Multivariate Time Series Evapotranspiration Forecasting using Machine Learning Techniques

(Article begins on next page)

Multivariate Time Series Evapotranspiration Forecasting using Machine Learning Techniques

UNIVERSITÀ DI TORINO

[1]Chalachew Muluken Liyew, [1]Rosa Meo, [2]Elvira Di Nardo, and [3]Stefano Ferraris

[1]Dipartimento di Informatica, Università degli Studi di Torino, Italy; [2]Dipartimento di Matematica "G. Peano", Università degli Studi di Torino, Italy; [3]DIST. Politecnico di Torino and Università degli Studi di Torino, Italy

INTRODUCTION

- Actual evapotranspiration (AET) is the loss of water by evaporation (from soil and water bodies) and by transpiration from the plants.
- When there is low rainfall and high evapotranspiration, 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.
- AET can be measured directly from a lysimeter device but it is not always possible.
- AET estimation by FAO-56PM equation is difficult because some features could be missing.
- We designed AET forecasting models by application of Machine Learning models using as inputs the most available meteorological variables.
- This work aims to assess AET forecasting models at a specific site: Cogne (Valle d'Aosta, Italy).
- We compared the results of AET forecasting models by application of several Machine Learning (ML) techniques: SARIMAX, LSTM, GRU, CNN, SVM, and RF).

SYSTEMATIC LITERATURE REVIEW

- Five electronic databases (Google scholar, Wiley online library, ACM digital library, Elsevier, and IEEEexplore) have been visited
- We retrieved 1854 papers of which 27 papers are selected using exclusion criteria
- Table 1 shows the most used features in the prediction of ET
- Table 2 shows the most used methods and evaluation parameters.

Table 1: Most used features on 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 (univariate TS)	7
Sunshine duration	4

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

ML methods	# of papers	Evaluation measures	# of papers
LSTM	11	RMSE	23
SVM	9	MAE	17
ANN and CNN	7	R2	12
RF	6	MSE	5
SARIMA	4	---	---

MATERIALS AND METHODS

- The dataset is collected from Cogne site, Italy (1.534m altitude, 45° 36′31.47″N7 °21′21.68″*E* of latitude and longitude)
- The missing values existed randomly and were imputed or predicted using a linear regression algorithm.
- Then the dataset was normalized into the [0, 1] interval.
- To select the relevant features for the prediction models, we applied correlation, tolerance and VIF methods.
  The result is a selection of five relevant features.
- They are the inputs for the AET forecasting models: statistical (SARIMAX), classical Machine Learning (SVM and RF), and deep learning (LSTM, GRU, and CNN).
- The performance of the models is measured and compared.

RESULTS

Table 3: Tolerance, and VIF score of independent variable

Variables	VIF	Tolerance	Re-VIF	Re-Tolerance
Net solar radiation	10.959	0.091	---	---
Net CO2	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
RH	2.167	0.461	2.167	0.461
Wind Speed	1.927	0.519	1.751	0.571

Table 4: 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

Conclusions

Five features have high correlation to AET and are used as inputs to ML models (with tolerance greater than 0.1 and VIF less than 10).

The deep learning models slightly outperform the statistical and the classical ML methods.

Among the deep learning models, LSTM outperforms the other ones.

The coefficient of determination of the LSTM method is 87.47.

Contact: chalachewmuluken.liyew@unito.it

UNIVERSITÀ DI TORINO