Received 26 October 2025, accepted 11 December 2025, date of publication 17 December 2025, date of current version 23 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3645705

APPLIED RESEARCH

Implementation of Long-Term Forecasting Models in Sugarcane for Agricultural Planning and Yield Goals Setting

JOEL MORALES LEMUS [1,2], DAVID MERA [1], JOSÉ LUIS QUEMÉ DE LEÓN [3], AND JOSÉ MANUEL COTOS [1]

[1]Computer Graphics and Data Engineering (COGRADE) Research Group, Department of Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[2]Agricultural Department, Azucarera Nacional, Aguadulce, Coclé 02040, Panama
[3]Guatemalan Center for Research and Training in Sugarcane (CENGICAÑA), Escuintla 05002, Guatemala

Corresponding authors: Joel Morales Lemus (joelestuardo.morales@rai.usc.es) and David Mera (david.mera@usc.es)

This work was supported in part by the Galician Government under Grant ED431B 2024/44 and in part bythe Universidade de Santiago de Compostela/Consorcio Interuniversitario do Sistema Universitario de Galicia (CISUG) for open access charge.

ABSTRACT This work aims to evaluate the efficiency of a long-term yield forecasting system for each minimum management unit (plot) at a sugar mill in Panama. Eleven teleconnections were used to forecast tons of cane per hectare (TCH) using Machine Learning (ML) with a lead time of 15-17 months before harvest. Three different ML models were trained, optimized, evaluated, and compared to select the best one for each plot. This process integrated individual plot-based predictions and measured overall efficiency. The Test results showed an average plot forecast of 79.00 TCH, while the average harvested TCH value was 78.59. Specifically, the plot-based models obtained an average Root Mean Square Error (RMSE) of 6.95 and a coefficient of determination (R^2) of 0.8018. Notably, this system contributes to achieving TCH forecasts by plot with greater efficiency than conventional estimation by the farm manager. The 2024 harvest was influenced by warm El Niño-Southern Oscillation (ENSO) conditions and was used as a case study to exemplify the system's applicability in managing work plans and adjusting them to the yield goals set by plot. To this end, individual plot yields were predicted in October 2022, resulting in a global average of 78.9 TCH. Based on the predicted results, a management plan was constructed and implemented in October 2022. This plan projected changes in management practices that would alter the expected production scenarios and unit costs. the cost per ton produced in the field decreased by 12.8 %, underscoring the effectiveness of the management strategies. The presented forecasting system could be implemented in other sugar mills, trained with local historical production data by plot, with the aim of being efficient in the use of inputs, minimizing environmental impact and taking actions that consider the effects of climate change on production.

INDEX TERMS Climate teleconnections, machine learning, sugarcane, yield forecasts.

I. INTRODUCTION

The sugarcane agroindustry in Panama serves as a vital economic engine. In the first half of 2024, it generated $464 million in exports, accounting for 9.4 % of the country's total export revenue [1]. This sector is crucial not only for international trade but also for economic development, particularly in rural areas. It creates jobs and stimulates related activities, such as transportation, logistics, and services.

This agroindustry contributes approximately 1 % to Panama's Gross Domestic Product (GDP), solidifying its role as a strategic pillar of the national economy. Domestic sugar demand is met by four sugar mills, with Azucarera Nacional being the largest. With 6,500 hectares of sugarcane fields, this mill faces the challenge of maintaining international competitiveness. Its strategy focuses on reducing production costs, optimizing agro-input usage, and mitigating

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva.

environmental impacts. Its main goal is to lower production unit costs below the international sugar price. By moving away from traditional practices such as global budgeting and end-of-cycle performance analysis, Azucarera Nacional aims to enhance expenditure management based on realistic production expectations.

The primary objective of the standard management process for sugarcane production is to reduce costs compared to previous years. A traditional drawback is the disconnect between the economic budget and the 12-month crop work plan. This management process is further complicated by the extensive land area and the staggered start dates of the crop cycle (from January to March), which introduce significant geographic and temporal diversity.

In addition to these complexities, the sugarcane agroindustry faces numerous challenges related to climate variability. These challenges affect various stages of production, including planting, harvesting, transportation, milling, and commercialization.

To improve outcomes, several key changes to the standard management process are proposed: 1) Production forecasting for each minimum management unit (plot) will be advanced by 15 to 17 months before harvest, depending on the harvest month; 2) Strategic work plans and budgets will be prepared 3 months before the start of the crop cycle for fields harvested in January, and 5 months earlier for those harvested in March. These plans will be aligned with the most probable climatic scenario in order to maximize production while minimizing costs; 3) A robust system will be implemented for recording environmental conditions, operational activities, and crop responses throughout the cycle. This will enable adaptive management by allowing for timely adjustments to emerging needs, thus reducing the risk of ineffective interventions and unnecessary financial expenditure.

An approach based on the proposed changes would enable Azucarera Nacional to implement an integrated risk management system that relies on climate forecasting, production estimation, and continuous data collection. This innovative approach would facilitate precise agronomic decision-making, optimizing profitability at the field level while rationalizing input management. Ultimately, this strategy would enhance export competitiveness while promoting economic, environmental, and food security in Panama.

In line with the first step of this roadmap, the objective of this work is to measure the efficiency of a novel long-term, plot-based forecasting system using Machine Learning (ML) models with a time horizon of 15 to 17 months. This system seeks to identify the most probable production scenario, considering climatic potential. These forecasts enable proactive management decisions to optimize resource allocation, maximize yields, reduce unit production costs, and set precise targets for operational planning and budgeting.

The developed system was evaluated using the 2024 harvest as a case study. Production forecasts were used to anticipate potential unit costs and adjust crop management plans. This revealed the need for a larger budget and, consequently, enabled production costs to be reduced by achieving output values higher than the initial forecasts, thanks to the effective management of changes to the original plan. This was consistent with the second point of the previously described roadmap.

It is important to note that this approach can be adapted to different agroclimatic conditions and geographical areas, offering a valuable framework for enhancing sugarcane production elsewhere.

The structure of this article is as follows: the related work to establish the foundation for the presented methodology is reviewed in Sect. II. After that, the datasets used to develop the forecasting systems are described in Sect. III. The subsequent Sect. IV details the methodology applied to develop and compare the long-term forecasting methods. Sect. V presents and discusses the results of the experiments, including evaluation metrics that demonstrate quantitative improvements in crop yield and cost reduction. Finally, Sect. VI summarizes the benefits of the approach and highlights its potential for replication and contribution to sustainable agricultural practices.

II. RELATED WORK

Integrating long-term forecasting tools is crucial for the effective management of agricultural systems, enabling the adoption of proactive strategies in areas such as management practices, water resources, environmental considerations, and agricultural supplies [2], [3].

Previous research has explored various strategies to optimize management processes in sugarcane production, paying particular attention to the effects of climatic variability on crop yield. Everingham et al. [4] highlighted the importance of integrating seasonal climate forecasting systems, which improve decision-making throughout the value chain regarding irrigation planning and yield expectations. Tyagi et al. [5] introduced a time series model to forecast long-term sugarcane production in India, assessing the impact of climate change on crop yield. However, this model focuses on regional forecasts and does not account for variations at the parcel level or different harvest months.

Many studies correlate sea surface temperatures and atmospheric interactions in different regions of the world with their impact on local climate variability to estimate climatic conditions. This phenomenon is known as teleconnections [6], [7], [8], [9]. Specifically, Literature shows studies that use the El Niño-Southern Oscillation (ENSO) [10], [11], [12], [13] or combinations of various teleconnections such as Sea Surface Temperature (SST), Arctic Oscillation (AO),Antarctic Oscillation (AAO), North Atlantic Oscillation (NAO), and Indian Ocean Dipole (IOD) [14], [15], [16] to forecast the annual production of various staple crops such as rice, wheat, soybeans, sorghum, peanuts, and corn in countries or macro-regions within countries.

Typically, these studies do not consider the harvest month when selecting the optimal teleconnections. However, this directly influences the results [17].

This strategy has also been used for long-term sugarcane forecasts based on ENSO [18], [19], [20], [21], [22], [23], [24] or the integration of various teleconnections [25], [26].

These works have typically covered an entire harvest season in a country or specific regions and have identified relationships between teleconnections and changes in the Normalized Difference Vegetation Index (NDVI) and production.

Nevertheless, this common approach does not usually consider planting and harvesting months, nor does it provide specific forecasts for each plot within a harvest season. Consequently, the management and contextual peculiarities of each plot are neglected.

Existing studies typically generate global production predictions rather than plot-level forecasts, often overlooking the variability of harvest months for each parcel. This limitation restricts the ability to visualize the most probable scenario and develop specific work plans for each plot that leverage climatic potential or mitigate negative environmental impacts, and agricultural management plans in accordance with the productive potential of each plot, which are managed based on profit margins.

In this context, this work aims to produce differentiated forecasts that take into account the month in which the specific production cycle for each plot begins, with the objective of defining the management plan and setting expectations for production targets and costs. To achieve this, a total of eleven teleconnections will be integrated as independent variables.

To our knowledge, no previous approach has combined multiple teleconnections to offer long-term, plot-specific forecasts.

III. DATASET

Azucarera Nacional operates in the central region of Coclé, Panama. The company manages production through plots. The company oversees a total of 630 plots, each averaging 15 area in hectares (ha) in size. The distribution of these plots is detailed in FIGURE 1.

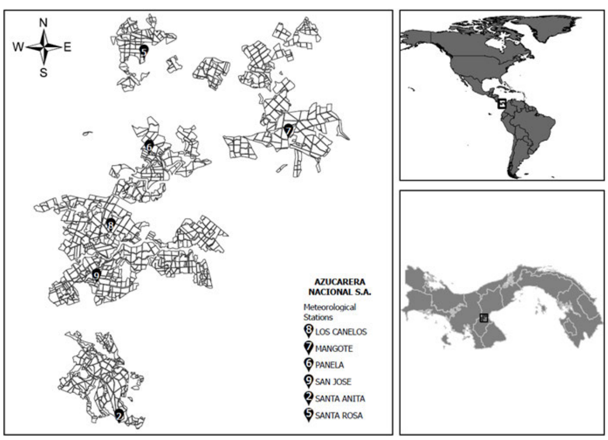


FIGURE 1. Location of the work area: Panama is located in the Americas (top-right). Azucarera Nacional is located in Panama (bottom-right). Distribution of Azucarera Nacional's production plots (left).

The datasets used for productivity analysis and to develop the forecasting models that support Azucarera Nacional's risk management strategy can be categorized into two main groups: 1) Internal data, developed by the company for private use, and 2) Public data, generated by third parties and openly accessible.

A. INTERNAL DATA

Azucarera Nacional provided a comprehensive harvest dataset covering the period from 2013 to 2023. This dataset included plot-level information such as ha, tons of cane per hectare (TCH), dollars per hectare ($/ha),dollars per hectare ($/TC), dollars per tons of sugar ($/TA), harvest date, planting date, number of harvests, and variety name. The resulting dataset was named DATA_YIELD.

B. PUBLIC DATA

The work also incorporated publicly available data on teleconnections in various regions, as shown in FIGURE 2. The specific teleconnections used are detailed below:

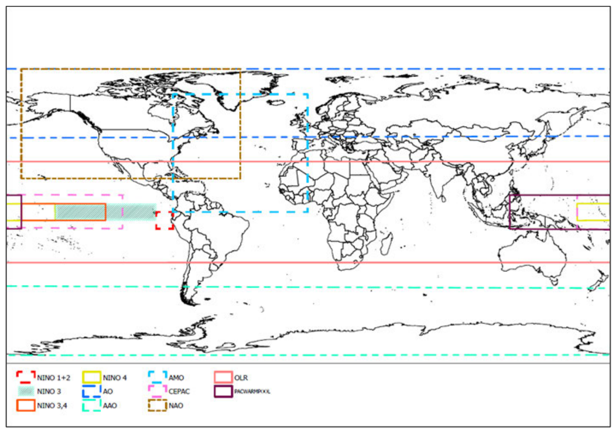


FIGURE 2. Monitoring areas of each teleconnection.

- AAO: A climate pattern characterized by variations in atmospheric pressure over the Antarctic region that influences weather patterns in the Southern Hemisphere.
- AO: The first mode of the Empirical Orthogonal Functions (EOF) analysis of monthly mean height anomalies at 1000 or 700 hectopascal (hPa), which affects weather patterns in the Northern Hemisphere.
- Atlantic Multidecadal Oscillation (AMO): A pattern of climate variability in the North Atlantic Ocean, characterized by non-smoothed, multi-decadal fluctuations in sea surface temperatures.
- Sea Surface Temperature in the Eastern Tropical Pacific (NIÑO 1+2): An index representing the sea surface temperature in the region of 0° - 10° S and 90° W, which is critical for understanding El Niño events.
- Sea Surface Temperature in the Central Tropical Pacific (NIÑO 3): An index representing the sea surface temperature in the region of 5° N - 5° S and 150° W - 90° W. This index is significant for climate variability assessments.

- Sea Surface Temperature in the Central-Western Tropical Pacific (NIÑO 3.4): An index that indicates the sea surface temperature in the region of 5° N - 5° S and 170° W - 120° W. It is often used in monitoring El Niño conditions.
- Sea Surface Temperature in the Western Tropical Pacific (NIÑO 4): An index referring to the sea surface temperature in the region of 5° N - 5° S and 160° E - 150° W. This index plays a role in climate dynamics.
- Warm Pool ENSO Index (CPAC): A climate index, also known as El Niño Modoki, representing variations in the warm pool of the Pacific Ocean that influence global weather patterns.
- NAO: A climate pattern in the North Atlantic that describes the fluctuations in the difference of atmospheric pressure at sea level between the Icelandic low and the Azores high.
- Outgoing Longwave Radiation (OLR): A measure of radiation emitted from the Earth's surface in the region of 5° N - 5° S and 160° E - 160° W. It provides insights into tropical convection and climate variability.
- Pacific Warm Pool (PACWARMPOOL): A climate feature characterized by warm water temperatures spanning from 10° N to 10° S and 120° E to 180° W. It plays a crucial role in tropical climate dynamics.

All these variables were integrated into a dataset, with values recorded for each month from October 2011 to July 2022. This database was named DATA_TELECOX.

It must be noted that the internal and the public data do not cover the same time period, as the aim is to make 15-17 month forecasts using teleconnection information. The monthly data were sourced from the National Oceanic and Atmospheric Administration (NOAA).

IV. METHODOLOGY

The internal and public datasets used in this work, referred to as DATA_YIELD and DATA_TELECOX, were merged into a single final dataset for developing the forecasting models. This new dataset, named DATA_WORK, was created by selecting the ID_plot, harvest year, harvest month, and TCH variables from the DATA_YIELD dataset. Furthermore, 10 monthly values of the linked teleconnections were added to these variables for each plot from the DATA_TELECOX dataset.

The selected 10-month period for each teleconnection begins in October two years prior to the harvest year and ends in July of the previous harvest year. Therefore, the period begins 15 months before the harvest if it occurs in January and 17 months before if it occurs in March. Conversely, the period ends 5 months prior to the harvest if it occurs in January, or 7 months if the harvest is carried out in March.

Taking the 2023 harvest as an example, as illustrated in FIGURE 3, the teleconnection period begins in October 2021 and ends in July 2022.

The beginning of the selected teleconnection period is always established based on the first harvest month, as all the plot-based forecasts must be available simultaneously to develop an appropriate general management plan and budget. This way, all the necessary inputs and materials for crop management can be anticipated and acquired.

It is worth noting that only 10 months of the teleconnections are considered when making production forecasts because relevant institutions' teleconnection projections are only available until July, when forecasts are made in October.

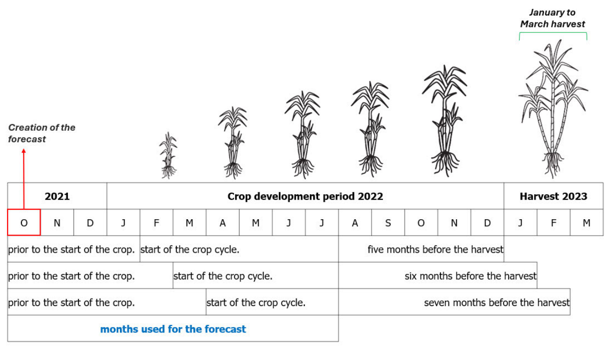


FIGURE 3. Months of the year considered in teleconnection-based forecasting, showing the lag period relative to the harvest month.

Several preprocessing data operations were applied to DATA_WORK in order to clean the dataset. First of all, missing data records were removed. Then, measurements from plots with anomalous yield values outside the 20-200 TCH range were eliminated. Finally, records related to plots with fewer than 9 years of TCH historical data were discarded. Consequently, 458 plots were used in the dataset, representing 73 % of the total dataset.

Once the DATA_WORK set was cleaned, it was divided into two subsets. The first one, which included data from 2013 to 2022, was used to train and validate the models, excluding 2021. This year was excluded due to significant deviations in crop management practices compared to historical trends, as shown in FIGURE 7. The 2021 data were severely limited by the effects of the pandemic, resulting in a 45 % reduction in fertilization and a 60 % reduction in weed control. The second subset contained parcel data from 2023 and was used to Test the models.

It is important to note that the observations within each subset were grouped according to their parcel of origin, as a specific model was developed for each plot of land. The methodological approach is outlined in FIGURE 4.

This work evaluated three distinct models for each plot: Random Forest (RF), Extreme Gradient Boosting (xgbTree), and Support Vector Machine with Radial Kernel (svmRadial). To predict the TCH for each plot, all available teleconnection variables were used as independent variables.

A hyperparameter optimization process was carried out for each plot-based model using a grid search technique and a 9-fold cross-validation approach. The evaluated hyperparameters are shown in TABLE 1.

The best model with optimal hyperparameter tuning was selected for each plot. This approach allowed us to analyze

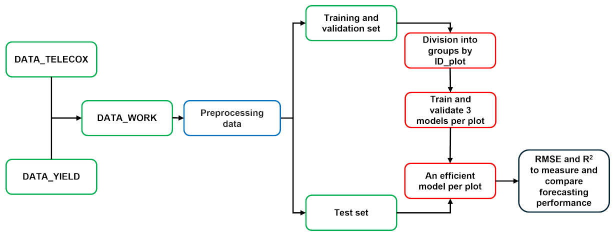


FIGURE 4. Process flow: Data integration and subset generation (green); preprocessing data (blue); model training, validation, and testing, with best model selection by plot (red).

how frequently each hyperparameter was selected for each plot-based model.

During this work, a SHapley Additive exPlanation (SHAP) analysis [27] was conducted to classify the importance of the independent variables for each plot. The analysis used the variation of the SHAP value relative to the variable with the highest value as a criterion. Then, the plot-based models, trained with specifically selected hyperparameter configurations, were evaluated using the Test data.

Finally, the selected plot-based models were compared with two baselines established using the current widely accepted approaches: 1) estimation based on farm manager experience, and 2) estimation based on accumulated rainfall [28].

TABLE 1. Hyperparameters evaluated via grid search as well as the frequency of the selected hyperparameter value (rounded frequency in parentheses). The following hyperparameters were kept for all training processes according to the model: RF_ntree=500, xgbTree_nrounds=100, xgbTree_colsample_bytree=1, xgbTree_lambda=1, xgbTree_alpha=0, xgbTree_scale_pos_weight=1 and svmRadial_scale=True.

Model	Evaluated hyperparameters	Hyperparameter frequency selection per plot
RF	mtry: 1, 4, 6, 8, 10	1 (40 %), 4 (60 %)
	max_depth: 10, 20, 30	10 (100 %)
	min_samples_leaf: 1, 5, 10	1 (100 %)
xgbTree	eta: 0.1, 0.2, 0.3	0.2 (20 %), 0.3 (80 %)
	max_depth: 2, 4, 6, 8	6 (90 %), 8 (10 %)
	min_child_weight: 0.5, 1, 2	1 (70 %), 2 (30 %)
	gamma: 0, 1, 2	0 (10 %), 1 (90 %)
	subsample: 0.5, 1	1 (100 %)
svmRadial	cost: 0.01, 0.1, 1, 10	1 (100 %)
	gamma: 0.01, 0.1, 1, 10	0.1 (70 %), 1 (30 %)
	epsilon: 0.1, 0.5, 1	0.1 (100 %)

A. HARVEST IMPROVEMENT SCENARIO

The purpose of developing a forecasting system is to create a reference tool to optimize agricultural management. Thus, the developed system was used in October 2022 to forecast production for 2024.

In this realistic scenario, it should be noted that the predictions were obtained using teleconnection projections, since the system was executed 15-17 months prior to harvest.

Based on these forecasts, several changes were implemented to optimize the harvest. The impact of these changes implemented in the management plan and budget was evaluated on the unit cost indicators of production for the 2024 harvest.

V. RESULTS AND DISCUSSION

The models were evaluated using criteria that minimize the Root Mean Square Error (RMSE) and maximize the coefficient of determination (R^2). TABLE 2 shows the results obtained using the cross-validation approach, taking into account the best-optimized model for each plot. TABLE 1 shows the frequency with which hyperparameters were selected for each plot. The harvest forecast results indicated a weighted average (by the area of each plot) of 76.32 TCH considering all plots, with a difference of only 0.25 TCH from the actual value.

The RMSE and R^2 averages were 1.19 and 0.9959, respectively. These values indicate that the model effectively captures the variability of historical data and reveals a global fit of the model system to the validation data. Teleconnections explain the variability of production, which results from the applied climatic conditions and agronomic management.

Therefore, the forecast allows for anticipating a possible production scenario under a management approach similar to that of the analyzed years, in future environmental conditions that can be projected comparably.

TABLE 2. Forecasting averaged metrics derived from cross-validation and testing phases.

Data Set	TCH Real	TCH Forecast	RMSE	R^2
Cross-validation (2013-22)	76.57	76.32	1.19	0.9959
Test (2023)	78.59	79.00	6.95	0.8018

Once the optimized models were selected through the cross-validation process, they were evaluated using the Test set corresponding to the 2023 harvest. Although a reduction in forecast efficiency was anticipated, the difference between the actual and predicted TCH averages was only 0.41. The actual average harvested value of TCH for this period was 78.59, while the forecast was 79.00. These results yielded an RMSE of 6.95 and an R^2 of 0.8018, as shown in TABLE 2.

A notable factor influencing the Test forecasting accuracy was the introduction of a novel crop variety during the 2023 harvest. This event led to significant production increases and, consequently, reduced prediction precision in certain plots. Despite this challenge, the plot-based models demonstrated robust generalization capabilities when presented with previously unseen data in the Test set.

Table 3 shows how often each of the three evaluated models was chosen using the plot-based selection method. The xgbTree model was selected most frequently, with a 95 % selection rate, which indicates its superior predictive performance for plot yields.

In contrast, the svmRadial model was selected only 4 % of the time, and the RF model was selected in 1 % of cases. This variability underscores the importance of plot-specific model selection for accurate point forecasting. It suggests that a single, universally applied model may not achieve optimal predictions across all plots. Additionally, even when the selected model was the same, each plot obtained a different

TABLE 3. **Plot-based model selection frequency using the proposed methodology.**

Model	Frequency
RF	1 %
svmRadial	4 %
xgbTree	95 %

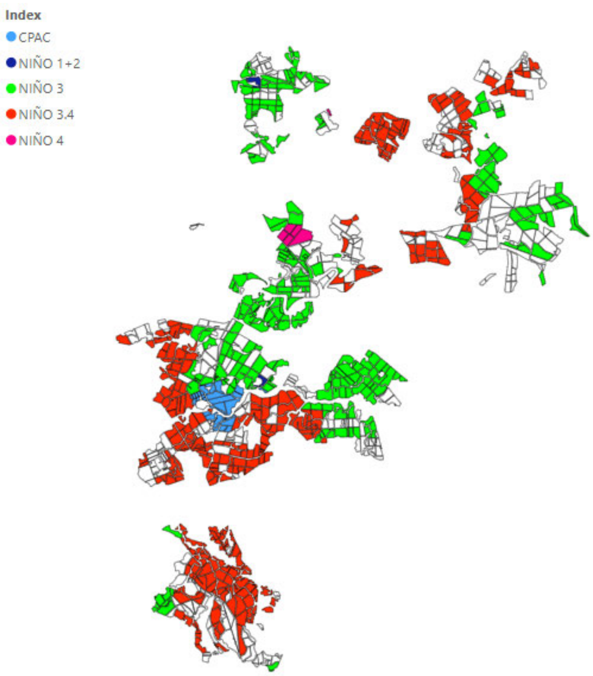


FIGURE 5. **Most relevant feature for each selected plot-based xgbTree model.**

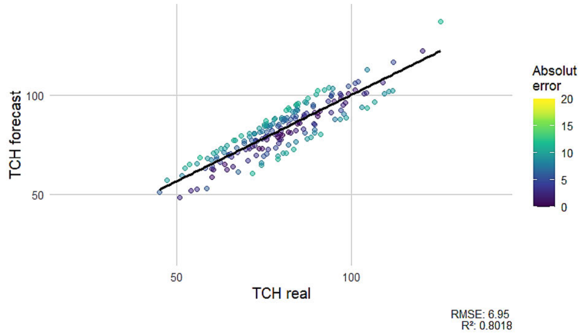


FIGURE 6. **Scatter chart of the observed and predicted TCH by plot between 15 and 17 months before harvest, utilizing the Test set from the 2023 harvest season.**

version of it due to the training process and hyperparameter optimization procedure.

The xgbTree model was selected in 435 plots out of a total of 458. The models linked to these plots were used to quantify how frequently the variables were classified as relevant, based on the SHAP value. A variable was considered important if its value varied by less than 50 % compared to the most important variable for each plot, accounting for the total number of plots where that variable was deemed relevant. The variables that appeared most frequently, in hierarchical order, were: NIÑO 3 (100 %), NIÑO 3.4 (100 %), NIÑO 1+2 (98 %), CPAC (87 %), NIÑO 4 (83 %), PACWARMPOOL (77 %), AMO (74 %), AAO (31 %), NAO (21 %), AO (17 %), and OLR (3 %).

FIGURE 5 shows the geographical distribution of the most relevant features for which the xgbTree model was selected. This distribution groups those teleconnections that most effectively explain the variation in production associated with climatic potential.

Spatial clustering sectors are observed that may provide a better understanding of the interaction between the environment and the crop, as well as characteristics such as soil type in relation to different teleconnections. These grouping sectors may be attributed to more homogeneous areas, which cause the crops to respond similarly to specific climatic conditions and to certain more specific teleconnections.

The Test results for each plot are illustrated in FIGURE 6. This figure demonstrates that the real yields of TCH varied between 48 and 162, with a forecast absolute error ranging from 0 to 12 TCH. The analysis of the 2023 data reveals a robust performance of the model in estimating future yields, taking into account the crop's potential with technological management practices similar to those employed in historical series.

The results of the optimized plot-based models were compared with two baselines using the 2023 harvest dataset, as shown in FIGURE 7. The first baseline is a comparison with the farm manager's estimation, which was conducted by plot and averaged to give a total of 82.4 TCH. By contrast, our approach predicted an average of 79.0 TCH, which closely aligns with the actual observed average value of 78.59 TCH. Specifically, the RMSE of the farm manager's estimation (11.16 TCH) was approximately 60.6 % higher than that of the presented approach (6.95 TCH). This indicates greater efficiency than the estimation made by the farm manager. Moreover, it should be noted that the farm manager's estimation was made only six months before the harvest. In contrast, our forecasting models were developed to predict 15-17 months before the harvest.

The second baseline relates the amount of accumulated rainfall during the elongation stage to the total yield. It should be noted that the crop requires approximately 1,100 mm of rainfall during this phase, and excessive or insufficient rainfall reduces production. Using this rainfall-based approach, the estimation was 79.5 TCH. While this is better than the one made by the farm manager, it is still worse than that generated by the presented forecasting system. However, it should be highlighted that this rainfall-based method was based on actual rainfall measurements. Any forecast generated using rainfall estimations will have a higher margin of error. Furthermore, the rainfall-based estimation is global, which prevents its use for plot-specific planning.

A. HARVEST IMPROVEMENT RESULTS

In 2022, the developed prediction system was used to forecast the 2024 harvest. This was achieved by using

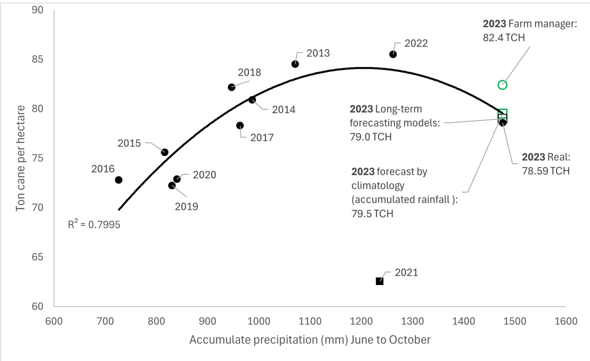


FIGURE 7. Comparison of the proposed forecasting system against two baselines: 1) the farm manager's estimation and (2) a model based on the correlation between accumulated rainfall during the elongation stage and total yield.

teleconnection projections, with the aim of optimizing agronomic management and maximizing harvest yields.

It should be noted that the long-term expectation in 2022 was that the 2024 harvest crop would develop under WARM ENSO conditions, as published by International Research Institute for Climate and Society (IRI). The SST anomaly index values exceeded 0.5 from March and reached 2.0 in December 2023. The total precipitation forecast was 1,421 mm, whereas the actual recorded precipitation was 1,315 mm. Between January and October 2023, an increase in the average temperature of +0.35°C was observed, alongside a 4.5 % increase in total energy (W/m^2) and a 54 % reduction in precipitation compared to the previous year.

In these conditions, a long-term production forecast was made in October 2022 for the growing season beginning in January 2023 and concluding with the harvest between January and March 2024. This forecast estimated a yield of 78.9 TCH, as illustrated in FIGURE 8.

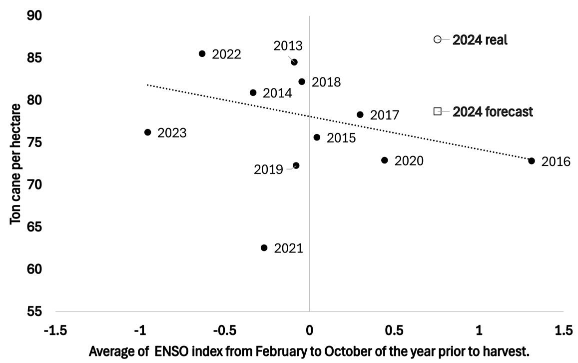


FIGURE 8. The average production in TCH from 2013 to 2023 is shown alongside the average ENSO index during the elongation period. Agronomic management was not carried out in 2021 due to the pandemic. The forecast for 2024 is shown alongside the actual TCH, taking into account changes to the management plan based on system predictions.

A budget and production $/TC were developed based on a TCH forecast of 78.9 under conventional management in 2024. Linked costs indicate an expected increase of 17 % above the target unit cost, which should be 5 % lower than in 2023. In light of this scenario, a specific management plan was designed for each plot that took into account the water balance and irrigation needs. On average, an additional 1.5 irrigations were estimated compared to the previous year, but an extra 2 irrigations were actually implemented. The objective of this plan was to manage the crop in order to maximize the forecasted energy increase and exceed historical production levels under similar climatic conditions.

Before implementation, the risk of failing to achieve the production and profitability objectives was assessed, taking into account the probable variation in $/TC. Two possible scenarios were proposed: 1) maintaining the budgeted $/ha and not increasing irrigation efforts, which would result in a production of 78.9 TCH, but with a $/TC cost 12 % higher than the previous year and 17 % above the target; or 2) increasing the cost relative to the budgeted $/ha by modifying crop management, thereby achieving an increase in projected production according to the forecast, with a $/TC close to the required target.

Consequently, in 2024, compared to the previous year, the $/TC index decreased by 12.8 %, which was 7.8 % lower than the target, and the $/TA index decreased by 9.7 %, relating to the production costs of standing sugarcane (excluding harvesting and manufacturing costs).

VI. CONCLUSION AND FUTURE WORK

Implementing a long-term yield forecasting system based on ML to estimation TCH by plot has proven to be efficient in predicting sugarcane yields. The developed system uses 11 climatic teleconnections with a seasonality of 15 to 17 months before harvest.

Forecasting results for the 2023 harvest, trained and validated using data from 2013 to 2022, indicated a real TCH of 78.59 and a predicted value of 79.00, with an RMSE of 6.95, demonstrating the system's accuracy by plot. The modeling framework has successfully integrated environmental variations, reflecting the crop's potential under similar management conditions. This is crucial for long-term planning and resource management.

Forecasting a yield of 78.9 TCH for the 2024 harvest, influenced by warm ENSO climatic conditions, highlights the importance of considering climatic factors in agricultural planning.

Having a prior production scenario enabled the development of a differentiated management plan and budget for the 2024 harvest, resulting in altered final productivity compared to the initial forecast. Despite the budget being increased to accommodate the modified plan, a 12.8 % reduction in the $/TC index was achieved compared to the previous year, demonstrating the effectiveness of differentiated management strategies. This suggests that anticipating possible scenarios and adapting management practices is fundamental to achieving cost efficiency and maximizing yield.

The presented methodology is generalizable, allowing it to be applied to other sugar mills in any region. To do so, the linked models must be trained using the historical yield data of the mill in question at the plot level. This enables predictions tailored to the productive potential of the environmental context and management practices. Consequently, the forecasting system could serve as part of a long-term planning tool of the mill.

Future research could involve evaluating additional ML algorithms and ensemble techniques to improve prediction accuracy. Furthermore, it would be beneficial to incorporate variables such as variety type and to consider technological or varietal changes among the variables. It is also crucial to integrate direct climatic data and cultivation management practices to strengthen forecasts in various agricultural management scenarios.

ACKNOWLEDGMENT

The authors thank Azucarera Nacional, Panama, for providing access to the information used in this work.

REFERENCES

[1] M. de Economía y Finanzas. (2024). *Informe Económico Y Social: Primer Trimestre De 2024*. [Online]. Available: https://www.mef.gob.pa/wp-content/uploads/2025/01/MEF-DAES.-Informe-Economico-y-Social-Primer-semestre-2024.pdf

[2] P. T. Ghazvinei, H. H. Darvishi, A. Mosavi, K. B. W. Yusof, M. Alizamir, S. Shamshirband, and K.-W. Chau, "Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 738–749, Jan. 2018.

[3] P. Ma, N. Zhang, Y. Yang, Z. Wang, G. Li, and Z. Fu, "Forecast of sugarcane yield in chongzuo, Guangxi—LSTM model based on fusion of trend yield and meteorological yield," *Agronomy*, vol. 14, no. 11, p. 2512, Oct. 2024.

[4] Y. L. Everingham, R. C. Muchow, R. C. Stone, N. G. Inman-Bamber, A. Singels, and C. N. Bezuidenhout, "Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts," *Agricult. Syst.*, vol. 74, no. 3, pp. 459–477, Dec. 2002.

[5] S. Tyagi, S. Chandra, and G. Tyagi, "Climate change and its impact on sugarcane production and future forecast in India: A comparison study of univariate and multivariate time series models," *Sugar Tech*, vol. 25, no. 5, pp. 1061–1069, Oct. 2023.

[6] M. Dobrynin, D. I. V. Domeisen, W. A. Müller, L. Bell, S. Brune, F. Bunzel, A. Düsterhus, K. Fröhlich, H. Pohlmann, and J. Baehr, "Improved teleconnection-based dynamical seasonal predictions of boreal winter," *Geophys. Res. Lett.*, vol. 45, no. 8, pp. 3605–3614, Apr. 2018.

[7] F. Chen, E. Du, H. Jia, Y. Chen, and L. Wang, "Lagged effects of atmospheric circulation teleconnections on agricultural drought prediction in China," *Int. J. Digit. Earth*, vol. 18, no. 1, Aug. 2025, Art. no. 2528628.

[8] W. U. Hassan, M. A. Nayak, and M. F. Azam, "Intensifying spatially compound heatwaves: Global implications to crop production and human population," *Sci. Total Environ.*, vol. 932, Jul. 2024, Art. no. 172914.

[9] P. E. Teodoro, C. A. da Silva Junior, R. C. Delgado, M. Lima, L. P. R. Teodoro, F. H. R. Baio, G. B. de Azevedo, G. T. de Oliveira Sousa Azevedo, A. de Andréa Pantaleão, G. F. Capristo-Silva, and C. U. Facco, "Twenty-year impact of fire foci and its relationship with climate variables in Brazilian regions," *Environ. Monitor. Assessment*, vol. 194, no. 2, p. 90, Feb. 2022.

[10] K. Kumar, K. R. Kumar, R. Ashrit, N. R. Deshpande, and J. Hansen, "Climate impacts on Indian agriculture," *Int. J. Climatol.*, vol. 24, no. 11, pp. 1375–1393, 2004.

[11] P. Kongsurakan, T. Kato, X. Li, T. Nakagawa, and A. Suzuki, "Yield dynamics of major crops in Thailand for the period 1918–2023: Regional patterns of growth, stagnation, and climate sensitivity," *Field Crops Res.*, vol. 333, Nov. 2025, Art. no. 110094.

[12] R.-Y. Wang, L. Xu, T. Li, P. Yin, W. Zhu, and Q. Han, "Temporal-spatial characteristics of climate drought and its effects on grain yield in Hunan province, China," *Frontiers Environ. Sci.*, vol. 12, Oct. 2024, Art. no. 1489662.

[13] S. Shukla, F. Zaheer, A. Hoell, W. Anderson, H. Jayanthi, G. Husak, D. Lee, B. Barker, S. Pervez, K. Slinski, C. Justice, J. Rowland, A. L. McNally, M. Budde, and J. Verdin, "ENSO-based outlook of droughts and agricultural outcomes in Afghanistan," *Weather Climate Extremes*, vol. 45, Sep. 2024, Art. no. 100697.

[14] R. Bhatla, S. Bhattacharyya, S. Verma, R. K. Mall, and R. S. Singh, "El Nino/La Nina and IOD impact on kharif season crops over western agro-climatic zones of India," *Theor. Appl. Climatol.*, vol. 151, nos. 3–4, pp. 1355–1368, Feb. 2023.

[15] S. Yan and H. Chen, "The significant influence of the sea surface temperature anomalies over North Atlantic and the maritime continent on maize yield in Northeast China," *Atmos. Res.*, vol. 314, Mar. 2025, Art. no. 107806.

[16] S. Peethani, M. Pathakoti, A. M. Kheir, A. Singh, and A. Govind, "Variability of ENSO teleconnections indices, and its impacts on Moroccan agriculture," *J. Atmos. Solar-Terrestrial Phys.*, vol. 265, Dec. 2024, Art. no. 106386.

[17] H. Ren, X. Li, P. Liu, Y. Zhou, T. Zhang, and X. Li, "Impact of Indian ocean sea surface temperature anomalies on winter wheat yields in China," *Theor. Appl. Climatol.*, vol. 156, no. 8, p. 419, Aug. 2025.

[18] O. Castro and H. Monterroso. (2016). *Memoria Presentación De Resultados De Investigación Zafra*. [Online]. Available: http://unfccc.int/resource/docs/convkp/convsp.pdf

[19] Y. L. Everingham, N. E. Stoeckl, J. Cusack, and J. A. Osborne, "Quantifying the benefits of a long-lead ENSO prediction model to enhance harvest management—A case study for the herbert sugarcane growing region, Australia," *Int. J. Climatol.*, vol. 32, no. 7, pp. 1069–1076, Jun. 2012.

[20] Y. Everingham, C. Baillie, G. Inman-Bamber, and J. Baillie, "Forecasting water allocations for bundaberg sugarcane farmers," *Climate Res.*, vol. 36, pp. 231–239, Jun. 2008.

[21] A. Schepen, J. Sexton, B. Philippa, S. Attard, D. E. Robertson, and Y. Everingham, "Downscaled numerical weather predictions can improve forecasts of sugarcane irrigation indices," *Comput. Electron. Agricult.*, vol. 221, Jun. 2024, Art. no. 109009.

[22] T. W. Mbhamali and M. R. Jury, "Climate-sensitivity of sugarcane yield in the Southeastern Africa lowlands," *Int. J. Climatol.*, vol. 41, no. 8, pp. 4187–4200, Jun. 2021.

[23] P. C. Sentelhas and A. B. Pereira, "El Niño–Southern oscillation and its impacts on local climate and sugarcane yield in Brazil," *Sugar Tech*, vol. 21, no. 6, pp. 976–985, Dec. 2019.

[24] G. De, S. Rolim, V. B. Moreto, and A. P. Pantano. *Influence of El Niño and La Niña on Sugarcane Yield and Sucrose Production in northern São Paulo, Brazil*. [Online]. Available: https://www.researchgate.net/publication/283100879

[25] W. De castro Silva and H. A. Barbosa, "Avaliação da precipitação na produtividade agrícola da cana-de-açúcar: estudo de caso usina coruripe para as safras de 2000/2005," *Revista Brasileira de Geografia Física*, vol. 13, no. 3, pp. 1352–1366, 2021.

[26] N.-B. Chang, M. V. Vasquez, C.-F. Chen, S. Imen, and L. Mullon, "Global nonlinear and nonstationary climate change effects on regional precipitation and forest phenology in Panama, central America," *Hydrological Processes*, vol. 29, no. 3, pp. 339–355, Jan. 2015.

[27] M. N. Uddin, L.-Z. Li, B.-Y. Deng, and J. Ye, "Interpretable XGBoost–SHAP machine learning technique to predict the compressive strength of environment-friendly Rice husk ash concrete," *Innov. Infrastructure Solutions*, vol. 8, no. 5, p. 147, May 2023.

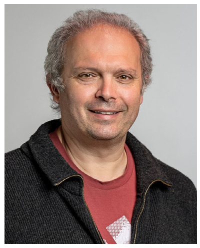
[28] N. G. Inman-Bamber and D. M. Smith, "Water relations in sugarcane and response to water deficits," *Field Crops Res.*, vol. 92, nos. 2–3, pp. 185–202, Jun. 2005.

JOEL MORALES LEMUS received the master's degree in economics and the master's degree in applied statistics from the University of San Carlos of Guatemala. He is currently pursuing the Ph.D. degree with the University of Santiago de Compostela, Spain. His doctoral thesis is titled "Predictive System of Productivity in Sugarcane Crops." He has experience in university teaching in applied statistics for agriculture. He has collaborated with Guatemalan Center for Training and Research on Sugarcane (CENGICAÑA) and has been the Head of Research at Ingenio Magdalena, Guatemala. He is the Head of research and production at Azucarera Nacional, Panama, and an Associate Researcher at the Central Pacific Agroecological Research Center of Panama (CIAPCP-AIP). He is an Agronomist. His focus is on transforming agricultural processes through the concept of agriculture 4.0. He implements advanced technologies for data-driven management and decision-making, aligning people, processes, and technology.

DAVID MERA received the Ph.D. degree in computer science from the Universidade de Santiago de Compostela (USC), in 2013. He is currently an Associate Professor with USC. His research interest includes the application of machine learning to industrial challenges. He has contributed significantly to his research field, authoring 13 articles in journals with a high impact factor and 25 conference papers. His innovative work is also reflected in two patents and two software patents. Bridging the gap between academia and industry, he actively leads several industry-driven machine learning projects, translating cutting-edge research into real-world applications. In 2013, he was awarded the Prestigious Marie Skłodowska-Curie Fellowship.

JOSÉ LUIS QUEMÉ DE LEÓN was born in Quetzaltenango, Guatemala, in June 1957. He received the degree in agricultural engineering from USAC, in 1982, the Master of Science degree in genetics from the Colegio de Postgraduados, Mexico, in 1989, and the Ph.D. degree in plant breeding from UAAAN, Mexico, in 2000.

Since 1982, he has professionally dedicated himself to agricultural research in maize (15 years) and sugarcane (since 2000), carrying out scientific and technological activities in the public, private, and academic sectors, mainly in the areas of genetic improvement and statistics, contributing to the sustainability and profitability of production systems in both crops. He served as the coordinator of the Maize Program, Institute of Agricultural Science and Technology (ICTA). He is currently a Plant Breeder with Guatemalan Center for Research and Training on Sugarcane (CENGICAÑA), where he has created 19 commercial CENGICAÑA-Guatemala (CG) sugarcane varieties, including CG02-163, which has had a significant impact on Guatemala's sugarcane agroindustry.He has published over 100 scientific articles, taught university courses with an emphasis on statistics at national and international levels, and has made significant contributions to methodological development in plant breeding. He is a member of the International Society of Sugar Cane Technologists (ISSCT), the Association of Sugar Technicians of Guatemala (ATAGUA), and the College of Agronomists of Guatemala. He has also trained diverse audiences in short courses on statistical analysis applied to agricultural research. He has received multiple national and international recognitions for his scientific and technological career and was a candidate for the Science and Technology Medal, the highest award granted in Guatemala in the fields of science, technology, and innovation.



JOSÉ MANUEL COTOS has been a Professor with the Department of Electronics and Computing, Universidad de Santiago de Compostela, since 1993. He is currently the Coordinator of the Computer Graphics and Data Engineering Research Group. He has participated in more than 20 research projects and in more than 50 contracts with companies and institutions, mostly related to the transfer of technology to the business sector. From 2009 to 2013, he was attached to the presidency of a university network for technology transfer, RedEmprendia. In addition, he was the Founding Partner and the Administrator of the spin-off Paralaxe, Multimedia and Virtual Systems SL, a spin-off company of the Institute of Technological Research, University of Santiago de Compostela, that was dedicated to the development of multimedia and virtual reality computer systems. He is involved in the implementation of machine learning algorithms applied to industrial processes.