Leveraging Satellite Data With Machine and Deep Learning Techniques for Corn Yield and Price Forecasting

Florian Teste, Hugo Gangloff, Mathilde Chen, Philippe Ciais, and David Makowski

Abstract— This research introduces a new method for predicting changes in corn yield and price, critical for food security. Instead of relying on difficult-to-access pre-harvest production data, our approach uses satellite-derived gross primary production (GPP) data and dimension-reduction techniques to forecast national corn yield and price changes. We conducted case studies in the U.S., Malawi, and South Africa to validate this approach, analyzing predictors from annual GPP variations during peak growing seasons. We applied dimension-reduction strategies, such as spatial averaging and empirical orthogonal functions (EOFs). Additionally, we used deep learning techniques such as autoencoders (AEs) and variational autoencoders (VAEs) to extract meaningful features from the high-dimensional GPP datasets. These features were then used as predictors in yield and price classification models based on generalized linear models (GLMs) and ElasticNet. We also considered a neural network model trained to predict yield and price variations from GPP input data directly. We evaluated the model performances using metrics such as area under the curve (AUC), brier skill score (BSS), and Matthew's correlation coefficient (MCC). Our results indicate that dimension-reduction techniques based on AEs and VAEs provided better predictive capabilities across all three countries compared to EOF, particularly in Malawi, where the BSS increased from 0.26 with EOF to 0.81 with AE for yield and from -0.004 with EOF to 0.77 with VAE for price. This study demonstrates that integrating open-access satellite data with dimension-reduction techniques can significantly improve crop forecast accuracy, providing an accessible tool to enhance agricultural management and food security.

Index Terms— Autoencoder (AE), classification, commodity price, crop yield, dimension reduction, neural network, remote sensing.

I. INTRODUCTION

OVER the past two decades, the global food system has undergone significant transformation, characterized by increased agricultural output, enhanced transportation infrastructure, and a shift toward market liberalization. These changes have collectively improved food availability across various regions. However, despite these advancements, issues related to equitable distribution persist, leading to ongoing food insecurity and malnutrition, particularly in low-income countries [9].

The incidence of undernourishment, a measure of chronic food deprivation, has been on a worrying upward trend since 2017 [9]. In 2022, an estimated 691 to 783 million individuals experienced hunger [9]. Projections indicate that global food demand will increase by 35% to 56% between 2010 and 2050 [31].

According to FAO food balance sheets, corn is one of the most consumed staple crops globally, contributing significantly to daily per capita caloric intake.

In Sub-Saharan Africa (SSA), corn is particularly crucial in ensuring food security [21]. For instance, in Malawi, corn cultivation occupies 61% of the total agricultural land and, according to Mazunda and Droppelmann [20] and Ecker and Qaim [7], accounts for 60% of the country's total food production and approximately 45% of its total food consumption, underscoring the pivotal role of corn in ensuring food security in the region.

The intricate relationship between crop yields and commodity prices has profound implications for global food security and economic stability. Accurate forecasting of crop yields and commodity prices at the national scale is essential for various stakeholders, including policymakers, traders, farmers, and consumers [28].

Research into agricultural commodity price forecasting has mainly relied on traditional econometric models such as autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and vector autoregression (VAR) [1], [5], [17]. Although these models perform well for short-term forecasts (especially daily forecasts), machine learning models have gained in popularity in recent years. Studies by Zelingher and Makowski [36], Xu and Zhang [33], and Ouyang et al. [22] highlight the potential of these algorithms to produce accurate predictions by capturing complex patterns within the data. These models have proven effective in forecasting commodity prices based on predictors related to annual crop yield or production variations [35], [36].

However, acquiring reliable crop yield and production data can pose serious challenges, particularly for predicting price changes before harvest. In this study, we predict changes in crop yield and prices, several months before harvest, using

only remote sensing data available during the growing season. In order to summarize high-dimensional remote sensing data, we compare a large range of dimension-reduction techniques, from the traditional principal component analysis (PCA) to the more recent variational autoencoder (VAE) method.

The synthetic gross primary production (GPP)-based variables generated by these dimension-reduction techniques are then used as predictors in classification models forecasting yield and price changes, several months before crop harvest. As the model predictors rely exclusively on freely available remote sensing data, our approach can be easily used in practice.

We assessed the ability of our approach to deliver forecasts of both yield and price variations in three contrasting countries: the United States—the world's leading corn producer and exporter—and two SSA countries, Malawi and South Africa, where corn is a major staple crop.

II. Materials and Methods

A. Data

1) Corn Price: For the USA, we obtained global monthly prices for corn (U.S. No. 2 yellow free on board (FOB) Gulf of Mexico) from the World Bank's commodity markets database [3], measured in U.S. dollars nominal price per metric ton. We focused on prices in October, the main harvest month critical to the market. For South Africa, we used the agricultural statistics provided by the reports of the South African Grain Information Service (SAGIS) [24]. These reports provide annual producer prices per marketing year for yellow corn, with data dating back to 1931. The Food and Agriculture Organization (FAO) supplied the annual corn prices received by producers in Malawi [8]. These producer prices reflect the price received by the farmers per marketing year, with data available from 1996 onward.

In this study, the corn price Fig. 1 is noted as p_t, where t represents the year. The difference in price compared to the previous year is expressed as

$$\Delta p_t = p_t - p_{t-1}. \tag{1}$$

Additionally, we defined a binary variable, Δp_t^b, which equals one in the case of a price increase ($\Delta p_t^b = 1$ when $\Delta p_t > 0$) and zero otherwise

$$\Delta p_t^b = \begin{cases} 1 & \text{if } \Delta p_t > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The quantities Δp_t and Δp_t^b were computed from 1982 to 2018 for the U.S. and South Africa and from 1996 to 2018 for Malawi.

2) Corn Yield: The United States, the leading global corn producer, accounted for 30% of the total corn production in 2023. A significant 80% of this production is concentrated within the U.S. corn belt region, which includes 13 Midwestern states: Illinois (IL), Indiana (IN), Iowa (IA), Kansas (KS), Kentucky (KY), Michigan (MI), Minnesota (MN), Missouri (MO), Nebraska (NE), North Dakota (ND), Ohio (OH), South Dakota (SD), and Wisconsin (WI). We have obtained comprehensive historical data on annual U.S. corn yields, measured in

hectograms per hectare, from 1961 to 2022 from the FAO [8]. Similarly, we obtained the same period's historical yield data for Malawi and South Africa from the FAO.

In our study, the annual corn yield Fig. 2 at time t is represented by yield_t. To analyze the yearly yield variations, we calculated the yield difference between consecutive years, denoted as Δyield_t, as follows:

$$\Delta \text{yield}_t = \text{yield}_t - \text{yield}_{t-1}. \tag{3}$$

Based on these differences, we also defined a binary variable, Δyield_t^b, equal to one in case of a yield increase ($\Delta \text{yield}_t > 0$) and zero otherwise

$$\Delta \text{yield}_t^b = \begin{cases} 1 & \text{if } \Delta \text{yield}_t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

3) Corn Mask: To focus on the corn-producing region within the U.S. corn belt, we employed the USDA National Agricultural Statistics Service's National Cropland Frequency Layer (2022) to construct a corn mask. This layer, derived from land cover data spanning 2008 to 2022 [30], pinpoints crop-specific planting frequency and is presented in a raster format with a 30-m resolution. We initially filtered out non-corn pixels from the raster layer to isolate the corn-growing areas. Subsequently, we excluded pixels where corn was not cultivated for at least five consecutive years during 2008–2022. Finally, we aggregated the remaining pixels at a spatial resolution of 0.05°.

Obtaining similar crop cover data can be challenging in regions such as Malawi and South Africa due to resource, infrastructure, and technology limitations. Becker-Reshef et al. [4] partly addressed this global data gap by providing crop-type maps that illustrate the spatial distribution of major crops at a 0.05° resolution. Each pixel in these rasters is associated with a percentage representing the area covered by a particular crop type. Building upon this data, we developed a corn mask by applying a 10% threshold, which means we excluded pixels where corn contributes less than 10% of the area.

However, applying a corn mask introduces no-data pixels, disrupting the neural network's convolution operation by altering the image's spatial structure and excluding valuable information beyond the crop's extent. To determine the actual interest of using crop masks, we also trained and tested our models without crop masks.

4) Remote Sensing Products: The development of advanced sensors and platforms, facilitated by the availability of extensive global datasets, has dramatically improved our understanding of agricultural practices and their environmental impact. Recent research leveraging satellite-based GPP data has been instrumental in monitoring and assessing crop yield and pricing (e.g., [11], [27]). Satellite-derived GPP provides valuable information on vegetation dynamics, demonstrating how agricultural practices and climatic conditions influence crop growth and development.

Among the various GPP estimation datasets available, the Global Land Surface Satellite (GLASS) program [19] is well-suited for tracking changes in land vegetation cover over time. The program offers high-resolution (0.05°) images every eight days from 1982 to 2018. While other well-established
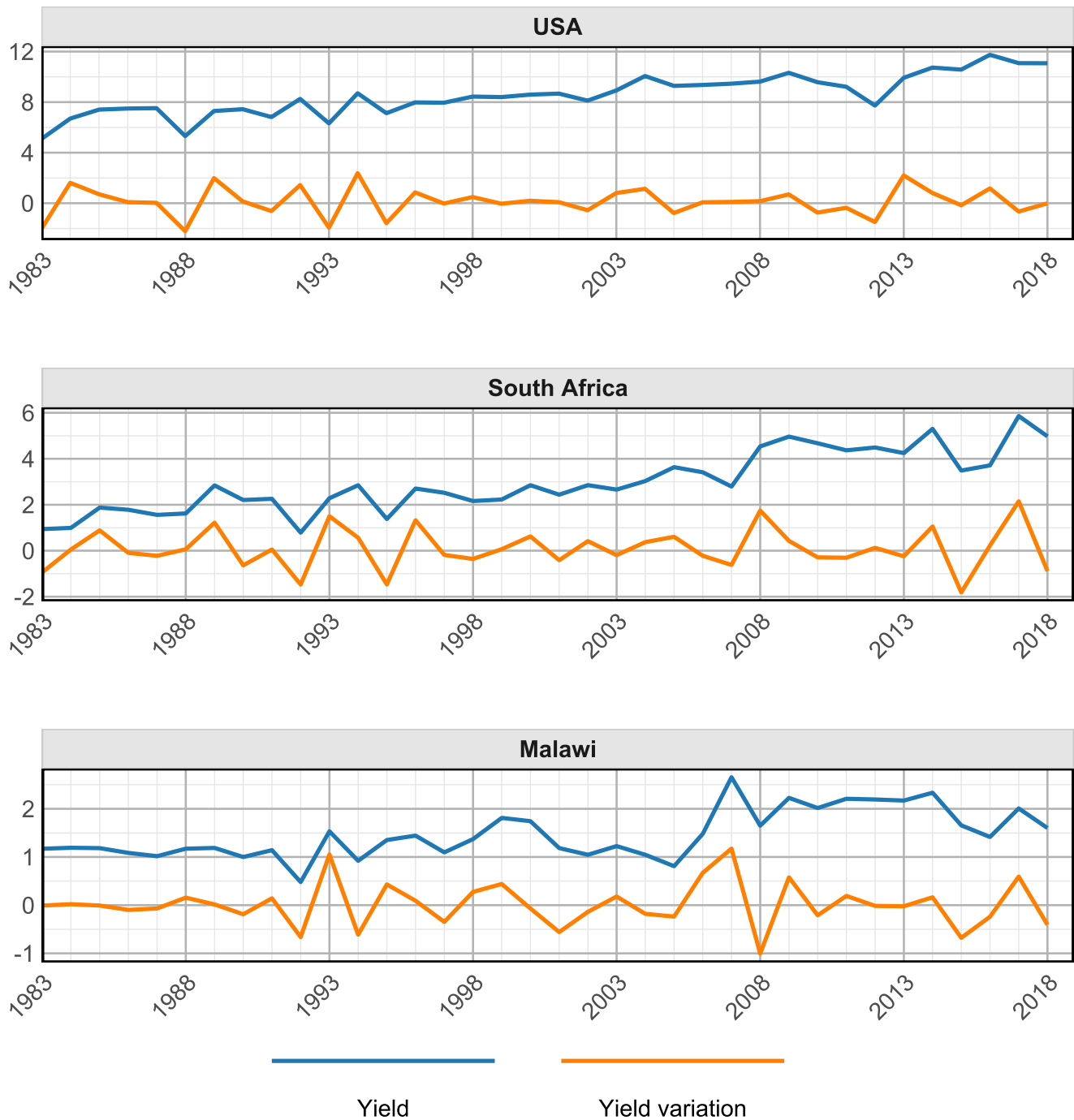
Fig. 1. Time series data of corn yield (t/ha) from 1983 to 2018 for the U.S., South Africa, and Malawi. The blue line represents the corn yield, while the orange line illustrates the interannual variation in corn yield.

satellite-based GPP products, such as MODIS and BESS [13], offer a finer spatial resolution (1 km), and FLUXCOM (0.5°) [15] provides superior temporal resolution with daily data, the GLASS GPP dataset offers longer time series. The extensive historical records from 1982 to 2018 provided by GLASS make this product particularly relevant to our analysis, guaranteeing correct model calibration.

For our analysis, we compiled the GLASS GPP products on a monthly basis. We applied a specific crop mask (see Section II-A3) to the U.S. corn belt, South Africa, and Malawi. The GPP product is denoted as GPP_{dsmt}, where d is the day, m is the month, s represents the pixel, and t is the year. The number of pixels s ranges from $\approx 122\,000$ (U.S.) to $\approx 10\,000$ (Malawi). In order to challenge our forecasting approach, we used the GPP products available three months before harvest, that is, July for the U.S. and March for Malawi and South Africa.

The high-dimensional nature of remote sensing data raises specific issues, as evidenced by the growing number of publications on the subject [18]. In Section II-B, we discuss several strategies to reduce the spatial dimension of the GPP maps derived from GLASS data.
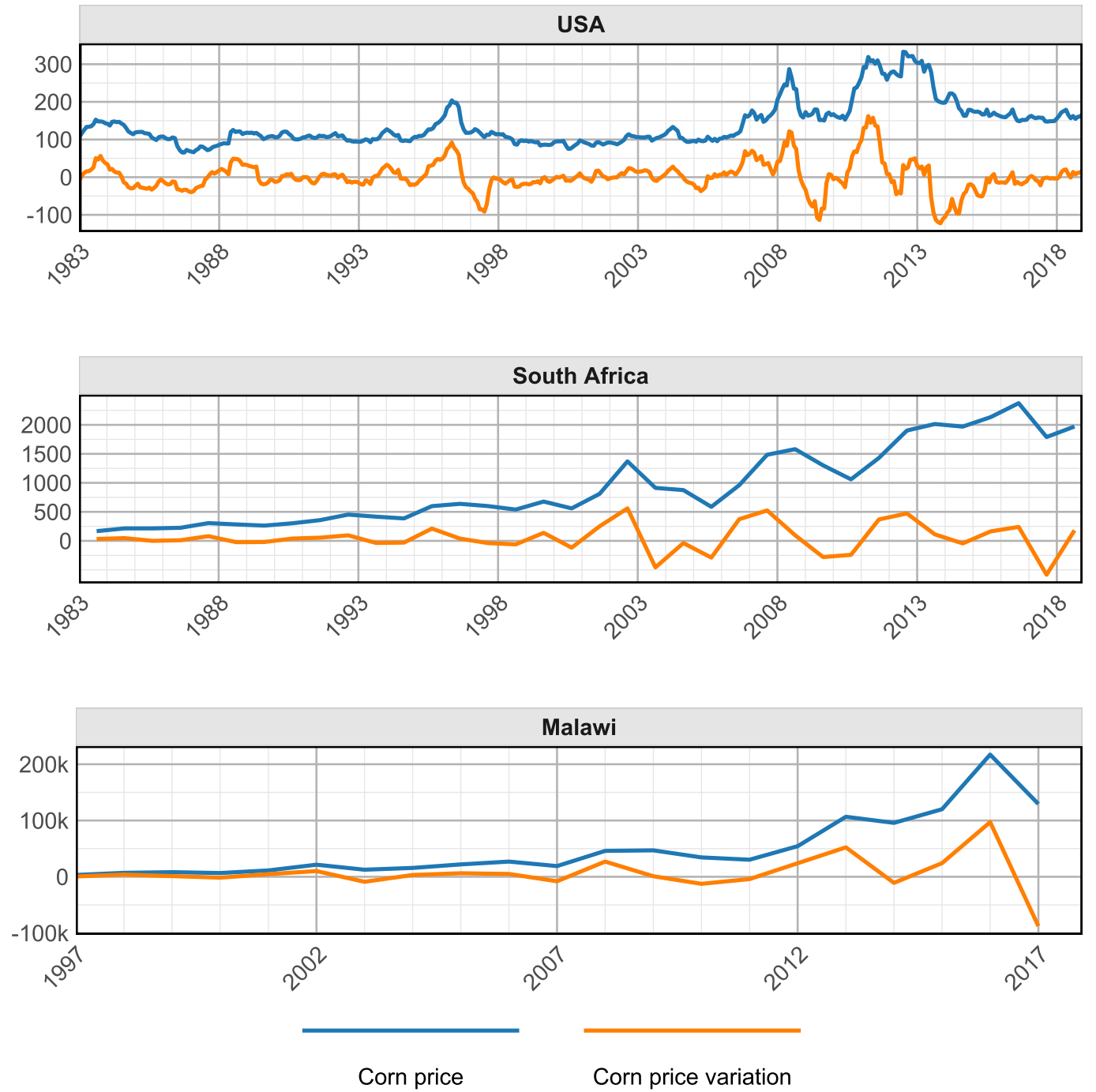
Fig. 2.   Time series data includes monthly corn prices from 1983 to 2018 for the U.S. (No. 2 yellow, FOB U.S. Gulf ports; USD/t), annual producer corn prices for South Africa (ZAR/t), and annual producer prices for Malawi from 1997 to 2017 (MWK/t). The blue line represents the corn price, while the orange line illustrates the interannual variation in corn prices.

B. Dimension Reduction

Dimension reduction involves converting data observations into a meaningful representation with a lower dimension.

In our study, we compared the capabilities of five distinct approaches. The first four produce one or a few synthetic features and then include them as inputs in classification models. These methods are the spatial mean, empirical orthogonal function (EOF), autoencoder (AE), and β-VAE. Additionally, we employed a convolutional neural network combining dimension reduction and classification simultaneously.

1) Spatial Average: The initial approach to dimension reduction of our image is straightforward. It simply relies on the spatial averaging of the annual eight-day GPP difference (Figs. 12 and 13), denoted as ΔGPP_{dsmt} and defined as

$$\Delta\text{GPP}_{dsmt} = \text{GPP}_{dsmt} - \text{GPP}_{dsm(t-1)}. \tag{5}$$

Here, d represents GPP for a specific eight-day period, m the month, s the pixel, and t the year. The monthly average of the yearly difference is computed as

$$\Delta\text{GPP}_{smt} = \frac{1}{D}\left(\sum_{d=1}^{D} \Delta\text{GPP}_{dsmt}\right). \tag{6}$$

Note that, as mentioned above, the month m is selected to be three months before the main harvest month in each country.

Finally, the monthly spatial average of the yearly difference is computed as

$$\Delta \overline{\text{GPP}}_{mt} = \frac{1}{S}\left(\sum_{s=1}^{S} \Delta \text{GPP}_{smt}\right). \quad (7)$$

2) Empirical Orthogonal Function: The EOF analysis is a dimensionality reduction technique based on PCA commonly employed in various scientific disciplines, particularly in environmental sciences. It serves as a valuable tool for identifying and characterizing intricate patterns of variability within a complex dataset [10]. The fundamental principle of EOF analysis lies in decomposing a high-dimensional dataset into a set of orthogonal functions known as EOFs. These EOFs are constructed to capture the maximum variance within the original dataset, effectively reducing the dimensionality while preserving the essential information. The construction of EOFs involves the computation of eigenvectors and eigenvalues. Each eigenvector represents an EOF, and its corresponding eigenvalue quantifies the proportion of variance it explains. The EOFs are ranked in descending order of their eigenvalues, with the first EOF capturing the largest portion of the variance, followed by the next EOF, and so on.

The EOF decomposition of a spatiotemporal matrix $\mathbf{X}_m$, representing the monthly ΔGPP_{smt}, with the temporal dimension (t) as rows and the number of pixels (s) as columns, is executed utilizing the singular value decomposition (SVD) method [10]

$$\mathbf{X}_m = \mathbf{A}_m \Lambda_m \mathbf{U}_m^{\mathrm{T}}, \quad x_{mts} = \sum_{k=1}^{E_m} a_{kmt}\lambda_{km}u_{kms}. \quad (8)$$

This matrix includes as many rows as years of observations (36 here) and as many columns as pixels (from $\approx 10\,000$ to $\approx 122\,000$, depending on the country).

The SVD method breaks down the matrix $\mathbf{X}_m$ for each month m into three distinct matrices: an orthogonal matrix $\mathbf{A}_m$ (dimensions $T_m \times T_m$, where T_m is the total number of years for month m), encompassing the principal components (PCs); a second orthogonal matrix $\mathbf{U}_m^{\mathrm{T}}$ (dimensions $S \times S$, where S represents the total number of pixels), encapsulating the EOFs or spatial patterns; and a diagonal matrix Λ_m containing the singular values for month m. The orthogonal matrices $\mathbf{A}_m$ and $\mathbf{U}_m^{\mathrm{T}}$ independently capture the temporal and spatial variations present in the original datasets for each month. The primary EOF for each month m is the mode—that is, a linear combination of S pixels—that accounts for the largest proportion of variance for that month. Each subsequent EOF for a given month m captures a decreasing amount of the total variance. The maximum number of EOFs for each month, denoted as E_m, is the smaller dimension—either the total number of years represented by rows (T_m) or the total number of spatial pixels represented by columns (S), that is, $E_m = \min(T_m, S)$. For a more detailed understanding of the principles and methodologies of EOF/PCA, the reader is directed to consult Jolliffe [14] and Wilks [32]. For illustration, Fig. 3 presents the EOF-SVD analysis results obtained in
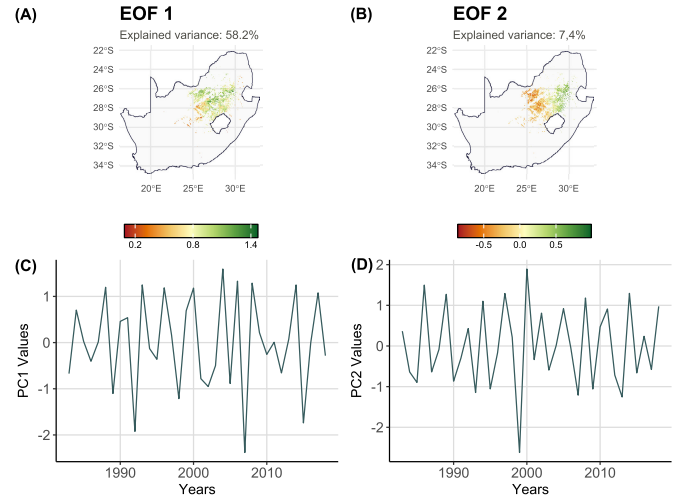

Fig. 3. Panel (A) presents the first EOF obtained from the EOF-SVD analysis of the March GPP (year-by-year) variation across South Africa, with a corn mask applied. This EOF accounts for a substantial 58.2% of the total variance in the dataset. The corresponding first PC, showing the temporal dynamics, is depicted in Panel (C). Similarly, Panels (B) and (D) display the second EOF and its associated PC, respectively. This second EOF explains 7.4% of the variability.

South Africa. In particular, Panels (A) and (B) depict the first and second EOFs, respectively, which account for 58.2% and 7.4% of the total variance of the GPP variation in March. Panels (C) and (D), on the other hand, showcase the corresponding PCs for each EOF. These panels illustrate the temporal variations of the vegetation index, specifically the annual gross primary productivity (GPP) variation in March.

3) Autoencoder: An AE (Hinton et al., [12]) is a type of unsupervised model that efficiently embeds data within a space of reduced dimension. The structure of an AE is composed of two elements: an encoder function, $g_\varphi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$, parameterized by φ, responsible for compressing input data into a lower-dimensional latent space ($d_z < d_x$), and a decoder function, $f_\theta : \mathbb{R}^{d_z} \to \mathbb{R}^{d_x}$, parameterized by θ, which uses the encoded information to reconstruct the original inputs. The low-dimensional latent space, learned for the input x in the bottleneck layer, is represented as $\mathbf{z} = g_\varphi(x)$. On the other hand, the decoder is responsible for reconstructing the original input data $\mathbf{x}' = f_\theta(g_\varphi(x))$. The optimization process on (θ, φ) aims at decoding the original input from the reduced representation provided by the encoder; in other words, it aims at learning the identity function under the compression constraint. The differences between the original vector x and the reconstructed vector $\mathbf{x}'$ are usually measured using the root mean square error (RMSE), and the values of θ and φ are determined to minimize the RMSE utilizing an optimization algorithm. Here, x represents a map of ΔGPP_{dsmt}, that is, the spatial distribution of GPP time variation in a given country (U.S., South Africa, or Malawi). The original dimension of x is thus of several $10\,000$ pixels. AE is used to reduce the dimension to no more than a few dozen latent features ($\mathbf{z}$). See Section II-E for more details.

The AE architectures for each country can be found in Tables I–III.

TABLE I
β-VAE AND AE ARCHITECTURE, U.S. H. REPRESENTS THE SIZE OF THE LATENT SPACE

Module	Layer
Encoder	Input
	Conv2D
	ReLU
	BatchNorm
	Conv2D
	ReLU
	BatchNorm
	Flatten
	Dense (units: H)
Decoder	Input (H)
	Dense
	ReLU
	Conv2DTranspose
	ReLU
	BatchNorm
	Conv2DTranspose
	Sigmoid
	Output

TABLE III
β-VAE AND AE ARCHITECTURE, MALAWI. H. REPRESENTS THE SIZE OF THE LATENT SPACE

Module	Layer
Encoder	Input
	Conv2D
	ReLU
	BatchNorm
	Conv2D
	ReLU
	BatchNorm
	Flatten
	Dense (units: H)
Decoder	Input (H)
	Dense
	ReLU
	Conv2DTranspose
	ReLU
	BatchNorm
	Conv2DTranspose
	Sigmoid
	Output

TABLE II
β-VAE AND AE ARCHITECTURE, SOUTH AFRICA. H. REPRESENTS THE SIZE OF THE LATENT SPACE

Module	Layer
Encoder	Input
	Conv2D
	ReLU
	BatchNorm
	Conv2D
	ReLU
	BatchNorm
	Flatten
	Dense (units: H)
Decoder	Input (H)
	Dense
	ReLU
	Conv2DTranspose
	ReLU
	BatchNorm
	Conv2DTranspose
	Sigmoid
	Output

4) Variational Autoencoder: The VAE, introduced by Kingma and Welling [16], extends the traditional AE by incorporating stochastic sampling and regularization, thus allowing for more expressive latent space representations. Unlike a conventional AE, which yields deterministic encodings, the encoder in a VAE maps each input to a set of parameters defining a probability distribution over the latent space, resulting in a more structured and continuous representation. Moreover, the VAE can reconstruct input data and generate new data points by sampling from the learned latent space representations.

Unlike the AE, the VAE is a generative model designed to learn compact representations of high-dimensional data in a lower-dimensional latent space. It comprises two primary components: a probabilistic encoder and a probabilistic decoder. Given that x is the set of input variables and z a set of latent variables of smaller dimension, the probabilistic encoder, typically implemented as a neural network, provides an approximation $q(z \mid x; \phi)$ for the true posterior $p(z \mid x; \theta)$. Meanwhile, the probabilistic decoder provides $p(x \mid z; \theta)$, the likelihood function parameterized by θ.

Mathematically, the VAE defines a joint distribution over input variables x and corresponding latent variables z as $p(x, z; \theta) = p(x|z; \theta)p(z)$, where $p(x|z; \theta)$ is the likelihood function parameterized by θ, and $p(z)$ denotes the prior distribution, typically assumed to be a standard normal distribution $\mathcal{N}(0, I)$. A reparameterization trick is employed to ensure the differentiability of the training objective with respect to both the encoder and decoder parameters, θ and ϕ. This allows for the use of standard gradient descent techniques during the training process. It involves adding a random noise term, typically drawn from a standard normal distribution. This leads to the formulation of the evidence lower bound (ELBO) given by

$$\mathcal{E}(\theta, \phi; x) = -\mathrm{KL}(q(z \mid x; \phi) \| p(z)) \\ + \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I)}\left[\log(p(x \mid z; \theta))\right]. \quad (9)$$

The approximate posterior distribution $q(z \mid x; \phi)$ is typically parameterized as $\mathcal{N}(\mu(x; \phi), \sigma^2(x; \phi))$, where μ and σ are outputs of the encoder neural network. The reparameterization trick involves $z = \mu(x; \phi) + \epsilon\sigma(x; \phi)$, with ϵ sampled from a standard normal distribution. This enables the model to backpropagate gradients through the sampling process, facilitating end-to-end training of the VAE. During training, the objective

is to minimize the negative of the ELBO, where the second term corresponds to the negative log-likelihood, serving as a reconstruction loss, and the first term, the Kullback–Leibler (KL) divergence, acts as a regularization term, promoting the alignment of the approximate posterior distribution with the prior distribution. The model learns meaningful latent representations by maximizing the ELBO while generating data points that closely match the observed data distribution.

Higgins et al. [12] introduced an extension to the VAE model known as β-VAE to enhance the disentanglement of latent representations. The key innovation lies in adjusting the weighting of the KL divergence term in the VAE objective function.

As defined by Higgins et al. [12], disentangled representations exhibit sensitivity to individual generative factors while remaining invariant to others. The VAE's default prior, often a Gaussian with a diagonal covariance matrix, inherently leads to independent dimensions resembling disentangled features. The introduction of the parameter β in β-VAE serves as a crucial modulator for learning constraints, influencing the capacity of the latent information channel. With a β value greater than 1, the model enforces a closer match between the approximate posterior distribution $q(z \mid x; \phi)$ and the prior distribution $p(z)$. The model is pushed toward a more efficient and disentangled latent representation, especially effective when the data contains independent underlying factors of variation. The original VAE framework corresponds to $\beta = 1$.

In a β-VAE, the loss becomes as follows:

$$\mathcal{E}^{\beta}(\theta, \phi; x) = -\beta \mathrm{KL}(q(z \mid x; \phi) \| p(z))$$
$$+ \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I)}\big[\log(p(x \mid z; \theta))\big]. \quad (10)$$

Similar to the AE, x represents a map of $\Delta \mathrm{GPP}_{dsmt}$, in a given country (U.S., South Africa, or Malawi). The β-VAE architectures for each country can be found in Tables I–III.

C. Classification Based on Generalized Linear Models

Generalized linear models (GLMs) based on logit regression were implemented to predict the sign annual variations in corn yield and price (i.e., increases or decreases relative to the previous year) using GPP-related predictors derived from the dimension-reduction techniques mentioned earlier (i.e., spatial average of interannual GPP variations, EOFs, and latent variables derived from AE or VAE).

A simple GLM was implemented to forecast yield and price changes (increase versus decrease in year t compared to year t-1) from spatial averages of interannual GPP variations, using a logit model whose parameters were estimated by maximum likelihood. As this model included a single input, it did not require any input selection step.

Additional GLMs were developed to predict yield and price changes from EOFs or latent variables derived from AE or VAE. These models included as many predictors as the number of EOFs or latent variables available. In order to optimize the number of predictors, the parameters of these GLM were estimated using an ElasticNet algorithm. ElasticNet is a regularization technique used to select the most relevant

input variables and thus reduce model complexity. It combines the attributes of least absolute shrinkage and selection operator (LASSO) and ridge regression [29], [38]. ElasticNet accomplishes this by incorporating a penalty term into the regression model, encompassing both L1 (LASSO) and L2 (ridge) penalties. The parameter α in ElasticNet governs the balance between the L1 and L2 penalties, ranging from 0 to 1. At $\alpha = 1$, ElasticNet behaves similar to LASSO, employing the L1 penalty. Conversely, it behaves similar to ridge regression at $\alpha = 0$, applying the L2 penalty. Intermediate values between 0 and 1 offer a blend of L1 and L2 penalties, enabling ElasticNet to handle correlated predictors, thereby ensuring model stability and effective variable selection. Although ElasticNet is based on a linear model and is a relatively old method, it seems well-suited here, given the limited number of years available. Moreover, ElasticNet was already used to forecast crop yields in previous studies [2], [25] and thus seems to be a natural choice. Here, we tested five values of α in the set 0, 0.25, 0.5, 0.75, 1. Separate models were developed for each country, using GPP data collected in July for the USA and in March for South Africa and Malawi.

The GLM computing the probabilities of price increase from the EOF-PCs were defined as follows:

$$\Pr\big(\Delta p_t^b = 1 \mid a_{1mt}, \ldots, a_{Emt}\big) = \frac{e^{\beta_0 + \sum_{k=1}^{E} \beta_k a_{kmt}}}{1 + e^{\beta_0 + \sum_{k=1}^{E} \beta_k a_{kmt}}}. \quad (11)$$

In this equation

1) $\Pr\big(\Delta p_t^b = 1 \mid a_{1mt}, \ldots, a_{Emt}\big)$ represents the probability that the price increase indicator Δp_t^b equals 1 (indicating a price increase) given the EOF-PCs $a_{1mt}, \ldots, a_{Emt}$.
2) β_0 is the intercept term in the regression model.
3) β_k indicates the coefficients associated with the EOF-PC a_{kmt}.
4) a_{kmt} indicates the EOF-PCs for $k = 1, \ldots, E$, where E is the total number of components, m is the month index, and t denotes the time index.

Similarly, for computing the probabilities of yield increase, we used the following model:

$$\Pr\big(\Delta \mathrm{yield}_t^b = 1 \mid a_{1mt}, \ldots, a_{Emt}\big) = \frac{e^{\beta_0 + \sum_{k=1}^{E} \beta_k a_{kmt}}}{1 + e^{\beta_0 + \sum_{k=1}^{E} \beta_k a_{kmt}}}. \quad (12)$$

For the VAE and the AE, if we denote the latent vectors as z_{it}, where i represents the ith latent feature and t denotes the year index, the GLM can be defined similarly

$$\Pr\big(\Delta p_t^b = 1 \mid z_{1t}, \ldots, z_{Ht}\big) = \frac{e^{\beta_0 + \sum_{i=1}^{H} \beta_i z_{it}}}{1 + e^{\beta_0 + \sum_{i=1}^{H} \beta_i z_{it}}} \quad (13)$$

$$\Pr\big(\Delta \mathrm{yield}_t^b = 1 \mid z_{1t}, \ldots, z_{Ht}\big) = \frac{e^{\beta_0 + \sum_{i=1}^{H} \beta_i z_{it}}}{1 + e^{\beta_0 + \sum_{i=1}^{H} \beta_i z_{it}}}. \quad (14)$$

H represents the dimensionality of the latent space (H was set successively to 1, 2, ..., 36, where 36 is the number of years under consideration in the analysis, reflecting as well the number of PCs generated by the EOF decomposition). The terms β_i are the coefficients to be estimated.

The GLM computing price and yield increase probabilities from the spatial average of the yearly difference of vegetation

TABLE IV
2D-CNN ARCHITECTURE, U.S. H. REPRESENTS
THE SIZE OF THE LATENT SPACE

Module	Layer
Encoder	Input
	Conv2D
	ReLU
	BatchNorm
	Conv2D
	ReLU
	BatchNorm
	Flatten
	Dense (units: H)
Classification	Input (H)
	Dense
	ReLU
	BatchNorm
	Dense
	Sigmoid

TABLE V
2D-CNN ARCHITECTURE, SOUTH AFRICA. H. REPRESENTS
THE SIZE OF THE LATENT SPACE

Module	Layer
Encoder	Input
	Conv2D
	ReLU
	BatchNorm
	Conv2D
	ReLU
	BatchNorm
	Flatten
	Dense (units: H)
Classification	Input (H)
	Dense
	ReLU
	BatchNorm
	Dense
	ReLU
	BatchNorm
	Dense
	Sigmoid

TABLE VI
2D-CNN ARCHITECTURE, MALAWI. H. REPRESENTS
THE SIZE OF THE LATENT SPACE

Module	Layer
Encoder	Input
	Conv2D
	ReLU
	BatchNorm
	Conv2D
	ReLU
	BatchNorm
	Flatten
	Dense (units: H)
Classification	Input (H)
	Dense
	ReLU
	BatchNorm
	Dense
	ReLU
	BatchNorm
	Dense
	Sigmoid

indices ($\overline{\Delta\text{GPP}_{mt}}$) were defined as follows:

$$\Pr\big(\Delta p_t^b = 1 \mid \overline{\Delta\text{GPP}_{mt}}\big) = \frac{e^{\beta_0 + \beta_1 \overline{\Delta\text{GPP}_{mt}}}}{1 + e^{\beta_0 + \beta_1 \overline{\Delta\text{GPP}_{mt}}}} \quad (15)$$

$$\Pr\big(\Delta\text{yield}_t^b = 1 \mid \overline{\Delta\text{GPP}_{mt}}\big) = \frac{e^{\beta_0 + \beta_1 \overline{\Delta\text{GPP}_{mt}}}}{1 + e^{\beta_0 + \beta_1 \overline{\Delta\text{GPP}_{mt}}}}. \quad (16)$$

D. Convolutional Neural Network Classification Model

We have designed a 2-D convolutional neural network (2D-CNN) (see Tables IV–VI) for yield and price classification, leveraging ΔGPP_{dsmt} variables as inputs without any dimension-reduction pre-processing. We used this approach to directly predict yield and price variations from the pixel-by-pixel GPP input data. The number of predictors of 2D-CNN was, thus, equal to the number of pixels considered in each country. The network included linear layers and a sigmoid activation function at the last layer to predict yield and price variations. Model training was performed using binary cross-entropy (BCE), a popular loss function commonly used in binary classification tasks. It assesses the disparity between the predicted probability distribution and the actual class labels. It measures the error or deviation between the model's predictions and the ground truth, penalizing the model for significant deviations. A BCE close to 0 indicates that the model's predictions closely match the true class probabilities. The formulation of BCE is as follows:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{T} \sum_{t=1}^{\text{T}} \big(y_t \log\big(\hat{p}_t\big) + (1 - y_t) \log\big(1 - \hat{p}_t\big)\big). \quad (17)$$

In this equation, $y_t \in \{0, 1\}$ represents the binary class label for the tth prediction in year t. The value 1 corresponds to the positive class (indicating an increase in price or yield), and 0 corresponds to the negative class (indicating a decrease in price or yield). The term $\hat{p}_t \in [0, 1]$ denotes the predicted probability that the tth prediction belongs to the positive class, while $(1 - \hat{p}_t) \in [0, 1]$ represents the probability that the tth prediction belongs to the negative class. The symbol T represents the total number of predictions.

E. Model Evaluation

For models such as the AE, the β-VAE, and the 2D-CNN, it is crucial to determine the optimal size of the latent space, further denoted as H. We defined a search space where $z \in \mathbb{R}^H$, $H \in [1, 36]$. The maximum dimension of the latent space was set to 36 because we considered 36 years of yield and price data (from 1983 to 2018). Moreover,
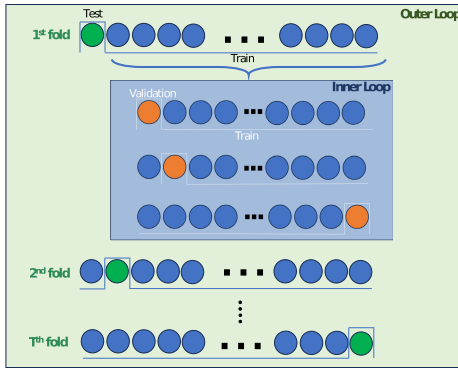
Fig. 4. This figure depicts the nested LOYO cross-validation procedure. This two-loop process features an inner loop nested within an outer loop. The inner loop serves to optimize the model's hyperparameters, while the outer loop evaluates the performance of the model's predictions using the optimized hyperparameters.

in the β-VAE model, selecting an appropriate β is crucial for the model's performance. We tested several values for this hyperparameter, $\beta \in \{0.001, 0.01, 0.1, 1, 2, 5, 10\}$.

To find the best values for H and β, we employed a double nested leave-one-year-out (LOYO) cross-validation strategy (see Fig. 4). This approach involved two levels of data partitioning:

1) *Outer Loop:* We divided the data covering T years into two sets: a test set for one year (t) and a training set containing the remaining $T - 1$ years.
2) *Inner Loop:* We performed another LOYO split within the training set. We then trained models with each combination of hyperparameters (36×7) on the inner training set and evaluated their performance on the inner validation set.

The combination of H and β that yielded the best performance on the inner loop was then used to predict yield and price variations for the outer loop test set. This approach ensured robust hyperparameter selection by minimizing overfitting and providing a more reliable estimate of generalizability.

The area under the curve (AUC) was used to assess the model's ability to distinguish between two classes (here, yield/price increase versus yield/price decrease). An AUC value of one signifies a perfect classification, while the value of 0.50 implies that the model's performance is no better than random guessing.

The brier skill score (BSS) was used to evaluate the models' performance compared to a baseline model (a constant GLM). It is computed using the following formula:

$$BSS = 1 - \frac{BS}{BS_{ref}} \qquad (18)$$

where

1) BS   brier score of the new model;
2) BS_{ref} brier score of a reference model;
3) In our case, this reference model is a GLM that predicts a constant probability of yield/price increase based on the entire dataset.

The brier score (BS) measures the accuracy of probabilistic predictions and is defined as the mean squared error applied to predicted probabilities

$$BS = \frac{1}{N} \sum_{t=1}^{N} (f_t - o_t)^2 \qquad (19)$$

where

1) N total number of forecasting instances; f_t forecast probability of yield/price increase, for instance t;

2) o_t actual outcome of the event at instance t (1 if the event occurs, 0 otherwise).

The BSS is a metric that evaluates the performance of a forecasting model in comparison to a reference or baseline model. It ranges from $-\infty$ to 1, where a BSS of 1 signifies perfect skill relative to the reference forecast. A BSS of 0 indicates no improvement over the reference model, while a negative BSS suggests that the new model performs worse than the reference model.

Finally, the Matthews correlation coefficient (MCC) was used to assess the performance of binary classifiers, especially in the context of imbalanced datasets. It incorporates information about true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It is calculated as the Pearson product-moment correlation coefficient between actual and predicted binary classifications. The MCC ranges from -1 to 1, with a value of -1 indicating perfect misclassification, 1 signifying perfect classification, and 0 implying that the classifier's performance is equivalent to random guessing

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}. \qquad (20)$$

TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

F. Implementation

We built the AE, β-VAE, and 2D-CNN using the PyTorch package (version 2.1.0) and trained them on a workstation equipped with a CUDA-enabled NVIDIA GPU (NVIDIA GeForce RTX 4070 Laptop). Additionally, the workstation featured an AMD Ryzen 9 7845HX processor and 32 GB of DDR5 RAM operating at a clock speed of 5200 MHz.

The models based on spatial averages were fitted using the GLM function in R [23]. The most relevant PCs and latent vectors were selected using the ElasticNet method implemented in the GLMnet package in R [26].

The number of available yield data points is inherently limited due to the annual collection of a single yield measurement, reflecting the single corn crop cycle each year.

III. RESULTS AND DISCUSSION

A. Yield

Fig. 5 presents the model evaluation results for the U.S., South Africa, and Malawi, assessed using three metrics: AUC,
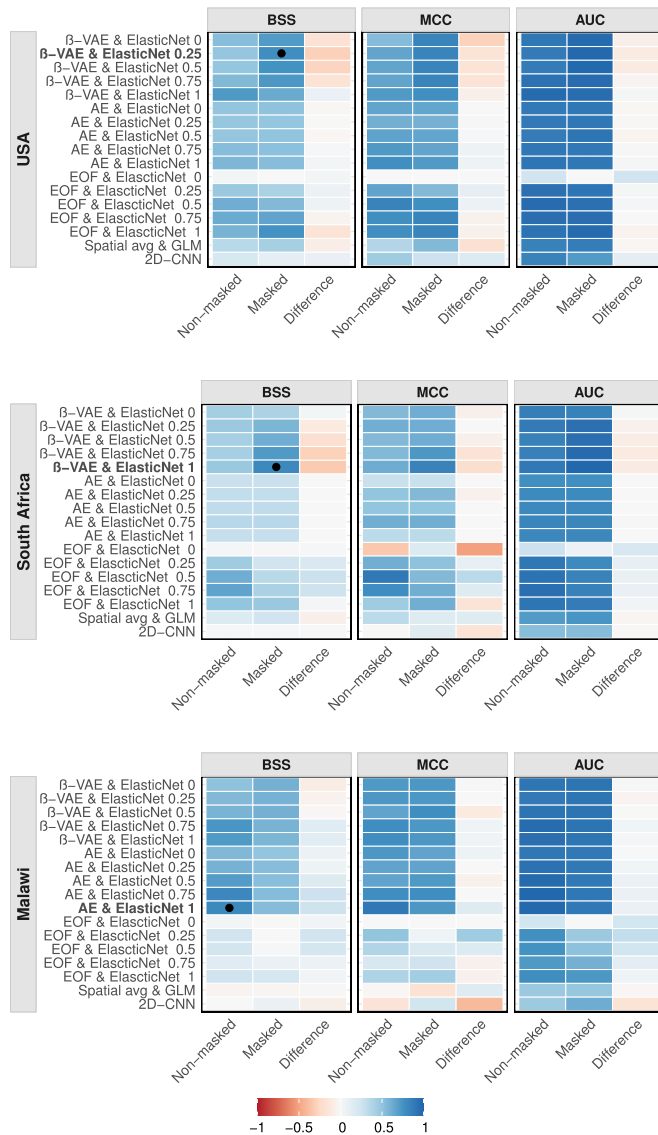
Fig. 5. Model performances when predicting changes in corn yield (increase or decrease compared to the previous year) across the U.S., South Africa, and Malawi. Each row corresponds to a specific forecasting method, combining a dimension-reduction technique and a prediction model. The columns show different evaluation metrics: AUC, MCC, and BSS. Scores are color-coded, with red indicating lower scores and blue indicating higher scores. The numerical values on the y-axis for ElasticNet represent the parameter α used in the penalty term, where $\alpha = 1$ corresponds to LASSO and $\alpha = 0$ corresponds to ridge regression. Black dots mark the best model according to the BSS.
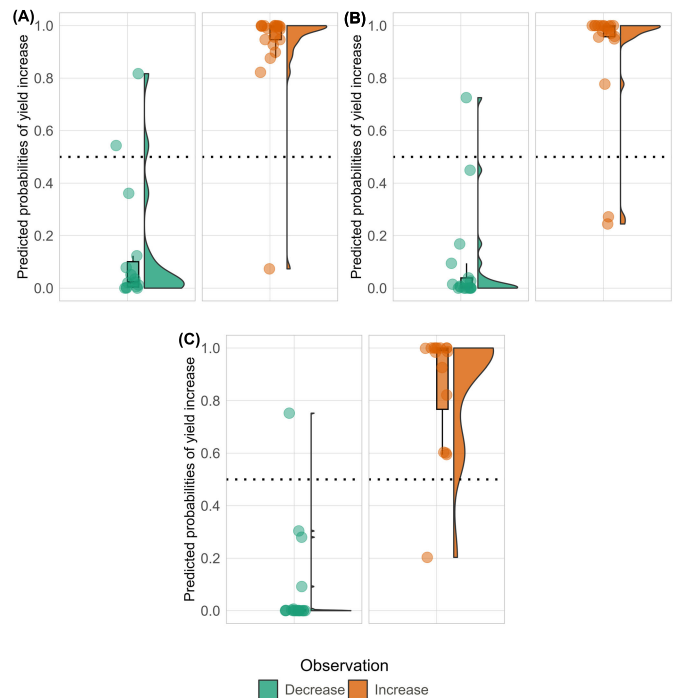

Fig. 6. Predicted probabilities of yield increase in case of actual yield increase (brown) or decrease (green); each point represents an individual instance. The Panels (A)–(C) correspond to the best models selected for the U.S., South Africa, and Malawi. The selected models are the ElasticNet with predictors derived from (A) ΔEOF, (B) AE, and (C) AE.

MCC, and BSS. The evaluation encompasses both masked and unmasked GPP data.

The approach based on the non-masked β-VAE ElasticNet ($\alpha = 1$, i.e., LASSO) achieved comparable results, although slightly less accurate, yielding a BSS of 0.72, an MCC of 0.71, and an AUC of 0.97. The approach using masked GPP and ElasticNet ($\alpha = 1$) with features derived from EOF (ΔEOF) also demonstrated commendable results, with a BSS of 0.59, an MCC of 0.77, and an AUC of 0.92. The ElasticNet performance was suboptimal, when set with $\alpha = 0$ and using EOF-derived features. When $\alpha = 0$, ElasticNet behaves similar to ridge regression. Ridge regression does not perform feature selection, that is, it does not set the coefficients to null values

but it shrinks some coefficients toward zero. If many EOF features are not relevant predictors, ridge regression is not able to remove them from the model. In contrast, while the 2D-CNN achieved relatively high AUC values (0.83 with the non-masked GPP and 0.71 with the masked GPP), it exhibited low BSS values of 0.21 and 0.12, respectively. The high AUC values suggest the model correctly classified instances as increasing or decreasing with specific classification thresholds. However, the low BSS values revealed a poor calibration of the probabilistic predictions.

The models combining β-VAE and ElasticNet demonstrated the best overall performance in South Africa. Specifically, the masked GPP β-VAE-ElasticNet achieved the highest results with an ElasticNet parameter α of 1 (LASSO), yielding a BSS of 0.8, an MCC of 0.84, and an AUC of 0.99. It is also interesting to note that the performance (based on the BSS scores) slightly increases with the value of the α parameter. The α parameter in ElasticNet controls the degree of mixing between ridge regression (when $\alpha = 0$) and LASSO (when $\alpha = 1$). Values of α between these extremes result in a blend of the two types of regression. The scores were the lowest with $\alpha = 0$ and increased gradually with each increment in α. This suggests that, among the dozens of synthesized features extracted by the β-VAE, only a few were useful predictors in the ElasticNet regression. Fig. 6(B) further illustrates the good performance of the model in South Africa. In case of an actual yield increase, the model predicted a median yield increase probability of 0.99. Conversely, in case of an actual yield decrease, the model predicted a median probability of yield increase very close to zero, equal to 0.004.
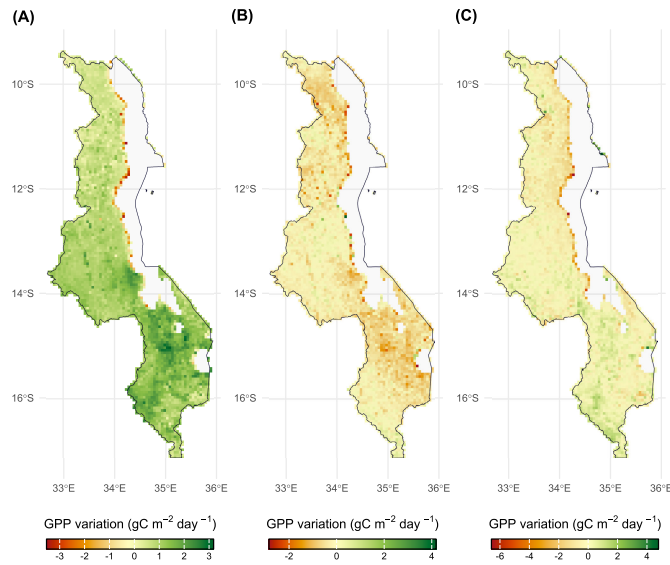
Fig. 7. Maps depicting the annual differences in March GPP for Malawi across three years. Panel (A) 2006 versus 2005. Panel (B) 2007 versus 2006. Panel (C) 1985 versus 1984. Greener colors represent positive differences, indicating an increase in GPP, while yellow to red hues indicate values closer to zero or negative differences.

The ElasticNet model based on the EOF synthesized features also delivered good results, with slightly better performance with non-masked GPP. The EOF with ElasticNet ($\alpha = 0.75$) based on the non-masked GPP achieved a BSS of 0.67, an MCC of 0.78, and an AUC of 0.94. Similar to the results in the U.S., the ElasticNet model based on the EOF features with $\alpha = 0$ underperformed compared to other EOF-based models with higher α values.

In Malawi, the most effective strategy was the ElasticNet model (with $\alpha = 0.1$) utilizing features synthesized by AE. It demonstrated exceptional performance, boasting a BSS of 0.8, indicating a significant enhancement over the baseline model. Furthermore, it attained an AUC of 0.98 and an MCC of 0.89, revealing very good classification accuracy. The β-VAE model also exhibited good results, achieving an AUC of 0.94. However, its BSS and MCC were comparatively lower, at 0.62 and 0.70, respectively.

Fig. 6(C) further confirms the good performances of AE. The model predicted a probability of a decrease much lower than 0.50 for all instances of actual yield decrease, except for one. Similarly, for cases of actual yield increase, the model predicted a probability of increase much higher than 0.50 for all instances, with only one exception. The single misclassified instance in the case of an actual yield increase corresponded to the year 2007, which was characterized by one of the most prolific corn productions in Malawi, with an estimated yield of 26 547 hg/ha (according to FAOStat), marking an increase of 79% compared to 2006.

Fig. 7(B) shows that the GPP values in 2007 were slightly lower than in 2006. Given that our approach relies solely on satellite GPP, this explains why the model did not predict the actual yield increase despite its substantial proportions. This above-average production was probably due to excellent conditions at the end of the season, not captured by the GPP map used three months before harvest [6], [34]. Similarly,
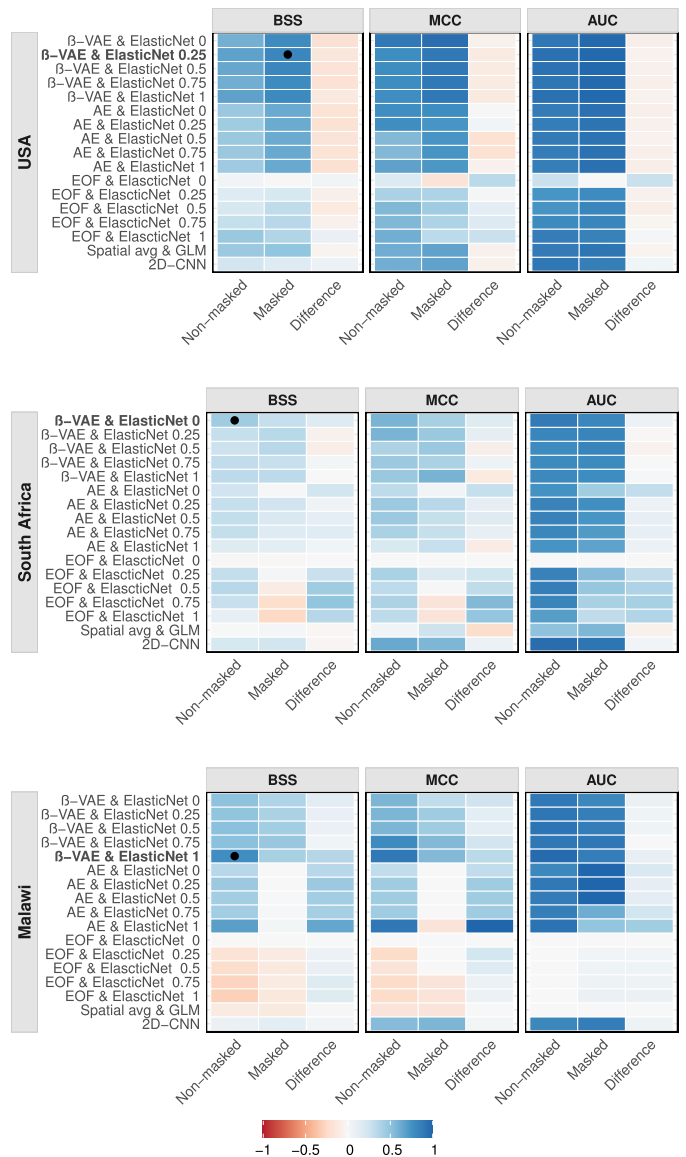

Fig. 8. Model performances when predicting changes in corn price (increase or decrease compared to the previous year) across the U.S., South Africa, and Malawi. Each row corresponds to a specific forecasting method, combining a dimension-reduction technique and a prediction model. The columns show different evaluation metrics: AUC, MCC, and BSS. Scores are color-coded, with red indicating lower scores and blue indicating higher scores. The numerical values on the y-axis for ElasticNet represent the parameter α used in the penalty term, where $\alpha = 1$ corresponds to LASSO and $\alpha = 0$ corresponds to ridge regression. Black dots mark the best model according to the BSS.

in 1985, the model predicted a yield increase despite an observed yield decrease. This discrepancy can once again be attributed to the slightly higher GPP values in 1985 compared to 1984, as shown in Fig. 7(C).

An interesting finding is that employing specific crop masks is not always beneficial. Fig. 5 shows that the BSS value was higher without crop masks for most models in the U.S., South Africa, and Malawi, and the best performance was achieved with non-masked GPP in Malawi. Despite their frequent use in remote sensing studies for yield prediction, the effectiveness of crop masks remains subject to certain limitations. Our study revealed no systematic performance improvement when employing crop masks in regions such as the U.S.,
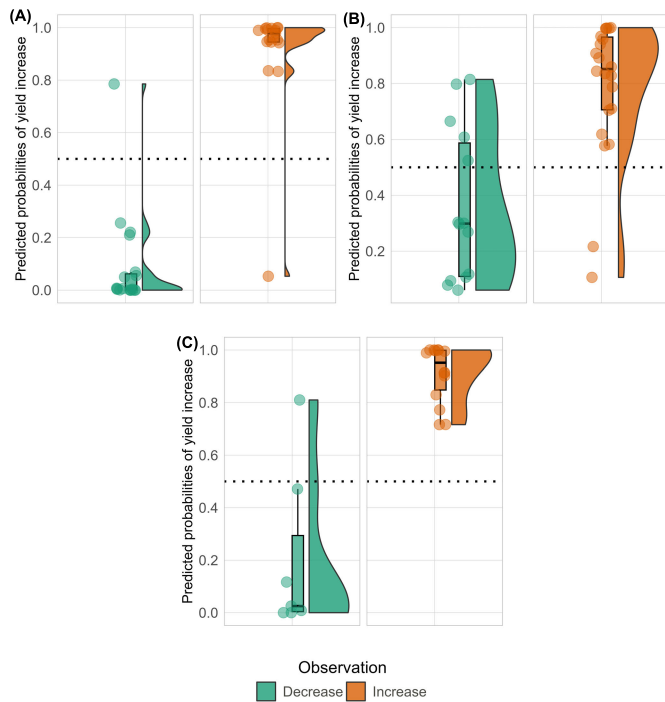
Fig. 9. Predicted probabilities of price increase in case of actual yield increase (brown) or decrease (green). The Panels (A)–(C) correspond to the best models selected for the U.S., South Africa, and Malawi, respectively. The selected models are (A) β-VAE, (B) β-VAE, and (C) β-VAE.

Fig. 11. Maps depicting the annual differences in March GPP for Malawi. Panel (A) 1999 versus 1998. Panel (B) 2000 versus 1999. The key corn-producing regions of Kasungu, Lilongwe, and Mchinji in the western part of the country are highlighted with black borders. Greener colors represent positive differences, indicating an increase in GPP, while yellow to red indicate values close to zero or negative differences.

Canada. This lack of benefit from crop masks is frequent for in-season yield predictions, given the limited availability of up-to-date annual crop maps during each specific growing season. Several factors contribute to the potential limitations of crop masks. First, their static nature and inherent inaccuracies can introduce noise. They often fail to capture variations in croplands throughout the season. The logistical challenge of regularly acquiring dynamic masks throughout the growing season could outweigh their benefits. Furthermore, irrespective of the method used for their creation, crop masks may contain errors, such as including non-target areas or excluding actual crops.

B. Price

Fig. 8 presents the evaluation results for price variations. The results indicate that the β-VAE approach consistently performed well across all three countries. Specifically, the β-VAE approach achieved the best results among all tested U.S., South Africa, and Malawi tested methods.

For instance, in the U.S., the ElasticNet model ($\alpha = 0.25$) using features extracted from masked GPP with β-VAE, along with non-masked GPP, achieved a BSS of 0.82, an MCC of 0.89, and an AUC of 0.98. The excellent classification performance is confirmed in Fig. 9(A), where the model perfectly classified the increase instances and nearly perfectly classified the decrease instances. The clustered predicted probabilities (the points in the figure) are near 0 for price decrease cases and near 1 for price increase cases.

Interestingly, the parameter α value for price prediction did not seem to severely impact the results for the AEs and VAEs using ElasticNet, as the scores tended to fall within a
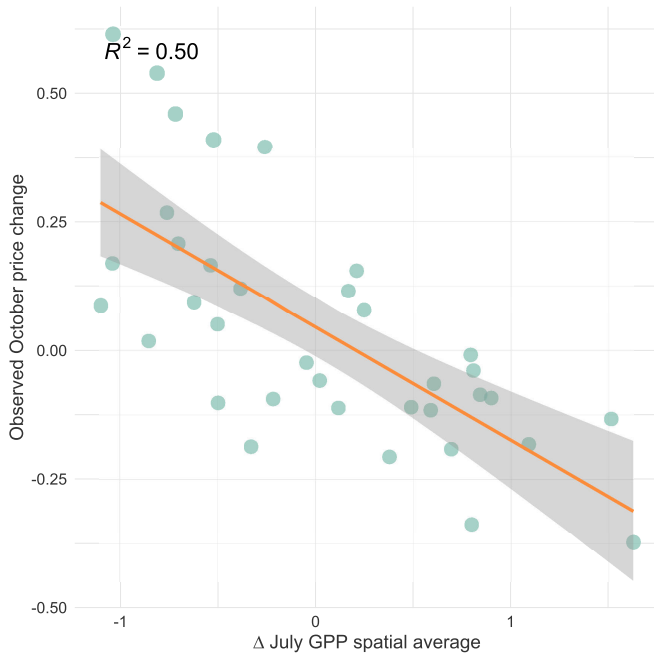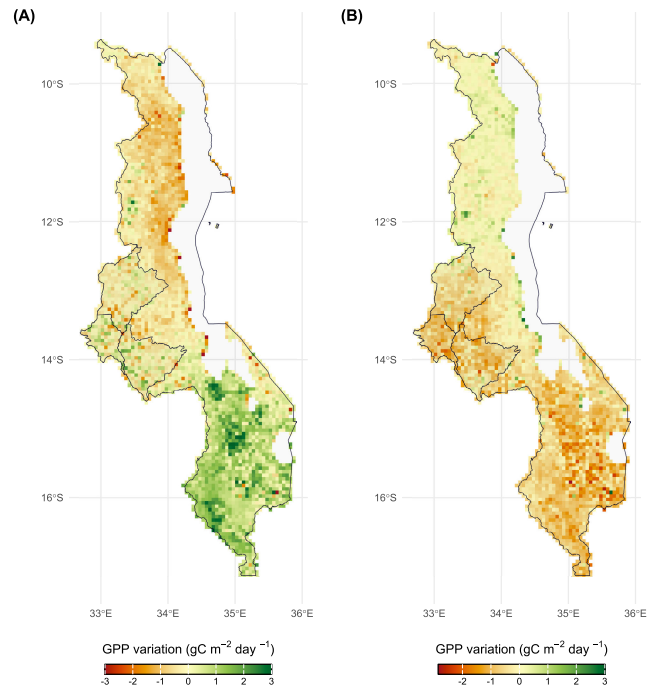

Fig. 10. Scatter plot showing the relationship between the spatial average value of July ΔGPP and the observed October price change. Green dots represent the data points, and the orange line indicates the best-fit linear regression. The shaded gray area represents the 95% confidence interval of the regression line.

South Africa, and Malawi, and we even observed a decline in performance in some cases.

Similar findings are documented in the literature; for example, when using crop masks, Zhang et al. [37] reported no enhancement in yield prediction for major corn areas in
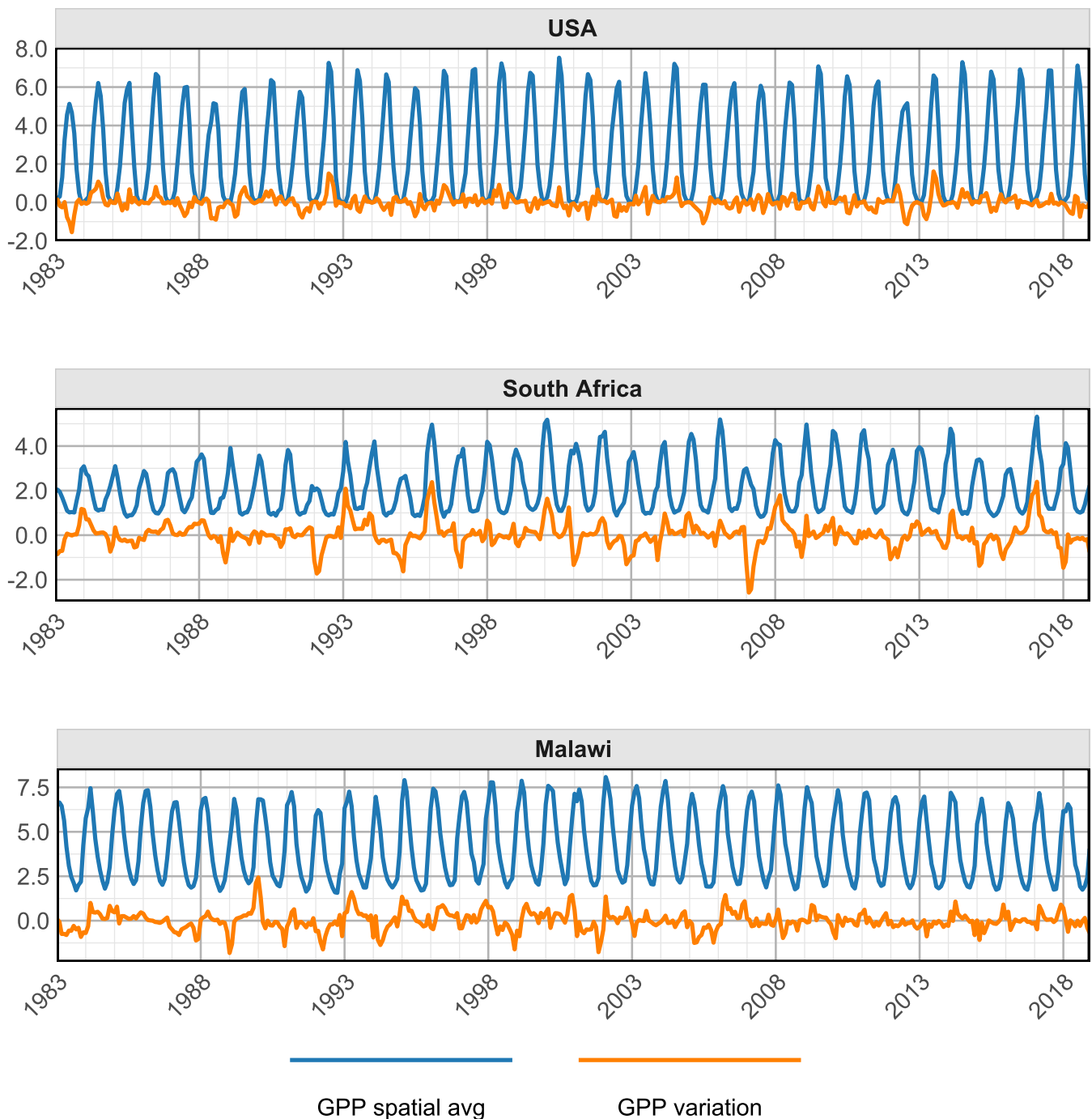
Fig. 12. Time series of the monthly spatial average and interannual variation of GPP, masked using a specific crop (corn) mask, from 1983 to 2018 for the U.S., South Africa, and Malawi. The blue line represents the spatial average of GPP, while the orange line illustrates the interannual variation in GPP.

narrow range. The BSS ranged from 0.79 with $\alpha = 0.75$ to 0.82 with $\alpha = 0.25$ for the VAEs, and from 0.63 with $\alpha = 0.25$ to 0.64 with $\alpha = 0.75$ for the AEs. Crop masks had a more substantial impact on the results. As shown in Fig. 8, the performance scores were often lower with non-masked than with masked data in the U.S. Conversely, for approaches based on EOF-extracted features, the values of α had an effect on BSS results, with higher BSS values obtained for higher α values.

Interestingly, the simple approach based on the GLM using the spatial average of the ΔGPP yielded relatively good results

in the U.S. This good performance can be attributed to the strong relationship between U.S. corn yield (particularly in the corn belt region) and the global corn price. Several studies have highlighted that variations in U.S. corn yield significantly influence global corn prices [36].

As an illustration, Fig. 10 shows the annual variations in GPP for July against the yearly variation in October corn prices. The value of R^2 reached 0.5, suggesting that about 50% of the variability in October corn prices was explained by changes in July GPP. Such a relationship indicates that higher GPP in July (which reflects better-growing conditions and
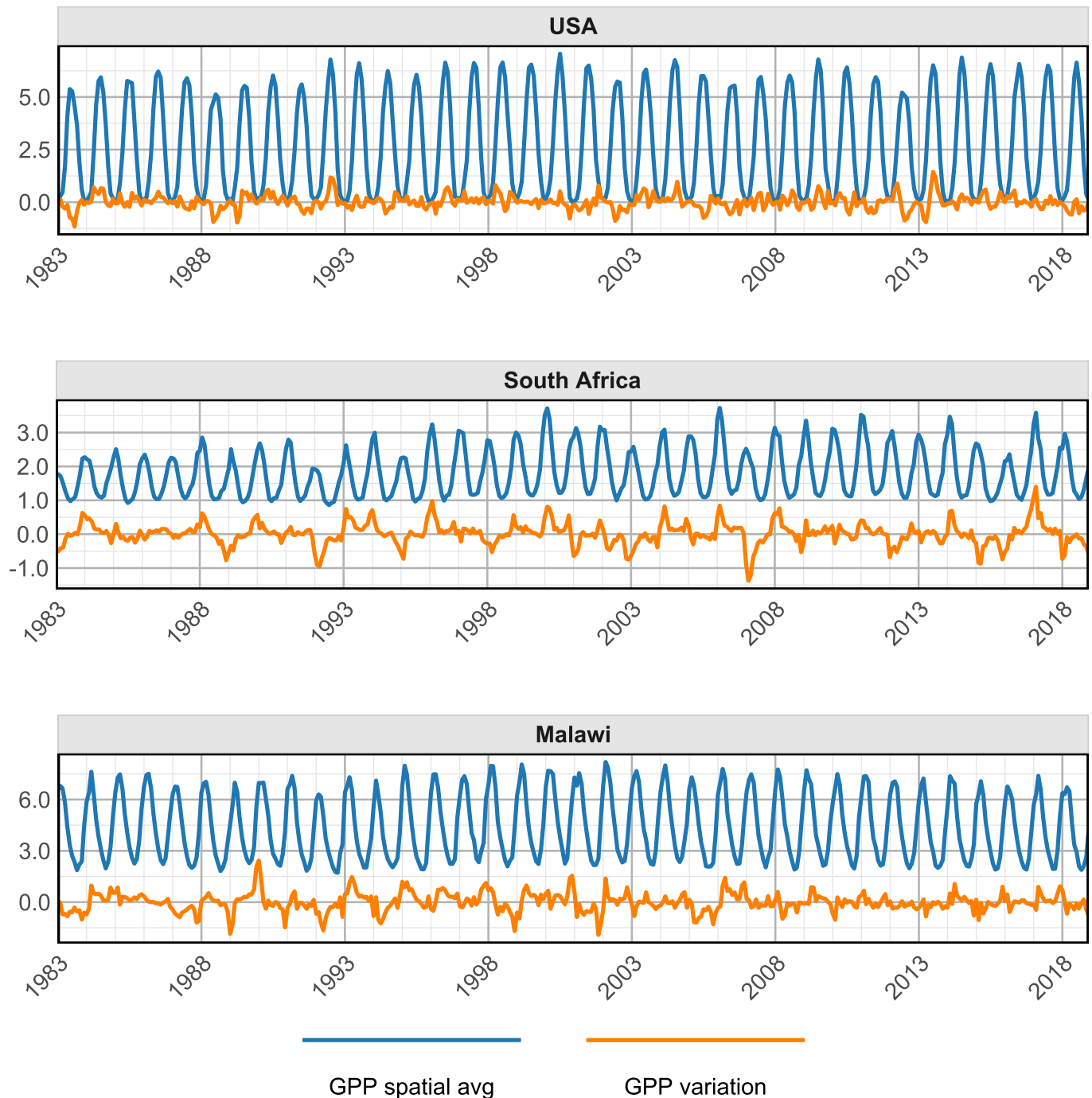
Fig. 13. Time series of the monthly spatial average and interannual variation of GPP from 1983 to 2018 for the U.S., South Africa, and Malawi, without applying a crop mask. The blue line represents the spatial average of GPP, while the orange line illustrates the interannual variation in GPP.

potentially higher yields) may contribute to lower corn prices in October due to an anticipated increase in corn supply and vice versa.

In Malawi, the approaches based on AEs and the β-VAE outperformed the other methods, that is, the EOF, the ΔGPP spatial average, and the 2D-CNN, as shown in Fig. 8. Additionally, in Malawi, the scores tended to be higher for predictions derived from non-masked GPP. The ElasticNet ($\alpha = 1$), using features extracted from the β-VAE, stands out with an excellent AUC of 0.97 and a high MCC of 0.69, highlighting its superior classification performance. The model also

achieved a high BSS of 0.77, indicating superior performance compared to a constant reference model.

Fig. 9(C) further confirms the remarkable prediction performance of the β-VAE with ElasticNet ($\alpha = 1$) approach in Malawi by depicting the predicted probabilities of price increases for years of actual price increase and decrease separately. Notably, the predicted probabilities of price increases were much higher in the case of an actual price increase. The median probability of a price increase was 0.95 when there was an actual price increase, but only 0.03 in the case of an actual price decrease, demonstrating the model ability

to discriminate between actual price increases and decreases effectively.

Fig. 9(C) illustrates that the β-VAE approach almost flawlessly discriminates between price increases and decreases, with only one instance of misclassification. This anomaly occurred in 2000 when the ElasticNet model erroneously forecast a price increase instead of a decrease. A potential explanation for this discrepancy can be discerned in Fig. 11, showcasing the annual GPP variation from 1999 to 2000. The figure delineates a stark contrast between the North and South regions, with a decline in GPP values observed in the South compared to the previous year, while no strong variation was reported in the North. The primary corn-producing regions (Kasungu, Lilongwe, Mchinji), situated in the western part of the country as depicted in Fig. 11, also experienced a drop in GPP values in 2000 compared to 1999. As our approach relies on GPP data, the decline in GPP values in the main producing region led to a predicted price increase resulting from an anticipated reduction of corn production, although the price had actually decreased in 2000. Nevertheless, despite this single misclassification, the model did not lead to any other classification error.

In South Africa, the ElasticNet ($\alpha = 0$, i.e., ridge) using features extracted from the non-masked GPP with the β-VAE approach emerged as the top performer, achieving a BSS of 0.44, an MCC of 0.58, and an AUC of 0.88 (see Fig. 8). Overall, model performances were lower in South Africa than in the U.S., and this result is confirmed in Fig. 9(B), where the predicted probabilities of price increase appear to be less contrasted between cases of actual price increases and decreases in South Africa compared to the other two countries. Fig. 8 indicates that the approach based on non-masked data significantly outperforms the masked one. Performance gains are particularly striking for the AE and β-VAE models.

While our focus has been on using satellite-based GPP, socio-economic factors such as regional conflicts or sociological crises can also influence corn production and prices. Including these factors could enhance the predictive power of our methodology. However, incorporating socio-economic factors may introduce new sources of uncertainty as well. Reliable data sources and appropriate indices or proxies for these factors might not be readily available, making them difficult to access early enough in the season for useful predictions. Moreover, some of these factors impact vegetation characteristics (e.g., poor crop management due to conflict reduces cropland cover, leading to decreased GPP), so their effects are already indirectly incorporated into our GPP-based predictors and accounted for by our classification models. Nonetheless, exploring the benefits of directly integrating socio-economic factors for predicting yield and price variations in the future would be valuable.

IV. CONCLUSION

This research introduces a novel methodology for predicting corn yield and price variations at the national scale, relying on satellite-derived GPP data and dimension-reduction techniques. Our approach is very promising for improving large-scale crop yield and price forecasts several months before harvest.

The study reveals that dimension-reduction strategies are useful to improve model accuracy. Neural network techniques based on AE and VAE demonstrated superior predictive capabilities across all three studied countries, in particular compared to simpler techniques based on EOFs and standard spatial GPP averages. Specifically, VAE showed the best results in the U.S. and South Africa, while AE performed better in Malawi.

This study highlights the potential of integrating satellite data with advanced dimension-reduction techniques to improve the accuracy of crop forecasts. As the proposed approach requires only open-access data, it could easily be used by many stakeholders to improve agricultural commodities' management, aiming to increase food security.

REFERENCES

[1] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and artificial neural networks models for stock price prediction," *J. Appl. Math.*, vol. 2014, pp. 1–7, Mar. 2014.

[2] J. Ansarifar, L. Wang, and S. V. Archontoulis, "An interaction regression model for crop yield prediction," *Sci. Rep.*, vol. 11, no. 1, p. 17754, Sep. 2021.

[3] W. Bank, "Commodity markets," World Bank 2024 Commodity Market 'Pink Sheet' Data, 2024. [Online]. Available: https://www.worldbank.org/en/research/commodity-markets#1

[4] I. Becker-Reshef et al., "Crop type maps for operational global agricultural monitoring," *Sci. Data*, vol. 10, no. 1, p. 172, Mar. 2023.

[5] J. Crespo Cuaresma, J. Hlouskova, and M. Obersteiner, "Agricultural commodity price dynamics and their determinants: A comprehensive econometric approach," *J. Forecasting*, vol. 40, no. 7, pp. 1245–1273, Nov. 2021.

[6] G. Denning et al., "Input subsidies to improve smallholder maize productivity in Malawi: Toward an African green revolution," *PLoS Biol.*, vol. 7, no. 1, Jan. 2009, Art. no. e1000023.

[7] O. Ecker and M. Qaim, "Analyzing nutritional impacts of policies: An empirical study for Malawi," *World Develop.*, vol. 39, no. 3, pp. 412–428, 2011.

[8] *Crop and Livestock Statistic*, FAO, Rome, Italy, 2023.

[9] *The State of Food Security and Nutrition in the World 2023: Urbanization, Agrifood Systems Transformation and Healthy Diets Across the Ruralurban Continuum*, FAO, Rome, Italy, 2023.

[10] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: A review," *Int. J. Climatol.*, vol. 27, no. 9, pp. 1119–1152, Jul. 2007.

[11] M. He et al., "Regional crop gross primary productivity and yield estimation using fused landsat-MODIS data," *Remote Sens.*, vol. 10, no. 3, p. 372, Feb. 2018.

[12] I. Higgins et al., "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. Int. Conf. Learn. Represent.*, 2016. [Online]. Available: https://openreview.net/forum?id=Sy2fzU9gl

[13] C. Jiang and Y. Ryu, "Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from breathing Earth system simulator (BESS)," *Remote Sens. Environ.*, vol. 186, pp. 528–547, Dec. 2016.

[14] I. T. Jolliffe, *Principal Component Analysis* (Springer Series in Statistics), 2nd ed., New York, NY, USA: Springer.

[15] M. Jung et al., "The FLUXCOM ensemble of global land-atmosphere energy fluxes," *Sci. Data*, vol. 6, no. 1, p. 74, May 2019.

[16] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[17] N. Kohzadi, M. S. Boyd, B. Kermanshahi, and I. Kaastra, "A comparison of artificial neural network and time series models for forecasting commodity prices," *Neurocomputing*, vol. 10, no. 2, pp. 169–181, Mar. 1996.

[18] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[19] S. Liang et al., "The global land surface satellite (GLASS) product suite," *Bull. Amer. Meteorol. Soc.*, vol. 102, no. 2, pp. E323–E337, Feb. 2021.

[20] J. Mazunda and K. Droppelmann, "Maize consumption estimation and dietary diversity assessment methods in Malawi," Int. Food Policy Res. Inst., Washington, DC, USA, MaSSP Tech. Rep. 11, 2012.

[21] F. Unsal, J. Spray, and C. Okou, "Staple food prices in sub-saharan Africa: An empirical assessment," *IMF Work. Papers*, vol. 2022, no. 135, p. 1, Jul. 2022.

[22] H. Ouyang, X. Wei, and Q. Wu, "Agricultural commodity futures prices prediction via long- and short-term time series network," *J. Appl. Econ.*, vol. 22, no. 1, pp. 468–483, Jan. 2019.

[23] *R: A Language and Environment for Statistical Computing*, R Core Team, Vienna, Austria, 2023.

[24] *Historic Whole Grain Info*, SAGI Service, Pretoria, South Africa, 2023.

[25] B. Sharif, D. Makowski, F. Plauborg, and J. E. Olesen, "Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark," *Eur. J. Agronomy*, vol. 82, pp. 11–20, Jan. 2017.

[26] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," *J. Stat. Softw.*, vol. 39, no. 5, pp. 1–13, 2011.

[27] F. Teste, D. Makowski, H. Bazzi, and P. Ciais, "Early forecasting of corn yield and price variations using satellite vegetation products," *Comput. Electron. Agricult.*, vol. 221, Jun. 2024, Art. no. 108962.

[28] R. S. Thompson, "Measuring the relationship between price and yield," in *Proc. Annu. Meeting Agricult. Appl. Econ. Assoc.*, Kansas City, Missouri, Jul. 2020. [Online]. Available: https://ageconsearch.umn.edu/record/304188?v=pdf

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

[30] *USDA National Agricultural Statistics Service Cropland Data Layer*, USDA-NASS, Washington, DC, USA, 2022.

[31] M. van Dijk, T. Morley, M. L. Rau, and Y. Saghai, "A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050," *Nature Food*, vol. 2, no. 7, pp. 494–501, Jul. 2021.

[32] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences* (International Geophysics Series), 2nd ed., Boston, MA, USA: Academic, 2006.

[33] X. Xu and Y. Zhang, "Soybean and soybean oil price forecasting through the nonlinear autoregressive neural network (NARNN) and NARNN with exogenous inputs (NARNNX)," *Intell. Syst. Appl.*, vol. 13, Jan. 2022, Art. no. 200061.

[34] S. Yang et al., "The influence of nitrogen fertilization on crop production and ecohydrology in an endorheic river basin," *J. Hydrol.*, vol. 625, Oct. 2023, Art. no. 130035.

[35] R. Zelingher and D. Makowski, "Forecasting global maize prices from regional productions," *Frontiers Sustain. Food Syst.*, vol. 6, Apr. 2022, Art. no. 836437.

[36] R. Zelingher and D. Makowski, "Investigating and forecasting the impact of crop production shocks on global commodity prices," *Environ. Res. Lett.*, vol. 19, no. 1, Jan. 2024, Art. no. 014026.

[37] Y. Zhang et al., "Effect of using crop specific masks on Earth observation based crop yield forecasting across Canada," *Remote Sens. Appl., Soc. Environ.*, vol. 13, pp. 121–137, Jan. 2019.

[38] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005.

Hugo Gangloff received the Ph.D. degree in statistical signal processing from the University of Strasbourg, Strasbourg, France, in 2020.

He is currently a Research Engineer with INRAE, Laboratory MIA—Paris Saclay. His research interests include optimization and inference in probabilistic models, physics-informed machine learning, and deep learning.

Mathilde Chen received the agronomist graduate degree from UniLaSalle Institute, Beauvais, France, in 2016, and the Ph.D. degree in agronomic sciences from Université Paris-Saclay and AgroParisTech, Palaiseau, France, in 2020.

Her main interests include analyzing epidemiological, agronomic, and climate data. During her research experience, she used several statistical methods (e.g., fixed and mixed-effect linear regression, survival analysis, and several machine learning algorithms such as decision trees and random forests).
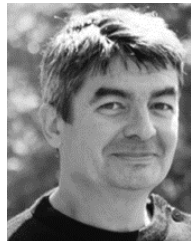
Philippe Ciais received the Ph.D. degree in atmospheric sciences from the University of Paris-VI, Paris, France, in 1991.

He has been a Post-Doctoral Researcher at NOAA, Boulder, CO, USA, and the Laboratory of Climate and Environmental Sciences (LSCE), CEA CNRS UVSQ, Gif-sur-Yvette, France. He has worked on the carbon cycle and greenhouse gas emissions with models and observations and contributed to several publications in this field.

Dr. Ciais is a member of the French and the Chinese Academy of Sciences.
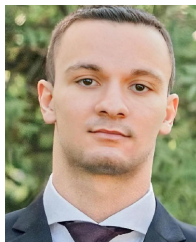
Florian Teste received the master's degree in remote sensing and geomatics from Aix-Marseille University, Marseille, France, in 2021. He is currently pursuing the Ph.D. degree in applied mathematics with the University of Paris-Saclay, Palaiseau, France, supported by Atos.

His research interests include developing innovative methods to predict agricultural crop price and yield variations, specifically for corn, using only satellite data. His key areas of interest include remote sensing, food security, agroecology, and environmental science.

David Makowski received the M.Sc. and Ph.D. degrees from AgroParisTech, Palaiseau, France, in 1996 and 2001, respectively.

He is a Senior Researcher at INRAE, University Paris-Saclay, Palaiseau. He manages scientific projects and supports research groups in analyzing complex datasets using statistical and machine learning methods. His main interests include statistical modeling applied to risk analysis, food safety, agroecology, environmental sciences, and climate change.