https://centaur-horizon.eu/

Forecasting Agricultural Drought Impact in Africa through Machine Learning and Earth Observation

Koen De Vos[1*], Sarah Gebruers[1], Jeroen Degerickx[1], Marian-Daniel Iordache[1], Jessica Keune[2], Francesca Di Giuseppe[2], Francisco Pereira[1], Hendrik Wouters[1], Else Swinnen[1], Koen Van Rossum[1], Laurent Tits[1]

1: Environmental Intelligence - Vlaamse Instelling voor Technologisch Onderzoek (VITO), Mol - Belgium

2: European Centre for Medium Range Weather Forecasts (ECMWF), Bonn - Germany

koen.devos@vito.be

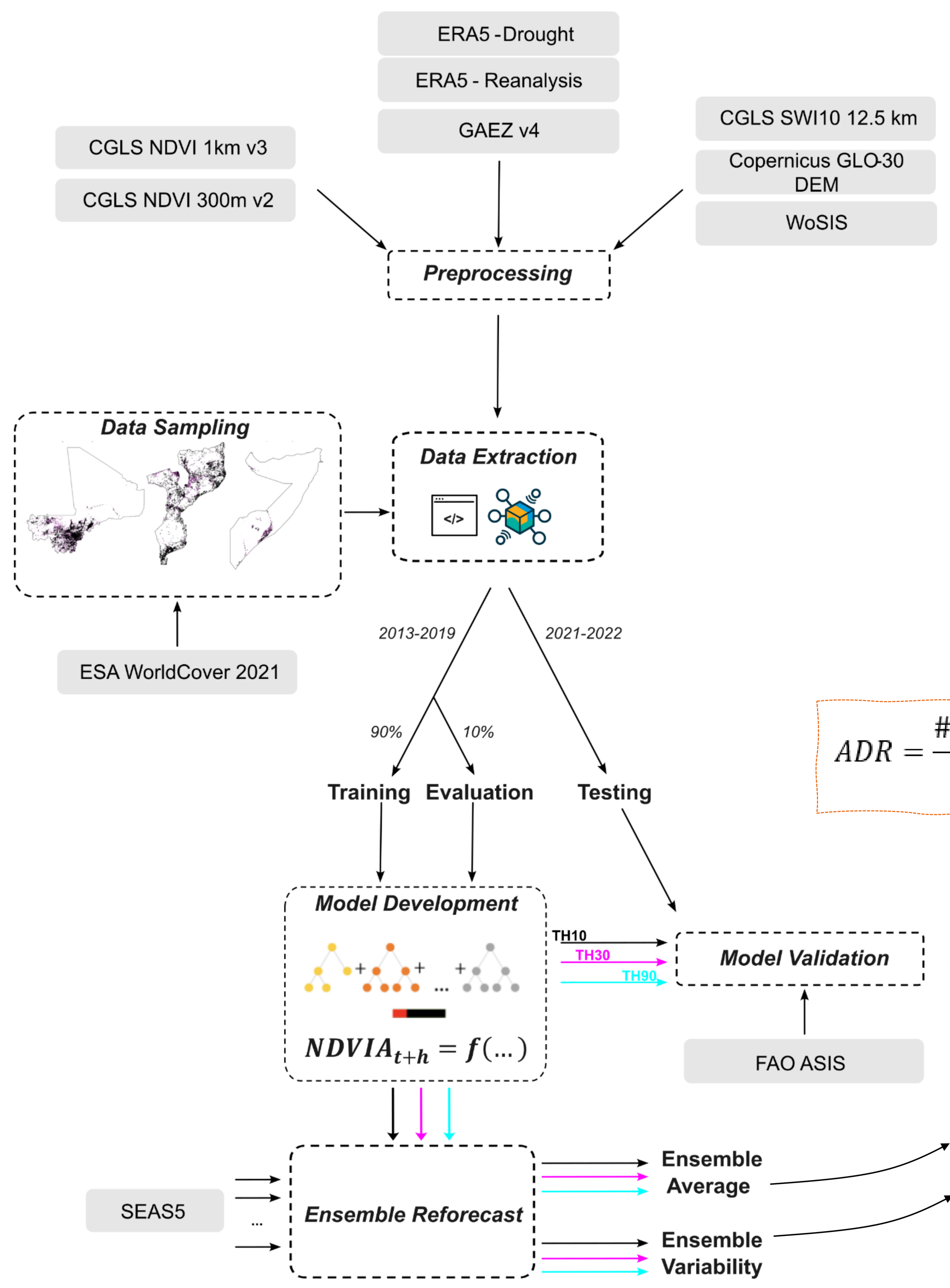
BACKGROUND

Agricultural droughts, driven by a deficit in root-zone soil moisture, pose challenges to food security and economic stability in Africa- which is simultaneously vulnerable to frequent droughts and strongly relies on rainfed agriculture. Recent droughts across the Horn of Africa (2021-2023) and Southern Africa (2023-2024) have resulted in widespread crop failures, livestock loss, and water scarcity which has affected millions of people.

In response, **Earth Observation** (EO) technologies have advanced in **near-real time** (NRT) monitoring of indices such as NDVI, SPI, and SPEI. Tools like FAO's Agricultural Stress Index System (ASIS) rely on NDVI anomalies to assess vegetation impacts of water stress. Other initiatives such as GEOGLAM and FEWSNET integrate EO with meteorological forecasts and expert assessments to monitor food insecurity risks.

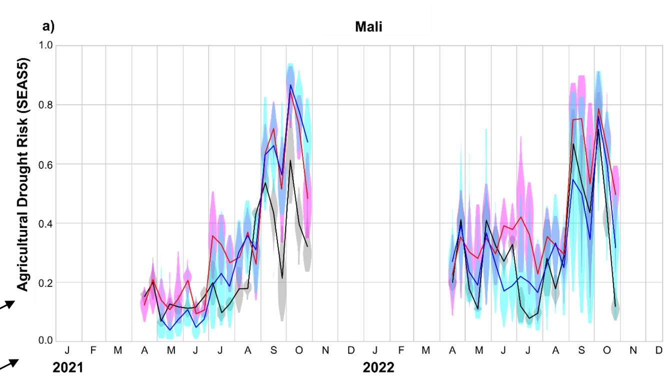
Despite these advancements, drought management in Africa remains predominantly **reactive**, with most EO cases focusing on post-event monitoring rather than forecasting. While efforts have grown, particularly in meteorological drought predictions or soil moisture forecasting through hydrological models, few systems forecast the impact of agricultural drought on vegetation at scale. **Machine learning** (ML) models show promise but have largely focused on crop-specific yield prediction or localized regions with no multi-country system in place that effectively provides a forecast of vegetation conditions.

This study attempts to fill that gap, by developing a model that integrates NRT Earth Observation data with meteorological forecasts to provide an estimate of vegetation conditions **10 days**, a month, and 3 months ahead of time. By providing predictive insights on vegetation impact, this approach sets a benchmark for further developments aimed at transforming drought management by enabling informed decision-making, resource allocation, and food security strategies across Africa's most vulnerable regions in a proactive manner.



$$NDVIA_{t+h} = f(\dots)$$

$$ADR = \frac{\#\,pixels\ in\ growing\ season\ with\ NDVIA\ below\ p15}{\#pixels\ in\ growing\ season}$$



$$NDVIA_{i,t} = \frac{NDVI_{i,t}}{NDVI_{CPS,t}}$$

METHODOLOGY

We developed a model framework to predict below-average NDVI anomalies using an array of both environmental and meteorological variables and tested it for **Mali, Mozambique, and Somalia**. In this study, NDVI serves as a proxy for vegetative productivity and is ideally fit for operational use in various monitoring products due to its widespread use and long, continuous archive. While we acknowledge the challenges of using NDVI in a (semi)-arid context, its broad applicability and familiarity across disciplines makes it a practical choice for this study over other indices. An overview of the processing chain is provided by the figure above.

In the pre-processing step, we resampled all data to match a spatial resolution of 1000m using a nearest neighborhood resampling in **OpenEO** and partitioned the case countries into Crop Production Systems (CPS) by clustering historical NDVI time series data into zones.

Next, a **Catboost Regressor Model** was developed using a combination of phenological, meteorological, and environmental predictors. The model was trained and evaluated using cropland data points from all three countries for the years 2013-2019 and was validated quantitatively by using data points for the years 2021-2022 and qualitatively by comparing our model outputs with the Agricultural Stress Index System by FAO.

Finally, we used an **ensemble** of meteorological reforecasts (SEAS5) to identify uncertainties in an operational setup. By doing so, we provide a new, robust tool for agricultural drought forecasting applications.

VALIDATION

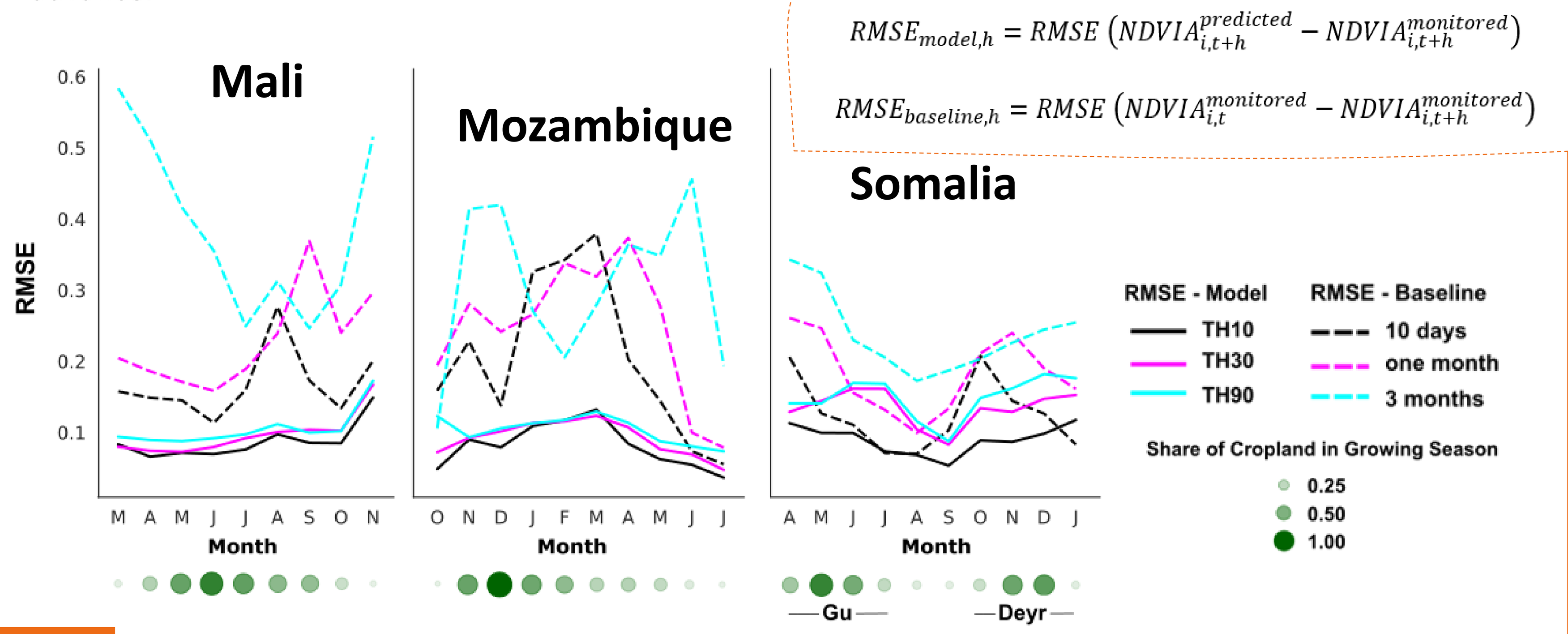
Quantitative

We tested how well our models perform by comparing the predicted NDVIA values for **2021-2022** with the actual NDVIA values that were monitored. To see how much better our models are than a situation where decision-makers rely solely on NRT NDVI, we also compared them to a baseline in which we calculated the RMSE between NDVIA values monitored at different timesteps separated from each other.

For almost all cases, **our framework outperforms the baseline**. Lowest RMSE is observed for predictions made 10 days in advance, but the largest gains in RMSE occur for the model predicting NDVI 3 months in advance.

$$RMSE_{model,h} = RMSE\left(NDVIA_{i,t+h}^{predicted} - NDVIA_{i,t+h}^{monitored}\right)$$

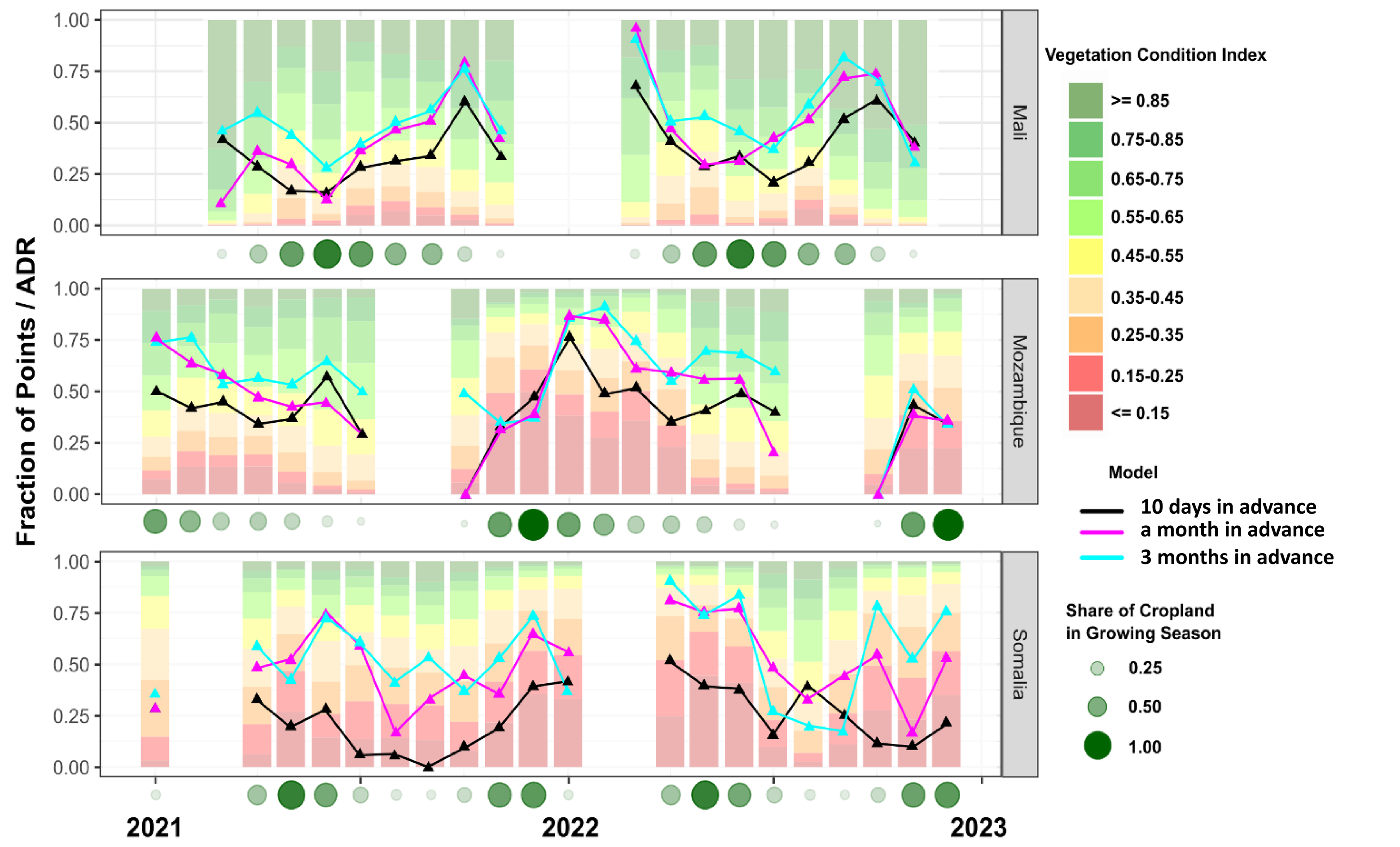
$$RMSE_{baseline,h} = RMSE\left(NDVIA_{i,t}^{monitored} - NDVIA_{i,t+h}^{monitored}\right)$$



Qualitative

We also conducted a qualitative comparison by evaluating the Agricultural Drought Risk (ADR), i.e. the proportion of cropland pixels predicted to experience agricultural drought impact per country. This ADR was compared to the Vegetation Condition Index (VCI) from FAO's ASIS system, which uses a similar spatial aggregation technique. The comparison with VCI is merely qualitative as values are not directly comparable but both provide a general indication of the impact on vegetation. Also, VCI cannot be considered ground truth as it is also a derived EO-product.

Our predicted values are comparable to VCI. However, one major difficulty is predicting vegetation changes during the transitions between seasons- coinciding with the **onset and end of the growing seasons**. For example, in Mali (October-November) and Somalia (June), our models often flag an increase in agricultural drought risk at the end of the growing season. This is likely due to early harvesting, which causes a sudden drop in NDVI, rather than a true agricultural drought impact. At the start of the season, our models also likely struggle to accurately predict the exact moment of crop emergence. This is because early-season growth depends not only on environmental conditions or antecedent vegetation which have been included in the model, but also on agronomic decisions- such as when farmers decide to plant their crops- which our models currently do not consider explicitly.



The Way Forward

Integrating our model into existing frameworks could enhance current drought management workstreams. By directly focusing on below-average NDVIA values, it complements existing soil moisture forecasting systems to provide a more comprehensive picture of anticipated agricultural drought impact.

Translating these insights into actionable strategies, however, poses challenges. Effective communication and capacity-building initiatives are essential to ensure that everyone can utilize our findings to inform them of their practices. Currently, our predictions are limited to croplands, but similar methods could/should be developed for grasslands. Combining our results with tools like crop type or irrigation maps could enable crop-specific drought impact forecasts. These advancements have the potential to feed into large-scale initiatives, further bridging the gap between forecasting and practical decision-making in agricultural management.

Key References

- Keune, J. et al. (2025) ERA5-Drought: Global drought indices based on ECMWF reanalysis. *Scientific Data 12(1), 616*
- Rojas, O. (2021) Next Generation Agricultural Stress Index System (ASIS) for Agricultural Drought Monitoring. *Remote Sensing 13(5), 959*
- Sheffield, J. et al. (2014) A drought monitoring and forecasting system for sub-Sahara African water resources and food security. *Bulletin of the American Meteorology Society, 95(6), 861-862*

partners:

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101082720 - CENTAUR