Ministry of Agriculture, Food and Rural Affairs
Republic of Korea

Application of an analytical framework to hindcast crop yield in major crop production regions in Mozambique

Kwang Soo Kim, Shinwoo Hyun, and Seok Ho Lee

Department of Agriculture, Forestry and Bioresources, Seoul National University

Abstract

Mozambique faces challenges in staple food crop production, which makes crop yield prediction vital for effective policy-making on food security. The analytic framework that integrates satellite data and crop growth simulations to forecast regional crop yield can aid policy makers. The objectives of this study were to apply the analytic framework to three major crop production regions in Mozambique including Gaza, Manica, and Nampula provinces for maize, soybean, and rice. The gridded crop growth simulations were performed using Decision Support System for Agrotechnology Transfer (DSSAT). A set of crop management scenarios were applied to the crop growth simulations. One of these simulations were identified to obtain crop yield hindcasts by cell comparing leaf area index data derived from the simulations and the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data. Crop yield hindcasts were obtained using a percentile of crop yield distribution using three preceding growing seasons. It was found that the percentile used for crop yield hindcasts differed by crop and province. The accuracy of maize and soybean yield hindcasts was within an acceptable range, e.g., < 20% of crop yield in growing seasons, whereas that of rice yield hindcasts was considerably high. Crop yield predictions were limited by the use of crop management scenarios such as cultivars and fertilizer application. Despite biases and limitations in representing real farming conditions, the framework provided insights into improving staple food crop production. It was also highlighted that detailed knowledge on crop management practices such as cultivar and fertilizer applications would improve the reliability of the analytic framework to predict crop yield in the major production regions in Mozambique.

Keywords: process-based model, similarity index, forecast, decision support system,

food security

CGIAR DIGITAL TRANSFORMATION

IFPRI ITC UNIVERSITY OF TWENTE. UNIVERSIDADE EDUARDO MONDLANE VERI TAS LUX MEA

1. Introduction

Mozambique has faced food security challenges due to natural disasters, conflict, and economic crisis (Matavel et al., 2022; Mabiso et al., 2014). In particular, natural disasters such as flooding and drought have led to the destruction of crops, which exacerbated food insecurity and financial strain for farm households (Manuel et al., 2021; Brida et., 2013). It would be advantageous to predict crop yield at regional scales, which would aid decision making on policies for food security (Singh et al., 2017). Such an effort would help farmers withstand and recover from the impact of climate disasters in Mozambique (Talacuece et al., 2016).

Application of satellite data would allow for prediction of crop yield before the given growing season ends (Liu et al., 2021). The vegetation index derived from the satellite data products indicate the growth conditions of crops in a large area (Huang et al., 2021). Empirical approaches such as neural networks have been used to predict crop yield at different scales (Nevavuori et al., 2019; Khaki and Wang, 2019). Weather data can also be integrated into a framework for crop yield forecast (Priya et al., 2023; Schwalbert et al., 2020), which would allow for a more comprehensive understanding of the conditions that affect crop growth and development.

Approaches based on satellite data often require a large amount of observation data to identify the relationship between crop yield and remote sensing data (Hara et al., 2021). These data requirements often include crop yield data at a farm level and crop distribution maps at a high spatial resolution (Nyéki et al., 2021). For example, deep neural networks have been developed to forecast crop yield using a large number, e.g., 2247 sites, of site-specific farm data for their training procedures (Khaki and Wang, 2019).

In developing countries, there is often a scarcity of crop yield data at national and regional levels. Financial and human resources dedicated to monitoring crop production are usually minimal, making it challenging to collect reliable data for crop yield (Grassini et al., 2015). For example, Mozambique has limited availability of observation data for crop yield as well as crop distribution (Araneda-Cabrera et al., 2021; You et al., 2009). This lack of reliable data hampers effective decision-making and policy design based on crop yield prediction using satellite data, which in turn impacts food security and economic growth.

Alternatively, the crop growth simulation models (CGSM) can be used for crop yield forecast using common knowledge on crop management in the region of interest (Therond et al., 2011; Lee et al., 2015). The CGSM mimics the response of crops to the given environmental and management conditions *in silico*. Once such a crop model is calibrated for local cultivars, it could be used to represent the likely outcome of crop management in the given region. In particular, crop yield can be forecasted once weather data in the future periods become available. Nevertheless, crop management data including spatial distribution of crop cultivars could be a key bottleneck to make the best use of CGSM due to their scarcity.

The analytic framework where both crop growth model and satellite data are used can complement the shortcomings of each approach for monitoring and prediction of crop yield. The objectives of this study were to examine reliability of the analytical framework for prediction of crop yield in Mozambique. In the present study, a procedure for the analytical framework was described briefly and followed by the outcome of the crop yield predictions in the major crop production regions in Mozambique. Such an analytical framework for spatial prediction of crop yield can be used to support policy-making processes for food security providing the outlook of crop production in a region. It can also be used to aid humanitarian assistance in the country of interest.

2. Materials and methods

2.1. Study area

The analytical framework was applied to major crop production regions in Mozambique including Gaza, Manica, and Nampula provinces (Fig. 1). Gaza province, which is a semi-arid region, depends on irrigation and rain-fed systems for maize and rice. Maize is widely grown as a staple food crop. Rice is also cultivated, mainly in low-lying areas with access to water resources. Manica province has more favorable agro-climatic conditions compared to Gaza, which makes it well-suited for cultivating both maize and rice. Nampula province is one of most agriculturally productive provinces where maize, rice and soybean are grown.

Reports of crop yield were obtained for the growing periods for maize, soybean and rice (Table 1). For maize and rice, crop yield data were available in Gaza, Manica and Nampula provinces for five growing seasons (Gioia, 2024). In Nampula, yield of soybean has been reported from 2015/16 to 2019/20 growing seasons. These data were used to evaluate the outcomes of the analytic framework by province.

2.2. Weather and soil input data

A set of gridded data was prepared to create crop yield maps for major food crops including maize, soybean, and rice using the spatial analytic framework. In the present study, the fifth generation ECMWF atmospheric reanalysis (ERA5) data were collected from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/). The ERA5 data includes hourly solar radiation, temperature, and precipitation data at a spatial resolution of 0.25°. These data were summarized by day to prepare weather input data for the crop growth simulations, e.g., daily solar radiation, maximum temperature, minimum temperature, and precipitation. In addition, bilinear interpolation was applied to these weather data at a 10 km resolution. The *terra* package of R, which has been used for geospatial data analysis (Hijmans et al., 2022), was used to perform spatial interpolation of weather data.

Figure 1. Provinces of Mozambique.

Table 1. Source of crop yield reports in major crop production provinces in Mozambique.

Crop	Provinces	Growing seasons	Data sources
Maize	Gaza, Manica, Nampula	2011/12 2013/14 2014/15 2016/17 2019/20	Gioia (2024)
Soybean	Nampula	2015/16 2016/17 2017/18 2018/19 2019/20	Anuario Estatistico Provincia de Nampula (2015-2020)*
Rice	Gaza, Manica, Nampula	2011/12 2013/14 2014/15 2016/17 2019/20	Gioia (2024)

* The annual reports were available at the website of Instituto Nacional de Estatstica, Mozambique (https://www.ine.gov.mz/)

The soil input data were prepared using the WISE soil database (https://www.isric.org/explore/wise-databases). These soil data include physical and chemical properties of soil by depth (Han et al., 2019). Maps of soil identification code were prepared to indicate the soil profiles in a given region.

2.3. Crop growth simulations

The gridded simulations of crop growth were performed in the production regions of interest in Mozambique. These simulations were performed from 2011/12 to 2023/24 growing seasons although the crop yield reports were available only for five growing seasons. These gridded simulations were used as inputs to the analytic framework to derive the distribution of crop yield hindcasts within a given province. The gridded crop growth simulations were performed using the automated gridded crop growth simulation support system (Kim et al., 2020).

In the present study, CERES-Maize, CropGro-Soybean, and CERES-Rice models were used to perform the gridded simulations of crop growth and development under a diverse set of crop management scenarios. These crop models are included in the Decision Support System for Agrotechnology Transfer (DSSAT), which supports crop growth simulations for a wide range of food and cash crops (Jones et al., 2003).

2.4. Crop management scenarios

Crop yield was predicted under a range of crop management scenarios for the growing seasons in the regions of interest (Table 2). To minimize information on the local cultivars, the default cultivars available from the DSSAT were used to perform crop growth simulations. In the present study, five maize cultivars and four soybean cultivars were chosen to represent the likely cultivars grown across Mozambique along the latitude. In addition, eight cultivars were used to perform crop growth simulations for rice. These cultivars were chosen among the indica rice cultivars.

The management scenarios for planting dates were set to represent the temporal patterns of crop growth and development under diverse conditions. The planting dates for maize and soybean were set to range from October 1st to November 30th each day. For rice, 13 planting dates were set for the same period at the interval of five days.

Both rain-fed and irrigated conditions were applied to the crop growth simulations. It was assumed that automated irrigation methods were available for maize production in some areas, which minimize the water stress providing the optimum amount of water. For the paddy rice growth simulation, auto irrigation was also used for the management options because rice is often grown in the low land areas where water availability supports paddy farming.

Different application rates of fertilizer were used to represent the input conditions in Mozambique by crop. It was assumed that no fertilizer was available to local farmers for maize. It was also assumed that soybeans were cultivated depending only on nitrogen fixation. On the other hand, crop growth simulations of rice were performed using a large quantity of crop residue, which was assumed to provide 45 N Kg ha^{-1}.

2.5. Hindcast of crop yield

Crop yield hindcasts were obtained using the analytic framework where both the outcome of crop growth simulations and time series of satellite data were used. These simulations derived from crop management scenarios were compared against the time series of the satellite data. The MODIS product (MOD15A2H) of LAI was used for the comparison procedures. The MODIS data were collected from the Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (https://ladsweb.modaps.eosdis.nasa.gov). The temporal change of LAI could be disrupted due to clouds and passing time. Thus, the time series data of the satellite data products were smoothed as suggested by Ban et al. (2017).

One of crop growth simulations was chosen to identify the likely management scenario for each cell of the satellite data product. The value of similarity index was

Table 2. The crop management scenarios for the crop growth simulations

Crop	Cultivars	Planting dates	Irrigation methods	Fertilizer application
Maize	2500-2600 GDD 2600-2650 GDD 2650-2700 GDD 2700-2750 GDD 2750-2800 GDD	1/Oct ~ 30/Nov	Rainfed, Auto irrigation	0 N kg / ha
Soybean	M GROUP 2 M GROUP 3 M GROUP 4 M GROUP 5	1/Oct ~ 30/Nov	Rainfed	0 N kg / ha
Rice	IR72 IR43 IR58 IR54 IR64 IR60 IR66 IR72X	1/Oct, 6/Oct, 11/Oct, 16/Oct, 21/Oct, 26/Oct, 31/Oct, 5/Nov, 10/Nov, 15/Nov	Rainfed, Reported dates (at the time of planting)	0 N kg/ ha, 45 N kg / ha (crop residue)

determined between LAI derived from crop growth simulations and remote sensing. In the present study, the structural similarity index (SSIM), which was suggested by Wang et al. (2004), was used to compare the LAI values from the beginning date of a growing season to the date of prediction. The crop management options and crop yield were derived to have the maximum value of the similarity index among the outcomes of crop growth simulations. The values of crop yield from the chosen simulations were pooled to represent the distribution of crop yield in the province of interest. When the SSIM index was less than 0.5, it was assumed that the given cell had no given crop in the present study.

2.6. Assessment of percentile for crop yield hindcasts

In the analytic framework, it was assumed that the crop growth simulations would represent likely crop yield under given environmental conditions. Because no actual data for crop management such as crop cultivar and planting date as well as land use, a bias correction was applied to the outcome of crop yield hindcasts. Due to limited availability of crop yield reports, a simple approach was used to determine crop yield hindcasts at the regional scale, e.g., province.

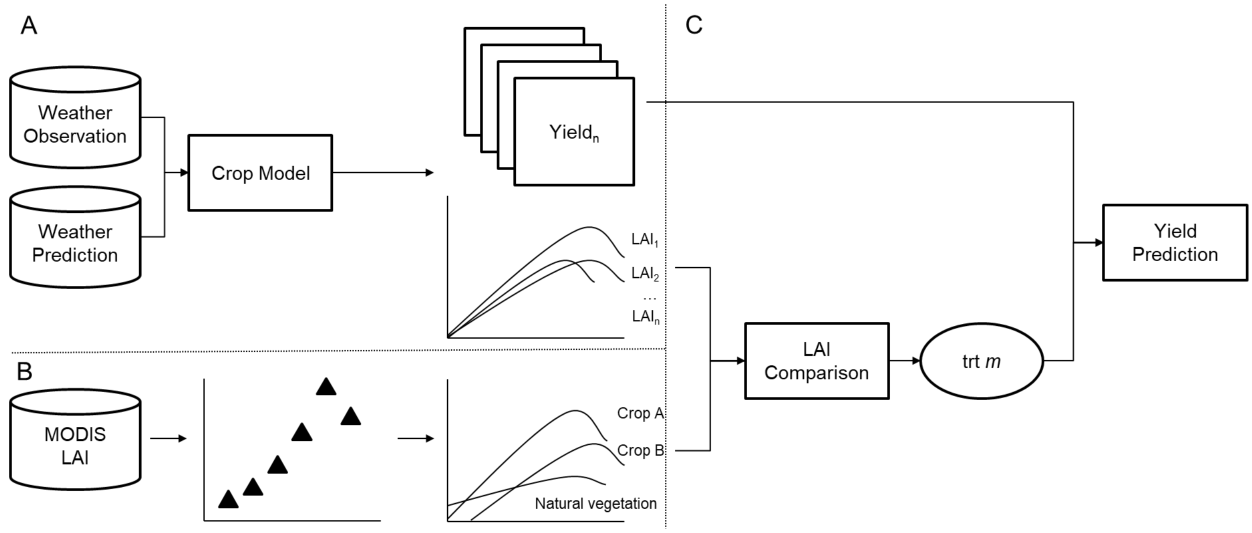

Figure 2. The crop yield prediction stages. The crop model was operated using gridded weather and soil data for *n* management scenarios (A). The time series of leaf area index (LAI) were compared against those of the MODIS LAI product by grid cell of the satellite data (B). The outcome of comparison was used to identify the likely crop management, which is denoted by trt *m* (C). The crop yield value under the management option of the trt *m* was pooled to represent crop yield at the given grid cell within a region of interest.

The distribution of crop yield was examined after one of the crop growth simulations was chosen using the similarity index. Crop yield at a range of percentile from the distribution was compared with reported yield using the previous growing seasons. In the present study, the three growing seasons were used to identify the percentile value to choose crop yield from the distribution. In the present study, the luminance index was used to compare crop yield from simulations and reports as follows (Wang et al., 2004):

$l(x,y) = \frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}$, (eq. 1)

where $\mu_x$is the average of a given variable *x*. The frequent value of quantile was determined to choose the likely crop yield for the other growing seasons during which no crop yield report was available including the 2023/24 growing season.

3. Results

3.1. Maize

In Gaza province, maize yield hindcasts had relatively small errors compared to the crop yield reports using the 5th percentile of crop yield for crop yield prediction (Fig. 3). For example, the absolute error rates for the 2011/12 and 2013/14 growing seasons

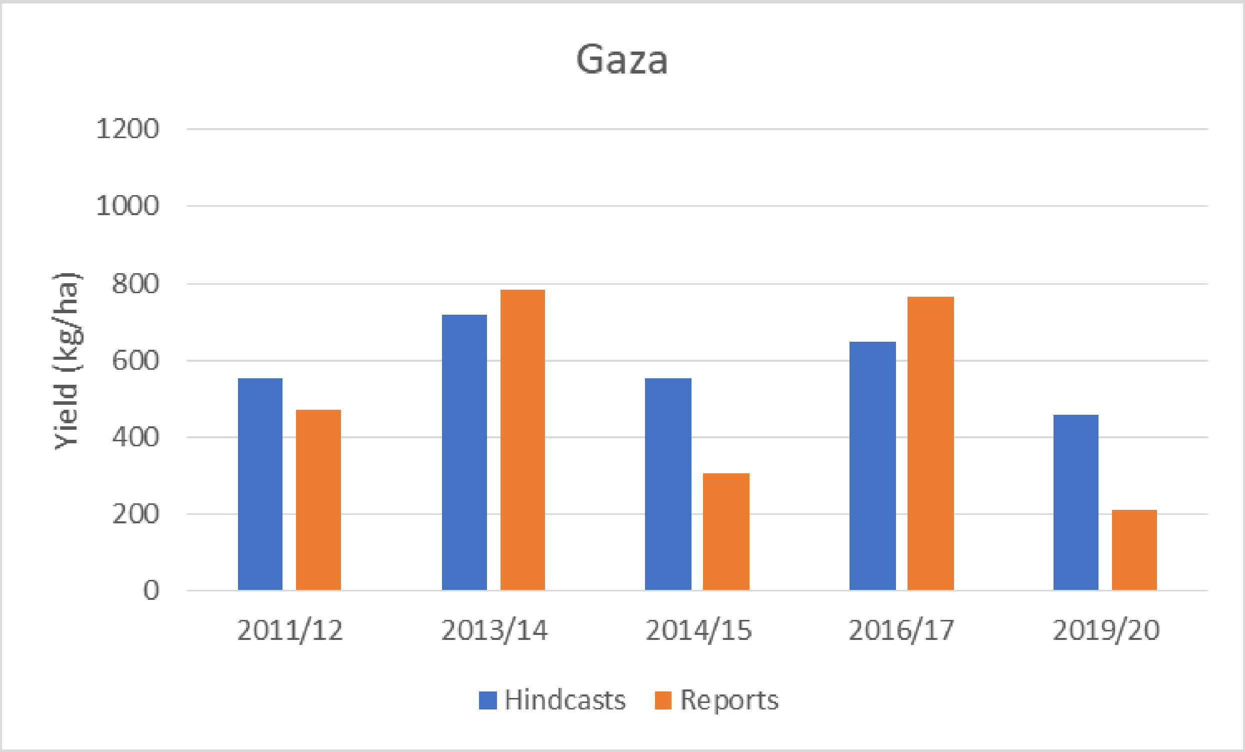

Figure 3. Comparison between reports and hindcasts of maize yield (kg / ha) in Gaza province, Mozambique.

were 17.6% and 8.7%, respectively. However, large errors occurred during the growing seasons especially when crop yield was relatively low. For example, yield was 308 kg/ha and 210 kg/ha in the 2014/15 and 2019/20 growing seasons where the error rate was 80.0% and 117.4%, respectively.

The errors of maize yield hindcast had less magnitude and variation at the 50 percentiles of crop yield values in Manica province (Fig. 4). The error rate of hindcasts ranged from 2.8% to 18.5% for the growing periods included in the present study. Still, variation of the crop yield hindcasts was less than that of crop yield reports.

Maize yield hindcasts had a smaller range of errors at the 95 percentiles of crop yield values in Nampula province (Fig. 5). The error rates differed by growing season, which ranged from 3.7% and 23.1%. Although the 95th percentile of crop yield values was used to predict crop yield, under-prediction error occurred in the 2016/17 and 2019/20 growing seasons.

The use of the analytic framework indicated that variation of maize yield would be common in the major production areas including Gaza, Manica and Nampula provinces (Fig. 6). Manica province had relatively large values of crop yield hindcasts, which reached above 1000 kg/ha in some growing seasons. In contrast, crop yield hindcasts remained lower compared to the other two provinces.

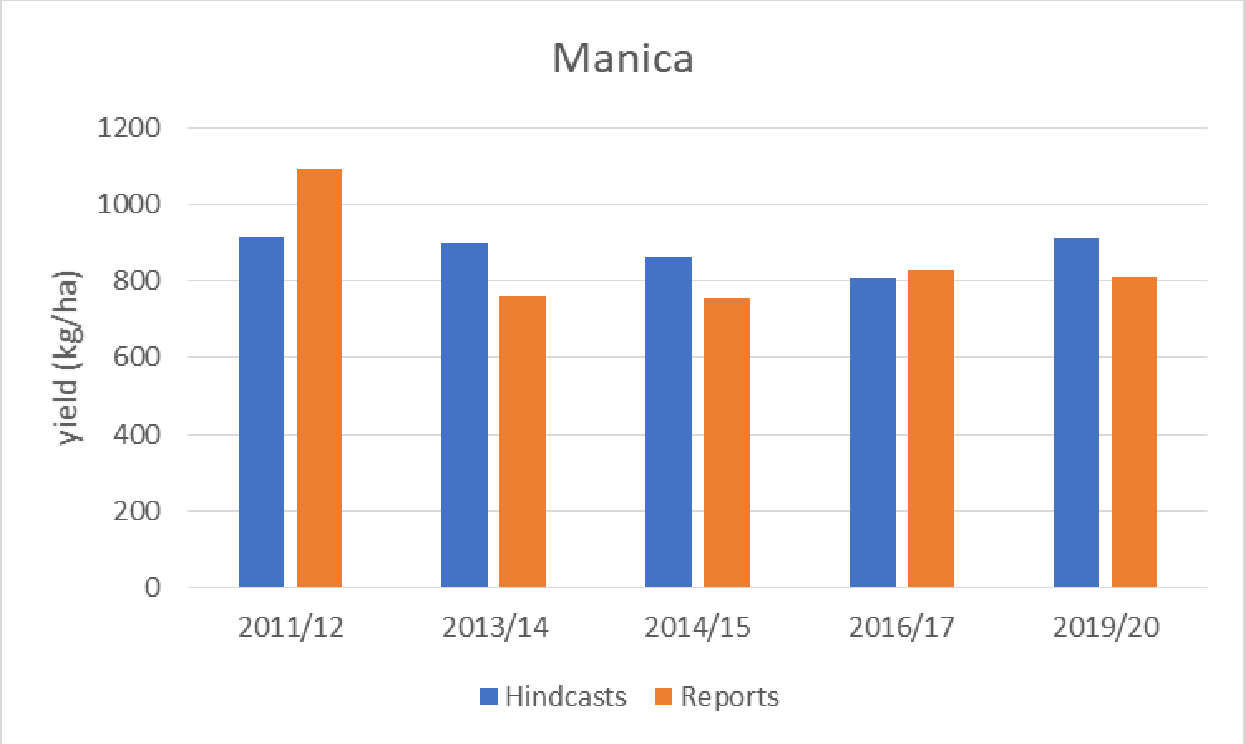

Figure 4. Comparison between reports and hindcasts of maize yield (kg / ha) in Manica province, Mozambique.

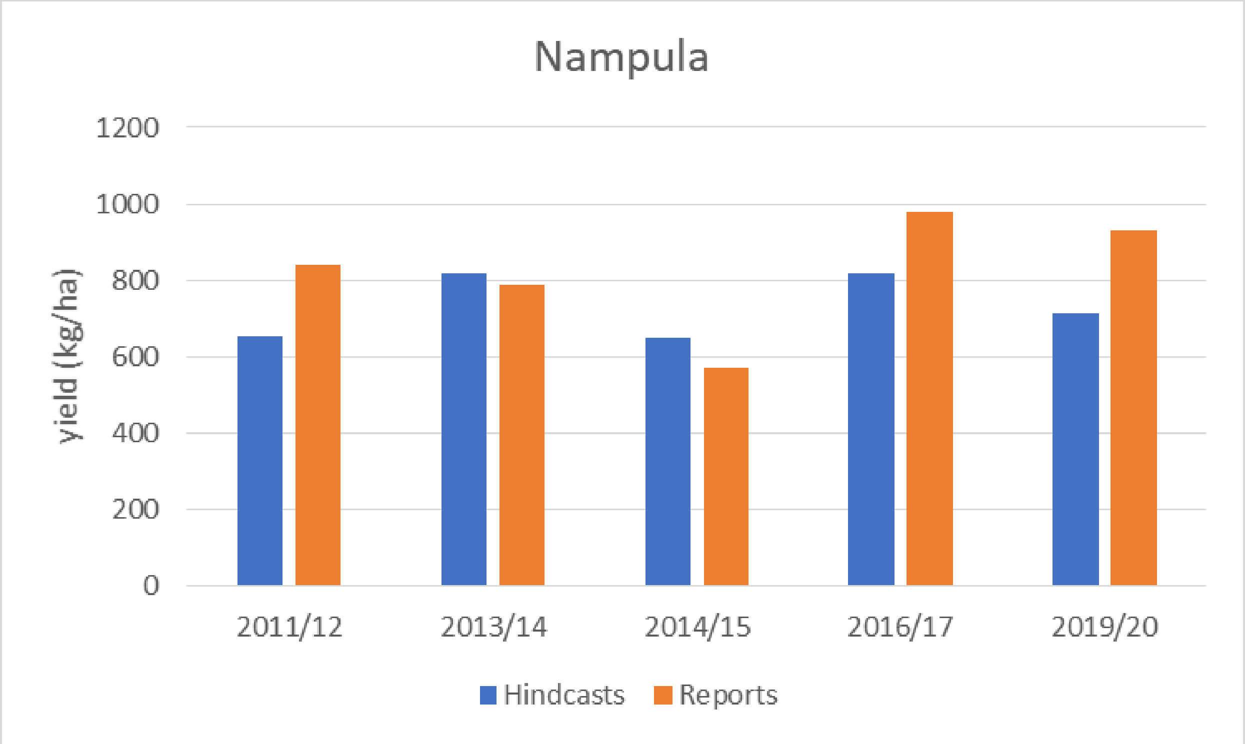

Figure 5. Comparison between reports and hindcasts of maize yield (kg / ha) in Nampula province, Mozambique.

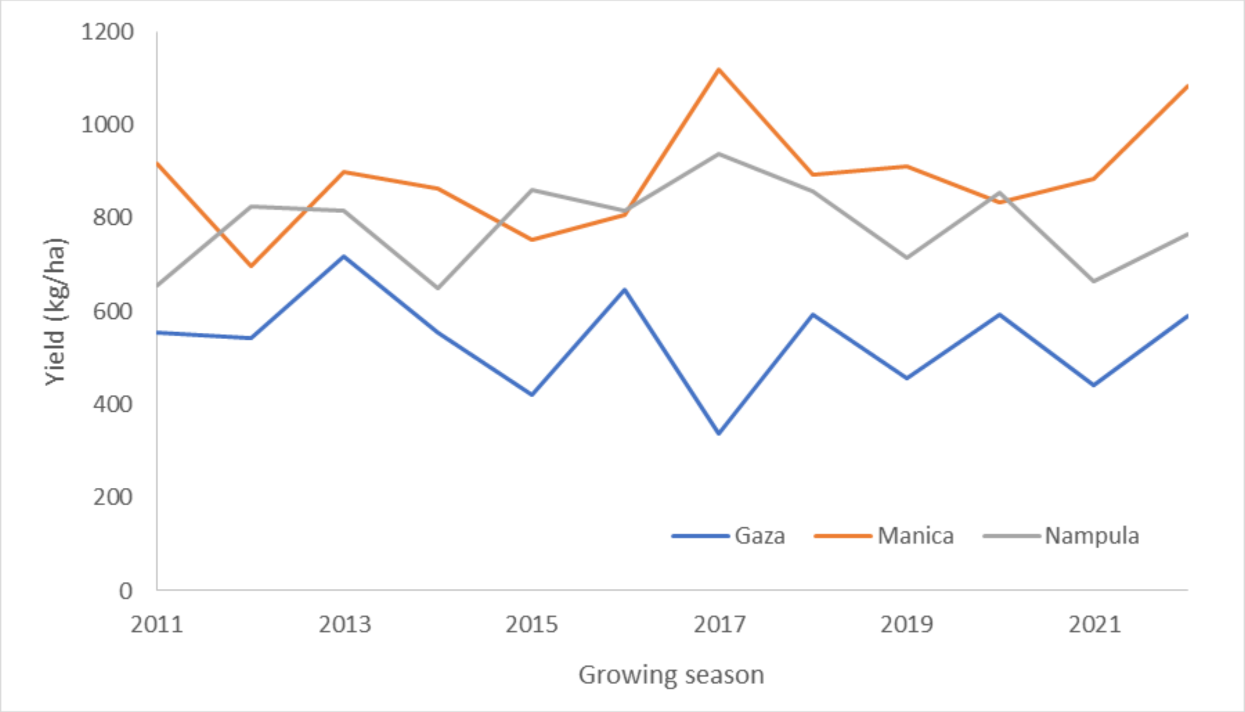

Figure 6. Time series of maize yield hindcasts in the provinces of (A) Gaza, (B) Manica, and (C) Nampula, Mozambique.

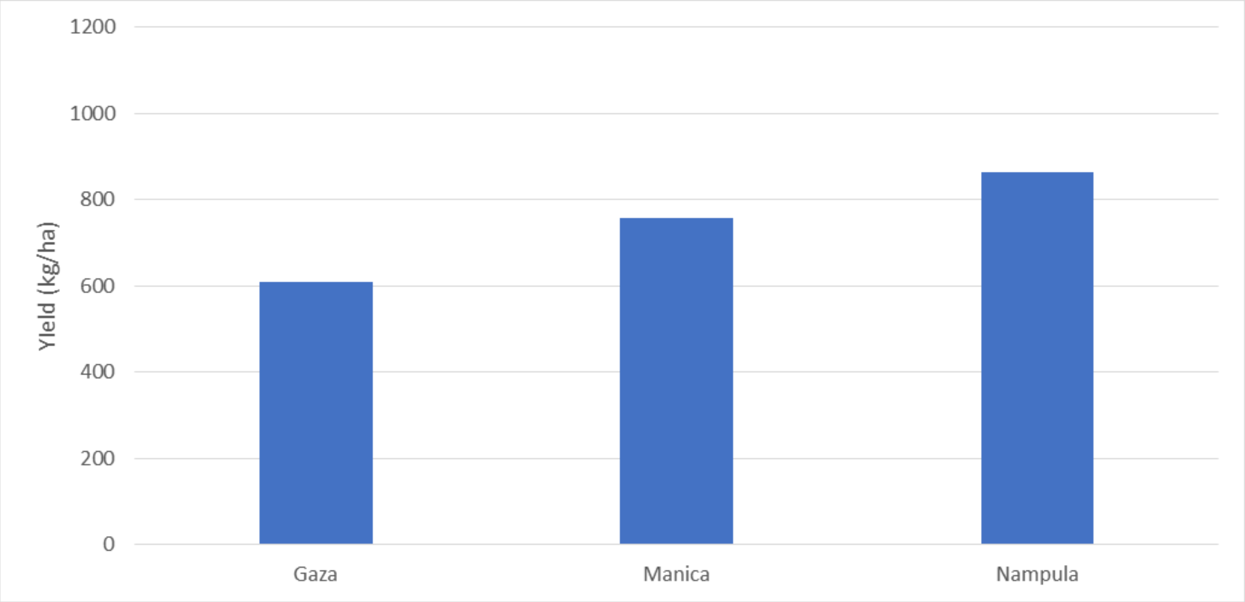

Figure 7. Forecasts of maize yield (kg / ha) in the major crop production provinces in Mozambique.

In the 2023/24 growing season, Nampula province had greater maize yield than the other provinces (Fig. 7). In Manica province, maize yield was relatively low compared to the other growing seasons. In contrast, Gaza province had relatively higher yield hindcat than the average values of maize yield hindicats during the last growing seasons.

3.2. Soybean

In Nampula province where yield of soybean was reported, crop yield hindcasts had relatively small errors at the 50 percentile of crop yield distribution (Fig. 8). In the 2018/19 and 2019/20 growing seasons, the error rates were 4.5% and 8.6%, respectively.

Crop yield hindcasts tended to increase over time in Nampula province (Fig. 9). On average, the crop yield hindcast was 1271 kg/ha from 2011/12 to 2022/23 growing seasons. Since the 2020/21 growing seasons, the crop yield hindcasts were >12% of the average yield hindcasts except for the 2022/23 growing season. In the 2023/24 growing season, the soybean yield hindcast was 1438 kg/ha.

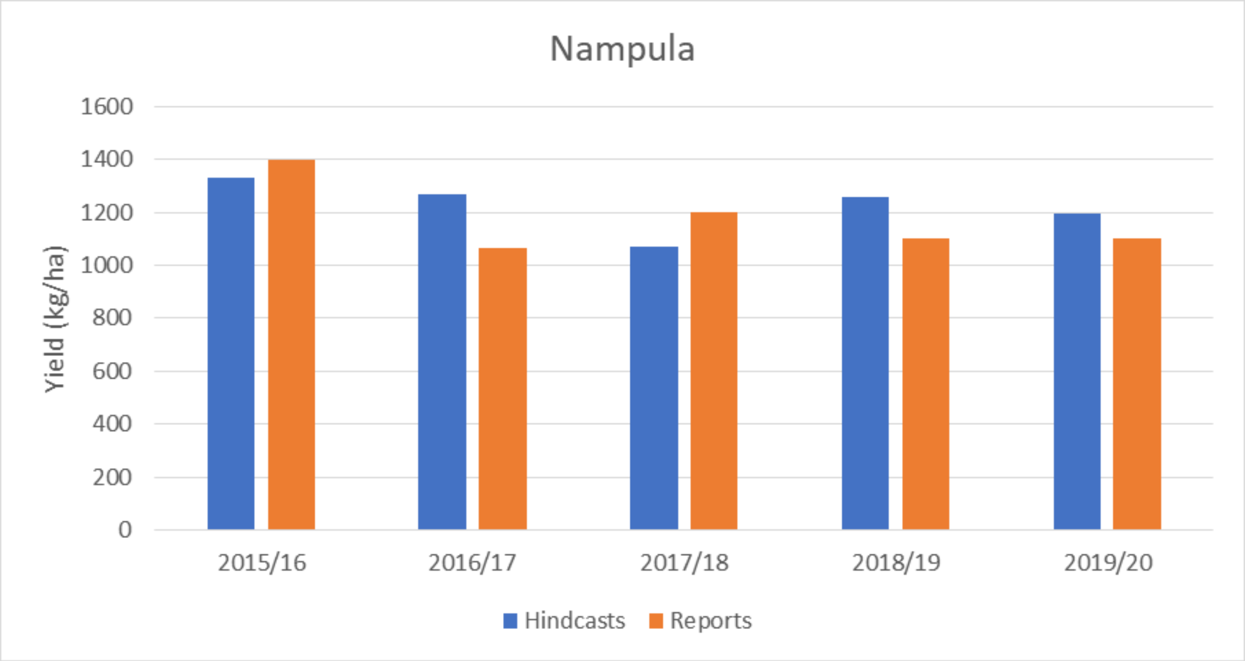

Figure 8. Comparison between reports and hindcasts of soybean yield (kg / ha) in Nampula province, Mozambique.

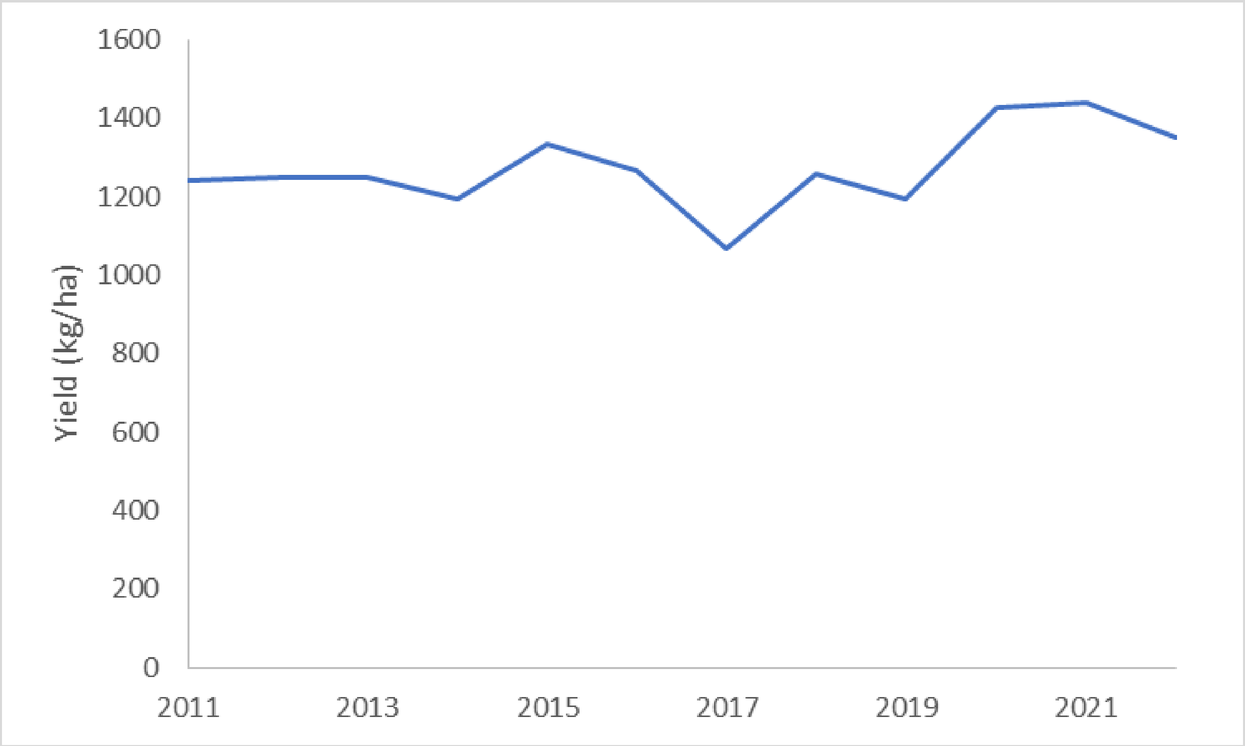

Figure 9. Time series of soybean yield hindcasts in the provinces of Nampula, Mozambique.

3.3. Rice

In Gaza province, the crop yield hindcasts at the 95th percentile had relatively good agreement with reported yield although a large error occurred in a couple of growing seasons (Fig. 10). For example, the absolute error rates of 2011/12, 2013/14, and 2014/15 growing seasons were 12.5%, 1.8%, 82.3%, respectively. Once the 95th percentile of crop yield was applied to 2016/17 growing seasons, the error rate was 27.1%. The growing season of 2019/20 had also relatively small errors, e.g., 20.1%.

Manica province had large variability in the accuracy of crop yield hindcasts compared to crop yield reports over the growing seasons when crop yield was predicated at the 95 percentile (Fig. 11). Except for the 2016/17 growing season, the error rates were considerably high, e.g., > 28.9%, for the growing seasons. The crop yield hindcasts were under-predicted in the 2011/12 and 2013/14 growing seasons whereas these were over-predicted in the 2014/15 and 2019/20 growing seasons.

In Nampula province, rice yield hindcasts were considerably larger than crop yield reports although the 5th percentile of crop yield values were used (Fig. 12). In particular, the over prediction errors were pronounced in the 2011/12 and 2014/15 growing seasons. In the 2016/17 and 2019/20 growing seasons, the error rates of the hindcasts were 45.6% and 35.4%, respectively.

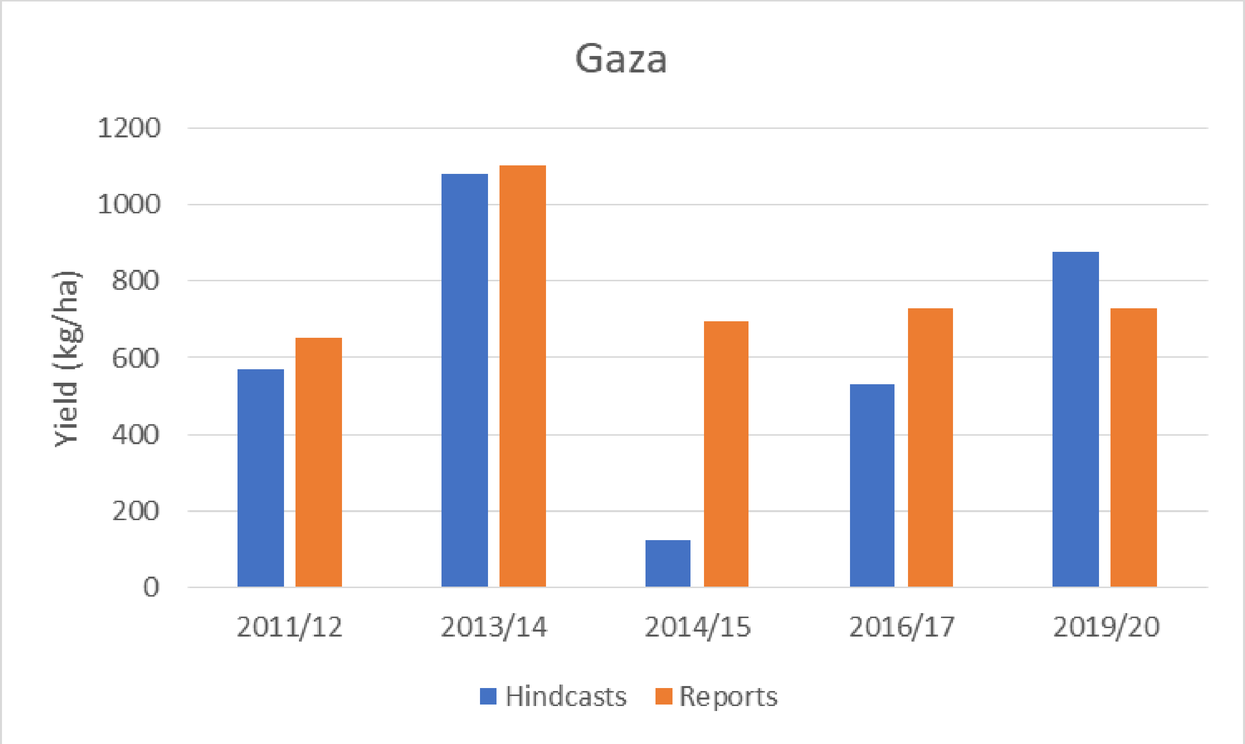

Figure 10. Comparison between reports and hindcasts of rice yield (kg / ha) in Gaza province, Mozambique.

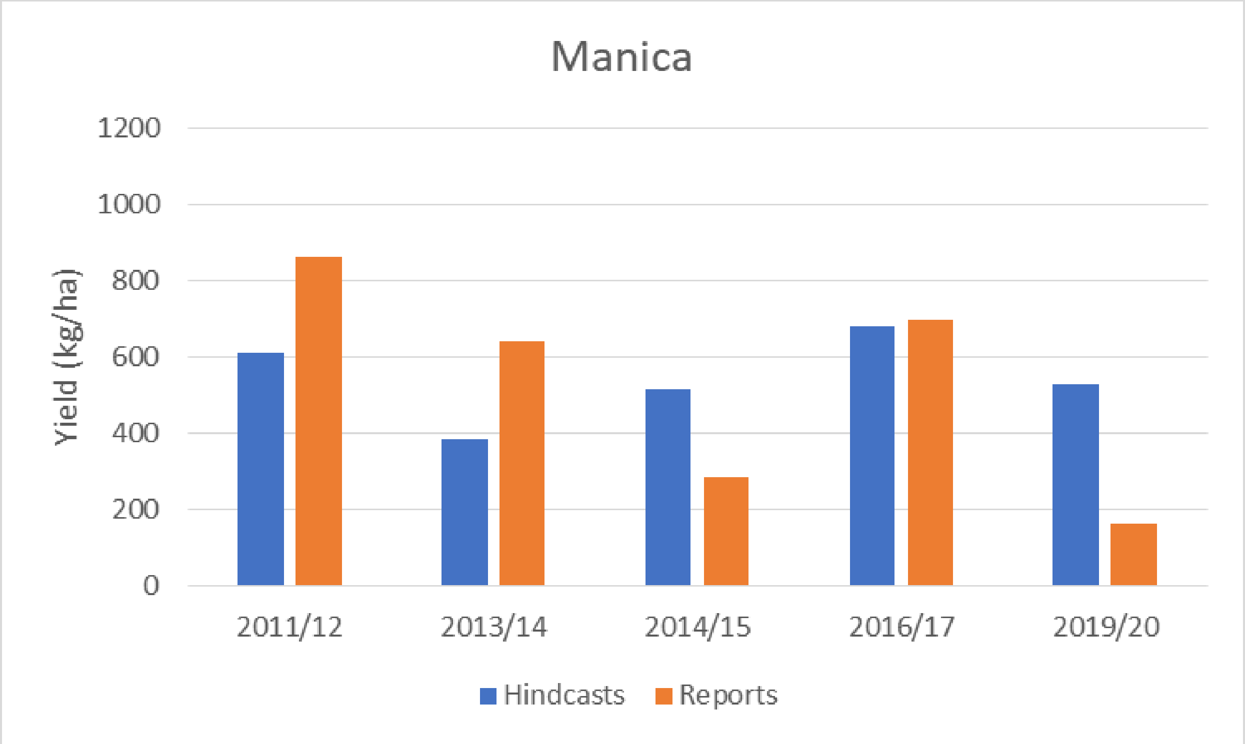

Figure 11. Comparison between reports and hindcasts of rice yield (kg / ha) in Manica province, Mozambique.

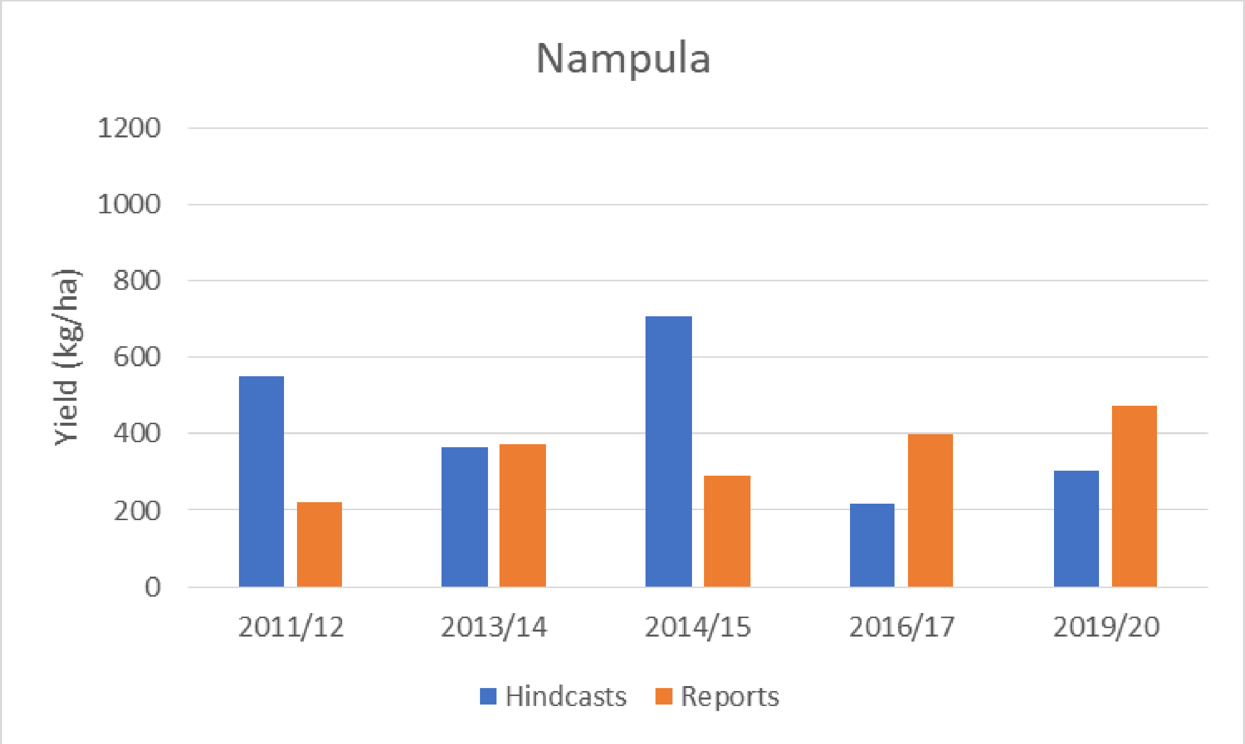

Figure 12. Comparison between reports and hindcasts of rice yield (kg / ha) in Nampula province, Mozambique.

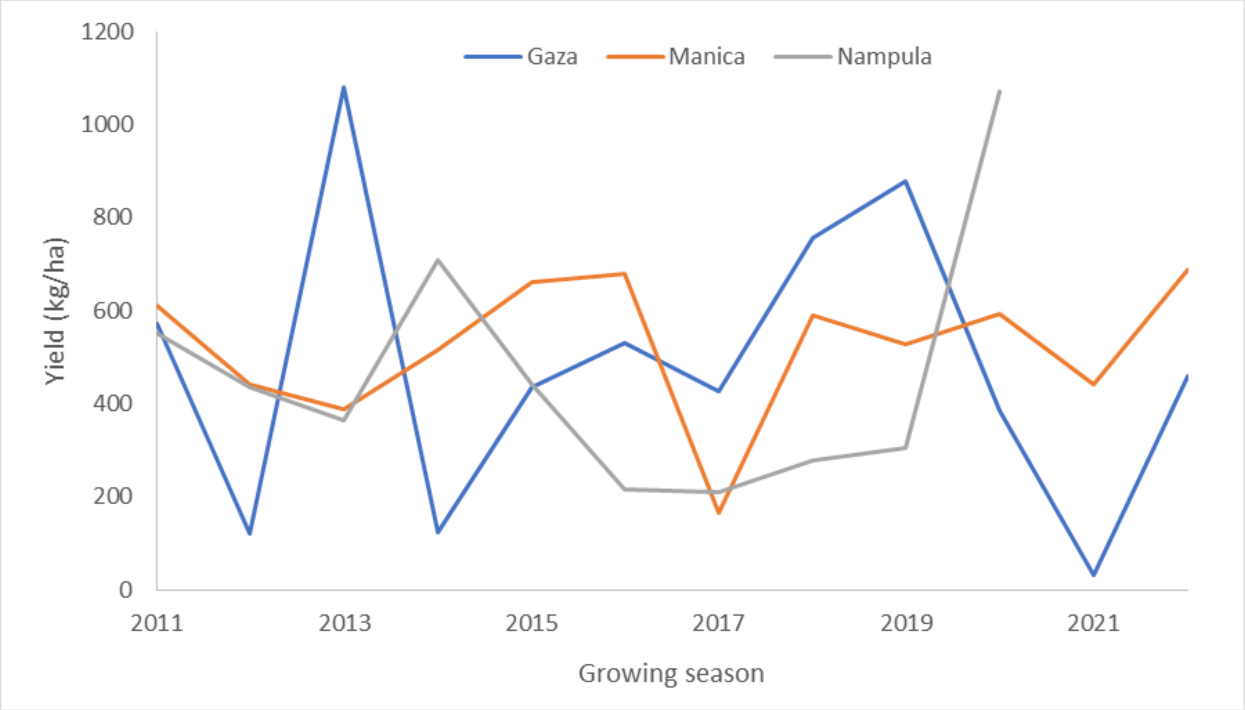

Figure 13. Time series of rice yield hindcasts in the provinces of Garza, Manica, and Nampla, Mozambique.

Time series of rice yield hindcasts were relatively similar between provinces on average although these differed by growing season (Fig. 13). There were the growing seasons where crop yield hindcasts were considerably low, e.g., < 200 kg / ha, mostly in Gaza province. In Nampula province, no yield was predicted in the 2021/22 growing season due to the few cells that had low value of similarity index.

In the 2023/24 growing season, spatial variability of rice yield hindcasts was considerably high among the provinces (Fig. 14). Rice yield hindcasts were considerably high in Gaza province, which was greater than 1000 kg ha^{-1}. In contrast, the crop yield hindcasts were relatively similar between Manica and Nampula provinces.

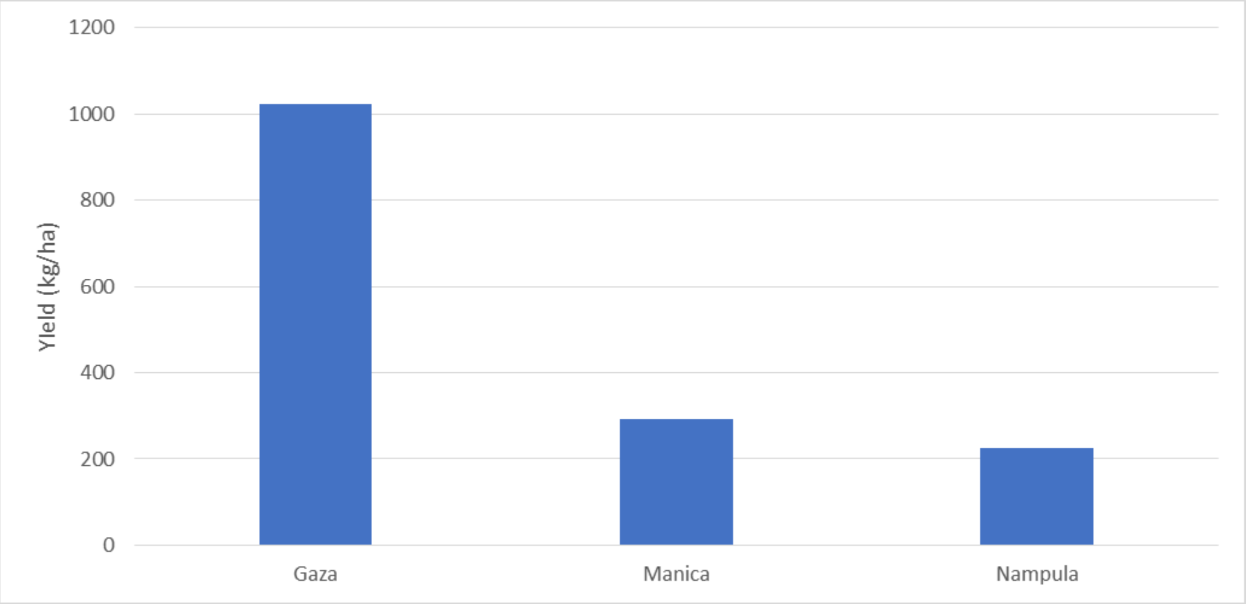

Figure 14. Rice yield hindcasts (kg / ha) in the major crop production provinces in Mozambique.

4. Discussion

Our results indicated that the analytical framework would become an alternative approach to forecast crop yield in Mozambique. The error rates of maize and soybean yield were within a marginal limit, e.g., 20%, although crop growth simulations were performed using scenarios rather than actual data for crop management practices such as crop cultivars, fertilizer rates, as well as land use. This suggested that the analytic framework could provide crop yield hindcasts with a reasonable accuracy in a region when more information on crop management practices becomes available in the given region. In particular, the use of local cultivars in the crop growth simulations would increase the chance to predict crop yield with greater accuracy. The cultivar parameter for local cultivars could have better representation of crop growth pattern and yield once the likely outcome of crop growth simulations was obtained from the comparison process between temporal change of crop growth derived from crop growth simulations and satellite data.

The percentile identified for prediction of crop yield tended to be consistent for the given crop and province, which suggested that the analytic framework had biases in crop growth simulations. It is inevitable that crop growth simulations have a large degree of errors due to uncertainty in representation of real crop management as well as environmental conditions. In particular, our results were based on crop management scenarios rather than actual crop management practices. Although it would be preferable to adjust the outcome of crop growth simulations to reality, the use of sophisticated procedures for bias correction was unwarranted due to limited availability of crop yield reports. Instead, the simple approach based on the percentile of crop yield within the distribution of crop growth simulations allowed for prediction of crop yield with acceptable accuracy except for rice in the present study.

This bias could result from the fact that landscapes in Mozambique would have a diversity in crops and vegetation likely due to dependence on subsistence farming (Debats et al., 2016). Subsistence farmers often choose intercropping for their farms due to its greater stability and efficiency than monocropping in terms of crop yield and land use, respectively (Drinkwater et al., 2021, Martin-Guay et al., 2018). It is likely that most growers may have a small field for production of food crops, which makes it challenging to characterize the time series of LAI for a specific crop (Manaze et al., 2020). Application of intercropping would make it more difficult to detect changes in crop growth even using satellite data at very high spatial resolution (Parra et al., 2022). Due to the absence of observation data for the fraction of farms where intercropping is used, the crop yield hindcasts based on monocropping conditions would be over-predicted due to competition between multiple crops grown in the given field compared to crop yield reports (Pierre et al., 2023). The mixture within crop canopy could also increase errors in the hindcast due to misidentification of crop management conditions using the analytic framework, which could have caused under-prediction errors. In the further studies, a survey on cropping systems in the region of interest would facilitate the development of an approach to correct biases caused by application of polyculture.

The analytical framework had large errors for rice yield, in part, likely due to the limited use of crop management scenarios. Application of fertilizer would be affordable mostly in commercial farms rather than subsistence farms. The price of chemical fertilizer can be prohibitive for a large fraction of growers in the given region (Benson and Mogues, 2018). Instead, crop residue can be used to improve crop yield by increasing soil organic matter and nutrients as well as soil moisture. Still, each farm may have a wide range of management practices, which would make it challenging to represent such practices using a small number of scenarios. For example, the green manure can be used as a cover crop, which would increase the sustainability of the farms (Meena et al., 2018). In addition, phosphorus contents have been recognized as one of limiting factors for crop production (Malhotra et al., 2018; Dzotsi et al., 2010). These practices cannot be added to the crop management scenarios due to limited computing resources for the gridded simulations.

In the crop growth simulations, the value of crop yield was obtained under the assumption that no damages by pests and diseases were taken into account (Yao et al., 2007). The spatial resolution of MODIS data would be too coarse to detect damages caused by such factors in individual fields (Gao et al., 2017). Thus, over-prediction errors would occur during the growing seasons where crops were damaged by these biotic stresses. In addition, abiotic stresses such as flooding often cause crop damages, which cannot be simulated using crop models. These hinted that the analytic framework could be improved by adding other procedures to account for biotic and abiotic stresses in forecasts of crop yield.

Our hindcasts provided insight on crop management to increase staple food crop production in Mozambique. In particular, high levels of crop yield could be attained using a large amount of fertilizer in Mozambique (JICA, 2022). For example, the crop yield hindcasts often failed to exceed 1000 kg / ha especially for maize and rice because little amount of fertilizer was applied to the crop growth simulations although crop yield could be higher than 7000 kg ha^{-1} in the crop growth simulations with fertilizer and irrigation applications at the rate used in the US and Korea. This suggested that the analytic framework could be revised to provide a crop yield predictions before the heading or flowering dates under a range of crop yield scenarios based on varying fertilizer management conditions. This would support local farmers in planning crop management in the later part of a growing season.

5. References

Araneda-Cabrera, R. J., Bermúdez, M., & Puertas, J. (2021). Assessment of the performance of drought indices for explaining crop yield variability at the national scale: Methodological framework and application to Mozambique. Agricultural Water Management, 246, 106692.

Ban, H. Y., Kim, K. S., Park, N. W., & Lee, B. W. (2016). Using MODIS data to predict regional corn yields. Remote Sensing, 9(1), 16.

Benson, T., & Mogues, T. (2018). Constraints in the fertilizer supply chain: evidence for fertilizer policy development from three African countries. Food Security, 10, 1479-1500.

Brida, A. B., Owiyo, T., & Sokona, Y. (2013). Loss and damage from the double blow of flood and drought in Mozambique. International Journal of Global Warming, 5(4), 514-531.

Debats, S. R., Luo, D., Estes, L. D., Fuchs, T. J., & Caylor, K. K. (2016). A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sensing of Environment, 179, 210-221.

Drinkwater, L. E., Midega, C. A., Awuor, R., Nyagol, D., & Khan, Z. R. (2021). Perennial legume intercrops provide multiple belowground ecosystem services in smallholder farming systems. Agriculture, Ecosystems & Environment, 320, 107566.
Dzotsi, K. A., Jones, J. W., Adiku, S. G. K., Naab, J. B., Singh, U., Porter, C. H., & Gijsman, A. J. (2010). Modeling soil and plant phosphorus within DSSAT. Ecological Modelling, 221(23), 2839-2849.
Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., ... & Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sensing of Environment, 188, 9-25.
Grassini, P., van Bussel, L. G., Van Wart, J., Wolf, J., Claessens, L., Yang, H., ... & Cassman, K. G. (2015). How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Research, 177, 49-63.
Gioia, F.S. 2024. TIA/IAI harmonized dataset: strengths and limitations. In: Agricultural development in Mozambique: trends, challenges and opportunities, IGM programme, June 11, 2024.
Han, E., Ines, A. V., & Koo, J. (2019). Development of a 10-km resolution global soil profile dataset for crop modeling applications. Environmental modelling & software, 119, 70-83.
Hara, P., Piekutowska, M., & Niedbała, G. (2021). Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data. Land, 10(6), 609.
Hijmans, R. J., Bivand, R., Forner, K., Ooms, J., Pebesma, E., & Sumner, M. D. (2022). Package ‘terra’. Maintainer: Vienna, Austria.
Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. Journal of Forestry Research, 32(1), 1-6.
JICA (Japan International Cooperation Agency). (2022). Republic of Mozambique project for improvement of rice productivity in Zambezia province (ProAPA) Work Completion Report.
Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., ... & Ritchie, J. T. (2003). The DSSAT cropping system model. European journal of agronomy, 18(3-4), 235-265.
Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in plant science, 10, 621.
Kim, J., Park, J., Hyun, S., Fleisher, D. H., & Kim, K. S. (2020). Development of an automated gridded crop growth simulation support system for distributed computing with virtual machines. Computers and electronics in agriculture, 169, 105196.
Lee, C. K., Kim, J., & Kim, K. S. (2015). Development and application of a weather data service client for preparation of weather input files to a crop model. Computers and Electronics in Agriculture, 114, 237-246.

Liu, Y., Kim, J., Fleisher, D. H., & Kim, K. S. (2021). Analogy-Based Crop Yield Forecasts Based on Temporal Similarity of Leaf Area Index. Remote Sensing, 13(16), 3069.
Mabiso, A., Cunguara, B., & Benfica, R. (2014). Food (In) security and its drivers: insights from trends and opportunities in rural Mozambique. Food Security, 6, 649-670.
Malhotra, H., Vandana, Sharma, S., & Pandey, R. (2018). Phosphorus nutrition: plant growth in response to deficiency and excess. Plant nutrients and abiotic stress tolerance, 171-190.
Mananze, S., Pôças, I., & Cunha, M. (2020). Mapping and assessing the dynamics of shifting agricultural landscapes using google earth engine cloud computing, a case study in Mozambique. Remote Sensing, 12(8), 1279.
Manuel, L., Chiziane, O., Mandhlate, G., Hartley, F., & Tostão, E. (2021). Impact of climate change on the agriculture sector and household welfare in Mozambique: an analysis based on a dynamic computable general equilibrium model. Climatic Change, 167, 1-18.
Matavel, C., Hoffmann, H., Rybak, C., Steinke, J., Sieber, S., & Müller, K. (2022). Understanding the drivers of food security among agriculture-based households in Gurué District, Central Mozambique. Agriculture & Food Security, 11(1), 7.
Martin-Guay, M. O., Paquette, A., Dupras, J., & Rivest, D. (2018). The new green revolution: sustainable intensification of agriculture by intercropping. Science of the total environment, 615, 767-772.
Meena, B. L., Fagodiya, R. K., Prajapat, K., Dotaniya, M. L., Kaledhonkar, M. J., Sharma, P. C., ... & Kumar, S. (2018). Legume green manuring: an option for soil sustainability. Legumes for soil health and sustainable management, 387-408.
Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. Computers and electronics in agriculture, 163, 104859.
Nyéki, A., Kerepesi, C., Daróczy, B., Benczúr, A., Milics, G., Nagy, J., ... & Neményi, M. (2021). Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods. Precision Agriculture, 22, 1397-1415.
Parra, L., Mostaza-Colado, D., Marin, J. F., Mauri, P. V., & Lloret, J. (2022). Methodology to Differentiate Legume Species in Intercropping Agroecosystems Based on UAV with RGB Camera. Electronics, 11(4), 609.
Pierre, J. F., Singh, U., Ruiz-Sánchez, E., & Pavan, W. (2023). Development of a Cereal–Legume Intercrop Model for DSSAT Version 4.8. Agriculture, 13(4), 845.
Priya, S. K., Balambiga, R. K., Mishra, P., & Das, S. S. (2023). Sugarcane yield forecast using weather based discriminant analysis. Smart Agricultural Technology, 3, 100076.
Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agricultural and Forest Meteorology, 284, 107886.

Singh, K., McClean, C. J., Büker, P., Hartley, S. E., & Hill, J. K. (2017). Mapping regional risks from climate change for rainfed rice cultivation in India. Agricultural systems, 156, 76-84.

Talacuece, M. A. D., Justino, F. B., Rodrigues, R. D. Á., Flores, M. E. P., Nascimento, J. G., & Santos, E. A. (2016). Modeling of soybean under present and future climates in mozambique. Climate, 4(2), 31.

Therond, O., Hengsdijk, H., Casellas, E., Wallach, D., Adam, M., Belhouchette, H., ... & Van Ittersum, M. K. (2011). Using a cropping system model at regional scale: Low-data approaches for crop management information and model calibration. Agriculture, Ecosystems & Environment, 142(1-2), 85-94.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

Yao, F., Xu, Y., Lin, E., Yokozawa, M., & Zhang, J. (2007). Assessing the impacts of climate change on rice yields in the main rice areas of China. Climatic Change, 80, 395-409.

You, L., Wood, S., & Wood-Sichra, U. (2009). Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. Agricultural Systems, 99(2-3), 126-140.

Supplementary Data

1. Maize yield hindcast (kg/ha) by province

Growing Season	Gaza		Manica		Nampula	
	Reports	Hindcasts	Reports	Hindcasts	Reports	Hindcasts
2011/12	471	553.8	1094	917.2	838	654.3
2012/13		543.3		697.9		825.4
2013/14	785	717	758	897.9	788	816.8
2014/15	308	554.4	757	864.1	572	648.4
2015/16		419.2		753.4		860
2016/17	764	646.7	829	805.8	978	816.5
2017/18		337.4		1119.9		936.4
2018/19		592.8		891.7		857.8
2019/20	210	456.6	811	911.4	929	713.9
2020/21		593.4		832.6		853.7
2021/22		442.5		885.2		664.5
2022/23		588.5		1082.1		764.1
2023/24		610.6		757.8		863.5

2. Soybean yield hindcast (kg/ha) by province

Growing Season	Nampula	
	Reports	Hindcasts
2011/12		1242.1
2012/13		1247.0
2013/14		1246.6
2014/15		1195.2
2015/16	1400	1331.3
2016/17	1067	1267.4
2017/18	1200	1067.9
2018/19	1100	1255.3
2019/20	1100	1194.6
2020/21		1424.7
2021/22		1437.1
2022/23		1348.9
2023/24		1438.6

3. Rice yield hindcast (kg/ha) by province

Growing Season	Gaza		Manica		Nampula	
	Reports	Hindcasts	Reports	Hindcasts	Reports	Hindcasts
2011/12	653	571.7	861	612.0	221	552.1
2012/13		121.4		441.4		435.9
2013/14	1101	1081.5	640	386.8	374	364.8
2014/15	696	123.1	285	514.2	291	707.6
2015/16		436.8		659.8		442.5
2016/17	729	531.1	696	677.8	398	216.4
2017/18		426.5		166.4		209.2
2018/19		755.6		589.6		279.0
2019/20	728	879.4	164	527.3	471	304.1
2020/21		384.6		593.4		1070.0
2021/22		31.3		440.6		-
2022/23		460.1		687.7		260.6
2023/24		1022.0		291.9		223.9