South Africa

Using machine learning to map bioclimatic zones and crop yields in water-scarce conditions

Case study prepared by:

H. Mugiyo, V.G.P. Chimonyo, R. Kunz, M. Sibanda, L. Nhamo, C. Ramakgahlele Masemola, Tsitsi Bangira, A.T. Modi and T. Mabhaudhi

Focus

Sub-Saharan Africa has severe dry spells associated with the annual Southern Oscillation, including the El Niño-Southern Oscillation (ENSO), causing worldwide precipitation and temperature anomalies. Since 1900, 80% of severe droughts in the region have been connected to El Niño episodes and a record-breaking ENSO-induced drought in 2015/2016 affected agricultural, water, food and nutrition security (Yan *et al.*, 2023)we investigated their interactive effects on tropical net ecosystem productivity (NEP.

South Africa is characterized by a mild, temperate climate, where only a small proportion of land (10.3%) is considered arable land, for agriculture. It is a water-scarce country with 61% of landmass receiving less than the minimum rainfall to support successful, rainfed farming. Evidence shows climate change has increased drought frequency and severity in the country. Drought, therefore, is a significant threat to crop production, water resources and, more importantly, food and nutrition security.

Mapping drought zones

Droughts can be meteorological, agricultural, hydrological or socioeconomic. Agricultural drought includes complicated soil water stress, vegetation growth and precipitation loss. Mapping drought-prone zones and predicting drought severity are crucial to reducing the impacts of shocks related to water scarcity. But given there are four types of drought, and given there are 150 available drought indices tracking multiple variables, it has proved difficult to map drought risk zones accurately and effectively. Standard indices can be used, but they require additional observations to compute weights, and although data mining strategies solve some limitations, these also have constraints.

Machine learning models can mitigate for these issues and be used to develop accurate drought data. Therefore, this study integrates existing indices, machine learning and data mining strategies in its model generation.

Matching under-utilized crop species

Sustainable farming practices can alleviate some of the impacts of drought. These practices include utilizing crops that successfully balance yield with environmental concerns, human health and general well-being. Recent research has focused on planting neglected and under-utilized crop species (NUS). These are typically wild and/or cultivated species that were once favoured, but have since been overlooked (Bvenura and Afolayan, 2015; Chivenge *et al.*, 2015)

NUS have lower yields, but they are nutritious, inexpensive and readily accessible. Most importantly, they are resistant to a variety of stresses (e.g. heat, salinity, drought), and thrive with minimal attention and fewer pesticides or fertilizers than commercial crops (Akinola *et al.*, 2020; Mabhaudhi *et al.*, 2019; Mohd Nizar *et al.*, 2021).

It is important to grow these species without affecting existing major crops, which are typically grown in arable and productive lands. Identifying drought-prone areas is also important since NUS can thrive in marginal lands – as compared to maize and wheat which require larger areas and more intensive farming.

This case study, therefore, uses an integrated 'hybrid' model that draws on existing indices and employs machine learning to identify bioclimatic zones with high rainfall variability in water-scarce conditions (drought risk zones). The resultant data is then used to match NUS with the selected, appropriate zones.

Team

The team included researchers at the University of KwaZulu-Natal, Pietermaritzburg, South Africa; the International Maize and Wheat Improvement Centre, Harare, Zimbabwe; the University of the Western Cape, Bellville, South Africa; the Water Research Commission of South Africa; the Council for Scientific and Industrial Research, Pretoria, South Africa; and the International Water Management Institute, Accra, Ghana.

Methods and models

The study used the Vegetation Drought Response Index (VegDRI); a 'hybrid' index that gleans data from existing climate tracking indices with regards rainfall (the Standardized Precipitation Index (SPI)), temperature (the Temperature Condition Index (TCI)), and vegetation (the Vegetation Condition Index (VCI)), to show the bioclimatic zones in South Africa that have both high rainfall variability and little water.

Rigorous machine learning techniques were used in the development of VegDRI. First, historical satellite climate data (1981–2019) was integrated with land use and cover maps of South Africa to generate five scales of drought, ranging from 'very severe' to 'no drought'. After that, a machine learning algorithm, the Classification and Regression Tree (CART) (Breiman, 2001) was used to produce a new dataset and

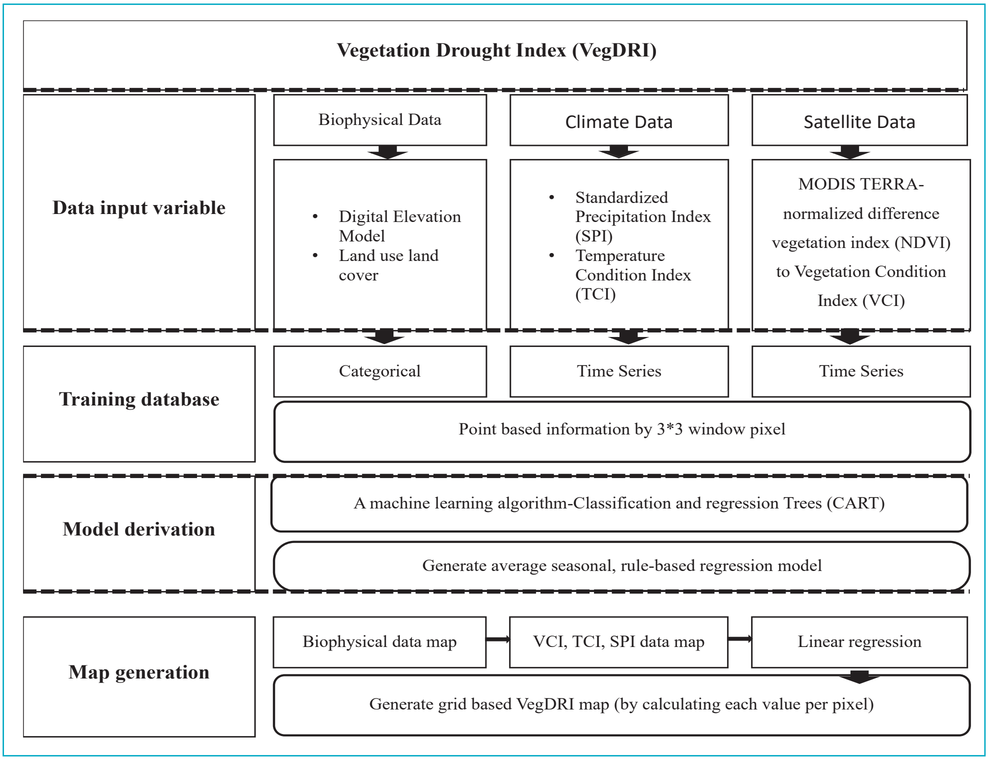

Figure 1: Flow chart of machine learning steps involved in generating the Vegetation Drought Response Index (VegDRI)

create map graphics. 80% of that resultant dataset was used for training and 20% for validation of the training model (a typical split to ensure accuracy in machine learning models). The dataset was then randomly sampled and split into calibration and validation datasets. This procedure was implemented 100 times to evaluate the stability of the model. The methodology is visually represented in Figure 1.

Drought index evaluation

Average sorghum yields obtained at the district level from official sources, were used to validate the results obtained from the mapping exercise. Farming households in the district were randomly selected as project focus areas. VegDRI, VCI, TCI, and SPI were then correlated against drought-tolerant crop yield data. The predictive accuracy of the drought risk maps was then computed from a pixel-by-pixel comparison using weighted Kappa statistics. The Kappa statistic is used to test how far the data collected in a given study represents the variables measured (Heikkinen *et al.*, 2006) In this case it was used to measure agreements between dry zones and sorghum-growing areas.

Results

Precipitation evaluation

Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) precipitation data was highly correlated with observed, in-situ weather data across all weather stations used in South Africa. Based on these results, CHIRPS datasets can be used confidently for agricultural drought analysis.

VegDRI

The VegDRI map (Figure 2) shows the locations and variations of drought intensity in South Africa. The five scales of intensity were classified as very severe drought (16%), severe drought (34%), moderate drought (38%), slight drought (11%) and no

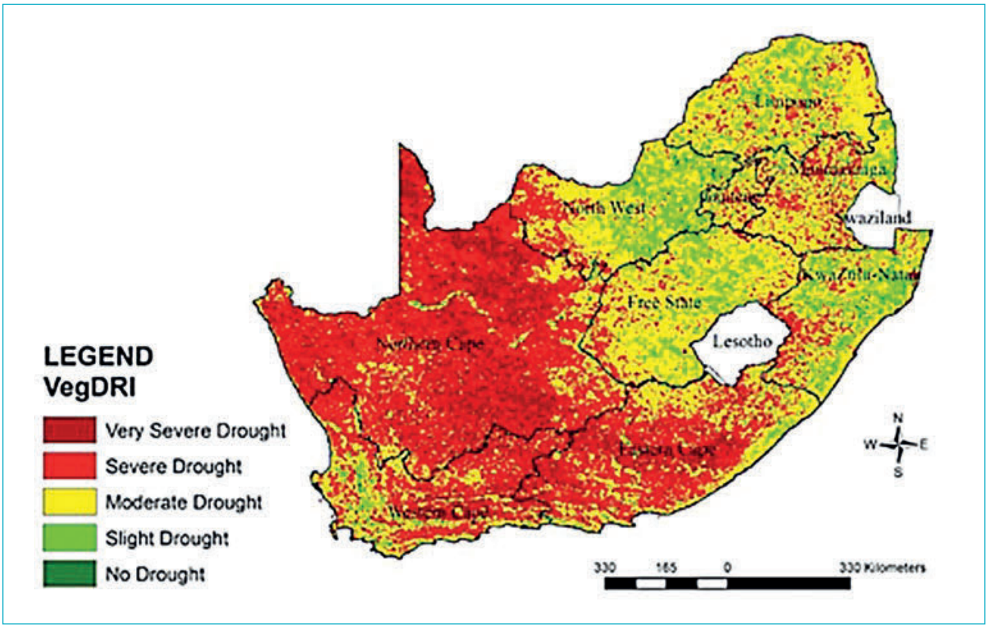

Figure 2: Average seasonal vegetation drought response index

drought (1% of South African agricultural land). Drought was very severe over the Northern Cape and Eastern Cape provinces, indicating acute water scarcity in the region. Moderate to no drought conditions were reported from the central province to the eastern provinces.

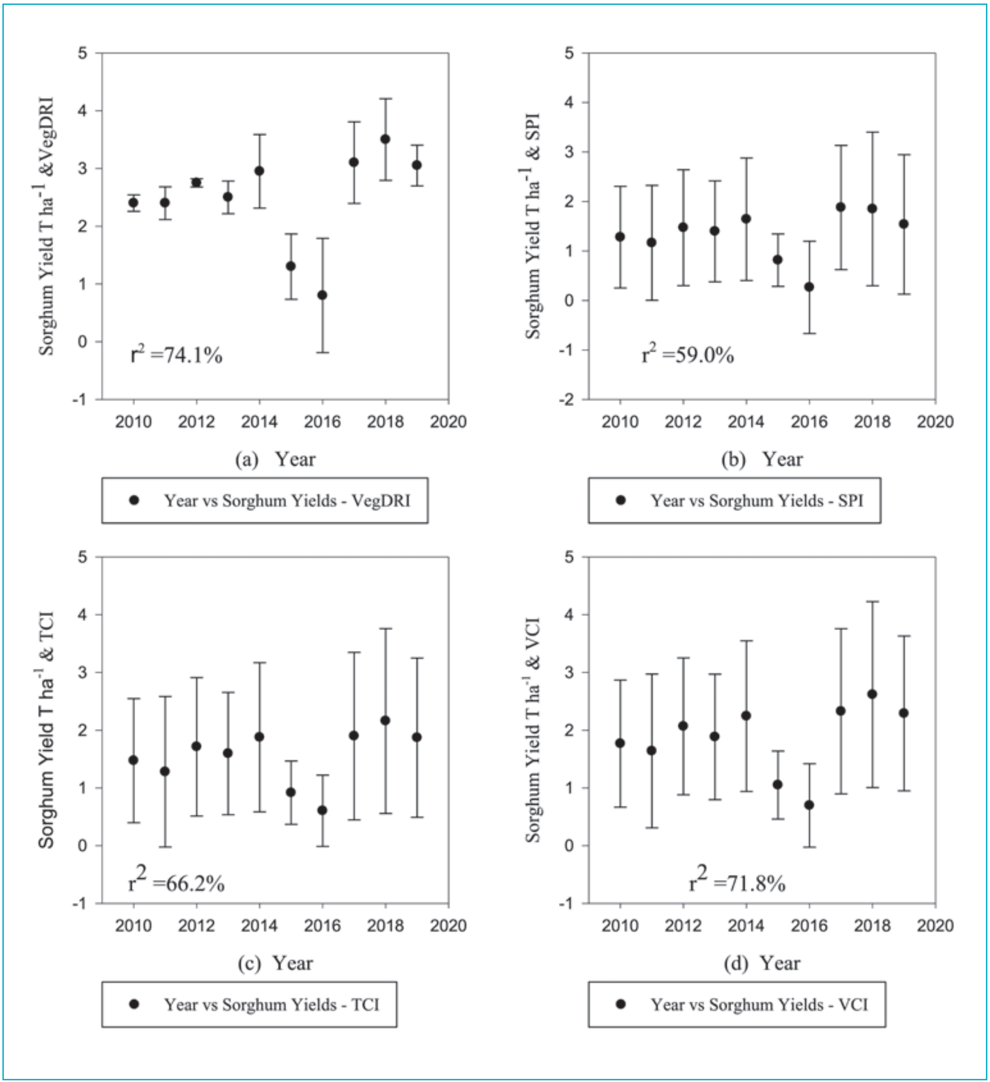

Figure 3: Showing the correlation between district sorghum yields and the drought indices used in the study for the period 2010 to 2019: (a) Vegetation Drought Response Index (VegDRI), (b) Standardized precipitation Index (SPI), (c) Temperature Condition Index (TCI), and (d) Vegetation Condition Index (VCI)

Drought index evaluation

Figure 3 illustrates the performance of the four indices (VegDRI, SPI, VCI, and TCI) when predicting sorghum yields using data from 2010 to 2019. VegDRI proved most successful at predicting yields (74.1% accuracy). All the indices, however, responded accurately to low rainfall in the 2015/16 agricultural season, which recorded the lowest sorghum yield in the research dataset. Overall, the three indices VegDRI, VCI, and TCI, performed systematically better than the precipitation-based SPI.

Figure 4 shows the accuracy of VegDRI to identify sorghum yields compared to VCI, TCI, and SPI using the Kappa statistic. Confidence intervals (±) indicate uncertainty in reported measurements, and in this case, confidence is 95%. The highest Kappa coefficients were observed between VegDRI and VCI, followed by TCI, meaning these indices were better at identifying yields than SPI, which had the lowest agreement. Again, the highest Kappa coefficients were observed in the 2015 agricultural season, which had the lowest rainfall, and which recorded the lowest sorghum yield.

The results suggest that VegDRI could map bioclimatic zones that are under stress and areas of high rainfall variability in South Africa. By integrating traditional drought

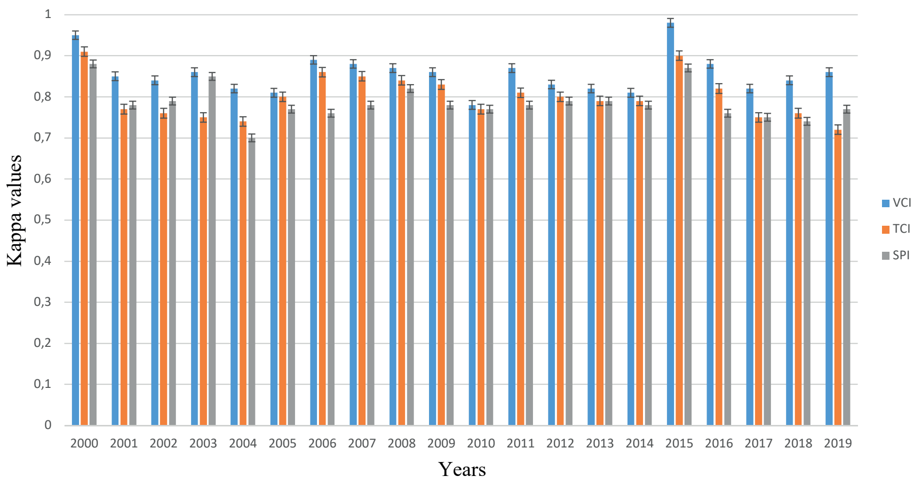

Figure 4: Kappa statistic compares the accuracy with which VegDRI can predict sorghum yields (2000 to 2019) versus the accuracy of VCI, TCI and SPI. Error bars (±) indicate 95% confidence intervals, meaning the results have 95% certainty of accuracy

indicators (VCI, TCI and SPI), the South Africa VegDRI map can select crops within bioclimatic zones, justify disaster management actions, identify livestock production risk zones and assess fire risk zones.

End-users

The intended end-users of this study are farmers, extension agronomists, researchers, non-governmental organizations (NGOs) and private sector companies, such as insurance companies and banks, who all need to understand and develop drought resilience strategies. Similarly, local and national decisionmakers, who need to improve drought response and mitigation.

Lessons learned

1. Water-stressed bioclimatic (drought risk) zones must be identified to inform crop management strategies and thus improve food security in South Africa's marginal lands;
2. The most effective index for identifying agricultural drought risk zones is VegDRI, although a combination of the VCI, TCI and SPI indices can detect drought risk, effectively;
3. Of the four indices analysed, three of them predicted sorghum yields with varying levels of accuracy: VegDRI (74%), VCI (72%), TCI (66%). SPI was the least successful at yield prediction at 59% accuracy (Figure 4);
4. Overall, VegDRI-based agricultural drought assessment is better at capturing water-stress, drought risk and yields.

Current maps can help with the planning and management of sustainable strategies in water-stressed and high rainfall areas. But drought impacts vary as much as their causes, and so validating and operationalising bioclimatic zone maps is crucial. Therefore, this study shows that the VegDRI hybrid index could be used to enhance agricultural support systems such as drought risk maps for early warning systems, crop yield forecasting models and water resource management tools.

Limitations

The study used biophysical factors to assess water-stressed bioclimatic zones that have high rainfall variability. These physical factors are inter-connected with the socioeconomic context of drought (i.e. those affected and their specific vulnerabilities).

In future, innovative methods are needed to integrate and model these socioeconomic factors.

Implications of the drought risk maps for crop production

Drought zone mapping is essential to drought management and integrated climate risk management. It helps policy makers and agriculturists plan and recommend sustainable agriculture production by identifying drought-prone areas. This aligns with the R4 Rural Resilience Initiative framework (2011), which helps vulnerable farmers adapt to climate risks by adopting sustainable intensification and climate-smart strategies.

Conclusion

The hybrid drought index, VegDRI, was developed using machine learning to characterize bioclimatic zones in South Africa with high rainfall variability and water scarcity. The resulting VegDRI outperformed other drought indices in identifying water-stressed zones and was verified using normalized sorghum crop yield data. A correlation test with the normalized sorghum crop yield data proved the index's applicability. VegDRI can be extended to more sub-Saharan African regions using climate, satellite and biophysical data. Future research could include hydrological, soil water, evapotranspiration and socioeconomic factors to improve drought management activities. This research suggests using VegDRI in agricultural decision support systems for crop yield forecasting, drought risk mapping and water resource management.

Acknowledgements

The authors would like to thank the Water Research Commission of South Africa, the uMngeni Resilience Project and the Sustainable and Healthy Food Systems (SHEFS) Programme, supported by the Wellcome Trust.

Bibliography

Akinola, R., Pereira, L.M., Mabhaudhi, T., de Bruin, F.M., Rusch, L. A review of indigenous food crops in Africa and the implications for more sustainable and healthy food systems. Sustain. 2020;12, 1–30. https://doi.org/10.3390/SU12083493

Breiman, L. Random forests. Mach. Learn. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 2001;5–32.

Bvenura, C., Afolayan, A.J. The role of wild vegetables in household food security in South Africa: A review. Food Res. Int. 2015. https://doi.org/10.1016/j.foodres.2015.06.013

Chivenge, P., Mabhaudhi, T., Modi, A.T., Mafongoya, P. The potential role of neglected and underutilized crop species as future crops under water scarce conditions in Sub-Saharan Africa. Int. J. Environ. Res. Public Health. 2015;12, 5685–5711. https://doi.org/10.3390/ijerph120605685

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geogr. 2006;30, 751–777. https://doi.org/10.1177/0309133306071957

Mabhaudhi, T., Chimonyo, V.G.P., Hlahla, S., Massawe, F., Mayes, S., Nhamo, L., Modi, A.T. Prospects of orphan crops in climate change. Planta. 2019;250, 695–708. https://doi.org/10.1007/s00425-019-03129-y

Mohd Nizar, N.M., Jahanshiri, E., Tharmandram, A.S., Salama, A., Mohd Sinin, S.S., Abdullah, N.J., Zolkepli, H., Wimalasiri, E.M., Mohd Suhairi, T.A.S.T., Hussin, H., Gregory, P.J., Azam-Ali, S.N. Underutilized crops database for supporting agricultural diversification. Comput. Electron. Agric. 2021;180, 105920. https://doi.org/10.1016/j.compag.2020.105920

R4 Rural Resilience Initiative. World Food Prgramme (WFP). 2011. https://www.wfp.org/r4-rural-resilience-initiative

Yan, R., Wang, J., Ju, W., Goll, D.S., Jain, A.K., Sitch, S., Tian, H., Benjamin, P., Jiang, F., Wang, H. Interactive effects of the El Niño-Southern Oscillation and Indian Ocean Dipole on the tropical net ecosystem productivity. Agric. For. Meteorol. 2023;336, 109472. https://doi.org/10.1016/j.agrformet.2023.109472