HarvestStat Africa: Application to Crop Yield Forecasting for Improved Food Security in Sub-Saharan Africa

Donghoon Lee[1], Weston Anderson[2], Frank Davenport[3], Shrad Shukla[3], Eric Yoon[3], Barnali Das[3], Greg Husak[3], and Chris Funk[3]

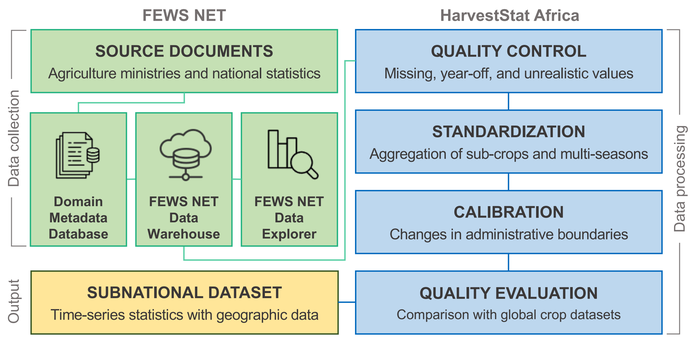
[1]University of Manitoba, [2]University of Maryland, College Park, [3]University of California, Santa Barbara
Correspondence to: Donghoon.Lee@umanitoba.ca

Can Subnational Crop Statistics Enhance Food Security Analyses?

Crop production statistics are essential for understanding yield gaps, production trends, and the effects of climate variability and change on food systems. However, the absence of harmonized subnational data creates significant barriers to applications such as crop yield forecasting and early warning systems. In this study, we (A) introduce the ***HarvestStat Africa*** dataset—a ready-to-use subnational crop statistics that enables more accurate assessments of climate impacts on crop production across Sub-Saharan Africa—and (B) demonstrate its practical value through a case study on in-season crop yield forecasting.

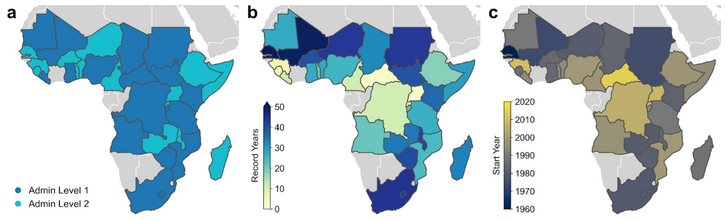
(A) Development of HarvestStat Africa

The data collection and processing workflow for ***HarvestStat Africa***[1] begins with USAID's **Famine Early Warning Systems Network (FEWS NET)** Data Warehouse, which serves as a central repository for agricultural statistics. Data from various sources, including government reports and agricultural agencies, are integrated, standardized, and geolocated. The metadata includes geocodes, crop classifications, season details, and production systems, ensuring comprehensive tracking across changing administrative boundaries.

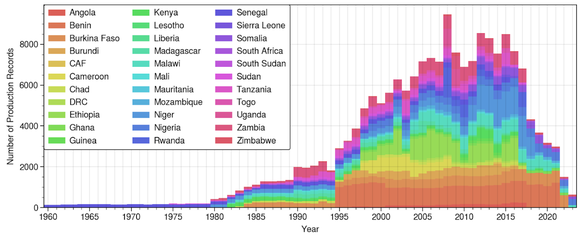


Then, the *HarvestStat Africa* framework applies **quality control** to identify anomalies, standardizes crop statistics across seasons and types, and calibrates data to account for boundary changes. The final step involves **quality evaluation**, comparing the dataset with global benchmarks to ensure accuracy. The resulting subnational dataset links time-series crop statistics to geographic boundary data, supporting robust analyses.

In total, **33 countries** have been included, comprising 18 with data at administrative level 1 and 15 at administrative level 2.

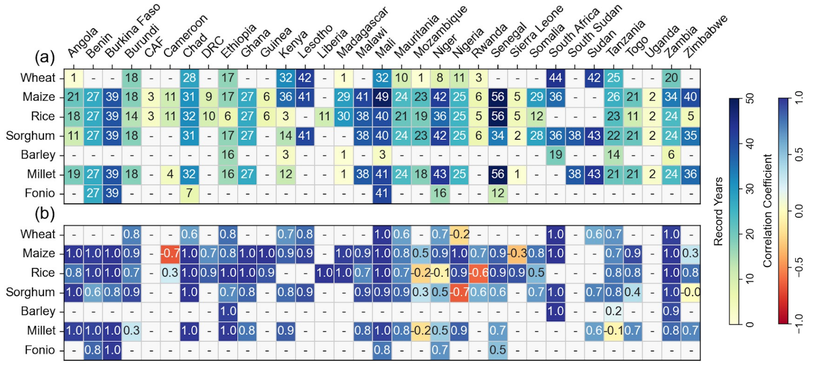


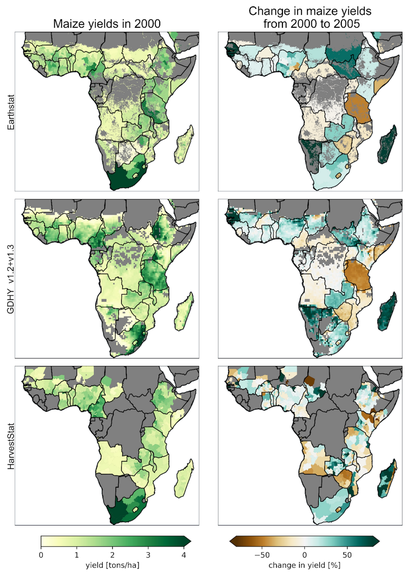
The *HarvestStat Africa* dataset includes **546,605 records**, primarily spanning from 1980 to 2022, detailing crop production, harvested areas, and yields for **90 crop types**, including key cereals such as wheat, maize, rice, sorghum, barley, millet, and fonio.



A notable increase in record volume began in the early 2000s, driven by improved data availability and reduced gaps.

The *HarvestStat Africa* dataset **aligns strongly with FAOSTAT**, with median correlation coefficients of 0.78 across all crops and 0.83 for grains. The dataset's strong correlations affirm its reliability for diverse crops and regions.



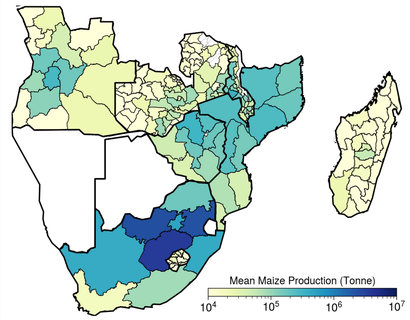


The *HarvestStat Africa* is the only dataset exclusively focused on subnational crop data for Africa, offering **higher spatial and temporal resolution** than others.

Unlike other global datasets[2,3], which rely on national data and methods such as vegetative health indices to estimate yields, *HarvestStat Africa* avoids infilling missing years, preserving the dataset's integrity for time-series analysis.

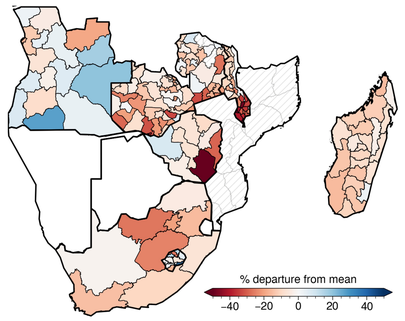
The *HarvestStat Africa*'s approach ensures data suitability for advanced applications like panel regression models.

(B) In-season Crop Yield Forecasting



In Southern Africa, the typical planting season begins in Oct to Nov, with harvesting concluding around Apr to May. For the **2023–24 season**, a strong **El Niño** event is anticipated, a phenomenon historically associated with severe droughts in Southern Africa, potentially threatening maize yields and exacerbating food insecurity in the region. Proactive measures and adaptive strategies are essential to mitigate these impacts.

Using historical climate, vegetative data, and maize yield records, along with machine learning models, we forecasted maize yields for the 2023-24 season in Southern Africa. Our November hindcast achieved **68% accuracy**, demonstrating promising predictive capabilities early in the season.



During the growing season, we **continuously update our forecasts** with near real-time Earth Observation data, providing essential information to FEWS NET.

Takeaways

We introduce ***HarvestStat Africa***, the largest open-source, harmonized, and ready-to-analyze subnational crop statistics dataset, designed to advance food security analyses and food security analyses while addressing critical challenges in agricultural research and policymaking.

The ***HarvestStat*** is an open-source project: github.com/HarvestStat

Check our Data and Paper!

Data

Paper

References

1. Lee, D., Anderson, W., Chen, X., Davenport, F., Shukla, S., Sahajpal, R., Budde, M., Rowland, J., Verdin, J., You, L., Ahouangbenon, M., Davis, K. F., Kebede, E., Ehrmann, S., Justice, C., & Meyer, C. (2024). HarvestStat Africa – Harmonized subnational crop statistics for Sub-Saharan Africa. EarthArXiv.
2. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., & Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. Nature Communications, 3, 1293.
3. Iizumi, T., & Sakai, T. (2020). The global dataset of historical yields for major crops 1981–2016. Scientific Data, 7, 97.

Acknowledgements

This work is partially supported by the Canada Research Chairs Program, the University of Manitoba Startup Grant, and the USGS Grant #G21AC00026.