Invention Journal of Research Technology in Engineering & Management (IJRTEM)
ISSN: 2455-3689
www.Ijrtem. com Volume 4 Issue 1 ǁ January - February 2020 ǁ PP 01-05

Estimation of soil temperature and soil heat flux using machine learning techniques

[1,]Rahim Najafzadeh, [2,]Mehmed Gozaloghlu
[1,2,](Department of Environmental Science, College of agriculture/Ahar Azad University, Iran)

ABSTRACT : Soil moisture and soil temperature are key parameters in the land surface-atmosphere interaction. Modeling the soil moisture and soil temperature are needed for quantifying seasonal and annual variation of physical characteristics of soil. Models are useful tools to understand the behaviors and processes in the real world, and to make inferences about the future. In this study data driven models are used to model soil moisture and soil temperature. The data driven approach builds relationships between a set of independent and dependent variables, without worrying too much about the underlying processes, using statistical methods and machine learning (ML) techniques. However, the process-based models are built on well-established mathematical and physical models. ML, in general, is the scientific study of statistical algorithms and models, and artificial intelligence which is used to extract knowledge from data by learning from data without being explicitly programmed. ML is developing a prediction model which is optimized by providing training data, for which the correct output is already known. In this research the performance of ML models, including 1) multiple linear regression, 2) Ridge and Lasso regression, 3) Principal Component Regression (PCR), and 4) Support Vector Regression (SVR) in soil temperature, and soil heat flux estimation were evaluated. SVR model could predict soil temperature, and soil heat flux better than the other models.

KEYWORDS: Heat flux, Land surface models, Machine learning, Noah-MP

I. INTRODUCTION

Models are useful tools to understand the behaviors and processes in the real world, and to make inferences about the future. Basically, there are two modelling approaches including 1) data driven models and 2) process-based models (Ourang et. al 2017, JALALI at al. 2014). The first approach builds relationships between a set of independent and dependent variables, without worrying too much about the underlying processes, using statistical methods and machine learning (ML) techniques. On the other way, the process-based models are built on well-established mathematical and physical models. ML, in general, is the scientific study of statistical algorithms and models, and artificial intelligence which is used to extract knowledge from data by learning from data without being explicitly programmed. In the other words, ML is developing a prediction model which is optimized by providing training data, for which the correct output is already known. ML is combination of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning (JALALI et al. 2014). ML models specifically Artificial Neural Network (ANN) can estimate the weather parameters such as global solar radiation and soil parameters accurately (Barzin, Shirvani, and Lotfi 2017). In the recent years, many investigations applied ML techniques to estimate weather parameters such as land surface temperature (Tan et al. 2019), global parameters in LSMs (Chaney et al. 2016), and precipitation forecasts (Du, Liu, and Liu 2018).

The land surface models, as a process-based models, have been used in numerical weather prediction (NWP) models and in intra-seasonal to interannual climate predictions. The land surface is a critical element in the Earth-Atmosphere and climate system, which control the partitioning of energy at the earth surface between sensible and latent heat fluxes (Pitman 2003). The physical characteristics of the land surfaces control the energy and moisture balances. In addition, the vertical heterogeneity of the soil and vegetation affect the exchange of energy and water between the land surface and the atmosphere and make the earth surface hydrology process extremely complex (Orth, Dutra, and Pappenberger 2016). The land surface models (LSMs) have remarkably developed during the recent decade and have become more comprehensive models (Li et al. 2018, Damirchelli, at al. 2014). These models simulate soil moisture, soil temperature, skin temperature, snowpack depth, snowpack water equivalent, canopy water content, and the energy flux and water flux terms of the surface energy balance and surface water balance. LSMs is coupled with the Weather Research and Forecast (WRF) model that provides multi-options for atmospheric physical processes. However, despite an improvement in LSMs, the performance of the LSM is still a challenging task. As the model's complexity increase, and more parameters involve,

The uncertainty considerably increase and make it difficult to calibrate the model in different climate conditions. The LSMs use atmospheric data such as temperature, moisture, precipitation, solar radiation, and pressure forcing from the surface layer scheme, the radiation scheme, and the microphysics convective scheme all together with the land's state variables and land-surface properties, to estimate heat and moisture fluxes over land (Cai et al. 2014). There are many different versions of LSMs such as the thermal diffusion scheme (TDS), rapid update cycle (RUC) (Damirchelli, at al 2014), and Noah and Noah with multi-parameterization (Noah-MP). The Noah-MP has two version including standalone and coupled with WRF model. In both versions, the forcing fields are wind speed (m/s), wind direction (degree), air temperature (Kelvin), relative humidity (%), surface air pressure(millibar), incoming solar radiation(w/m2), incoming longwave radiation, and precipitation (mm). Noah-MP provide the following outputs: soil moisture (%), soil temperature(K), skin temperature(k), snow depth(m), snow water equivalent(mm), canopy water content, and surface energy(J), water, and CO2 fluxes. To calibrate the LSMs and address the uncertainties FLUXNET data used, which is an international network of sites that measure the land surface exchanges of carbon, water and energy using the eddy covariance technique (Williams et al. 2009).

NWP models has long been a difficult task for land surface parameters prediction due to the complexity of the atmospheric dynamics. Mathematical equations used to forecast the weather are sensitive to initial conditions; that is, slightly perturbed initial conditions might yield very different forecasts. Another challenge in NWP models is the role of model parameter uncertainty (Ghanbari, Gh, and Mohammad HadiFarahi,), particularly at unmonitored sites (Chaney et al. 2016). The primary objective of this study is to evaluate the performance of ML models, including 1) multiple linear regression, 2) Ridge and Lasso regression, 3) Principal Component Regression (PCR), and 4) Support Vector Regression (SVR), soil temperature, and soil heat flux estimation.

II. DATA AND METHODS

In this study, wind speed (m/s), wind direction (degree), air temperature (Kelvin), specific humidity (%), air pressure (millibar), shortwave solar radiation (w/m2), longwave solar radiation (w/m2) and precipitation (mm) data were used as a predictor variables to estimate soil temperature and heat flux. Soil temperature at two different depth including -2cm and -5cm, and soil heat flux were used as a response variable. These data were collected in the Bondville, Illinois FLUXNET station every 30 minutes overt January to June, 1998. The FLUXNET (https://fluxnet.fluxdata.org/data/) is a global network of micrometeorological tower sites that use eddy covariance methods to measure the exchanges of carbon dioxide, water vapor, and energy between terrestrial ecosystems and the atmosphere. More than 500 tower sites around the world are operating on a long-term basis. Figure 1 illustrate the correlation matrix of the input and output variables. It shows that the input and output variables are highly correlated. For instance, the correlation coefficient between Air temperature and soil temperature (-2cm and -5cm) are higher than 80 percent.

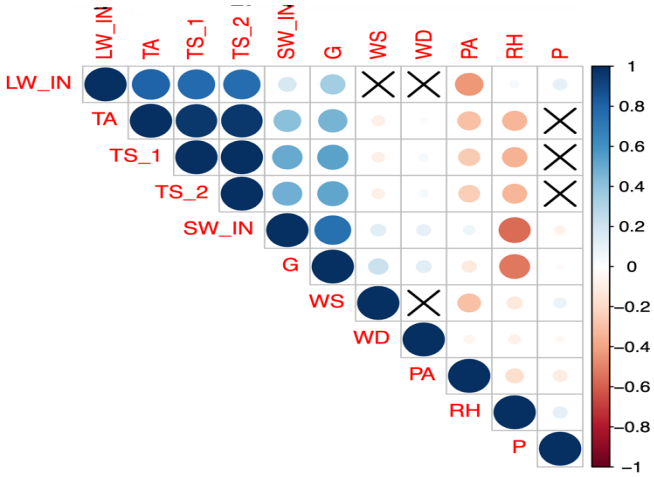

Figure 1. Input Variables Correlation

Since some variables are highly correlated, therefore, these data set could be as a good example to examine the statistical modeling approaches such as Ridge and Lasso regression, and PCR. However, Considering the nonlinear relationships, it will be shown that support vector regression performs better prediction on these data set (Pathak, Rohit, et al., 2018).

SVM offers very high accuracy classification method to separates data points using a hyperplane with the largest amount of margin. That is why an SVM classifier is also known as a discriminative classifier. Generally, SVM is considered to be a classification approach, it but can be employed in both types of classification and regression problems. It can easily handle multiple continuous and categorical variables. Similarly, PCA is multivariate statistical technique which is widely used in the atmospheric sciences. Principal component Analysis reduces a data set containing a large number of variables to a data set containing fewer new variables (Ghanbari, Gh, and Mohammad HadiFarahi, 2014). These new variables are linear combinations of the original ones, and these linear combinations are chosen to represent the maximum possible fraction of the inconsistency contained in the original data. In case of strong collinearity between the independent variables, such as Noah-MP forcing data, the estimated regression model would be unstable. To solve this problem, the PCR is used instead of regression model principal components as input variables in the regression modeling.

III. RESULTS AND DISCUSSION

In this study Ridge and Lasso regression models were used to find the best linear relationship between the weather observations such as air temperature, air pressure, relative humidity, solar radiation, and wind speed and direction. R-square and root mean square error statistics were used to evaluate the performance of the models. The R-square (R2) statistic, provides a measure of how well the model is fitting the real observation (Armin, et al. 2017 and R Damirchelli, et al. 2014). R2 varies between 0 and 1, and a number close to 1 does clarify the observed variance in the response variable. Generally, this statistic increase as more variables are included in the model. Therefore, Adjusted r2 is the preferred measure as it adjusts for the number of variables considered. In these models, the value of adj-R2 have shown in the below table. Based on the results, the linear regression model can estimate soil temperature at -2cm and -5cm accurately. However, the value of adj-R2 and Root Mean Square Error (RMSE) illustrate that the linear regression model is not an appropriate model to estimate soil heat flux. Table 1. Shows the root mean square (RMSE) and R-square values of estimated soil moisture and soil temperature at 5cm and 10 cm underground respectively. Fig. 1 shows the scatter plot of the estimated and observed 5sm soil temperature at the Bandville station during the study period.

Table 1.

Linear Models	R-square	RMSE(K)
Linear regression – 5cm soil temperature	0.95	2.35
Linear regression – 10cm soil temperature	0.93	2.53
Linear regression – 5cm soil moisture	0.66	19.15

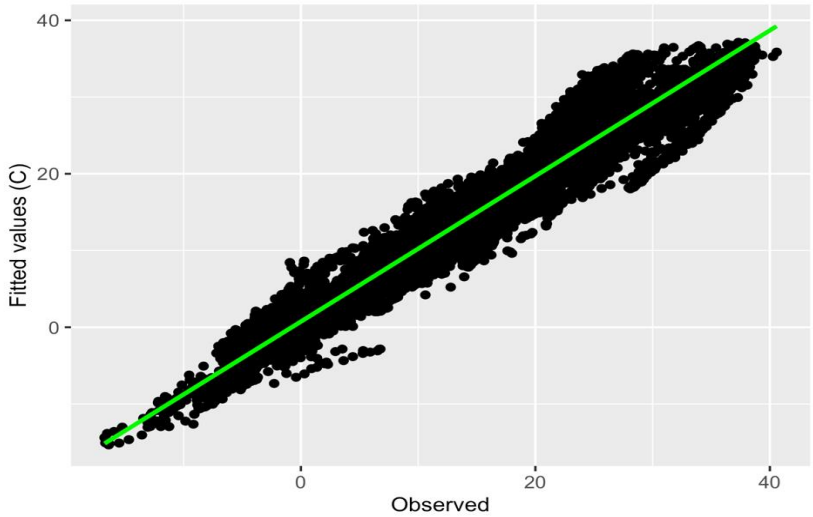

Figure 2. Estimated and observed soil moisture

As illustrated at this figure, the linear regression model could estimate the soil temperature accurately. However, because of the linear relationship between the predictands, principal component regression was used to capture the information of the weather observations and to estimate soil moisture and soil temperature in Bandville

weather station. Table 2. indicates the cumulative percentages of six eigenvalues, or variance of each principal component, and the cumulative percentages of variances that accounted for the principal components. For example, the first four principal components for the Bandville station explain 93.34% of the total variance of the predictor variables.

Table 2. Cumulative percentage from PC1 to PC6 variance in all

PCR	PC. 1	PC. 2	PC. 3	PC. 4	PC. 5	PC. 6
	83.05	85.10	90.28	93.34	94.43	96.78

Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). To estimate soil physical characteristics, the Support Vector Regression (SVR) model were used which is uses the same principles as the SVM for classification, with only a few minor differences. Results of SVR model shows that this method do better estimation than the other models used in this project. For instance, in the linear regression, Ridge regression, and PCR models the RMSE values of soil heat flux were around 19.5 (J/s), but in the SVR model RMSE significantly reduced to 14.6 (J/s).

IV. CONCLUSION

In this project the performance of four models including linear regression mode, Ridge/Lasso regression, PCR and SVM were evaluated to predict soil temperature and soil heat fluxes. As a result, the same results obtained from all models. However, SVR model could predict soil temperature, and soil heat flux better than the PCR and ridge and lasso regression models. In comparison to the linear fitting, shrinking methods could not improve the model performance. From this study, we have reached meaningful result that may be of interest of the weather forecasting model developers. The ML-simulated soil moisture can be used is the land surface models such as Noah MP model.

V. ACKNOWLEDGEMENTS

This work is supported by the SANA Science group, award FSR13AJ47H grant 31257087. The authors would like to thank Fluxnet network data providers and Dr. Salehi for his help in model setup, and for sharing the data, and for his help in the calculation of modeled.

REFERENCES

1. A Krenker, J. Bešter, and A. Kos, Introduction to the Artificial Neural Networks, In: Suzuki K (Ed), Artificial Neural Networks: Methodological Advances and Biomedical Applications, InTech, (2011), 1–18.http://www.intechopen.com/books/artificial-neural-networks-methodological-advances-and-biomedical-applications/introduction-to-the-artificial-neural-networks.
2. Barzin, Razieh, Amin Shirvani, and Hossein Lotfi. "Estimation of daily average downward shortwave radiation from MODIS data using principal components regression method: Fars province case study." International agrophysics 31, no. 1 (2017): 23-34.
3. C Xitian et al., Assessment of Simulated Water Balance from Noah, Noah-MP, CLM, and VIC over CONUS Using the NLDAS Test Bed, Journal of Geophysical Research Atmospheres, 119(24), (2014), 13–751.
4. Ch Nathaniel W., J. D. Herman, M. B. Ek, and E. F. Wood. Deriving Global Parameter Estimates for the Noah Land Surface Model Using FLUXNET and Machine Learning." Journal of Geophysical Research 121(22): 13,218-13,235.
5. Du Jinglin, Y. Liu, and Zh. Liu. Study of Precipitation Forecast Based on Deep Belief Networks, Algorithms 11(9), 2018, 1–11.
6. J Zavisa, R. Gall, and M. E. Pyle, NCAR technical note Scientific Documentation for the NMM Solver, (2010), https://opensky.ucar.edu/islandora/object/technotes:490.
7. Li Fuqin et al. Net and Solar Radiation Relations Over I R R I G a T E D Field, Atmospheric Research 4(2), (2018), 1–10
8. O Rene, E. Dutra, and F. Pappenberger, Improving Weather Predictability by Including Land Surface Model Parameter Uncertainty, Monthly Weather Review, 144(4), (2016), 1551–69.

9. Pathak, Rohit, Razieh Barzin, and Ganesh C. Bora. "Data-driven precision agricultural applications using field sensors and Unmanned Aerial Vehicle." International Journal of Precision Agricultural Aviation 1, no. 1 (2018).
10. O Armin, S. Pilehvar, M. Mortezaei, and R. Damircheli, Effect of aluminum doped iron oxide nanoparticles on magnetic properties of the polyacrylonitrile nanofibers, Journal of Polymer Engineering, 37, (2017), 135-141.
11. P, A. J. et al., The Evolution of, and Revolution in, Land Surface Schemes Designed for Climate Models, International Journal of Climatology, 23(5), (2003), 479–510.
12. T Jiancan et al., Deep Learning Convolutional Neural Network for the Retrieval of Land Surface Temperature from AMSR2 Data in China, Sensors 19(13), 2019, 2987.
13. Gh Ghanbari, M. H. Farahi, Optimal control of a delayed HIV infection model via Fourier series, Journal of Nonlinear Dynamics, (2014).
14. M Williams, et al. Improving Land Surface Models with FLUXNET Data, Biogeosciences 6(7), (2009), 1341–59.
15. Z Jalali, et al., Improving biocompatibility of the polyacrylonitlienanofibrous scaffolds, (2014).
16. R Damirchelli, et al., Electrospun Nanofibrous scaffolds based on alginate for skin tissue engineering., (2014).