
Comparison between the Holt-Winters and SARIMA Models in the Prediction of NDVI in an Arid Region in Kenya using Pixel-wise NDVI Time Series

Mwana Said Omar^a and Hajime Kawamukai^{a*}

^aDepartment of Policy and Management Informatics, Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan | *Corresponding author

Email: kawamukai@ai.u-hyogo.ac.jp

Abstract

Desertification is major issue in arid and semi-arid lands (ASAL) with devastating environmental and socio-economic impacts. Time series analysis was applied on 19 years' pixel-wise monthly mean Normalized Difference Vegetation Index (NDVI) data. The aim of this study was to identify a time series model that can be used to predict NDVI at the pixel level in an arid region in Kenya. The Holt-Winters and Seasonal Auto Regressive Integrated Moving Average (SARIMA) models were developed and statistical analysis was carried out using both models on the study area. We performed a grid search to optimise and determine the best hyper parameters for the models. Results from the grid search identified the Holt-Winters model as an additive model and a SARIMA model with a trend autoregressive (AR) order of 1, a trend moving average (MA) order of 1 and a seasonal MA order of 2, with both models having a seasonal period of 12 months. It was concluded that the Holt-Winters model showed the best performance for 600×600 pixels (MAE = 0.0744, RMSE = 0.096) compared to the SARIMA model.

Keywords: Arid, Desertification, Kenya, MODIS NDVI, SARIMA, The Holt-Winters model.

1. Introduction

ASALs are susceptible to droughts and flooding and are at an increased risk of desertification, which in turn threatens livelihoods, food security and biodiversity (IUCN, 2020; UNDP, 2013). Desertification is defined as land degradation in drylands (ASAL and dry sub-humid areas) as a result of climate change and anthropogenic activities resulting in long term loss in biological productivity and ecological integrity (UNCCD, 1994; Mirzabaev, et.al. 2019; Huang, et.al 2015). Drylands cover 46.2% of the earth's surface and are home to 3 billion people (Mirzabaev, et.al. 2019) including the arid and semi-arid (ASAL) regions. The arid areas contain barren and sparse vegetation mostly covered with shrubland and grassland while the semi-arid and dry-humid regions are covered with grassland (Huang, et.al 2015). Desertification hotspots denoted by a decline in vegetation productivity extended to 9.2% of drylands between 1980s and 2000s, with a projected increase due to climate change and increased human activities (Mirzabaev, et.al. 2019). The ASAL constitutes over 80% of the total land in Kenya and accounts to 30% of the country's population who are engaged in pastoralism and small scale agriculture (IUCN, 2020). According to studies in 1997, 64% of Kenya's land likely experienced moderate desertification while 23% of the land was at risk of severe to very severe desertification (Macharia, 2004). A recent study by Gichenje and Godinho, (2018) found that 21.6% of the country experienced persistent negative trends in vegetation during 1992 – 2015 period.

Limited and precise field data for vegetation production (Chamailé-Jammes and Fritz, 2009) has led to drawbacks in monitoring and forecasting vegetation (Wang et. al. 2018), thus, remote sensing has become an indispensable tool that provides information about ecological processes and time series of different ecological variables can be obtained from such data (Fernández-Manso, et.al., 2011). Remotely sensed vegetation indices such as NDVI are widely used to detect greening and browning trends (de Jong, et.al. 2011) and can reveal regions with thriving or stressed vegetation.

Time series models commonly used in predictive analysis in econometrics such as the Holt-Winters and SARIMA models have been increasingly used for monitoring ecological and environmental variables.

Fernández-Manso, et.al., (2011) developed a SARIMA model and elaborated a short-term forecast of the NDVI in each 10-day period using 10-days maximum value composite (MVC) bands of the NDVI obtained from National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) data. They concluded that time series models could be used for vegetation monitoring at regional level. Riaño, et.al., (2007) used 18 years' time series monthly NOAA- AVHRR and demonstrated that SARIMA model is a powerful tool for forecasting potential burned area with significant statistical results. Ji and Peters (2004) used the conterminous U.S 1-km AVHRR NDVI dataset obtained from the Earth Resources Observation Systems (EROS) Data Centre, U. S Geological Survey to design a vegetation forecast greenness (VGF) model based on the SARIMA model. They used the model to predict short-term vegetation status in cropland and grassland up to 12 weeks in advance, with higher R^2 achieved for shorter predictions. Mutti, et. al. (2020), modelled NDVI using a lower resolution MODIS MOD13A2 product, at 500m spatial resolution and compared the performance of the Holt-Winters and SARIMA models at desertification hotspots in Brazil. They concluded that the use of non-pixel-wise models limits their applicability in forecasting degradation patterns in vulnerable regions.

The above mentioned studies used coarse and lower resolution NDVI data. Thus, this study used the finer resolution MODIS MOD13Q1 Terra Vegetation Indices product, at a spatial resolution of 250m. We used 19 years of NDVI data from February 2000 to September 2019 to develop pixel-wise time series models based on the Holt-Winters and SARIMA models to predict vegetation in an arid region in Kenya. We analysed the temporal trends in NDVI during the study period and compared the performance of the Holt-Winters and SARIMA models in NDVI prediction using pixel-wise NDVI time series.

2. Materials and Methods

2.1 Study area

The study area was carried out in a semi-arid to arid region in Kenya,

Located in the eastern part of the country in the Middle Tana River Basin region (Figure 1). 600×600 pixels, covering an area of $22,500 \text{ km}^2$ was used in the study. The region is located at coordinates 0° and 2° S and 38° and 40° E . The elevation is below 1300 m (Knoop, et.al., 2012) and the region receives a bimodal rainfall below 800 mm (Baker, et.al., 2015) and an annual temperature between 24° C to 34° C . Pastoralist grazing land, dry land farming and dryland forestry are practised due to the semi-arid to arid nature of the region (Knoop, et.al., 2012). The region has a variety of vegetation predominately bushland, woodland and a few patches of forest (Baker, et.al., 2015). However, it faces land degradation due to uncontrolled sand mining, removal of trees and overgrazing (Knoop, et.al., 2012), making the region more vulnerable to desertification. In addition, further land degradation occurs due to charcoal production as an alternative means for survival among the pastoralist communities (Baker, et.al., 2015).

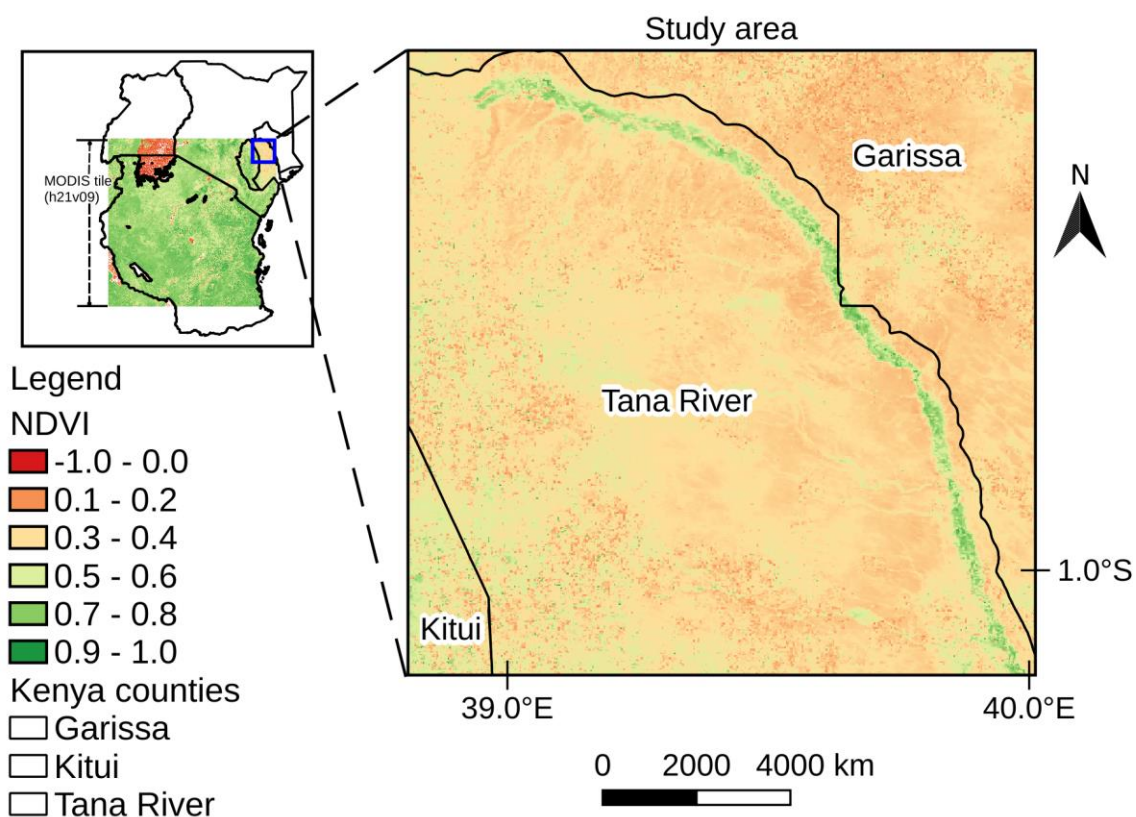


Figure 1. The study area (on the right side), corresponds to an area covering 600×600 pixels ($22,500 \text{ km}^2$) located in the Middle Tana River Basin in Kenya.

Inset map (on the left side) shows the East Africa countries (Kenya, Tanzania and Uganda) and the MODIS tile (h21v09) used.

2.2 Remote sensing data

NDVI data was obtained from the MOD13Q1 Terra Vegetation Indices product at 16-days composite period at a finer spatial resolution of 250m, recommended for pixel-wise modelling. The data is freely available and can be downloaded from the USGS EarthExplorer. A total of 452 images were used covering a period of 19 years, from February 2000 to September 2019. The data was derived from the MODIS tile: h21v09, which covers parts of Kenya, Tanzania and Uganda (inset map in Figure 1). 600×600 pixels were extracted from the top right corner of the tile to analyse and model the pixel-wise NDVI time series.

Monthly mean pixel-wise NDVI time series was created by calculating the average NDVI per month from the 16-days composite period. This was achieved by creating a pandas dataframe to create monthly time series data. We divided 80% of the mean monthly pixel-wise data into a training dataset which was used to train the models and 20% as a testing dataset to validate the models' predictions.

2.3 Time series models

Pixel-wise NDVI time series were modelled using the Holt-Winters and SARIMA models and was implemented using Python programming language (Python Software Foundation, 2020). Both models were implemented using the statsmodels library; a python module that provides classes and functions for the estimation of statistical models, conducting statistical tests and statistical data exploration (Perktold, et.al. 2020). We configured each model using a grid search (Brownlee, 2018a; 2018b) to automate and identify the hyperparameters resulting in the best performance for the Holt-Winters and SARIMA models that produced the best fit models to the pixel-wise NDVI time series data.

2.3.1 The Holt-Winters model

The Holt-Winters model developed by Holt (2004) and Winters (1960) also known as Triple Exponential Smoothing supports the trend and seasonal component in time series data. The model consists of three smoothing parameters associated with the level (α), trend (β) and seasonality (γ). The grid search returned the best parameters for the Holt-Winters model as add, add and 12 for the level, trend and seasonality parameters respectively, resulting in an additive Holt-Winters model, which was applied to the pixel-wise NDVI time series data. The additive model was used as the trend and seasonality in our data does not increase over time. In addition, previous studies applied the additive model with significant results (Mutti, et. al., 2020). The three smoothing equations associated with the additive Holt-Winters model for a series $\{x_t\}$ are as follows (Cowpertwait and Metcalfe, 2009):

$$a_t = \alpha(x_t - s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1}) \quad (1)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \quad (2)$$

$$s_t = \gamma(x_t - a_t) + (1 - \gamma)s_{t-p} \quad (3)$$

Where: a_t , b_t and s_t are the estimated level, trend and seasonal effect respectively at time t , p is the period, and α , β and γ are the smoothing parameters.

The forecasting equation for x_{n+k} made after the observation at time n is shown below (Cowpertwait and Metcalfe, 2009):

$$\hat{x}_{n+k|n} = a_n + kb_n + s_{n+k-p} \quad k \leq p \quad (4)$$

We imported the class: `statsmodels.tsa.holtwinters.ExponentialSmoothing` from the `statsmodels` library to implement the Holt-Winters model. We specified the configuration for the Holt-Winters model by defining the parameters as `seasonal_periods = 12`, `trend = 'add'`, `seasonal = 'add'`. The `seasonal_periods` denotes the number of periods in a complete seasonal cycle, in this case: 12, for the 12 months in a year,

The trend and seasonal defines the type of trend component and type of seasonal component respectively and can either be additive or multiplicative. The model was then fitted to the training data using the fit () function to select the smoothing parameters for the level, trend and seasonality. The forecast () function returns a one-step forecast and was used to make the predictions.

2.3.2 SARIMA model

The SARIMA model is an extension of the Auto Regressive Integrated Moving Average (ARIMA) model developed by Box et. al, (1994) to support seasonality in time series data. The SARIMA model requires few parameters to describe a time series which exhibit non-stationary within and across seasons and is represented as:

$$SARIMA(p, d, q)(P, D, Q) \quad (5)$$

Where (p, d, q) is the non-seasonal part of the model, (P, D, Q) is the seasonal part of the model and s is the length of the seasonal period (Durdu, 2010). The model is expressed as shown in equation (6):

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (6)$$

where: p is the order of the non-seasonal AR, d is the number of regular differencing, q is the order of non-seasonal MA, P is the order of the seasonal AR, D is the number of seasonal differencing, Q is the order of seasonal MA, Φ is the seasonal AR parameter of order P , Θ is the seasonal MA parameter of order Q .

We defined the SARIMA model by calling an instance of the class SARIMAX from the statsmodels library and also defined the model parameters (from the result of the grid search) of order (1, 0, 1) and seasonal_order = (0, 0, 2, 12). The order argument defines the AR parameters, differences and MA parameters of the model as 1, 0, and 1 respectively, while the seasonal_order argument denotes the (P, D, Q, s) order of the seasonal component of the model, with the length of the seasonal period as 12, for the 12 months' yearly data.

Similarly to the Holt-Winters model, we fitted the model to the training data using the fit() function before finally making the predictions using the forecast() function. Detailed explanation of the functions and classes can be found in the statsmodels module (Perktold, et.al. 2020).

Figure 2 shows the summary of the flowchart used in the study.

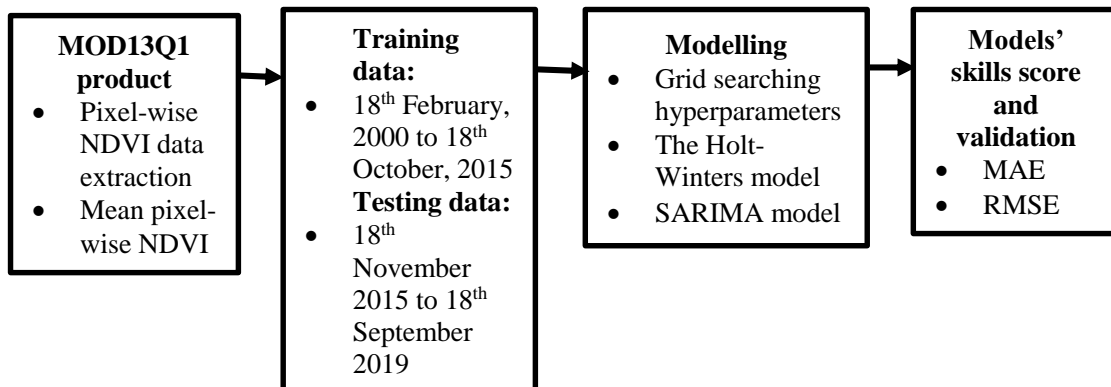


Figure 2. Flowchart of the study

2.4 Models' skills score

The skills score for the Holt-Winters and SARIMA models was evaluated using the mean absolute error (MAE) and root mean square error (RMSE), which are common metrics used in model evaluations. The metrics were implemented using scikit-learn, a python module that provides tools for predictive data analysis (Pedregosa, et.al. 2011). Both metrics were imported from the sklearn.metrics module.

MAE is the arithmetic average of the absolute values of the differences between the members of each pair. MAE is zero if the forecasts are perfect and increases as discrepancies between the forecasts and the observations become larger (Wilks, 2011). RMSE represents the individual quadratic differences between the observed and adjusted data (Camelo, et. al., 2018). MAE and RMSE are expressed as:

$$MAE = \frac{1}{n} \sum_{k=1}^n |y_k - o_k| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2} \quad (8)$$

Where (y_k, o_k) is the k th of n pairs of forecasts and observations

3. Results and Discussion

3.1 NDVI time series trend

Figures 2 (a) and 2 (b) shows the 1×1 NDVI time series for the training, testing and predicted values using the Holt-Winters and SARIMA models respectively. The models were tested on the sample data from September 2015 to September 2019 and a 6 months' lead prediction was made. The observed NDVI ranged between 0.0 to 0.6, with the lowest observed value at 0.0, indicating that the region experienced scarce vegetation during that year. The SARIMA model performed poorly compared to the Holt-Winter model as it underestimates the predictions as the series consistently increased over time.

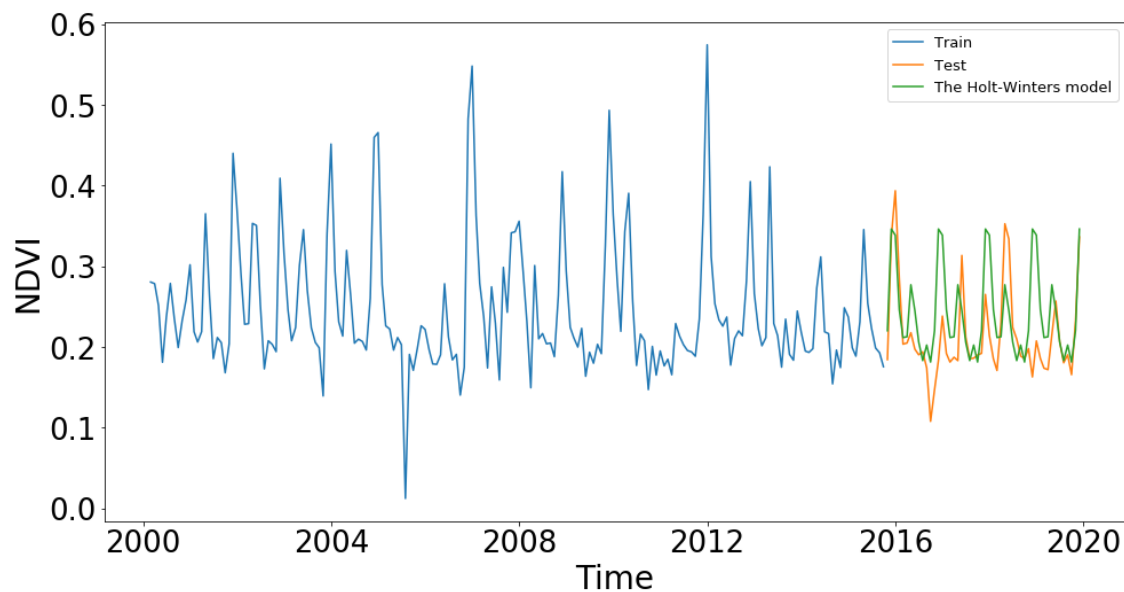


Figure 3 (a). 1×1 pixels monthly mean NDVI time series. The 6 months' predictions were performed using the Holt-Winters model

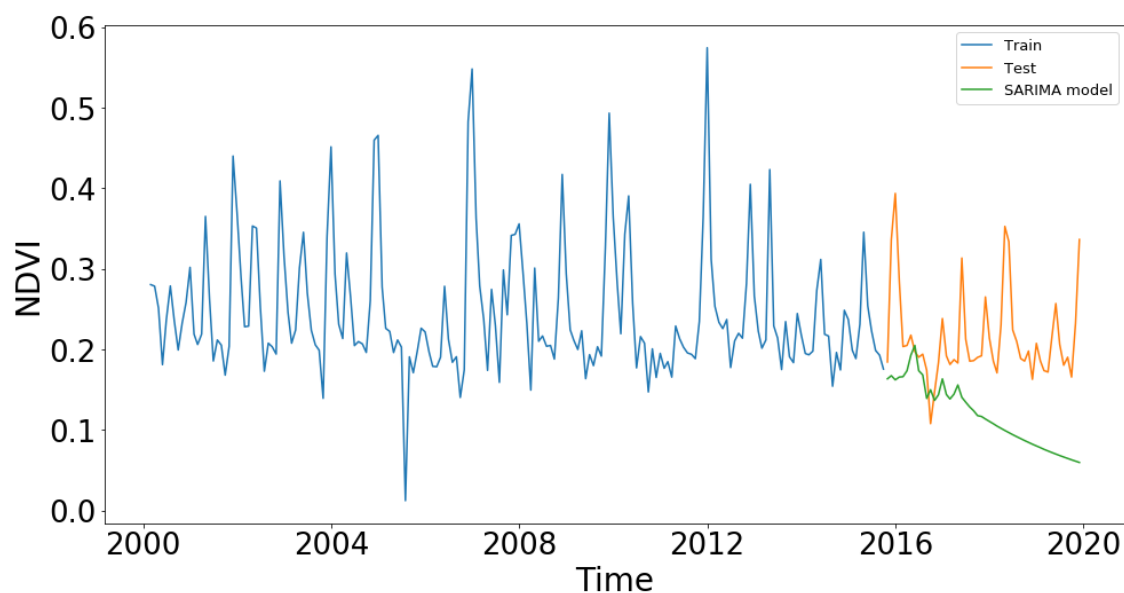


Figure 3 (b). 1×1 pixels monthly mean NDVI time series. The 6 months' predictions were performed using the SARIMA model

3.2 Models evaluation

For this study, we chose the models which had the best performance from the grid search. The models evaluation metrics for the mean pixel-wise NDVI are presented in Table 1 and Table 2 for the 1×1 and 600×600 pixels respectively. The performance of the Holt-Winters model was best compared to the SARIMA model for both 1×1 and 600×600 pixels.

Table 1. Models evaluation for the monthly mean pixel-wise NDVI for 1×1 pixels

Model	MAE	RMSE
SARIMA(1, 0, 1)(0, 0, 2)12	0.0997	0.1185
The Holt-Winters additive	0.0443	0.0037

Table 2. Models evaluation for the monthly mean pixel-wise NDVI for 600×600 pixels

Model	MAE	RMSE
SARIMA(1, 0, 1)(0, 0, 2) ₁₂	0.1112	0.1344
The Holt-Winters additive	0.0744	0.096

4. Conclusion

The aim of this study was to develop the Holt-Winters and SARIMA models for predicting NDVI in an arid region in Kenya. We analysed the temporal trends in NDVI during the study period and evaluated the performance of the Holt-Winters and SARIMA models in NDVI prediction using pixel-wise NDVI data. Results from the 1×1 and 600×600 pixels showed that the Holt-Winters model is more accurate than the SARIMA model based on the MAE and RMSE since it gives more significance to recent observations.

The approach proposed in this study covered a few pixels due to the high computational power and cost involved in pixel-wise modelling. Future research perspective would be to explore more pixels covering a larger area and compare the models performance in different desertification hotspots.

Acknowledgement

We are grateful to the Land Processes Distributed Active Centre (LPDAAC) for providing the NDVI data.

Disclosure statement

The authors declare no conflicts of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not for profit sectors.

References

- Baker, T., et.al. 2015. “Baseline Review and Ecosystem Services Assessment of the Tana River Basin, Kenya.” International Water Management Institute. Colombo, Sri Lanka. IWMI Working Papers. 165.
https://www.iwmi.cgiar.org/Publications/Working_Papers/working/wor165.pdf
- Box, G. E. P., G. M. Jenkins, G. C. Reinsel. 1994. “Time Series Analysis: Forecasting and Control.” Prentice Hall, Englewood Cliffs, NJ, USA.
- Brownlee, J. 2018a. “How to Grid Search Triple Exponential Smoothing for Time Series Forecasting in Python.” <https://machinelearningmastery.com/how-to-grid-search-triple-exponential-smoothing-for-time-series-forecasting-in-python/>
- Brownlee, J. 2018b. “How to Grid Search SARIMA Hyperparameters for Time Series Forecasting.” <https://machinelearningmastery.com/how-to-grid-search-sarima-model-hyperparameters-for-time-series-forecasting-in-python/>
- Camelo, H. D. N., et. al. 2018. “Innovative Hybrid Models for Forecasting Time Series Applied in Wind Generation based on the Combination of Time Series Models with Artificial Neural Networks.” Energy. 151, 347 – 357.
<https://doi.org/10.1016/j.energy.2018.03.077>
- Chamaillé-Jammes, S. and H. Fritz. 2009. “Precipitation-NDVI Relationships in Eastern and Southern African Savannas Vary Along a Precipitation Gradient.” Int. J. Remote Sens. 30(13), 3409 – 3422.
<https://doi.org/10.1080/01431160802562206>
- Cowpertwait, P. S. P, A. V. Metcalfe. 2009. “Introductory Time Series with R.” Springer, New York City, USA.
- de Jong, R., et.al. 2011. “Analysis of Monotonic Greening and Browning Trends from Global NDVI Time Series.” Remote Sens. Environ. 115, 692 – 702.
<https://doi.org/10.1016/j.rse.2010.10.011>
- Durdu, Ö. F. 2010. “Application of Linear Stochastic Models for Drought Forecasting in the Büyük Menderes River Basin, Western Turkey.” Stoch. Environ. Res. Risk. Assess. 24, 1145-1162. <https://doi.org/10.1007/s00477-010-0366-3>

- Fernández-Manso, A., C. Quintano and O. Fernández-Manso. 2011. “Forecast of NDVI in Coniferous Areas using Temporal ARIMA Analysis and Climatic Data at a Regional Scale.” *Int. J. Remote Sens.* 32 (6), 1595-1617.
<http://dx.doi.org/10.1080/01431160903586765>
- Gichenje, H., S. Godinho. 2018. “Establishing a Land Degradation Neutrality National Baseline through Trend Analysis of GIMMS NDVI Time Series.” *Land Degrad. Dev.* 29 (9), 2985 – 2997. <https://doi.org/10.1002/ldr.3067>
- Holt, C. C. 2004. “Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages.” *Int. J. Forecast.* 20 (1), 5 – 10.
<https://doi.org/10.1016/j.ijforecast.2003.09.015>
- Huang, J., et.al. 2015. “Global Semi-Arid Climate Change Over Last 60 Years.” *Clim. Dyn.* 46, 1131 – 1150. <https://doi.org/10.1007/s00382-015-2636-8>
- IUCN. 2020. “Kenya – Arid and Semi-Arid Lands.” International Union for Conservation of Nature. <https://www.iucn.org/restoration-initiative/projects/kenya-arid-and-semi-arid-lands>
- Ji, L. and A. J. Peters. 2004. “Forecasting Vegetation Greenness with Satellite and Climate Data.” *IEEE Geosci. Remote Sen. Lett.* 1(1), 3 – 6.
<https://doi.org/10.1109/LGRS.2003.821264>
- Knoop, L., F. Sambalino and F. Steenbergen. 2012.: *Securing Water and Land in the Tana River Basin: A Resource Book for Water Managers and Practitioners.* Wageningen, The Netherlands: 3R Water Secretariat.
- Macharia, P. 2004. “Gateway to Land and Water Information: Kenya.” Land and Water Development Division, Food and Agriculture Organization of the United Nations, Rome. <http://www.fao.org/land-water/en/>
- Mirzabaev, A., et. al. 2019. “Chapter 3: Desertification” in *Climate Change and Land: an IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems.*
https://www.ipcc.ch/site/assets/uploads/2019/11/06_Chapter-3.pdf

- Mutti, P. R., P. S. Lúcio., V. Dubreuil, and B. G. Bezerra. 2020. “NDVI Time Series Stochastic Models for the Forecast of Vegetation Dynamics over Desertification Hotspots.” *Int. J. Remote Sens.*, 41(7), 2759 – 2788.
<https://doi.org/10.1080/01431161.2019.1697008>
- Pedregosa, F., et. al. 2011. “Scikit-learn: Machine Learning in Python”. *J. Mach. Learn. Res.* 12, 2825 – 2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- Perktold, J., S. Seabold, and J. Taylor. 2020. “Statsmodels.” February, 2020.
<https://www.statsmodels.org/stable/index.html>
- Python Software Foundation. “Python.” <https://www.python.org/>
- Riaño, D. J. A. M. Ruiz, J. B. Martínez, S. L. Ustin. 2007. “Burned Area Forecasting using Past Burned Area Records and Southern Oscillation Index for Tropical Africa (1981 – 1999).” *Remote Sens. Environ.* 107 (4), 571 – 581.
<https://doi.org/10.1016/j.rse.2006.10.008>
- UNCCD. 1994. United Nations Convention to Combat Desertification in those Countries Experiencing Serious Drought and/or Desertification, Particularly in Africa. Paris, France: A/AC, 241-27, United Nations Convention to Combat Desertification. https://www.unccd.int/sites/default/files/relevant-links/2017-01/UNCCD_Convention_ENG_0.pdf
- UNDP. 2013. “Combating Desertification in Kenya: Emerging Lessons from Empowering Local Communities.” United Nations Development Programme. Nairobi, Kenya.
- Wang, J., et. al. 2018. “Optimization of the Time Series NDVI-Rainfall Relationship using Linear Mixed-Effects Modeling for the Anti-Desertification Area in the Beijing and Tianjin Sandstorm Source Region.” *Theor. Appl. Climatol.* 132, 1291 – 1301. <https://doi.org/10.1007/s00704-017-2144-3>
- Wilks, D. S. 2011. “Chapter 8 – Forecast Verification.” *Int. Geophys. Ser. in Statistical Methods in Atmospheric Sciences*, Vol. 100, pp. 301 - 394, Academic Press.
<https://doi.org/10.1016/B978-0-12-385022-5.00008-7>

Winters, P. R. 1960. “Forecasting Sales by Exponentially Weighted Moving Averages.” *Manage. Sci. INFORMS*. 6 (3), 324 – 342.

<https://doi.org/10.1287/mnsc.6.3.324>

Copyright © 2021 **Mwana Said Omar^a** and **Hajime Kawamukai^{a*}**, AJRSP. This is an open-access article distributed under the terms of the Creative Commons Attribution

License (CC BY NC). <https://doi.org/10.52132/Ajrsp/en.2231>