiding the multicollinear variable. Finally, five variables are considered as input for the models. In conclusion, we obtained five variables (Net CO_2, Sensible heat flux, Mean temperature, Relative humidity, and Wind speed) for training and evaluating the proposed models for this study. The result of this study showed that the LSTM and GRU models slightly outperform the SARIMAX model. In turn, the SARIMAX model outperforms the traditional Machine Learning models. Among the deep learning approaches, the LSTM model performs better than the other two deep learning methods, and the SVM demonstrates relatively diminutive performance in forecasting AET.

REFERENCES

[1] AA Masrur Ahmed, Ravinesh C Deo, Qi Feng, Afshin Ghahramani, Nawin Raj, Zhenliang Yin, and Linshan Yang. 2021. Hybrid Deep Learning for Week-Ahead Evapotranspiration Forecasting. (April 2021). https://doi.org/10.21203/rs.3.rs-424493/v1
[2] Rahim Barzegar, Mohammad Taghi Aalami, and Jan Adamowski. 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment* 34, 2 (February 2020), 415–433. https://doi.org/10.1007/s00477-020-01776-2
[3] Zhijun Chen, Zhenchuang Zhu, Hao Jiang, and Shijun Sun. 2020. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *Journal of Hydrology* 591 (December 2020), 125286. https://doi.org/10.1016/j.jhydrol.2020.125286
[4] Piedad María Cristiano, Daniel Alberto Pereyra, Sandra Janet Bucci, Nora Madanes, Fabian Gustavo Scholz, and G Goldstein. 2016. Remote sensing and ground-based measurements of evapotranspiration in an extreme cold Patagonian desert. *Hydrological Processes* 30, 24 (August 2016), 4449–4461. https://doi.org/10.1002/hyp.10934
[5] Xianming Dou and Yongguo Yang. 2018. Modeling evapotranspiration response to climatic forcings using data-driven techniques in grassland ecosystems. *Advances in Meteorology* 2018 (Apr. 2018). https://doi.org/10.1155/2018/1824317
[6] Patrícia de Oliveira e Lucas, Marcos Antonio Alves, Petrônio Cândido de Lima e Silva, and Frederico Gadelha Guimarães. 2020. Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks. *Computers and electronics in agriculture* 177 (October 2020), 105700. https://doi.org/10.1016/j.compag.2020.105700
[7] Mohamed M Fathi, Ayman G Awadallah, Ahmed M Abdelbaki, and Mohammed Haggag. 2019. A new Budyko framework extension using time series SARIMAX model. *Journal of hydrology* 570 (March 2019), 827–838. https://doi.org/10.1016/j.jhydrol.2019.01.037
[8] Lucas Borges Ferreira and Fernando França da Cunha. 2020. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Computers and electronics in agriculture* 178 (Nov. 2020), 105728. https://doi.org/10.1016/j.compag.2020.105728
[9] Chiou-Jye Huang and Ping-Huan Kuo. 2018. A deep CNN-LSTM model for particulate matter (PM2. 5) forecasting in smart cities. *Sensors* 18, 7 (10 July 2018), 2220. https://doi.org/10.3390/s18072220
[10] Vardaan Kishore and M Pushpalatha. 2017. Forecasting Evapotranspiration for Irrigation Scheduling using Neural Networks and ARIMA. Vol. 12. 10841–10847.
[11] Yufeng Luo, Seydou Traore, Xinwei Lyu, Weiguang Wang, Ying Wang, Yongyu Xie, Xiyun Jiao, and Guy Fipps. 2015. Medium range daily reference evapotranspiration forecasting by using ANN and public weather forecasts. *Water resources management* 29, 10 (19 May 2015), 3863–3876. https://doi.org/10.1007/s11269-015-1033-8
[12] Balduíno César Mateus, Mateus Mendes, José Torres Farinha, Rui Assis, and António Marques Cardoso. 2021. Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press. *Energies* 14, 21 (22 Oct. 2021), 6958. https://doi.org/10.3390/en14216958
[13] Paul CD Milly and Krista A Dunne. 2016. Potential evapotranspiration and continental drying. *Nature Climate Change* 6, 10 (06 June 2016), 946–949. https://doi.org/10.1038/nclimate3046
[14] Azlyna Senawi, Hua-Liang Wei, and Stephen A Billings. 2017. A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking. *Pattern Recognition* 67 (July 2017), 47–61. https://doi.org/10.1016/j.patcog.2017.01.026
[15] MR Sowmya, MB Santosh Kumar, and Sooraj K Ambat. 2020. Comparison of deep neural networks for reference evapotranspiration prediction using minimal meteorological data. In *2020 advanced computing and communication Technologies for High Performance Applications (ACCTHPA)*. IEEE, 27–33. https://doi.org/10.1109/ACCTHPA49271.2020.9213201
[16] Ammara Talib, Ankur R Desai, Jingyi Huang, Tim J Griffis, David E Reed, and Jiquan Chen. 2021. Evaluation of prediction and forecasting models for evapotranspiration of agricultural lands in the Midwest US. *Journal of Hydrology* 600 (Sept. 2021), 126579. https://doi.org/10.1016/j.jhydrol.2021.126579
[17] Axel Thomas. 2000. Spatial and temporal characteristics of potential evapotranspiration trends over China. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 20, 4 (march 2000), 381–396. https://doi.org/10.1002/(SICI)1097-0088(20000330)20:4%3C381::AID-JOC477%3E3.0.CO;2-K
[18] Lu Ye, Musaddak M Abdul Zahra, Najah Kadhim Al-Bedyry, and Zaher Mundher Yaseen. 2022. Daily scale evapotranspiration prediction over the coastal region of southwest Bangladesh: new development of artificial intelligence model. *Stochastic Environmental Research and Risk Assessment* 36, 2 (10 July 2022), 451–471. https://doi.org/10.1007/s00477-021-02055-4