Prediction of Drought based on NDVI – LST Relationship using Random Forest: Case Study of Garissa, Kenya

Mwana Said OMAR and Hajime KAWAMUKAI

Abstract: Forecasting of drought aids in decision making and it's essential for proper management and planning of resources. In this study, we propose a Random Forest (RF) algorithm to predict Normalized Difference Vegetation Index (NDVI) in Garissa, Kenya based on NDVI and Land Surface Temperature (LST) relationship. Performance of the RF was evaluated using root mean square error (RMSE) and mean absolute error (MAE). Results showed LST had a negative correlation with NDVI. The forecasting performance showed satisfactory accuracy and the method may be applied on larger areas.

Keywords: Drought, Random Forest, MODIS, Garissa, Kenya

1. Introduction

Drought is considered a costly hazard (Wilhite, 2000) as it erodes the assets of poor communities and undermines their livelihood strategies, culminating in a downward spiral of increasing poverty and food insecurity (Ndiritu, 2019). In addition, it affects water resources leading to reduced supply with adverse effects on agricultural production and socio-economic activities (Riebsame, et. Al, 1991).

Drought indexes based on remote sensing data such as NDVI have been commonly used to monitor agricultural drought (Peters et. al. 2002; Gu et. al. 2008) and can be used to identify vulnerable agricultural systems, understand past agricultural responses to drought and guide efforts to increase resilience to future drought (Nay, et.al. 2018).

LST is a crucial factor that increases the risk of droughts by causing increased evaporation and loss in soil moisture (Wilhite, et.al. 1985). Thus, combination of NDVI and LST provides a better understanding of droughts (Cai, et.al. 2011).

Numerous studies have used NDVI and LST in drought prediction and monitoring. Karnieli, et.al. (2010) combined NDVI and LST to monitor drought. Similar studies were carried out by Sun and Kaftatos (2007) as well as Hu, et.al. (2019).

The RF proposed by Breiman (2001) is an ensemble of decision tree algorithms that can be used for classification and regression problems (Brownlee, 2020). An RF fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control overfitting (Scikit-Learn Developers, 2020).

This study proposes an RF algorithm to predict drought based on the correlation between NDVI and LST. In contrast with the aforementioned studies, this study uses pixel-wise data in NDVI forecasting.

[1] Mwana Said Omar, [2] Hajime Kawamukai
*Mwana Said Omar (University of Hyogo)
〒672-8023 Himeji-Shi, Shirahamacho-Kou, 295-4 Sejyuru Shirahama 201
. Tel: 050-6865-8317. Email: saidmwana@gmail.com
**Hajime Kawamukai (University of Hyogo)
〒650-0047 Computational Science Building, 7-1-28, Minatojima Minami Machi, Chuo, Kobe, Hyogo, Japan
Tel: +81-78-303-1931. FAX: +81-78-303-1922. E-mail: kawamukai@ai.u-hyogo.ac.jp

2. Materials and Methodology

2.1 Study area

A section of Garissa county, Kenya was selected as a case study. The region covers an area of $4km^2$ and is located in the semi-arid region. The study area lies between latitudes 0° and 2° south and longitudes 38° and 40° east (Figure 1). The altitude of the region is between 20 m to 400 m above sea level (a.s.l). The temperature in the region is generally high ranging from 20° C to 38° C, with an average temperature of 36° C. The main economic activity is pastoralism (Government of Kenya, 2014).

2.2 NDVI and LST data

NDVI data was obtained from the MODIS sensor MOD13Q1 (v006) – Terra Vegetation Indices product at 250m spatial resolution and 16-day temporal resolution (Didan, 2015).

LST data was acquired from the MOD11A2 product at 1 km spatial resolution and 8-days temporal resolution (Wan, et.al., 2015).

The data covered approximately 19 years from February 2000 to September 2019. Both NDVI and LST data was extracted from the MODIS tile: h21v09. 8×8 pixels ($4km^2$) was extracted from the NDVI data and 2×2 pixels was extracted from the LST data. This was done in order to have a coordinate match at the pixel level.

2.3 Evaluation of the Random Forest

The performance metrics of the RF was evaluated using RMSE and MAE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \qquad (2)$$

where n is the number of observations, y_i is the observed value and $\hat{y}_i$ is the predicted value.

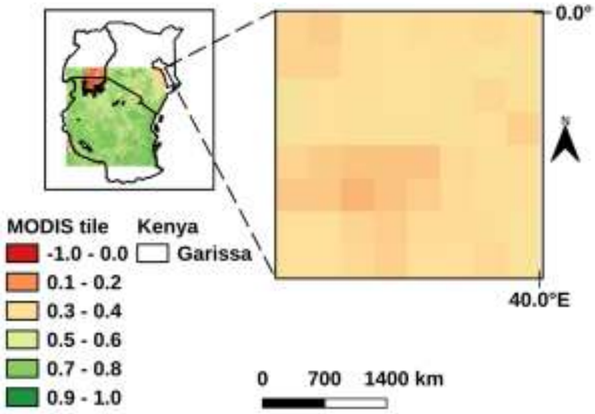

Figure 1: Location of the study area in Garissa county. Inset map shows East Africa countries and the MODIS NDVI tile used.

2.4 Experimental Procedure

We extracted pixel-wise data from NDVI and LST using python scripts. Monthly mean pixel-wise NDVI and LST was obtained by computing the monthly mean. A scale factor of 0.0001 was then applied to normalize the NDVI values between -1 to +1. For the LST data, a scale factor of 0.02 was applied and the temperature was converted from Kelvin to °C as shown in equation (3).

$$C = 0.02LST - 273.15 \qquad (3)$$

where: C is temperature in degree Celsius, 0.02 is the scale factor and LST is land surface temperature (LST_Day + LST_Night).

We converted the datasets into a pandas dataframe and created a csv file. Scikit-learn was used to implement train-test split, RF algorithm and the performance metrics. The train_test_split function was used to split the dataset into training and testing sets at a ratio of 80% to 20% respectively. Implementation of the RF algorithm was achieved using the RandomForestRegressor function. The sklearn.metrics function was then used to evaluate the performance of the algorithm.

3. Results

3.1 Time Series

Figure 2 (a) and (b) shows the time series of 1×1 pixel for LST and NDVI respectively.

The region is marked with high temperatures between the range of 30°C and 47°C as shown in Figure 2 (a). These high temperatures are due to the geographical location of the region.

Figure 2 (b) shows a dry area with poor vegetation. NDVI values ranged between 0.2 and 0.6 during the study period. The low NDVI values maybe due to the plants' phenological changes and climate patterns.

3.2 Correlation Analysis between NDVI and LST

The correlation between NDVI and LST is shown in Table 1 and using the scatterplot in Figure 3.

Table 1 shows that LST has a negative correlation with NDVI at -0.3301. The scatterplot shows negative correlation between LST and NDVI. This indicates that NDVI is high whenever LST is low. The small correlation value may be due to the nature of the data as shown in the time series in Figure 2 (a) and (b). The negative correlation is in agreement with previous studies (Sun and Kafatos, 2007; Guha and Govil, 2020).

3.3 Statistical Analysis

The results of the forecasting performance is shown in Table 2. The RF algorithm achieved satisfactory accuracy with RMSE of 0.0042 and MAE of 0.0483.

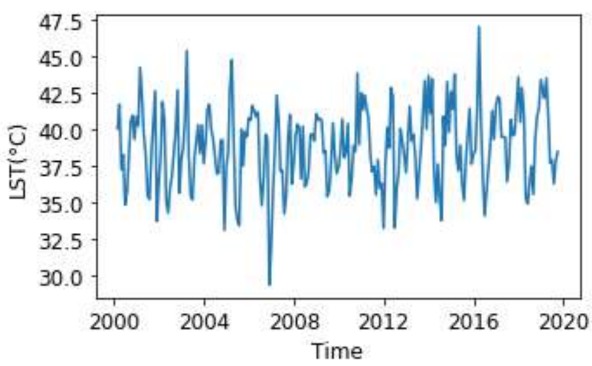

Figure 2 (a): 1×1 pixel LST time series from February 2000 to September 2019.

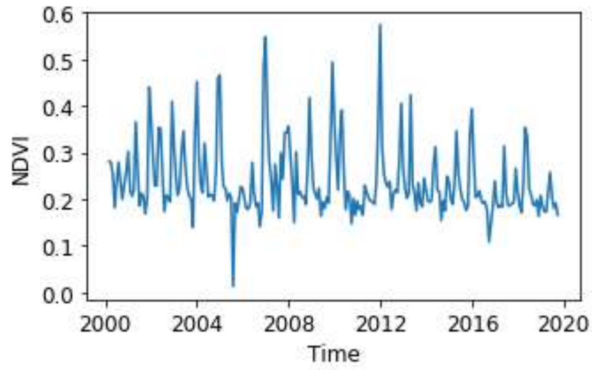

Figure 2 (b): 1×1 pixel NDVI time series from February 2000 to September 2019.

Table 1: Correlation coefficient (*r*) between LST and NDVI from 2000 to 2019

	LST	NDVI
LST	1.0000	-0.3301
NDVI	-0.3301	1.0000

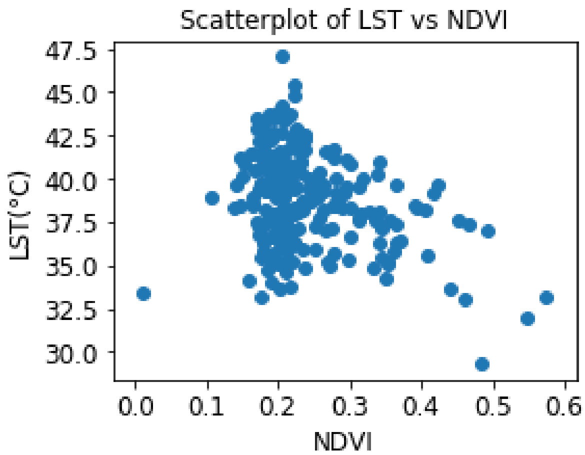

Figure 3. Scatterplot of LST versus NDVI from 2000 to 2019

Table2: Forecasting performance of the RF algorithm

RMSE	MAE
0.0042	0.0483

4. Conclusion

In this study, we propose a method of forecasting drought using MODIS NDVI and LST data through RF. NDVI and LST shows negative correlation. Future research can apply the proposed method on a larger area. The RF shows capability in predicting NDVI. The result of the correlation was influenced by the nature of the pixel data.

Acknowledgements

The authors acknowledge the support of the University of Hyogo in carrying out this study. We also thank the Land Processes Distributed Active Archive Center (LPDAAC) for providing the MODIS NDVI and LST data.

References

D. A. Wilhite. 2000. "Chapter 1 Drought as a Natural Hazard: Concepts and Definitions." *Drought Mitigation Center Faculty Publications, I: 3-18.* London. Routledge.

S. W. Ndiritu. 2019. "Drought Preparedness and Livestock Management Strategies by Pastoralists in Semi-Arid Lands: Laikipia North, Kenya." *Current Directions in Water Scarcity Research*, 2: 247-263.

W. E. Riebsame, et. al. 1991. "Drought and Natural Resources Management in the United States: Impacts and Implications of the 1987-1989 Drought." *Westview Press.* Boulder. Co.

A. J Peters, et. al. 2002. "Drought Monitoring with NDVI-Based Standardized Vegetation Index." *Photogrammetric Engineering & Remote Sensing* 68(1): 71-75.

Y. Gu, et. al. 2008. "Evaluation of NDVI and NDWI for Vegetation Drought Monitoring using Oklahoma Mesonet Soil Moisture Data." *Geophysical Research Letters* 35, L22401.

J. Nay, et. al. 2018. "A Machine Learning Approach to Forecasting Remotely Sensed Vegetation Health." *International Journal of Remote Sensing* 39(6): 1800-1816.

D. A. Wilhite and M. H. Glantz. 1985. "Understanding the Drought Phenomenon: The Role of Definitions." Water International, 10(3): 111-120

D. Sun and M. Kafatos. 2007. "Note on the NDVI-LST Relationship and the Use of Temperature-Related Drought Indices over North America." *Geophysical Research Letters*, 34, L24406

L. Breiman. 2001. "Random Forests." *Machine Learning*, 45: 5-32

J. Brownlee. 2020. "How to Develop a Random Forest Ensemble in Python." Accessed 2020-06-05. https://machinelearningmastery.com/random-forest-ensemble-in-python/

Scikit-Learn Developers. 2020. "3.2.4.3.1. sklearn.ensemble.RandomForestRegressor.". Accessed 2020-06-05. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

G. Cai, et.al. 2011. "Regional Drought Monitoring and Analyzing using MODIS Data- A Case Study in Yunnan Province." *Computer and Computing Technologies in Agriculture*, 243-251.

A. Karnieli, et.al. 2010. "Use of NDVI and Land Surface Temperature for Drought Assessment: Merits and Limitations." *Journal of Climate*, 23(3): 616-633

D. Sun and M. Kaftatos. 2007. "Note on the NDVI-LST Relationship and the Use of Temperature-Related Drought Indices over North America." *Geophysical Research Letters*, 34, L24406

X. Hu. Et. al. 2019. "Agricultural Drought Monitoring using European Space Agency Sentinel 3A Land Surface Temperature and Normalized Difference Vegetation Index Imageries." *Agricultural and Forest Meteorology*, 279: 107707

K. Didan. 2015. "MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2019-12-10 from https://doi.org/10.5067/MODIS/MOD13Q1.006

Z. Wan, et.al. 2015. "MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2019-12-10 from https://doi.org/10.5067/MODIS/MOD11A2.006

S. Guha and H. Govil. 2020. "An Assessment on the Relationship between Land Surface Temperature and Normalized Difference Vegetation Index." *Environment, Development and Sustainability*