Impacts of climate change on spatial wheat yield and nutritional values using hybrid machine learning

Ahmed M S Kheir, Osama a M Ali, Ashifur Rahman Shawon, Ahmed S Elrys, Marwa G M Ali, Mohamed A Darwish, Ahmed M Elmahdy, Ayman Farid Abou-Hadid, Rogerio de S Nóia Júnior, Til Feike

▶ To cite this version:

Ahmed M S Kheir, Osama a M Ali, Ashifur Rahman Shawon, Ahmed S Elrys, Marwa G M Ali, et al.. Impacts of climate change on spatial wheat yield and nutritional values using hybrid machine learning. Environmental Research Letters, 2024, 19, pp.104049. 10.1088/1748-9326/ad75ab . hal-04723843

ENVIRONMENTAL RESEARCH
LETTERS

LETTER • **OPEN ACCESS**

Impacts of climate change on spatial wheat yield and nutritional values using hybrid machine learning

To cite this article: Ahmed M S Kheir *et al* 2024 *Environ. Res. Lett.* **19** 104049

IOP Publishing Environ. Res. Lett. **19** (2024) 104049 https://doi.org/10.1088/1748-9326/ad75ab

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
25 May 2024

REVISED
22 August 2024

ACCEPTED FOR PUBLICATION
30 August 2024

PUBLISHED
9 September 2024

LETTER

Impacts of climate change on spatial wheat yield and nutritional values using hybrid machine learning

Ahmed M S Kheir[1,2,*], **Osama A M Ali**[3], **Ashifur Rahman Shawon**[1], **Ahmed S Elrys**[4,5,6], **Marwa G M Ali**[2], **Mohamed A Darwish**[7], **Ahmed M Elmahdy**[2], **Ayman Farid Abou-Hadid**[8], **Rogerio de S Nóia Júnior**[9] and **Til Feike**[1]

[1] Institute for Strategies and Technology Assessment, Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, 14532 Kleinmachnow, Germany
[2] Soils, Water and Environment Research Institute, Agricultural Research Center, Giza 12112, Egypt
[3] Department of Crop Science, Faculty of Agriculture, Menoufia University, 32514 Shebin El-Kom, Egypt
[4] Soil Science Department, Faculty of Agriculture, Zagazig University, Zagazig 44511, Egypt
[5] College of Tropical Crops, Hainan University, Haikou 570228, People's Republic of China
[6] Liebig Centre for Agroecology and Climate Impact Research, Justus Liebig University, Giessen, Germany
[7] Wheat Research Department, Field Crops Research Institute, Agricultural Research Center, Giza 12619, Egypt
[8] Faculty of Agriculture, Ain Shams University, Hadayek, shobra 11241 Cairo, Egypt
[9] LEPSE, Univ Montpellier, INRAE, Institut Agro Montpellier, Montpellier, France
[*] Author to whom any correspondence should be addressed.

E-mail: drahmedkheir2015@gmail.com and ahmed.kheir@julius-kuehn.de

Keywords: CMIP6, downscaled NEX scenarios, automatic machine learning, stacked ensemble model, uncertainty, nutritional concentrations

Supplementary material for this article is available online

Abstract

Wheat's nutritional value is critical for human nutrition and food security. However, more attention is needed, particularly regarding the content and concentration of iron (Fe) and zinc (Zn), especially in the context of climate change (CC) impacts. To address this, various controlled field experiments were conducted, involving the cultivation of three wheat cultivars over three growing seasons at multiple locations with different soil and climate conditions under varying Fe and Zn treatments. The yield and yield attributes, including nutritional values such as nitrogen (N), Fe and Zn, from these experiments were integrated with national yield statistics from other locations to train and test different machine learning (ML) algorithms. Automated ML leveraging a large number of models, outperformed traditional ML models, enabling the training and testing of numerous models, and achieving robust predictions of grain yield (GY) ($R^2 > 0.78$), N ($R^2 > 0.75$), Fe ($R^2 > 0.71$) and Zn ($R^2 > 0.71$) through a stacked ensemble of all models. The ensemble model predicted GY, N, Fe, and Zn at spatial explicit in the mid-century (2020–2050) using three Global Circulation Models (GCMs): GFDL-ESM4, HadGEM3-GC31-MM, and MRI-ESM2-0 under two shared socioeconomic pathways (SSPs) specifically SSP2-45 and SSP5-85, from the downscaled NEX-GDDP-CMIP6. Averaged across different GCMs and SSPs, CC is projected to increase wheat yield by 4.5%, and protein concentration by 0.8% with high variability. However, it is expected to decrease Fe concentration by 5.5%, and Zn concentration by 4.5% in the mid-century (2020–2050) relative to the historical period (1980–2010). Positive impacts of CC on wheat yield encountered by negative impacts on nutritional concentrations, further exacerbating challenges related to food security and nutrition.

1. Introduction

Food and nutrition security are the primary motivators for sustainable development by avoiding both visible and hidden hunger (Wood *et al* 2018, Willett *et al* 2019, Wang *et al* 2023). During the mid of 20th century the green revolution tripled the production outputs with the aid of various technological advances (Gould 2017). Such efforts have continued into the twenty-first century to produce more food and reduce hunger (Tilman *et al* 2011, Asseng *et al* 2018, Springmann *et al* 2018, Van Dijk *et al* 2021), but nourishment issues (malnutrition) still require greater attention under the global food security agenda. Significant attention should be focused on increasing crop production by 60% by 2050 to meet food demand and minimize hunger caused by high population growth (Godfray *et al* 2010, Asseng *et al* 2020). Another global crisis, hidden hunger, is a type of undernutrition associated with a lack of micronutrients, particularly iron (Fe) and zinc (Zn), affects 2 billion people (Wang *et al* 2023). Inadequate levels of Fe and Zn not only reducing yield by slowing down the rate of photosynthesis (Roosta *et al* 2018), but they can also adversely affect human health by leading to various ailments like tuberculosis, human immunodeficiency virus, and malaria (Zimmermann and Hurrell 2007). Crop grain protein contributes to wheat flour's nutritional quality, end-use value, and baking qualities (Shewry and Halford 2002). However, research on the interplay impacts of nitrogen as a source of grain protein, Fe, and Zn on wheat yield and quality has received little attention thus far. Furthermore, climate change (CC) has had a significant detrimental impact on crop output (Kheir *et al* 2019, Liu and Dai 2020, Zheng *et al* 2020, Li *et al* 2021, Abbas *et al* 2023, Watts *et al* 2023), with little studies on protein (Asseng *et al* 2019), while much less emphasis has been placed on Fe and Zn as key components in food and nutritional security (Nelson *et al* 2014, 2018, Tilman and Clark 2014).

The relationships between amounts and concentrations of nutritional values under CC remain uncertain, requiring further attention. These concentrations mainly depend on different combinations of soil, weather, genotype, and management practices (Triboi *et al* 2006, Yan *et al* 2022), which need to be considered on the respective approach exploring future impacts. This letter helps to address such gaps, reporting results on exploring the impacts of climate scenarios projected to 2050 that span a wide range of potential futures on wheat yield as well as on amounts and concentrations of protein, Fe and Zn.

To study the historical and future impacts of CC on crop production and nutritional values, an appropriate predictive approach is required. Process-based models have been widely utilized to study and assess the CC impact on crop growth and development from local to global scales (Chen *et al* 2023, Tan *et al* 2023, Kheir *et al* 2023b), but considerably less emphasis has been placed on nutritional values, despite the fact that this is an important part of food security (Haddad *et al* 2016). The main reason of this less attention is the need to develop such models in sub-routines by including additional factors such as Fe and Zn, which require time and high-quality experimental dataset from multiple environments. Machine learning (ML) approaches can fill this gap by linking inputs to responses, allowing flexible future predictions of nutritional values under CC scenarios. ML has shown robust performances in different applications, among them crop yield predictions (Kheir *et al* 2022, Udristioiu *et al* 2023, Hailegnaw *et al* 2024), and CC impacts (Prodhan *et al* 2022, Gao *et al* 2024) with limited studies on nutrient estimation (Ma *et al* 2021, You *et al* 2023). ML techniques can be either traditional (Attia *et al* 2022, Kheir *et al* 2023a) or automatic (Waring *et al* 2020, Xu *et al* 2022), which permits training a large number of algorithms at simultaneously in a short period and with less computing power. Automatic ML can compare and deploy high-performance ML models automatically (Kheir *et al* 2024), but it has still received less attention so far. Coupling both traditional and automatic routines may improve the prediction robustness and reduce uncertainty. For that, we used hybrid ML approaches (traditional and automated) to explore the past and future impacts of CC scenarios from the NASA Earth Exchange Global Daily Downscaled Projections (Thrasher *et al* 2022), from Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring *et al* 2016) on wheat yield and nutritional values and concentrations (protein, Fe and Zn) in arid regions. To our knowledge, this is the first paper investigate the impact of latest CC scenarios on crop production and nutritional values (i.e. protein, Fe and Zn) using hybrid ML approach.

The main objectives to achieve this goal are (I) investigate the impact of different Fe and Zn treatments on wheat production and grain quality for different genotypes cultivated in varied environments; (II) train and test hybrid ML technique to predict wheat yield, protein, Fe and Zn using multi-year national actual grain yield (GY) (t ha^{-1}), soil characteristics, weather conditions, and topography over Egypt; and (III) deploy the trained stacked ensemble of all models (the best model) to predict the historical (1980–2010) and the future (2020–2050) yield and nutritional values using three Global Circulation Models (GCMs) (GFDL-ESM4, HadGEM3-GC31-MM, and MRI-ESM2-0) and two shared socioeconomic pathways (SSPs) (SSP2-45 and SSP5-85).

2. Materials and methods

Since determining crop yield nutritional values necessitates controlled experiments under varied soil and weather circumstances, the methodology is divided into two main sections. The first section is about field experiments to explore the interrelationships between GY, yield attributes, and grain nutritional values for different cultivars grown in multiple locations and growing seasons. The second part is related to training, and testing hybrid ML algorithms using 5 years actual dataset of GY and N content in grains (2015–2019), and deploying the trained models for predicting yield and nutritional values in historical

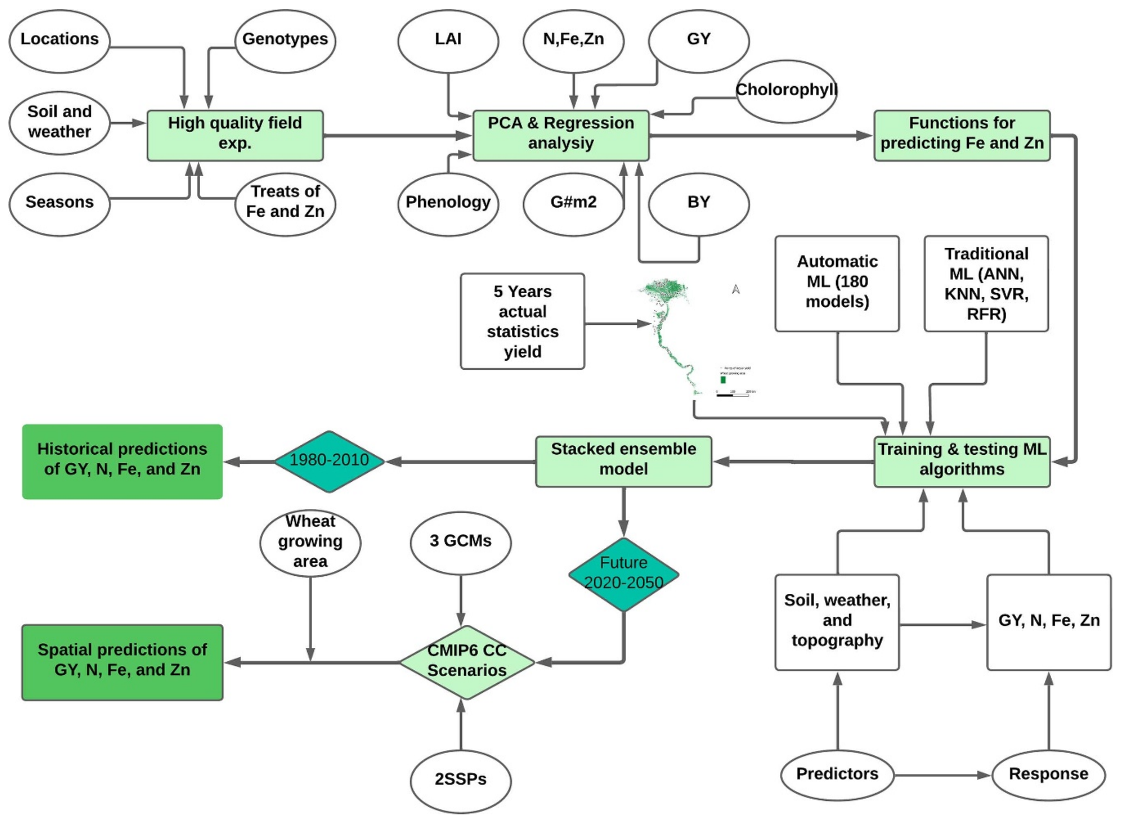

Figure 1. Flowchart summarizes the respective steps for exploring climate change's impacts on wheat grain nutrition (i.e. N, Fe and Zn). The field experiments included growing 3 new wheat cultivars (Giza171, Misr2 and Shandweel1), in three agroclimatic locations (Sakha, Menoufia and Luxor) (figure 2), and different growing seasons (2018/2019, 2019/2020 and 2020/2021) using four treatments with different forms of foliar application from Fe (200 ppm) and Zn (150 ppm) as control, EDTA, citric, and mineral. The recommended N fertilizer (180 kg N ha^{-1}) was added at three doses for all treatments. Regression analysis used to develop functions of estimating Fe and Zn from N. The spatial actual yields from statistics (GY) and N content in grains of 2400 sites over 5 years (2015–2019) were collected from country statistics (Ministry of Agriculture and Land Reclamation 2021) and used as response in machine learning algorithms (ML). Other responses as nitrogen content (N), iron (Fe) and Zn were calculated from regression analysis (supplementary figures 3 and 4). Soil physical and chemical dataset were extracted from ISRIC at 250 m resolution and bias corrected by observations of various location measurements. Daily climate data for 2400 locations during the 5 years included the variables of temperatures, and solar radiation were extracted from reanalysis ERA5 dataset and used along with soil dataset, elevation, and years as predictors in ML. The trained best estimator of ML algorithm was deployed to predict GY, N, Fe and Zn using CMIP6 climate scenarios two SSPs (SSP2-45 and SSP5-85) and three GCMs (MRI-ESM2-0, HadGEM3-GC31-MM and GFDL-ESM4) for historical (1980—2010), and future (2020–2050) time periods.

and future time periods using different CMIP6 climate scenarios. Detailed information of each section is explained here and, in the flowchart, (figure 1).

2.1. Field experiments and actual spatial yield

Three-field experiment were conducted over Egypt (area of interest), at different agroclimatic zones such as Sakha (high latitude, low elevation, and moderate temperature), Menoufia (middle delta), and Luxor (low latitude, high elevation and high temperature), more details are presented in supplementary file and S. Figure 1).

Actual yield from various places in Egypt was acquired from national statistics (Ministry of Agriculture and Land Reclamation 2021) for the years (2015–2019) and used in conjunction with the experimental dataset to train and test ML methods (figure 2). To maintain the consistency of the spatial dataset used to train the ML models, regression analysis was performed to create functions for estimating Fe and Zn elements from N, based on the relevant measurements from field experiments. Incorporating principal component analysis (PCA) into our analysis not only streamlines the dataset but also enhances the robustness and interpretability of the findings, identify and highlight the most influential variables (principal components) affecting wheat yield and nutritional content. This helps in understanding the underlying patterns and relationships within the data, providing clearer insights into how climate variables impact wheat production.

2.2. ML approaches (training and testing)

2.2.1. Dataset

A diverse dataset of soil, weather, and topography, as well as various responses such as GY, grain N content, grain Fe content, and grain Zn content, were utilized to train and test the ML techniques during a five-year period (2015–2019). For more than 2000 sites, soil properties (i.e. sand, silt, clay, soil organic carbon, pH, and bulk density) were downloaded from

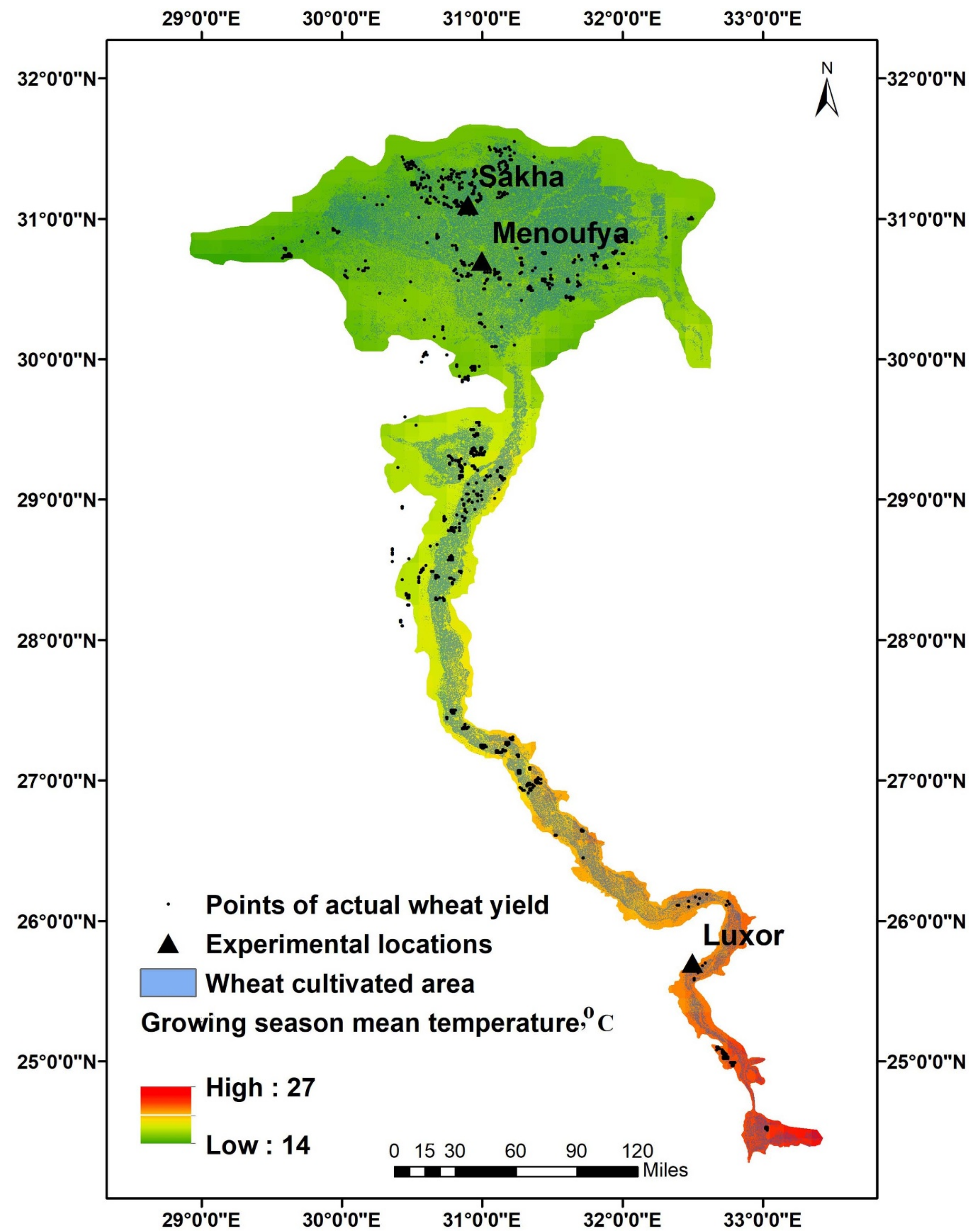

Figure 2. Wheat controlled field experiments (triangles), statistics points of actual yield (points), and temperature gradient along the Nile. Growing seasons mean temperature at 10 km grid resolution and averaged over last 10 years (2010–2020) for Egypt. The experimental locations were used for exploring the non-linear correlations between yield, nitrogen, iron and zinc content and concentration over three growing seasons, 2018/19, 2019/20 and 2020/21. Wheat cultivated area were extracted from winter crop type mapping in the region and used for masking the spatial distribution of yield and nutritional values.

International Soil Reference and Information Centre (ISRIC) dataset at 250 m resolution (www.isric.org/explore/isric-soil-data-hub) and calibrated with some observed dataset over the region. Topography dataset for the same points were downloaded by MODIS at 250 m resolution. Because the experimental sites involved just three locations, and ML trained over 2000 sites that require meteorological dataset for each site, the average maximum temperature, minimum temperature, and solar radiation were retrieved from the ERA5 global reanalysis at 10 km resolution (Hersbach *et al* 2020).

2.2.2. ML approaches

Two approaches were considered in the current work, including traditional ML and automated ML (AutoML), creating a hybrid approach.

Traditional ML and AutoML are two techniques to developing predictive models, but they differ in terms of their processes, complexity, and level of human intervention. For the traditional ML we developed four algorithms such as artificial neural networks (ANN) (van Klompenburg *et al* 2020), random forest (RF) (Xu *et al* 2020), Support Vector Regressor (SVR) (Paryani *et al* 2022), and k-nearest neighbors (KNN). The ANNs are computational models that mimic the network of neurons in the human brain. They consist of layers of nodes (neurons), with each layer connected to the next, and the output of a neuron y is calculated as:

$$y = f\left(\sum_{i=1}^{n} WiXi + b\right) \quad (1)$$

where, Xi are the input features, Wi are the weights, b is the bias, and f is the activation function (e.g. sigmoid, ReLU).

RF is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the individual trees, and the prediction for an input (X) is given by:

$$y = 1/T\sum_{t=1}^{T} ht(X) \quad (2)$$

where T is the number of trees, and ht is the prediction of the tth tree

The SVR is used for regression problems, aiming to find a hyperplane that best fits the data while maintaining a margin of tolerance ϵ, and the objective function for SVR is:

$$\frac{\min 1}{2||W||^2} + C\sum_{i=1}^{n} \max(0, |yi - (W.Xi + b)| - \varepsilon \quad (3)$$

where C is the regularization parameter, and ϵ\epsilon$\epsilon$ is the margin of tolerance.

KNN predicts the value of a new data point based on the average of the values of its KNN according the following equation:

$$y = 1/k\sum_{i=1}^{k} yi \quad (4)$$

where yi are the target values of the KNN.

In Automatic ML (H_2OAutoML) (Ledell and Poirier 2020), we automatically compared and deployed high-performance and large number of ML models. AutoML streamlines the process of applying ML by automating the se-lection and tuning of models. H_2O AutoML incorporates XGBoost Gradient Boosting Machines (GBM), H_2O GBM, RFs (Default and Extremely Randomized Tree), deep neural networks, and Generalized Linear Models. H_2O AutoML includes a wrapper for the popular XGBoost program, allowing us to use this third-party method. We utilized H_2OAutoML in our study which automates the model training process by evaluating multiple algorithms and their hyperparameters, ultimately selecting the best-performing model according to the following equation,

$$\text{Best Model} = \arg\max_{m \in M} \text{Performance } (m, D) \quad (5)$$

where M is the set of models and hyperparameters, D is the dataset, and Performance is a metric such as accuracy or root mean square error (RMSE).

For hybridization, we followed the ensemble learning approach, which combines predictions from multiple models to create a more robust and potentially more accurate model. This method leverages the strengths of different models to improve overall performance. Specifically, by combining predictions from AutoML models with those from traditional models, we harnessed the diverse strengths of both approaches, thereby enhancing the overall predictive performance and robustness of the final model. This integration led to a superior model that effectively capitalized on the strengths of both traditional and AutoML techniques. The best estimator (stacked ensemble model) of all models was considered for future predictions.

2.2.3. Data processing, feature selections and hyper parameter tuning

To avoid the impacts of different scaling, the scikit-learn StandardScaler algorithm uses the means and standard deviation of the variables was considered as the first stage of pre-processing in the data to standardize them. The developed data include only topography, weather, and soil which is considered enough for training and testing ML approaches with suitable fitting, thus feature selection is not necessary here. The default approach was used to test the ML models by randomly splitting the whole dataset to 80% for training and 20% for testing. To optimize the hyper-parameters of the ML algorithms (Zeng *et al* 2022) the grid search technique with a 5-fold cross-validation approach (Pedregosa *et al* 2011) is used. Grid search is a comprehensive strategy that can attempt to locate all potential hyperparameter combinations in order to secure the best decision (appendix A). Maintaining high performance and showcasing diversity are essential for choosing the ideal model. Despite needing more hyper-parameter tuning, neural networks often yield better performance, so we used them in both automated and conventional ML techniques.

2.2.4. Assessment of ML models

To assess the trained ML model, different equations were used and included determination coefficient (R^2), RMSE, relative bias (RB), and mean absolute error (MAE). The R^2 value shows the degree to which the fitted regression line and the data are similar. In multiple regression, it is sometimes referred to as the coefficient of determination or the coefficient of multiple determination. It can be defined as the proportion of the response variable's variance that a linear model can account for. It is always in the range of 0–1, with numbers closer to 1 denoting increased accuracy,

$$(R)^2 = 1 - \frac{\sum(Y - Y1)^2}{\sum(Y - Y1)^2} \tag{6}$$

where Y, $Y1$, and n are actual, estimated, and a number of data points respectively.

Bias in RB considers the influences on the outcome that happen at any point during the analytical process and are systematic over time. It is helpful when the absolute bias is proportionate to the analyte concentration, and it is intended to use the precise bias estimate at various concentration levels,

$$\mathrm{RB} = \frac{Y1 - X}{X} \times 100 \tag{7}$$

where $Y1$ is the estimated and X is the observed value.

The MAE provides a measure of the average magnitude of errors between predicted and true values, the lower value the robust model accuracy. The formula for calculating the MAE is as follows:

$$\mathrm{MAE} = \mathrm{n}1 \sum i = 1\mathrm{n}\left|X - Y1\right|. \tag{8}$$

MAE calculates the absolute difference between each predicted value X and its corresponding true value $Y1$, averages these absolute differences across all observations, and provides a single, interpretable value.

The RMSE is another commonly used metric to measure the accuracy of predictions by quantifying the average magnitude of errors. It is particularly useful for assessing the predictive performance of a model, and it is closely related to the mean squared error. The formula for calculating the RMSE is as follows:

$$\mathrm{RMSE} = \frac{\sqrt{\sum(X - Y)^2}}{n} \tag{9}$$

where n is the number of observations, X is the estimated and Y is the observation.

2.3. CC scenarios

To explore the future impact of CC on wheat yield and nutritional values, two SSPs (SSP245 and SSP585) and three GCMs (MRI-ESM2-0, HadGEM3-GC31-MM and GFDL-ESM4) for historical (1980–2010), and future (2020–2050) were used. We chose these three GCMs (MRI-ESM2-0, HadGEM3-GC31-MM, and GFDL-ESM4) due to their extensive use in studies focusing on the Middle East and North Africa region, including Egypt, albeit under CMIP5 scenarios (Asseng *et al* 2018). By applying these models with the most recent CMIP6 scenarios, we enhance the relevance and reliability of our projections and findings. Additionally, these models cover a range of climate sensitivities, offering a comprehensive spectrum of potential future climate scenarios. This diversity ensures that our analysis accounts for a wide array of possible impacts on wheat production and nutritional quality, thereby strengthening the robustness and applicability of our results. The scenarios were extracted from the downscaled NEX-GDDP-CMIP6 at 25 km resolution. The R package (RclimChange) was used to download daily GCM data from NCCS THREDDS NEX-GDDP-CMIP6.

2.4. Uncertainty and structural equation models

Uncertainty was calculated using the coefficient of variation by means and standard deviation between observations (Asseng *et al* 2013). To test the presence of collinearities among variables, the variance inflation factor (VIF) analysis was applied and variables with VIF $>$ 5 were excluded. We employed SEM to investigate how wheat GY is influenced by plant nutrient concentration (e.g. grain N and Zn) as well as environmental factors (e.g. latitude, slope, and minimum and maximum temperatures). The conceptual SEM included the direct impacts of grain N and Zn concentrations, latitude, slope, minimum temperature, and maximum temperature on wheat GY and the indirect impacts where latitude, slope, minimum temperature, and maximum temperature affected wheat GY via changing grain N and Zn concentrations. The conceptual SEM also included the indirect impacts of latitude on wheat GY and its N and Zn concentrations via altering minimum and maximum temperatures. The 'lavaan' package in R was applied to test the SEM, and the conceptual model was evaluated by the goodness- of-fit statistics [comparative fit index (CFI) = 0.999, and Tucker–Lewis index = 0.996, and RMSE of Approximation = 0.039].

3. Results

3.1. Wheat yield and nutritional values from controlled field experiments

The PCA revealed significant correlations between treatments (EDTA and Citric) and various wheat yield and nutritional values, including GY, biomass yield, and grain nutrient content (figure 3). Grain Fe and Zn were significantly correlated with each other but less so with grain N across different seasons, locations, and cultivars (supplementary figures 3 and 4). Regression analysis supported these findings (supplementary figures 5–7), enabling equations to estimate Fe and Zn from measured N with acceptable accuracy. Rising mean temperatures during the growing season reduced GY and contents of N, Fe, and Zn, with variability between cultivars and treatments. The cultivar

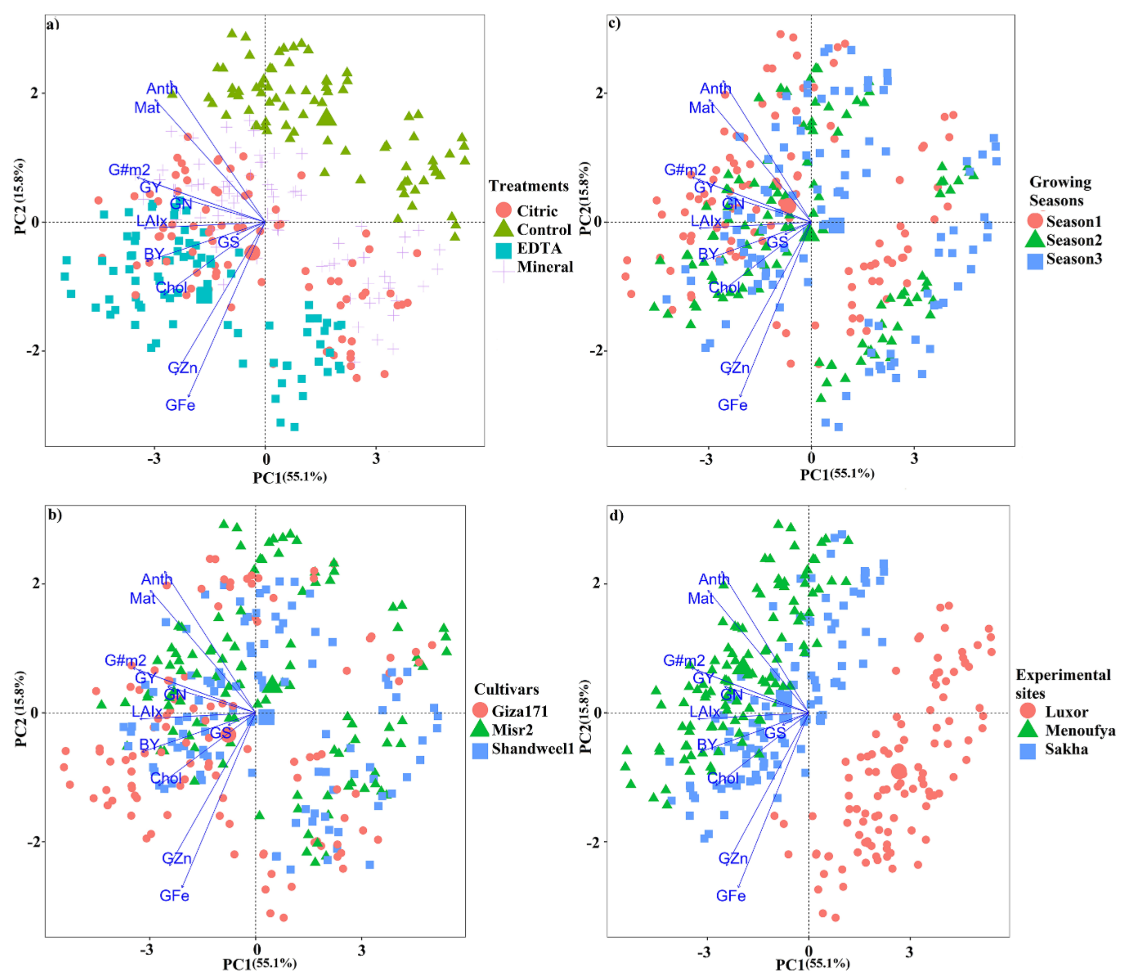

Figure 3. Principal component analysis (PCA) of wheat yield and yield attributes subjected to different variables. The variables include (a) three different treatments of Fe and Zn (EDTA, citric and mineral) for (b) three wheat cultivars (Giza171, Misr2 and Shandweel1) in (c) three successive growing seasons (2018/2019, 2019/2020 and 2020/2021), and (d) three locations (Sakha, Menoufia and Luxor). The figure contains the first two principal components, PC1 and PC2, and their respective scores explaining variation within the data. The variables included grain yield (GY), biomass yield (BY), grain number per m2 (G#m2), grain N content (GN), maximum leaf area index (LAIx), chlorophyll content (Chol), grain weight (GZ), grain Fe content (GFe), grain Zn content (GZn), anthesis date (Anth) and maturity date (Mat).

Giza171 and the use of Fe and Zn as EDTA mitigated the negative effects of rising temperatures.

3.2. Training and testing hybrid ML

The traditional ML algorithms including ANN, RF, SVR, and KNN showed robust training and validation to estimate GY, grain N content, grain Fe content and grain Zn content (supplementary figures 10–13). The trained ML had a higher R^2 (>0.65) and lower RMSE, RB, and MAE for predicting yield and nutritional values (N, Fe and Zn), whereas ANN and RFR outperformed SVR and KNN. On the other hand, H_2O AutoML presently provides the same automatic data preprocessing as all H_2O supervised learning algorithms. This contains automatic imputation, normalizing (when needed), and one-hot encoding for XGBoost models. H_2O tree-based models (GBM, RFs) accept categorical variable grouping, allowing categorical data to be handled natively. This also enables GPU-accelerated training. To achieve robust accuracy, we combined AutoML and traditional ML by incorporating the same traditional models into AutoML, resulting in a hybrid approach. The AutoML approach used approximately 180 models (appendix A), selecting the best 30 models with higher accuracy, and aggregating them into a single ensemble model, which attained the highest accuracy. Compared with traditional ML models, the stacked ensemble model from the hybrid approach showed higher accuracy of predicting wheat GY ($R^2 = 0.78$), grain protein content ($R^2 = 0.75$), grain Fe and Zn contents ($R^2 = 0.71$) (figure 4). This suggests that this approach has the potential to be used to predict wheat production and nutritional attributes in a variety of spatiotemporal ways even under CC, as current models are trained on diverse locations and years of data across Egypt.

3.3. CC impacts on wheat GY, and nutritional values

The trained stacked ensemble ML model was deployed to predict wheat GY, grain N content, grain Fe and Zn contents across the historical (1980–2010)

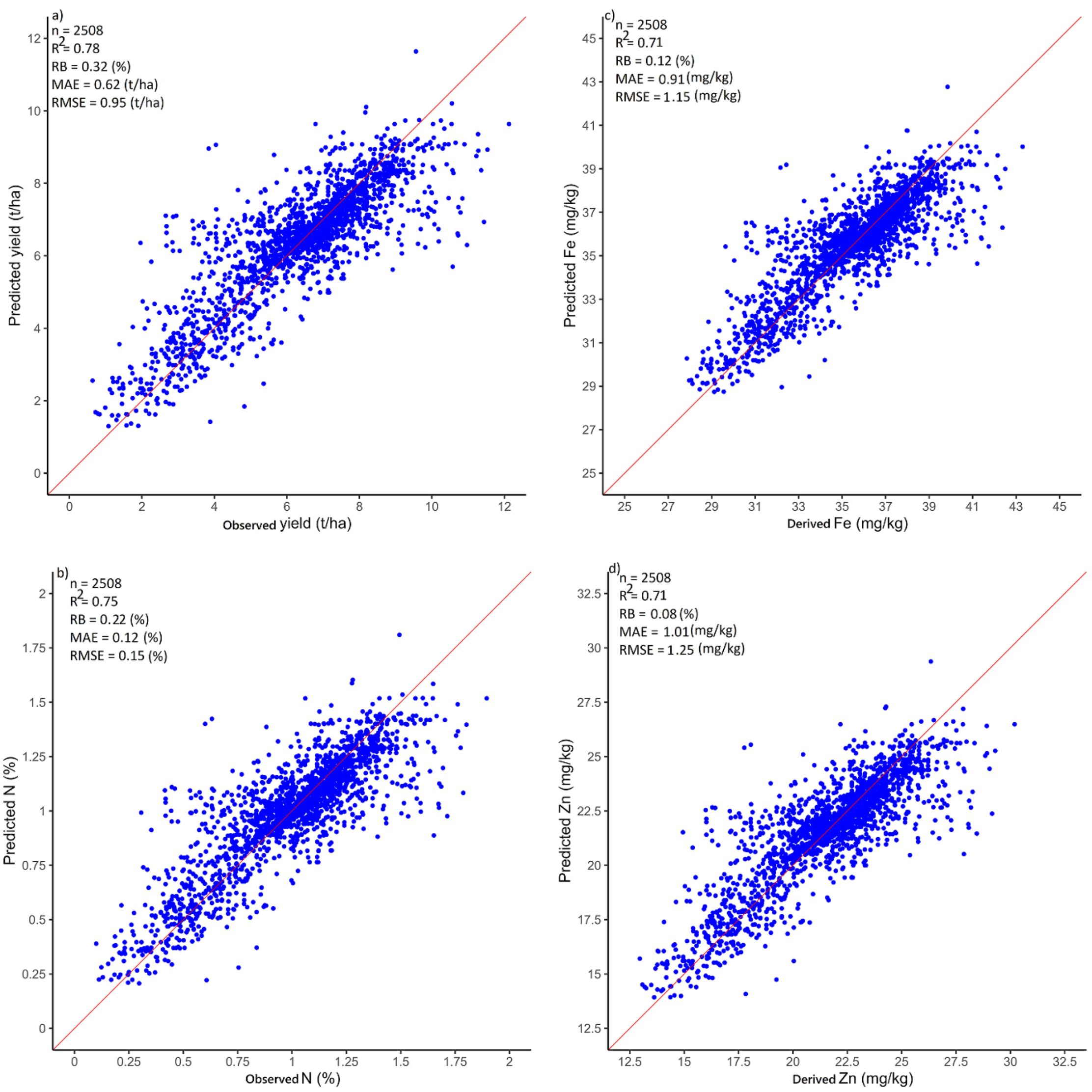

Figure 4. Predicted against observed wheat grain yield (a), wheat grain N content (b), wheat grain Fe content (c), and wheat grain Zn content (d) using the ensemble hybrid machine learning model. Across 2006 sites five years of actual yield (2015–2019) and nitrogen were used as response while soil, weather and topography were used as predictors. Values of Fe and Zn were estimated from N (supplementary figure 7) and used herein also as response for training ML models. The default training approach was used by randomly splitting 80% from the whole dataset for training and 20% for testing.

and future (2020–2050) time periods considering 3 GCMs and two SSPs. The concentration of grain protein, Fe and Zn were calculated from the main contents and GY and considered hereinafter in the analysis.

The SEM analysis revealed that grain N and Zn concentrations were the key factors controlling wheat GY, whereby wheat GY increased significantly with increasing grain N concentration but decreased significantly with increasing grain Zn concentration (figure 5). Wheat GY increased significantly with increasing the slope and decreasing the maximum temperature. Our SEM also showed that increased maximum temperature and decreased minimum temperature indirectly enhanced wheat GY by increased grain N concentration and decreased grain Zn concentration ($p < 0.001$). Higher latitude also increased wheat GY through reducing the maximum temperature, but also can decrease wheat GY through increasing grain Zn concentration directly or indirectly by increasing the minimum temperature ($p < 0.001$). The controlling factors in combination explained the 99%, 87%, and 24% variation of wheat GY, grain N concentration, and grain Zn concentration, respectively. In general, climatic change had a beneficial impact on wheat yield, protein yield, and protein concentration, while negatively affecting Fe and Zn concentrations (figures 6 and 7). The Shared Socioeconomic Pathway scenario, SSP8.5, which is distinguished by high population growth, slow economic development, and a high reliance on fossil fuels, resulted in a greater reduction in yield and protein content but a lower reduction in Fe and Zn concentrations than the scenario SSP45 (figure 7).

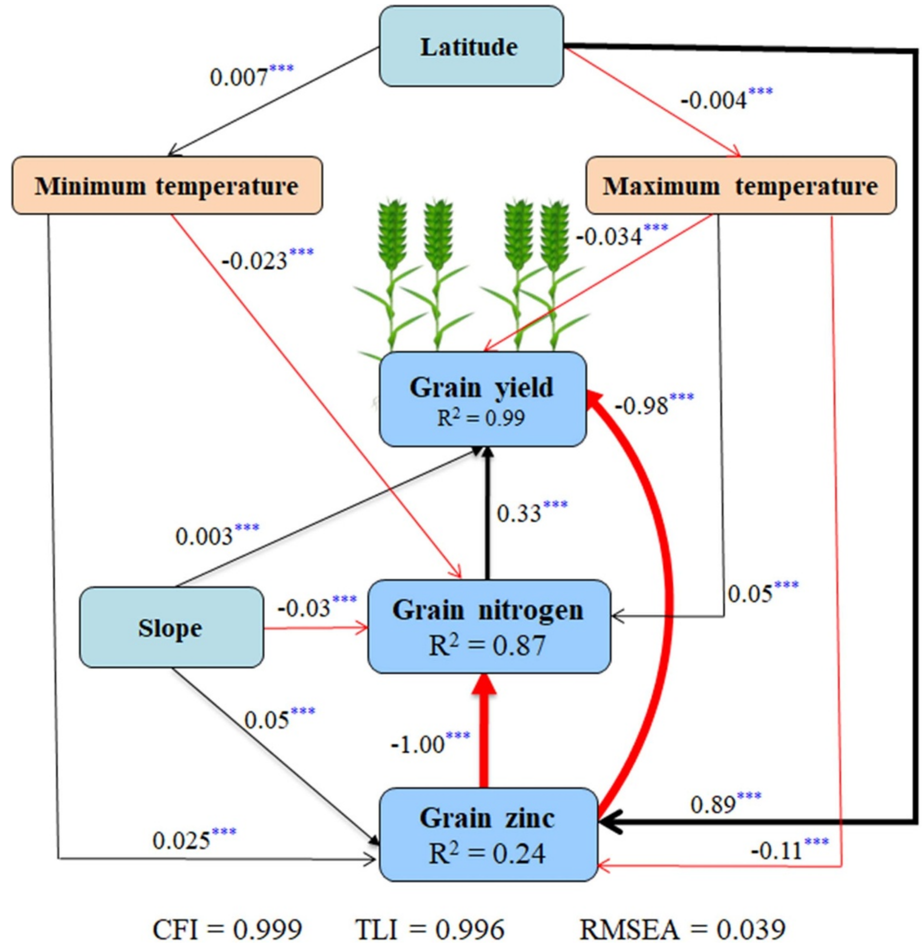

Figure 5. Structural Equation Model (SEM) depicting the multiple relations of latitude, maximum temperature, minimum temperature, slope, grain yield, grain nitrogen concentration and grain Zn concentration. The black and red lines are the positive and negative relationships, respectively. Arrows represent a directional influence of one variable on another. The numbers beside the arrows are standardized path coefficients. The thickness of the arrows is proportional to the magnitude of the standardized path coefficients. R^2 conditional (R^2 C) and R^2 marginal (R^2 M) indicate the combination from the random and fixed effect, and the amount of variation of the variable interpreted by all paths from the single fixed effects, respectively. The significance level was set at $\alpha = 0.05$, based on two-tailed tests. Data averaged over all GCMs, SSPs and Years for historical and future time periods.

Iron and zinc concentrations declined by 4.8% and 4.5% respectively under SSP45 and SSP85 (figure 7) with less uncertainty between GCMs under historical and future decades (figures 8–10). The overall uncertainty in yield, and concentrations of protein, Fe and Zn during the historical period (figure 8), was higher than that in the future under SSP85 (figure 10) scenario, passing by that in the SSP45 scenario (figure 9). During the historical decades, uncertainty ranged (15.6%–30.5%), (5.9%–6.7%), (15.8%–26.3%), and (12.5%–18.6%) for GY, protein concentration, Fe concentration, and Zn concentration respectively (figure 8). Such values decreased to (5.1%–11.5%), (3.7%–4.9%), (8.8%–15.6%), and (7%–11%) under SSP45 (figure 9), while SSP85 showed moderate uncertainty between historical and future SSP45 (figure 10). The spatial change maps (S. figure 18) reveal a pronounced reduction in GY, protein, and micronutrient concentrations (Fe and Zn) across key agricultural areas, particularly under the SSP5-85 scenario. Under both SSP2-45 and SSP5-85 scenarios, there is a consistent spatial pattern of decline in crop nutritional quality, with significant hot spots of nutrient loss concentrated in the southern and central regions, underscoring the potential impact of future CC on food security.

CC has a more variable impact on grain protein concentration, which is affected by both grain and protein yield (figure 11(a)). Impact of CC on the concentrations of Fe and Zn (figures 11(b) and (c)) was less variable than protein concentration. Distribution densities confirmed the positive impacts of CC on yield (figure 11(d)), and protein concentration (figure 11(f)), and the negative impacts on Fe concentration (figure 11(e)) and Zn concentration (figure 11(g)).

4. Discussion

At the center of the global agenda for food and nutrition security is the elimination of both overt and covert hunger (Wang *et al* 2023). However, the overall

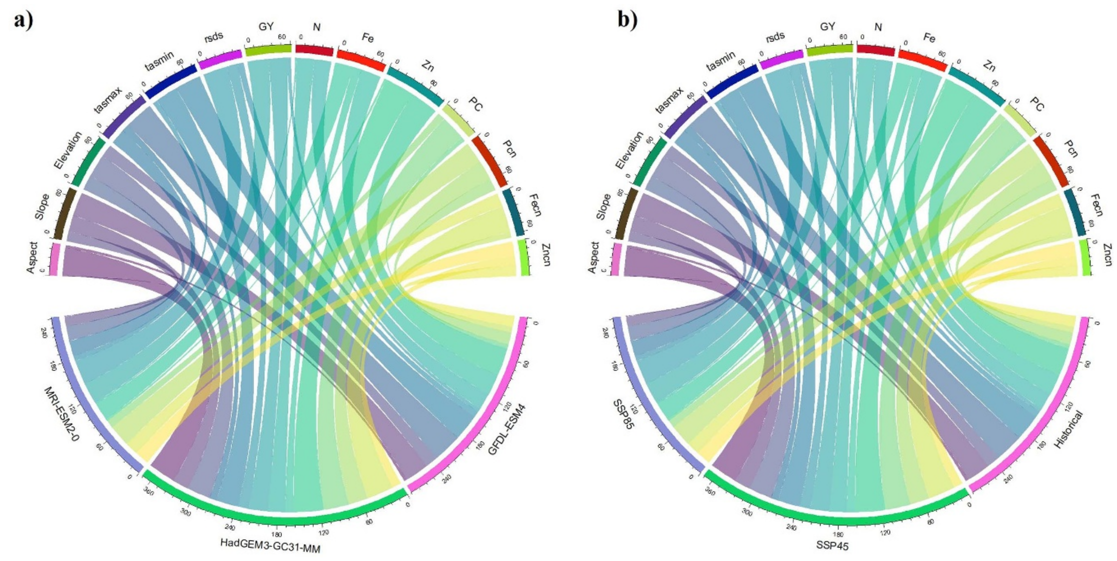

Figure 6. Circular chord diagram illustrating the impacts of climate change on wheat yield and nutritional content: This diagram visualizes the relationships between climate change variables and scenarios and wheat yield, protein yield, protein concentration, iron (Fe) concentration, and zinc (Zn) concentration. Panel (a) represents the GCMs, and panel (b) represents the SSPs scenarios. The arcs around the perimeter represent different variables, including climate variables (average maximum temperature [tasmax], minimum temperature [tasmin], solar radiation [rsds]), agronomic parameters (grain yield [GY], nitrogen [N]), and nutritional content (protein yield [PC], protein concentration [Pcn], Fe concentration [Fecn], Zn concentration [Zncn]). The color of each arc corresponds to a specific variable for easy identification. The bands connecting the arcs illustrate the relationships between these variables. The thickness and color intensity of the bands indicate the strength of these relationships. Positive correlations are shown by bands connecting climate variables to GY, PC, and Pcn, whereas negative correlations are shown by bands connecting climate variables to Fecn and Zncn. The diagram clearly shows that increased tasmax, tasmin, and rsds (climate variables) are positively correlated with higher GY (wheat yield), PC (protein yield), and Pcn (protein concentration). This is indicated by the strong, brightly colored bands connecting these variables. Conversely, the negative impacts on Fe and Zn concentrations are highlighted by the bands connecting tasmax, tasmin, and rsds to Fecn and Zncn. These bands indicate a reduction in these micronutrients as a result of increased temperature and solar radiation.

security of nourishment under CC remains uncertain. Few studies have investigated the nutritional value of wheat crops (Asseng *et al* 2019), focusing mainly on wheat protein and overlooking essential elements like Fe and Zn. This study emphasizes the impact of recent CMIP6 climate scenarios on wheat yield, protein, Fe, and Zn in Egypt through the mid-century. Field trials validated prediction tools by demonstrating the interrelationships between yield, yield attributes, phenology, and nutritional values. Foliar application of Fe and Zn improved wheat yield and quality, as Zn and Fe are vital for regulating cellular processes in plants (Zou *et al* 2012, Cakmak and Kutman 2018). To explore yield and nutritional values spatially, we supplemented experimental datasets with national statistics from 2015 to 2019.

Further, regression analysis was used to develop functions of predicting Fe and Zn from N in the other locations to be consistent with the yield dataset. In our initial analysis, we assumed a linear relationship between nitrogen (N) and zinc (Zn)/iron (Fe) concentrations, based on preliminary empirical evidence and existing literature (Caliskan *et al* 2008, Singh *et al* 2018) suggesting a correlation under controlled experimental conditions. To investigate the validity of this assumption further, we conducted additional analyses exploring potential non-linear relationships. Specifically, we employed polynomial regression models and compared their performance against the linear models by calculating and analyzing the respective residuals (supplementary figures 8 and 9). This approach enabled us to visualize residuals, deviations and inspect the differences between the models. Our findings revealed that while the polynomial regression models slightly improved predictive accuracy for certain datasets, the overall trends remained consistent with those captured by the linear models. The linear models provided a satisfactory fit under the controlled conditions of our experiments, capturing the essential dynamics without significant loss of generality. However, we acknowledge the potential limitations of assuming a linear relationship. This assumption may not reliably generalize across different spatial and temporal dimensions, especially in future scenarios. Therefore, we emphasize the importance of validating these relationships with more diverse datasets encompassing various regions and time periods as well as experiments under elevated CO_2. Such efforts will be crucial for confirming the robustness and applicability of our findings in broader contexts. While iron and zinc content share a strong linear relationship with protein content, our use of ML aimed to explore whether CC introduces non-linear effects or complex interactions that could influence nutrient content beyond protein levels alone. By modeling

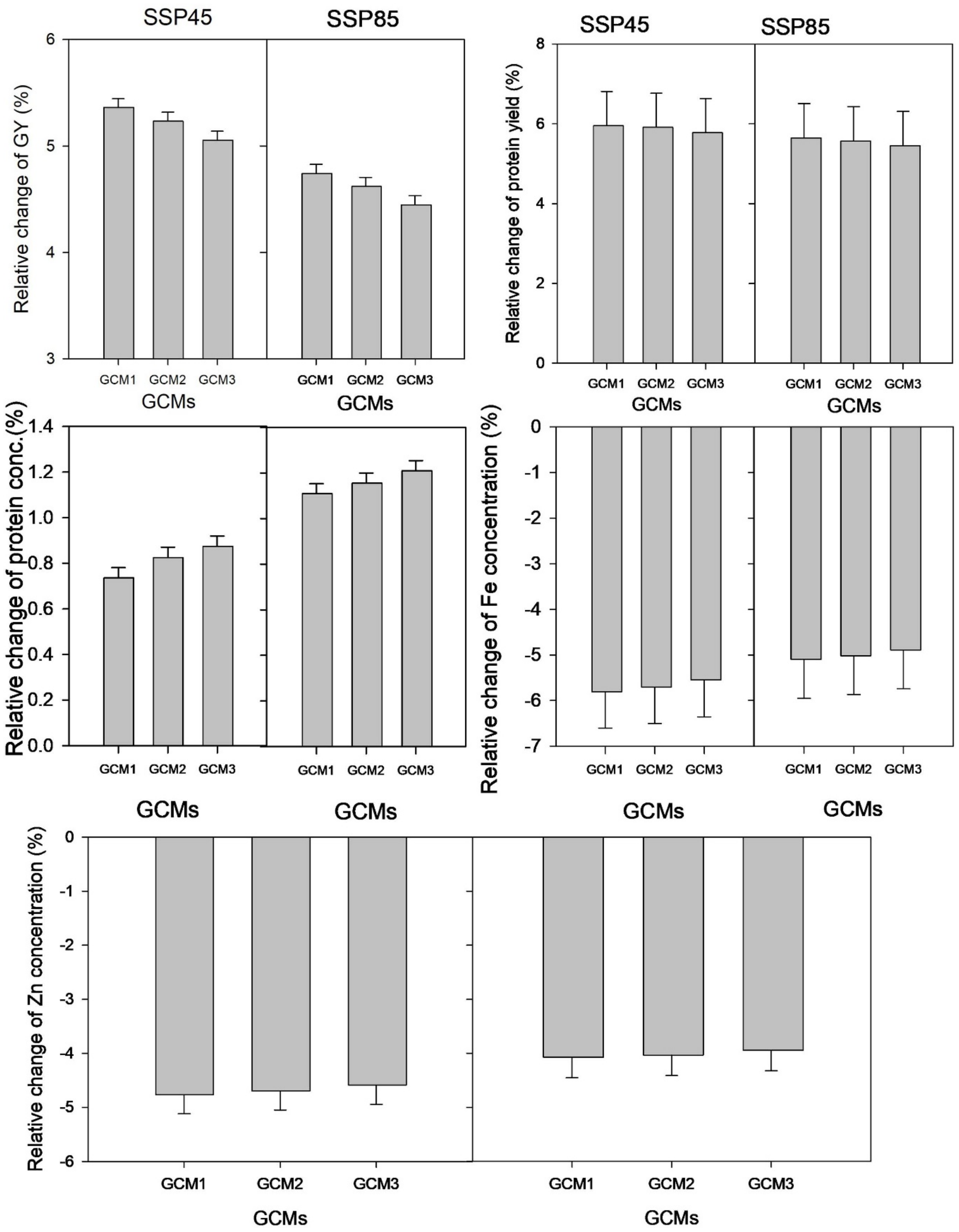

Figure 7. Changes of wheat grain yield, protein yield, protein concentration, iron concentration and zinc concentration in the mid century (2020–2050) relative to historical period under different GCMs as GFDL-ESM4 (GCM1), HadGEM3-GC31-MM (GCM2), and MRI-ESM2-0 (GCM3) and SSPs (SSP2-45, left and SSP5-85, right). Predictions were done using the stacked ensemble machine learning algorithm (best of family).

nutrient content directly, we aimed to capture these potential variations, providing a more comprehensive understanding of climate impacts on crop nutrition. This approach ensures that subtle but significant climate-induced changes in nutrient uptake are not overlooked.

CC impacts could be studied using different tools including crop models (Asseng *et al* 2020, Fan 2023) which showed robust predictions with less uncertainty. However, such studies focused only on crop production and protein, neglecting other important nutritional values such as Fe and Zn. The main reason for such limitation is that all crop models including only nitrogen factor, while Fe and Zn are absent in the model sub-routine. Furthermore, developing Fe and Zn models needs a global, high-quality dataset from diverse environments and cultivars, a time-consuming and costly process requiring global

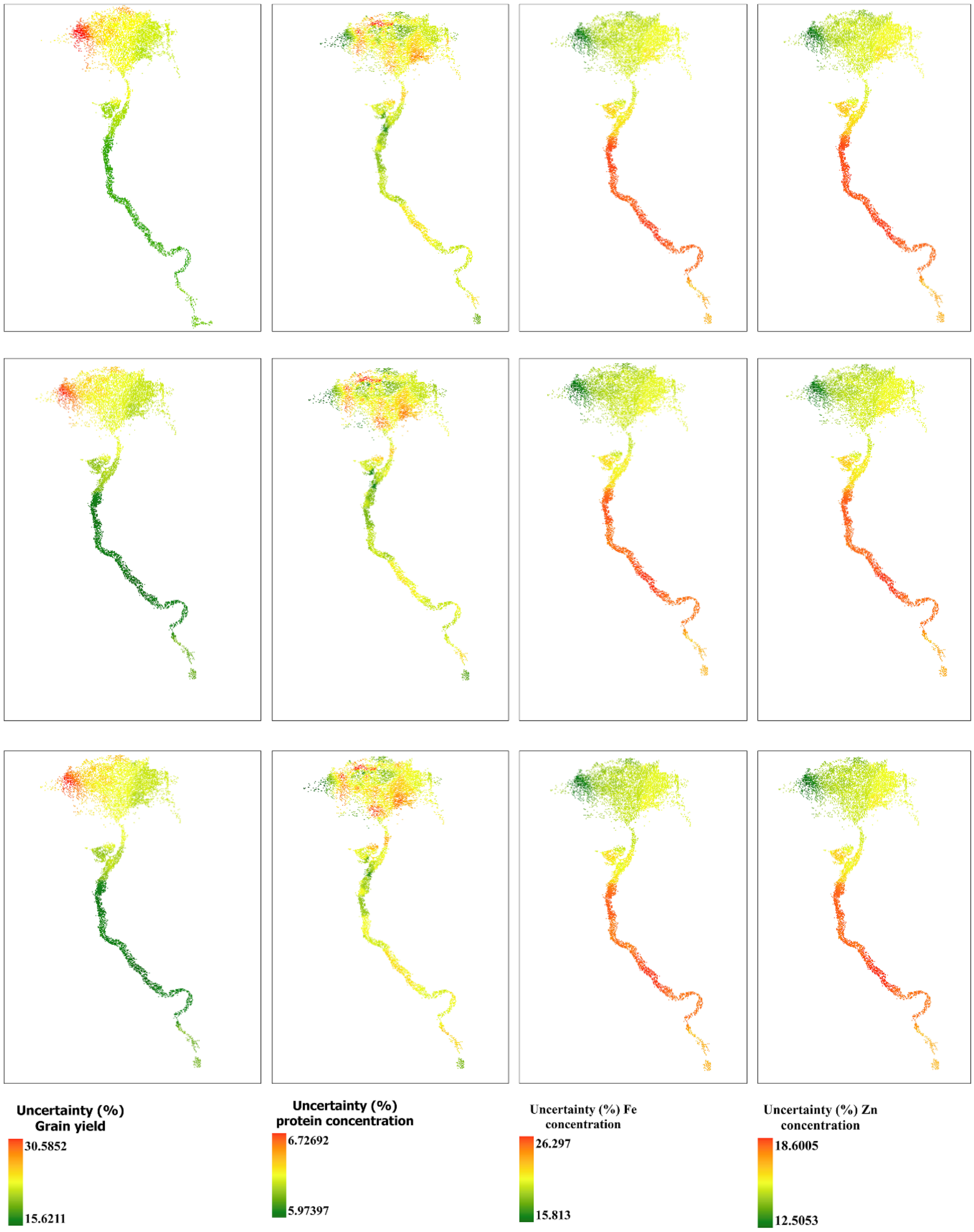

Figure 8. Uncertainty of predicting grain yield, protein concentration, Fe concentration, and Zn concentration during the historical period (1980–2010) under 3 GCMs, GFDL-ESM4 (upper row), HadGEM3-GC31-MM (middle row), and MRI-ESM2-0 (lower row).

funding. We bridged this gap by incorporating ML technologies that might be successfully employed for similar goals while saving time and money. Here, the trained stacked ensemble of ML from the hybrid approach demonstrated robust prediction of GY protein content, Fe content and Zn content. ML has recently been employed to predict crop yields under CC consequences (Grell *et al* 2021, Tsai *et al* 2021). Nonetheless, expanding the application of ML on investigating CC impacts on crop nutritional values has received less attention thus far.

Another merit of this work is employing the most recent climatic scenarios (CMIP6) in Egypt, rather than using CMIP5 previously (Ali *et al* 2020, Asseng *et al* 2018). CMIP6 represents a more recent set of simulations, covering a broader time have higher spatial resolution, providing more detailed information about regional climate patterns. Thus,

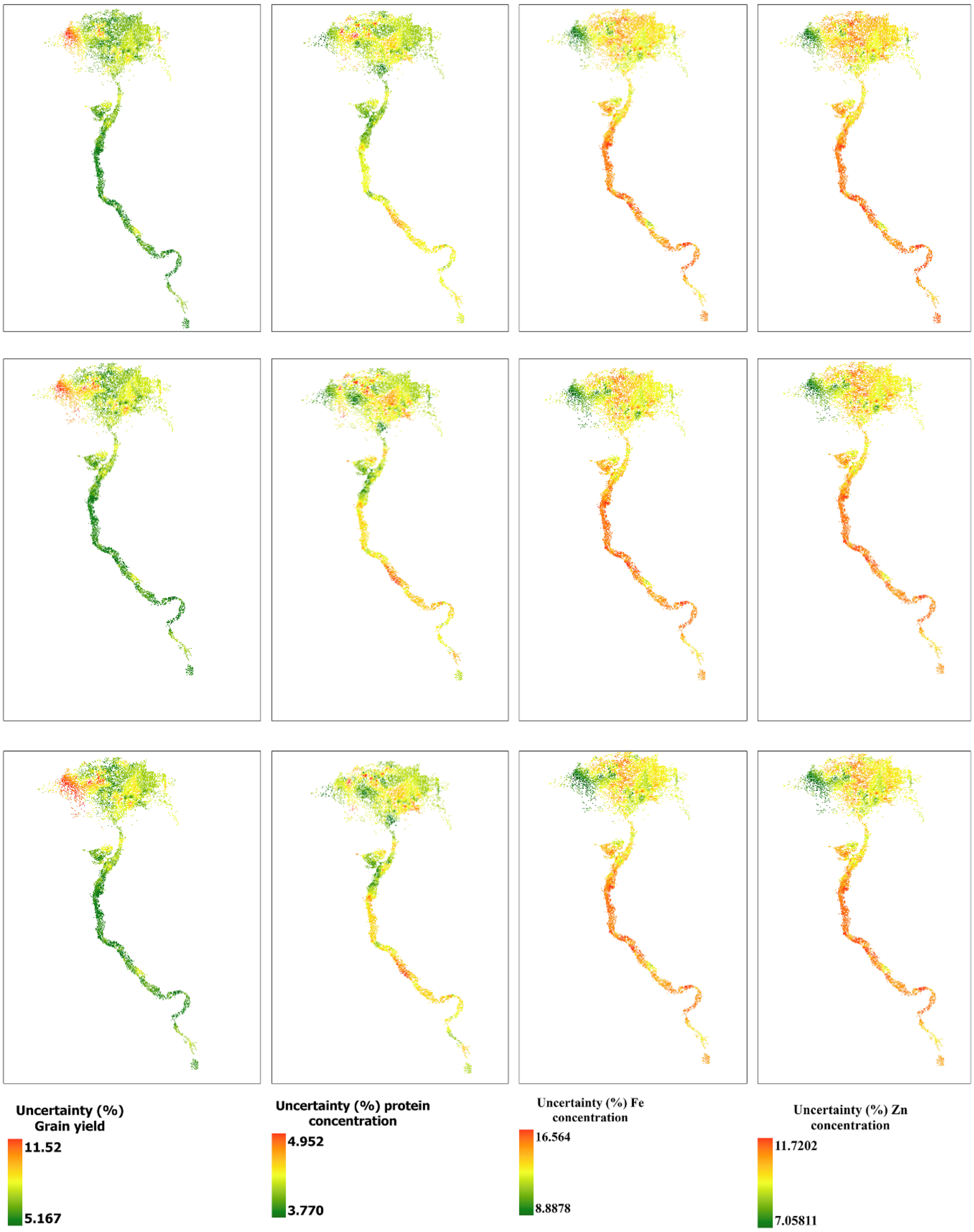

Figure 9. Uncertainty of predicting grain yield, protein concentration, Fe concentration, and Zn concentration during the mid-century decades (2020–2050) under 3 GCMs, GFDL-ESM4 (upper row), HadGEM3-GC31-MM (middle row), and MRI-ESM2-0 (lower row) and SSP2-45.

CMIP6 builds on CMIP5 with improved models, higher resolution, and a more integrated framework. These improvements aim to enhance our understanding of future CCs and their impacts (Tebaldi *et al* 2021). In Egypt, where spring-irrigated wheat is prevalent, prior CC studies based on CMIP5 scenarios indicated favorable benefits on wheat yield in the mid-century (2050) relative to baseline (Asseng *et al* 2018). The same trend has been observed here with CMIP6 scenarios, demonstrating an increase of wheat yield in response to CC. This aligns with global studies showing CC benefits for crop production, especially in irrigated and high-latitude regions (Rosenzweig *et al* 2014). Grain protein, Fe, and Zn concentrations, the ratio of grain quantities to GY, are significant characteristics influencing nutritional quality, although their behavior under CC remains unknown. Here, ML approach estimated an increase in GY and protein concentration. The increase in protein concentration under CC may be attributed

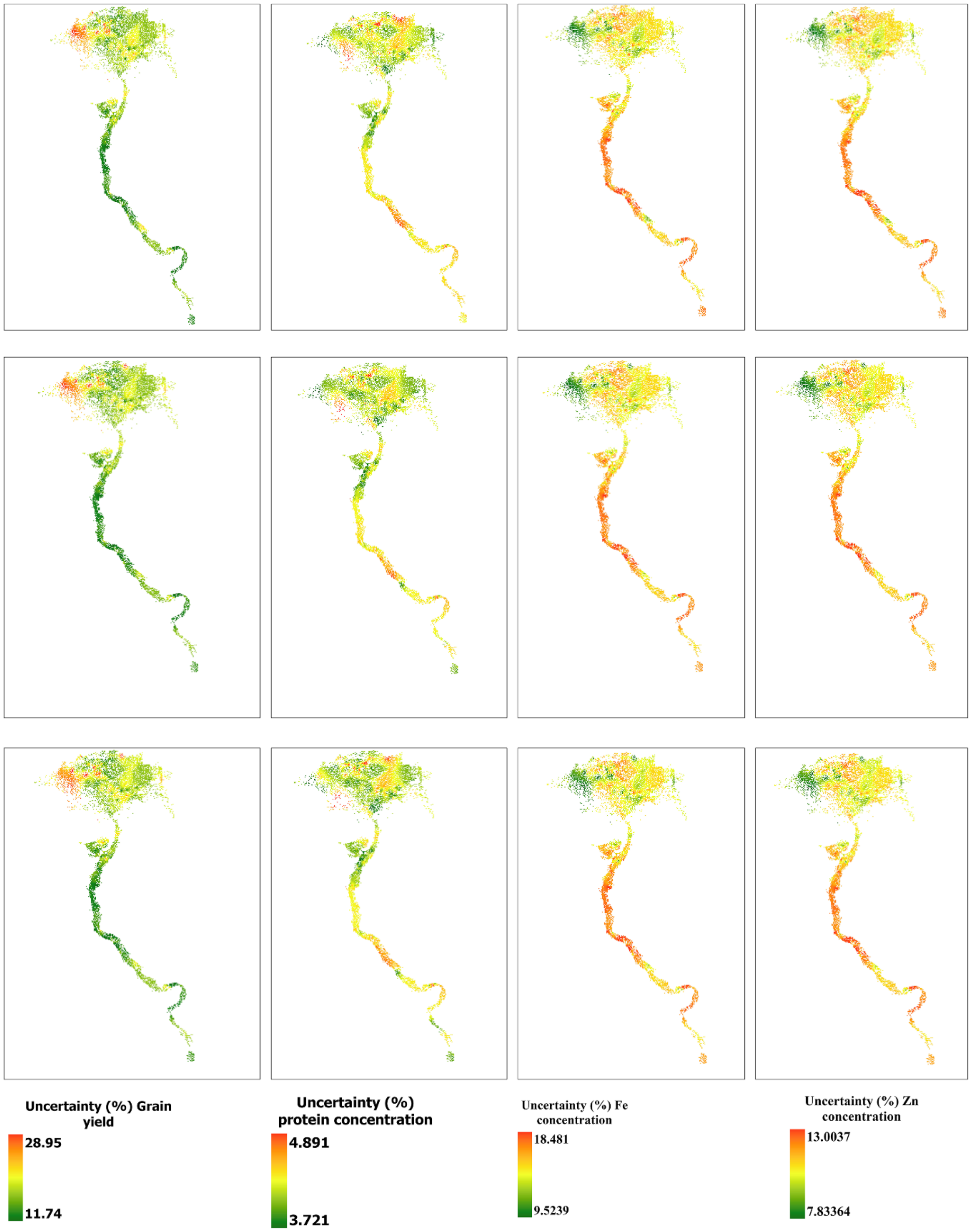

Figure 10. Uncertainty of predicting grain yield, protein concentration, Fe concentration, and Zn concentration during the mid-century decades (2020–2050) under 3 GCMs, GFDL-ESM4 (upper row), HadGEM3-GC31-MM (middle row), and MRI-ESM2-0 (lower row) and SSP5-85.

to reduced starch accumulation under rising temperatures (Triboi and Triboi-Blondel 2002). Other factors that influence grain protein concentrations include crop genotype, soil, crop management, atmospheric CO_2 concentration, and weather conditions. Accordingly, grain protein concentrations differ in spring wheat than winter wheat and in irrigated than rainfed wheat. In the present study, a spring irrigated wheat with unlimited N fertilization, consider a virtual adaptation which enhanced grain protein concentration under CC. Unlike, protein concentration which increased with CC, iron and zinc concentrations decreased with CC. A diluting effect, generated by a greater rise in grain production than in grain nutrient accumulation, might result in a drop in Fe and Zn concentrations (Oury *et al* 2006, Morgounov *et al* 2007, Fan *et al* 2008). The observed increase in wheat yields and the concurrent decrease

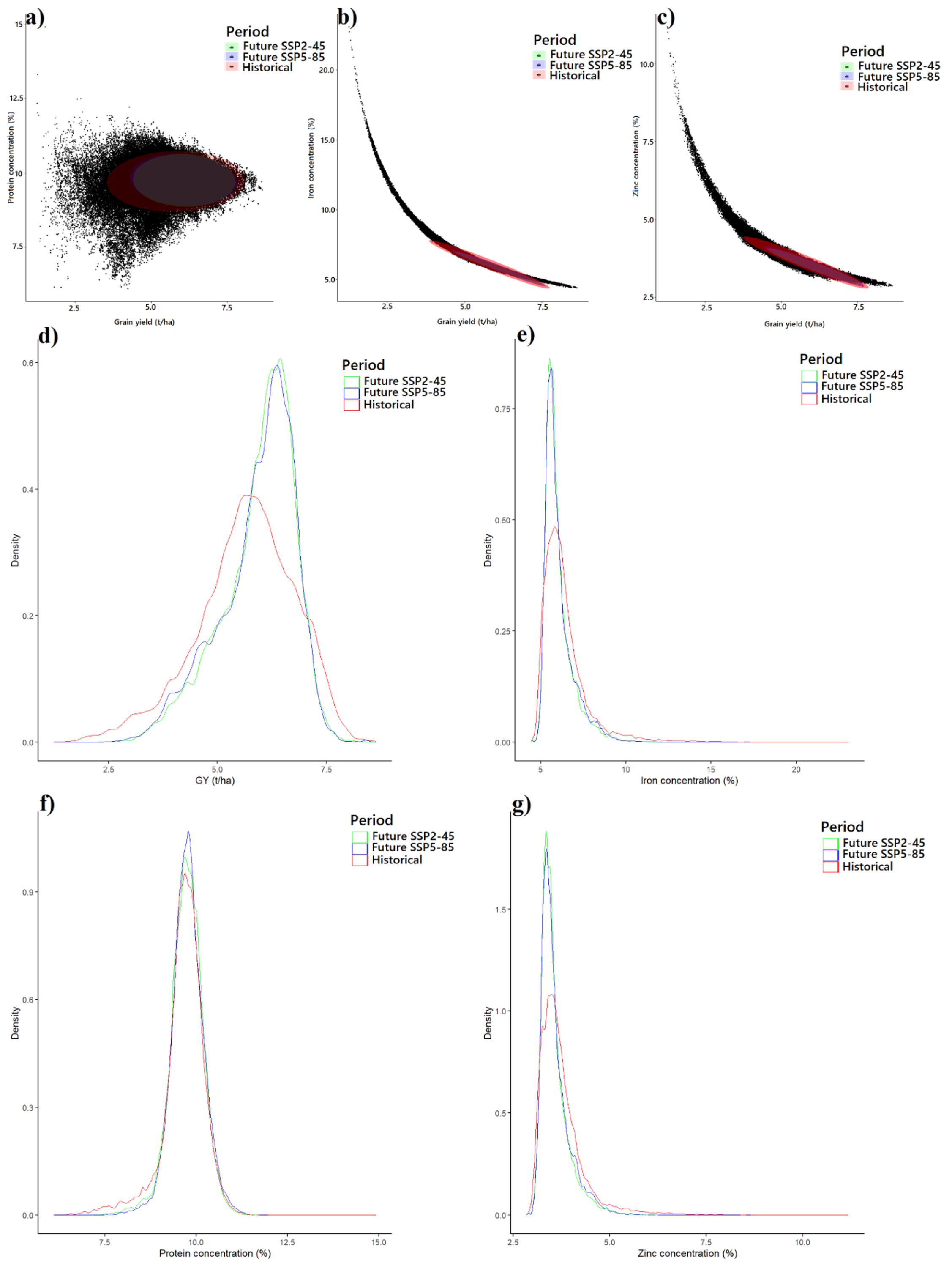

Figure 11. Impact of climate change on the relationship between grain yield and protein Concentration (a), grain yield and iron concentration (b), grain yield and zinc concentration (c). Projections of annual wheat grain yield and grain protein concentration are shown for baseline period 1980–2010 (light red), for SSP45 climate change impact in 2020–2050 (green) and for SSP85 climate change impact in 2020–2050 (pink). Medians across GCMs using the stacked ML model were plotted. The ellipses capture 95% confidence levels of data in each scenario or time. Distributions of values for grain yield (d), iron concentration (e), protein concentration (f), and zinc concentration (g).

in Fe and Zn concentrations under CC can be attributed to several physiological and ecological mechanisms. Elevated temperatures accelerate wheat growth, shortening the grain-filling period and reducing the time for nutrient accumulation in grains (Zahra *et al* 2023). Higher CO_2 levels enhance photosynthesis, carbon assimilation and biomass production, potentially increasing GYs but at the same time diluting micronutrient concentrations due to the 'CO_2 fertilization effect (Gojon *et al* 2023).' Soil microbial activity, essential for nutrient cycling, may also impacted by CC, altering Fe and Zn availability. Faster organic matter decomposition rates in warmer conditions may initially increase nutrient release but could lead to long-term depletion (Robinson *et al* 2022). Moreover, nutrient interactions, such as high nitrogen levels, can enhance growth while diluting micronutrient concentrations. To address these challenges, future research should focus on breeding nutrient-efficient wheat varieties, optimizing fertilization strategies, and improving soil management to ensure micronutrient availability. These strategies can help mitigate the negative impacts of CC on wheat nutritional quality. The ML uncertainty varied across locations, while the GCM uncertainty showed less spatial variation, due to the higher heterogeneity between locations in soil properties and temperature gradient.

The study highlights the impact of CC on key nutritional elements using ML but notes some limitations for future consideration. While increased atmospheric CO_2 can enhance photosynthesis and plant growth, its effect on wheat's nutritional value, especially Fe and Zn concentrations, remains uncertain and needs further investigation. CC increased wheat and protein yields but reduced Fe and Zn levels, requiring adaptation efforts like integrating Genotype × Environment × Management interactions. Combining process-based models with ML approaches could improve understanding of Fe and Zn dynamics in soil and plants, though this integration needs development to include these elements as currently done with N.

5. Conclusion

In the current study, coupling high quality observations with national dataset helped to develop different functions for wheat yield and nutritional values, developing a hybrid ML approach for future predictions. The hybrid ML approached was trained and tested using diverse dataset, showed robust prediction of wheat yield, protein, Fe and Zn concentrations under different CC scenarios. Although CC increased wheat and protein yields, it reduced Fe and Zn concentrations, putting further pressures on food security and nutrition. Further adaptations are required to enhance wheat nutritional values in a country suffering from rapid population growth and increasing food demand.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://zenodo.org/doi/10.5281/zenodo.11304509

Acknowledgments

We would like to thank Prof Senthold Asseng for his critical revision and comments on the first draft of the manuscript. We also thank Dr Ajit Govind for his review of the first draft of the manuscript.

ORCID iDs

Ahmed M S Kheir https://orcid.org/0000-0001-9569-5420
Osama A M Ali https://orcid.org/0000-0002-7121-4220
Ashifur Rahman Shawon https://orcid.org/0000-0002-0276-7497
Ahmed S Elrys https://orcid.org/0000-0003-0373-888X
Ayman Farid Abou-Hadid https://orcid.org/0000-0003-0548-7447
Rogerio de S Nóia Júnior https://orcid.org/0000-0002-4096-7588
Til Feike https://orcid.org/0000-0002-1978-9473

References

Abbas G, Ahmed M, Fatima Z, Hussain S, Kheir A M S, Erciṣli S and Ahmad S 2023 Modeling the potential impact of climate change on maize-maize cropping system in semi-arid environment and designing of adaptation options *Agric. For. Meteorol.* **341** 109674

Ali M G M, Ibrahim M M, El Baroudy A, Fullen M, Omar E-S H, Ding Z and Kheir A M S 2020 Climate change impact and adaptation on wheat yield, water use and water use efficiency at North Nile Delta *Front. Earth Sci.* **14** 522–36

Asseng S *et al* 2013 Uncertainty in simulating wheat yields under climate change *Nat. Clim. Change* **3** 827–32

Asseng S *et al* 2019 Climate change impact and adaptation for wheat protein *Glob. Change Biol.* **25** 155–73

Asseng S, Guarin J R, Raman M, Monje O, Kiss G, Despommier D D, Meggers F M and Gauthier P P G 2020 Wheat yield potential in controlled-environment vertical farms *Proc. Natl Acad. Sci.* **117** 19131–5

Asseng S, Kheir A M S, Kassie B T, Hoogenboom G, Abdelaal A I N, Haman D Z and Ruane A C 2018 Can Egypt become self-sufficient in wheat? *Environ. Res. Lett.* **13** 094012

Attia A, Govind A, Qureshi A S, Feike T, Rizk M S, Shabana M M A and Kheir A M S 2022 Coupling process-based models and machine learning algorithms for predicting yield and evapotranspiration of maize in arid environments *Water* **14** 3647

Cakmak I and Kutman U Á 2018 Agronomic biofortification of cereals with zinc: a review *Eur. J. Soil Sci.* **69** 172–80

Caliskan S, Ozkaya I, Caliskan M E and Arslan M 2008 The effects of nitrogen and iron fertilization on growth, yield and fertilizer use efficiency of soybean in a Mediterranean-type soil *Field Crops Res.* **108** 126–32

Chen C, Ota N, Wang B, FU G and Fletcher A 2023 Adaptation to climate change through strategic integration of long fallow into cropping system in a dryland Mediterranean-type environment *Scil Total Environ.* **880** 163230

Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58

Fan M-S, Zhao F-J, Fairweather-Tait S J, Poulton P R, Dunham S J and Mcgrath S P 2008 Evidence of decreasing mineral density in wheat grain over the last 160 years *J. Trace Elem. Med. Biol.* **22** 315–24

Fan Y 2023 Unequal effects of climate intervention on agriculture *Nat. Food* **4** 835–6

Gao Y, Dong K and Yue Y 2024 Projecting global fertilizer consumption under shared socioeconomic pathway (SSP)

scenarios using an approach of ensemble machine learning *Scil Total Environ.* **912** 169130
Godfray H C J, Beddington J R, Crute I R, Haddad L, Lawrence D, Muir J F, Pretty J, Robinson S, Thomas S M and Toulmin C 2010 Food security: the challenge of feeding 9 billion people *Science* **327** 812–8
Gojon A, Cassan O, Bach L, Lejay L and Martin A 2023 The decline of plant mineral nutrition under rising CO_2: physiological and molecular aspects of a bad deal *Trends Plant Sci.* **28** 185–98
Gould J 2017 Nutrition: a world of insecurity *Nature* **544** S6–S7
Grell M, Barandun G, Asfour T, Kasimatis M, Collins A S P, Wang J and Güder F 2021 Point-of-use sensors and machine learning enable low-cost determination of soil nitrogen *Nat. Food* **2** 981–9
Haddad L, Hawkes C, Webb P, Thomas S, Beddington J, Waage J and Flynn D 2016 A new global research agenda for food *Nature* **540** 30–32
Hailegnaw N S, Bayabil H K, Berihun M L, Teshome F T, Shelia V and Getachew F 2024 Integrating machine learning and empirical evapotranspiration modeling with DSSAT: implications for agricultural water management *Sci. Total Environ.* **912** 169403
Hersbach H *et al* 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049
Kheir A M S, Ammar K A, Amer A, Ali M G M, Ding Z and Elnashar A 2022 Machine learning-based cloud computing improved wheat yield simulation in arid regions *Comput. Electron. Agric.* **203** 107457
Kheir A M S, Elnashar A, Mosad A and Govind A 2023a An improved deep learning procedure for statistical downscaling of climate data *Heliyon* **9** e18200
Kheir A M S, Govind A, Nangia V, Devkota M, Elnashar A, Omar M E D and Feike T 2024 Developing automated machine learning approach for fast and robust crop yield prediction using a fusion of remote sensing, soil, and weather dataset *Environ. Res. Commun.* **6** 041005
Kheir A M S, Mkuhlani S, Mugo J W, Elnashar A, Nangia V, Devare M and Govind A 2023b Integrating APSIM model with machine learning to predict wheat yield spatial distribution *Agron. J.* **115** 3188–96
Kheir A, M S, El Baroudy A, Aiad M A, Zoghdan M G, Abd El-Aziz M A, Ali M G M and Fullen M A 2019 Impacts of rising temperature, carbon dioxide concentration and sea level on wheat production in North Nile delta *Sci. Total Environ.* **651** 3161–73
Ledell E and Poirier S 2020 H_2O AutoML: scalable automatic machine learning *7th ICML Workshop on Automated Machine Learning*
Li N, Li Y, Biswas A, Wang J, Dong H, Chen J, Liu C and Fan X 2021 Impact of climate change and crop management on cotton phenology based on statistical analysis in the main-cotton-planting areas of China *J. Clean. Prod.* **298** 126750
Liu Y and Dai L 2020 Modelling the impacts of climate change and crop management measures on soybean phenology in China *J. Clean. Prod.* **262** 121271
Ma P, Li A, Yu N, Li Y, Bahadur R, Wang Q and Ahuja J K 2021 Application of machine learning for estimating label nutrients using USDA Global Branded Food Products Database, (BFPD) *J. Food Compos. Anal.* **100** 103857
Ministry of Agriculture and Land Reclamation 2021 *Agricultural Bulletin Part 1. Reclamation* ed MOAAL (The Egyptian Economic Affairs Sector)
Morgounov A, Gómez-Becerra H F, Abugalieva A, Dzhunusova M, Yessimbekova M, Muminjanov H, Zelenskiy Y, Ozturk L and Cakmak I 2007 Iron and zinc grain density in common wheat grown in Central Asia *Euphytica* **155** 193–203
Nelson G C *et al* 2014 Climate change effects on agriculture: economic responses to biophysical shocks *Proc. Natl Acad. Sci. USA* **111** 3274–9
Nelson G *et al* 2018 Income growth and climate change effects on global nutrition security to mid-century *Nat. Sustain.* **1** 773–81
Oury F X, Leenhardt F, Rémésy C, Chanliaud E, Duperrier B, Balfourier F and Charmet G 2006 Genetic variability and stability of grain magnesium, zinc and iron concentrations in bread wheat *Eur. J. Agron.* **25** 177–85
Paryani S, Neshat A, Pourghasemi H R, Ntona M M and Kazakis N 2022 A novel hybrid of support vector regression and metaheuristic algorithms for groundwater spring potential mapping *Scil Total Environ.* **807** 151055
Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
Prodhan F A, Zhang J, Pangali Sharma T P, Nanzad L, Zhang D, Seka A M, Ahmed N, Hasan S S, Hoque M Z and Mohana H P 2022 Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach *Scil Total Environ.* **807** 151029
Robinson S I, O'gorman E J, Frey B, Hagner M and Mikola J 2022 Soil organic matter, rather than temperature, determines the structure and functioning of subarctic decomposer communities *Glob. Change Biol.* **28** 3929–43
Roosta H R, Estaji A and Niknam F 2018 Effect of iron, zinc and manganese shortage-induced change on photosynthetic pigments, some osmoregulators and chlorophyll fluorescence parameters in lettuce *Photosynthetica* **56** 606–15
Rosenzweig C *et al* 2014 Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison *Proc. Natl Acad. Sci.* **111** 3268–73
Shewry P R and Halford N G 2002 Cereal seed storage proteins: structures, properties and role in grain utilization *J. Experim. Botany* **53** 947–58
Singh B R, Timsina Y N, Lind O C, Cagno S and Janssens K 2018 Zinc and iron concentration as affected by nitrogen fertilization and their localization in wheat grain *Front. Plant Sci.* **9** 307
Springmann M *et al* 2018 Options for keeping the food system within environmental limits *Nature* **562** 519–25
Tan L *et al* 2023 Assessment of the sustainability of groundwater utilization and crop production under optimized irrigation strategies in the North China Plain under future climate change *Scil Total Environ.* **899** 165619
Tebaldi C *et al* 2021 Climate model projections from the scenario model intercomparison project (ScenarioMIP) of CMIP6 *Earth Syst. Dyn.* **12** 253–93
Thrasher B, Wang W, Michaelis A, Melton F, Lee T and Nemani R 2022 NASA global daily downscaled projections, CMIP6 *Sci. Data* **9** 262
Tilman D, Balzer C, Hill J and Befort B L 2011 Global food demand and the sustainable intensification of agriculture *Proc. Natl Acad. Sci.* **108** 20260–4
Tilman D and Clark M 2014 Global diets link environmental sustainability and human health *Nature* **515** 518–22
Triboi E, Martre P, Girousse C, Ravel C and Triboi-Blondel A-M 2006 Unravelling environmental and genetic relationships between grain yield and nitrogen concentration for wheat *Eur. J. Agron.* **25** 108–18
Triboi E and Triboi-Blondel A-M 2002 Productivity and grain or seed composition: a new approach to an old problem—invited paper *Eur. J. Agron.* **16** 163–86
Tsai W-P, Feng D, Pan M, Beck H, Lawson K, Yang Y, Liu J and Shen C 2021 From calibration to parameter learning: harnessing the scaling effects of big data in geoscientific modeling *Nat. Commun.* **12** 5988
Udristioiu M T, El Mghouchi Y and Yildizhan H 2023 Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning *J. Clean. Prod.* **421** 138496
Van Dijk M, Morley T, Rau M L and Saghai Y 2021 A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050 *Nat. Food* **2** 494–501

van Klompenburg T, Kassahun A and Catal C 2020 Crop yield prediction using machine learning: a systematic literature review *Comput. Electron. Agric.* **177** 105709

Wang X *et al* 2023 Global food nutrients analysis reveals alarming gaps and daunting challenges *Nat. Food* **4** 1007–17

Waring J, Lindvall C and Umeton R 2020 Automated machine learning: review of the state-of-the-art and opportunities for healthcare *Artif. Intell. Med.* **104** 101822

Watts M, Mpanda M, Hemp A and Peh K S H 2023 The potential impact of future climate change on the production of a major food and cash crop in tropical (sub)montane homegardens *Sci. Total Environ.* **865** 161263

Willett W *et al* 2019 Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems *Lancet* **393** 447–92

Wood S A, Smith M R, Fanzo J, Remans R and Defries R S 2018 Trade and the equitability of global food nutrient distribution *Nat. Sustain.* **1** 34–37

Xu J-X, Ma J, Tang Y-N, Wu W-X, Shao J-H, Wu W-B, Wei S-Y, Liu Y-F, Wang Y-C and Guo H-Q 2020 Estimation of sugarcane yield using a machine learning approach based on uav-lidar data *Remote Sens.* **12** 2823

Xu R-Z, Cao J-S, Ye T, Wang S-N, Luo J-Y, Ni B-J and Fang F 2022 Automated machine learning-based prediction of microplastics induced impacts on methane production in anaerobic digestion *Water Res.* **223** 118975

Yan H *et al* 2022 Crop traits enabling yield gains under more frequent extreme climatic events *Sci. Total Environ.* **808** 152170

You H, Zhou M, Zhang J, Peng W and Sun C 2023 Sugarcane nitrogen nutrition estimation with digital images and machine learning methods *Sci. Rep.* **13** 14939

Zahra N, Hafeez M B, Wahid A, Al Masruri M H, Ullah A, Siddique K H and Farooq M 2023 Impact of climate change on wheat grain composition and quality *J. Sci. Food Agric.* **103** 2745–51

Zeng H, Elnashar A, Wu B, Zhang M, Zhu W, Tian F and Ma Z 2022 A framework for separating natural and anthropogenic contributions to evapotranspiration of human-managed land covers in watersheds based on machine learning *Sci. Total Environ.* **823** 153726

Zheng J, Wang W, Ding Y, Liu G, Xing W, Cao X and Chen D 2020 Assessment of climate change impact on the water footprint in rice production: historical simulation and future projections at two representative rice cropping sites of China *Sci. Total Environ.* **709** 136190

Zimmermann M B and Hurrell R F 2007 Nutritional iron deficiency *Lancet* **370** 511–20

Zou C *et al* 2012 Biofortification of wheat with zinc through zinc fertilization in seven countries *Plant Soil* **361** 119–30