

Maize response patterns to soil and climate factors form the basis for predicting changes in its growing conditions under climate change

Y. Nykytiuk, O. Kravchenko

Polissia National University, Zhytomyr, Ukraine

Article info

Received 01.08.2025

Received in revised form
18.07.2025

Accepted 02.09.2025

Polissia National
University, 7 Staryi
Boulevard, Zhytomyr,
10008, Ukraine.
Tel.: +380-67-448-38-48.
E-mail:
r.alimovih@gmail.com

Nykytiuk, Y., & Kravchenko, O. (2025). Maize response patterns to soil and climate factors form the basis for predicting changes in its growing conditions under climate change. *Biosystems Diversity*, 33(4), e2553. doi:10.15421/012553

The present study elucidates the mechanisms by which soil and climatic factors determine the suitability of the Polissia and Forest-Steppe regions in Ukraine for the cultivation of maize, and presents predictive models of how these conditions will shift under global climate change. Spatial modelling was performed using CROPGRIDS v1.08 data (maize sowing density), the WorldClim v2.1 database (19 bioclimatic indicators), and SoilGrids v2.0 (nine chemical and physical soil parameters at 5–15 cm depth). The screening of climate variables was conducted through the utilisation of Principal Component Analysis and residual orthogonalisation techniques. By contrast, soil variables underwent a process of normalisation and standardisation. The maize-area response was Box–Cox transformed, and four regression approaches were fitted: Ordinary Least Squares (OLS), Ridge Regression, Generalised Additive Models (GAM), and a Random-Forest ensemble (RF). The most significant factor was determined to be soil reaction (pH 6.0–7.5), which ensured optimal nutrient availability; values outside this range resulted in element fixation into insoluble forms or leaching. In the context of soil properties, the sand content (30–40 %) was found to regulate drainage and moisture, the silt content (20–35%) was determined to maintain the water-air balance, the organic carbon (up to ≈ 30 g/kg) was found to enhance suitability until saturation, and the total nitrogen exhibited a near-linear positive effect. The key climatic predictors included residual components of annual mean temperature, seasonality and diurnal amplitude, and precipitation volume (300–600 mm/yr, with optima in both the wettest and driest months). The GAMs captured nonlinear “peak–plateau–decline” responses for pH, texture, and rainfall, whereas RF delivered the highest predictive accuracy ($R^2 = 0.96$; RMSE = 14.77; MAE = 6.73) by automatically modelling complex interactions. Linear models (OLS and Ridge) explained 60–64% of the variance. Based on the best-performing models, suitability maps were generated for three future periods (2021–2040, 2041–2060, 2061–2080) under low (SSP1-2.6) to high (SSP5-8.5) emission scenarios. Results indicate a mid-term decline in optimal areas, followed by partial long-term recovery driven by compensatory climate dynamics and adaptive measures. The practical significance lies in identifying narrow optimum ranges for soil and climatic factors, enabling targeted agronomic recommendations: localised liming, texture adjustment, sowing-date optimisation, hybrid selection, and irrigation management. The resulting models and suitability maps provide a scientific basis for evidence-based planning of adaptive maize-production strategies in the face of global climate change.

Keywords: crop suitability modelling; edaphic limitations; temperature-precipitation interactions; spatial regression; agro-ecological zoning; Shared Socioeconomic Pathways; Eastern European agriculture; land-use adaptation.

Introduction

Forecasting global climate change plays a fundamental role in ensuring food security, as it allows for an early assessment of potential shifts in agro-ecosystems, growing seasons and yields of key crops (Benitez-Alfonso et al., 2023). The formulation of agricultural policy, planning process, and decision making are all attempts to operate in a highly uncertain environment in the absence of a comprehensive understanding of future changes in temperature, precipitation distribution, frequency of extreme weather events, and greenhouse gas emissions (Hultgren et al., 2025). Forecasts help to identify regions with an increased risk of droughts or floods, predict the necessary crop rotations, determine the need for irrigation systems, and select varieties and hybrids adapted to new climatic conditions (Zymarioieva et al., 2024). The repercussions of climate change on the agricultural sector are multifaceted, impacting yields directly and consequently the logistics of storage and transportation of products, due to changes in growing conditions resulting in uneven flowering and ripening of crops and fluctuations in the volume and quality of the harvest (Yuan et al., 2024). Knowledge of future changes makes it possible to develop early warning systems and adjust harvesting, processing, and storage plans to minimise losses (Feng et al., 2023). Climate forecasts help integrate agricultural production with other sectors, such as water, electricity and ecosystem services, into a single system (Zymarioieva et al., 2022). This system allocates resources more efficiently and takes future challenges into account (Mykhailyuk et al., 2023). Forecasting climate change provides a foundation for developing

long-term sustainable strategies that address food security's economic, social, and environmental aspects (Charalampopoulos & Droulia, 2024). Forecasting stimulates innovations in crop breeding, precision agriculture, and agroforestry, and contributes to the development of adaptation and mitigation policies at international and national levels (Stefanovska et al., 2025). Ensuring a stable food supply in the face of global warming becomes difficult without proper modelling and analysis of climate trends (Lee et al., 2024).

Maize is central to global food security due to its versatility and high productivity (Tanumihardjo et al., 2020). It is one of the three primary sources of calories for billions of people, meeting their nutritional needs as a direct product (cereals, flour, oil) and as a significant component of livestock feed rations, indirectly supporting the supply of meat, milk, and eggs (Shiferaw et al., 2011). Corn hybrids are characterised by fast growth and high yields, even in arid conditions, making them a sustainable resource in unstable climates (Kapustian et al., 2021). The integration of contemporary agricultural technologies and intensive breeding methodologies has enabled the adaptation of corn to diverse soil and climatic conditions, thereby enhancing its resistance to pests, diseases, and extreme weather events (Drobotko et al., 2024). Maize is a valuable crop due to its relatively low production cost and ease of processing (Pichura et al., 2023). Its integration into crop rotations, irrigation systems, and precision farming is advantageous, as it reduces losses and optimises the use of limited resources (Yakovenko & Zhukov, 2021). In addition to its utilisation as a food and feed source, corn is increasingly employed as a feedstock for biofuels. This development drives economic growth in rural regions

and provides supplementary incentives for farmers (Landis et al., 2008). The importance of maize in the global food system lies in its contribution to dietary calories and its multilevel value creation from food and feed to bioenergy and industrial feedstocks, making it an indispensable element in sustainable food supply strategies around the world (Erenstein et al., 2022).

Understanding the mechanisms by which soil and climate factors affect maize growth and development is key to adequately predicting the impacts of global climate change (Olesen et al., 2011). Knowledge of specific physical and chemical processes in the soil (such as how pH fluctuations affect nutrient availability or how soil structure determines water holding capacity and aeration) can help predict which scenarios of rising temperatures and changing precipitation patterns will make certain areas less or more suitable for the crop (Agnolucci & De Lipsis, 2019). Understanding the response of maize to temperature and humidity gradients (monotonic and unimodal responses) allows to assess which temporal and seasonal climate changes will lead to an extended or shortened growing season (Koshchev et al., 2021), shifts in flowering and maturation, and risks of heat or water stress (Zhang et al., 2022). Integrating these soil and climate mechanisms into a unified model can identify synergistic effects (Nykytiuk et al., 2025). For instance, it can aid in elucidating how the escalation in drought frequency is amplified in texture-unstable soils, or how predictions of increased precipitation intensity can compensate for potassium deficiency in acidic soils (Masereka et al., 2019). This multidimensional approach to mechanistic explanation enables the quantification of losses or gains of productivity in different regions (Zhukov et al., 2022). Consequently, it facilitates the development of targeted adaptive strategies, such as adjusting farming practices and fertilisers, or breeding hybrids that best meet expected conditions. A comprehensive comprehension of the mechanisms underlying soil and climatic factors constitutes the foundation for the future management of global warming challenges in maize production (Rehman et al., 2024).

Nonetheless, there is a paucity of knowledge of how the response patterns of maize growing conditions vary at the landscape level and within physiographic zones (Nykytiuk et al., 2025). However, these scales are critical for accurate prediction of the impact of global warming (Nykytiuk & Kravchenko, 2024). Most studies focus on either local experimental sites or general climate models, without considering the spatial heterogeneity of agroecosystems. This limitation restricts their value for regional planning. The objective of the present study was twofold: firstly, to study the responses of maize growing conditions to soil and climatic factors, taking into account the landscape and physical and geographical features of Polissia and the Forest-Steppe of Ukraine; and secondly, to use the results obtained to predict the impact of global climate change.

Materials and methods

The global geospatial dataset Cropgrids v1.08 (Tang et al., 2024) was used to model the spatial distribution of areas under crop, which contains maps of 173 crops with a spatial resolution of 0.05° (~5.5 km) in the WGS-84 coordinate system. The data are available in NetCDF format and are distributed under the CC BY 4.0 open licence. From the analysed file CROPGRIDSv1.08_sunflower.nc, the crop area layer was selected, reflecting the area occupied by the respective crop (as opposed to the harv area variable, which records the harvested yield). The choice of the crop area variable was justified by its greater stability as an indicator of land use, especially when interpreting long-term trends. After the layer was imported, it was cropped to the boundaries of the study area using the polygon layer Region_2_outer.shp, which set the spatial mask. Normalisation was applied to avoid the influence of single abnormally high values in the image: All values above the 95th percentile were replaced with the values of this quartile. This allowed us to reduce the effect of statistical outliers in further model building and increase the reliability of spatial estimates. The processed layer was brought to a standard format and coordinate system with the rest of the environmental predictors used in the modelling and, if necessary, saved in GeoTIFF format.

The bioclimatic variables were used from the global WorldClim v2.1 dataset, which contains 19 derived climate indicators (BIO1-BIO19) describing average, extreme, and seasonal aspects of temperature and precipitation to characterise climate conditions. The data have a high spatial resolution of 2.5 arcmin (approximately 4.5 km at the equator) and are available in the public domain on www.worldclim.org (Tutova et al., 2025). The bioclimatic data were downloaded and preprocessed automatically using the `worldclim_global()` function from the `geodata` package. A stack of rasters for all 19 variables was obtained, cropped to the boundaries of the study region, and brought to the spatial parameters of the crop area layer (`crop_mask`) by reprojection and resampling using bilinear interpolation methods.

The SoilGrids v2.0 global array (ISRIC – World Soil Information) was used to describe soil conditions, which provides digital soil maps with a spatial resolution of 250 m for several physicochemical parameters in standard depth intervals (0–5, 5–15, 15–30 cm, etc.). The study involved nine variables from the top 5–15 cm layer: `bd05` (bulk density), `cfvo` (coarse skeleton content), clay, sand, silt, `phh2o` (water pH), nitrogen, `soc` (organic C), and `ocd` (organic C density). The data were downloaded using the `soil_world()` function of the `geodata` package, which directly accesses the SoilGrids REST service and automatically saves the GeoTIFF image to a temporary directory. For each variable, the same procedure was followed: (i) cropping to the boundaries of the region's polygon, (ii) masking to its contour, (iii) reprojection to the crop-area-layer coordinate system (WGS-84 lon/lat) with bilinear interpolation, and (iv) resampling to the same grid as the other predictors. The downloaded and preprocessed soil layers were combined into a single `soil_hist` stack and then added to the climate and land variables to form a matrix of primary predictors.

To assess the spatial patterns of crop cultivation, we used a set of bioclimatic variables (BIO1-BIO19) from the WorldClim v2.1 global climate dataset with a spatial resolution of 2.5' (~5 km), which covers the average long-term values of climate parameters for 1970–2000. Given the high multicollinearity among bioclimatic variables, the principal component analysis (PCA) method was applied to reduce redundant information and eliminate internal correlations. Based on the study of the spectrum of eigenvalues, four principal components with $\lambda > 1$ were selected, which retained the most significant contribution to the variation in climatic conditions. From each of these components, one variable with the highest load (the maximum absolute value of the coefficient in the inverse loadings matrix) was extracted, which allowed us to interpret the leading climate gradients based on fundamental variables rather than abstract components. The residual climate variables not included in the main set were orthogonalised with respect to the selected leading variables by building auxiliary linear models and obtaining regression residuals. In this way, a new set of independent climate predictors was formed that did not share common variation with the main climate gradients. This approach allowed us to reduce collinearity in the input dataset and ensure the correct interpretation of the contribution of individual factors to the models, particularly when building regressions and machine learning algorithms.

The Box-Cox transformation was applied to normalise the response variable and crop area distribution and improve compliance with the statistical assumptions underlying linear models. In its original form, this variable had an asymmetric, right-handed skewed distribution with a large number of zero or close to zero values, which negatively affected the modelling quality and led to a violation of homoscedasticity and normality of the residuals. To make the Box-Cox transformation possible, all variable values were shifted to the positive region by adding ϵ , calculated as half of the minimum non-zero value. Next, the `boxcox()` function from the MASS package was applied to the shifted vector in the range of $\lambda \in [-10; 10]$, which allowed us to select the optimal parameter of the transformation empirically, λ . If $\lambda \approx 0$, a logarithmic transformation was used; in other cases, a power function was used according to the formula:

$$y_{\text{Box-Cox}} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

The resulting transformed value of the variable (*harv_bc*) demonstrated better symmetry and reduced residual variance, which ensured model stability, increased forecast accuracy, and the acceptability of further application of both linear and nonlinear machine learning models.

All numerical variables in the study were standardised to ensure the mathematical stability of the models and increase the interpretability of the results. As the predictors had different units of measurement (e.g., temperature in degrees Celsius, sand content in per cent, land use in hectares), standardisation was performed to bring them to a single scale, with a zero mean and standard deviation of 1. This eliminated the influence of variables with significant variance, which could dominate the modelling process without scaling, especially in scale-sensitive methods such as Ridge Regression, Random Forest, or Principal Component Analysis (PCA). Standardisation is also a prerequisite for regularisation methods, based on comparing coefficient values that must be brought to the same conditions. This procedure makes it possible to interpret variables' importance in the models objectively, compare their contribution, and avoid statistical distortions. Standardisation was performed using the classic z-transform formula: subtracting the mean and dividing by the standard deviation. All predictors were standardised similarly, except for the response variable, which was subjected to a separate Box-Cox transformation to normalise its distribution.

To address the multicollinearity problem, a stepwise filtering of variables based on their mutual correlation was applied when building a linear regression model using the least squares method (Kayode Ayinde & Nwosu, 2021). Since high multicollinearity ($|r| > 0.7$ or 0.8) can lead to instability of regression coefficient estimates, inflated standard errors, and difficulties in interpreting the impact of individual factors, a variable selection procedure was implemented to minimise this problem. At the first stage, a pairwise Pearson correlation matrix was calculated for all standardised predictors. Then, from each pair of highly correlated variables ($|r| \geq 0.7$), only one was left - the one with the highest correlation with the dependent variable (i.e., the area under crop cultivation after Box-Cox transformation). This approach allowed us to preserve as many informative predictors for the model as possible while avoiding duplication of information due to strong multicollinearity. As a result, a subset of predictors was formed that ensures independence and allows classical linear regression without the risk of distortions associated with excessive interdependence of variables.

The linear regression in this study was based on a standard Ordinary Least Squares (OLS) model using a transformed response variable (*harv_bc*) obtained by applying the Box-Cox transformation to the area under crop cultivation. Before building the model, all predictors were standardised to a mean of 0 and a standard deviation of 1, which ensures comparability of coefficients and correct interpretation of the weight of each variable's influence. To avoid multicollinearity, a subset of predictors was selected based on the results of clustering the variables by the structure of correlations. One representative was chosen from each cluster of mutually correlated predictors, thus reducing information redundancy in the model.

After that, the model formula was formulated in the form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon,$$

where Y is the transformed response variable (crop area); β_0 is the free term; x_1, x_2, \dots, x_k are the selected independent variables; $\beta_1, \beta_2, \dots, \beta_k$ are the regression coefficients reflecting the effect of each predictor; ε is the random error covering unmodelled factors.

The model was built using the `lm()` function in R, and its parameters were estimated by minimising the sum of squared residuals. The adequacy of the model was checked based on standard metrics: coefficient of determination R^2 , root mean square error (RMSE), and mean absolute error (MAE), as well as through the observed versus predicted graph, which allows for assessing the accuracy visually.

Ridge Regression was used in this study to build a stable linear model in the presence of a large number of potentially correlated predictors. Ridge Regression minimises the following loss function:

$$mi n_{\beta} \left[\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right]$$

where y_i is the value of the dependent variable (area under crop after Box-Cox), x_{ij} is the value of the j -th predictor in the i -th observation (all standardised), β_j are the model coefficients, and λ is the regularisation parameter controlling the strength of the penalty.

The model was estimated based on standardised predictor variables and the transformed response variable obtained after applying the Box-Cox transformation to the area under crop cultivation. In contrast to classical least squares regression, ridge regression includes a regularisation term that penalises overly large model coefficients, reducing the risk of overfitting and increasing the stability of the estimates in case of multicollinearity. The model was built in R using the `glmnet()` function from the `glmnet` package, which implements regularised regression. The predictor variables were transformed into a matrix, and the response variable into a vector. The alpha parameter was set to 0, corresponding to ridge regression (as opposed to Lasso Regression, where $\alpha = 1$). To select the optimal value of the regularisation parameter λ , which controls the strength of the penalty, a cross-validation was performed using the `cv.glmnet()` function. After that, the model was retrained using the optimal λ (`lambda.min`), and the values of the response variable were predicted on its basis. The quality of the built model was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), which allowed us to compare its performance with alternative approaches. Ridge regression proved more accurate and stable than classical linear regression, especially when the predictors were highly correlated.

To consider the potentially non-linear effect of environmental predictors on the transformed value of the crop area (after Box-Cox transformation), a generalised additive model (GAM) was built using the `mgcv` package in R. In this model, the response variable (*harv_bc*) is modelled as a sum of smooth functions from each predictor. Formally, the model has the following form:

$$Y = \beta_0 + s_1(x_1) + s_2(x_2) + \dots + s_k(x_k) + \varepsilon,$$

where Y is the transformed response variable (crop area); β_0 is the free term; $s_j(x_j)$ is a smooth function (spline) of the j -th predictor; ε is a random error covering unmodelled factors.

Each predictor was included in the model as a cubic regression spline with the number of nodes $k = 8$, which balances flexibility and generalisation. The model was approximated using the Restricted Maximum Likelihood method (`method = REML`), reducing the overfitting risk. After building the model, standard quality metrics were evaluated: coefficient of determination R^2 , root mean square error (RMSE), and mean absolute error (MAE). This approach made it possible to consider the complex curvilinear relationships between agroecological conditions and the suitability of areas for crop production.

To model the spatial patterns of crop area distribution, a Random Forest (RF) model was built, which implements an ensemble regression approach by aggregating many independent tree solutions. Standardised predictors were used as input data, cleared of missing values and zero areas. Before building the model, all predictors were scaled (z-normalisation), and observations with missing values or zero responses were excluded from the analysis. We used the `randomForest()` function from the `randomForest` library in R to build the model. The parameters were set as `nree = 500` (the number of trees in the forest) and `mtry = \sqrt{p}`, where p is the number of independent variables (predictors). The model was evaluated on the whole training set using such metrics as the coefficient of determination R^2 , root mean square error (RMSE), and mean absolute error (MAE). In addition, the model estimated the importance of variables based on the criterion of relative decrease in accuracy (`%IncMSE`), which allowed us to identify the key predictors that have the most significant impact on the predicted cultivation area.

The best forecasting model was selected by comparing the accuracy of several approaches, including ordinary linear regression, Ridge Regression, Generalised Additive Model (GAM), and Random Forest. To ensure the objectivity of the estimation, the same data was used with pre-standardised predictors, cleared of missing values.

For the OLS and Ridge models, a 10-fold cross-validation was applied using the caret package, which reduces the risk of over-adaptation and provides a stable assessment of the models' generalisability. For the GAM and Random Forest models, the accuracy was calculated on the whole training set, given the complexity of internal cross-validation in these algorithms within a given amount of data. The evaluation was based on three generalised metrics: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The metrics were calculated by comparing the predicted and actual values of the target variable and plotting the agreement graphs to analyse the forecast quality visually. The best model was selected to provide the optimal balance between accuracy and stability of the results, i.e., it demonstrates the highest level of explained variance (R^2) with the lowest forecast errors (RMSE and MAE).

All predictive models were analysed to assess the specific contribution of each predictor to the outcome. For the linear model (OLS), we calculated standardised coefficients (β), which reflect the change in the dependent variable per unit of standard deviation of the predictor. The absolute values of these coefficients were then used to determine the relative contribution of each variable as a percentage of the sum of the absolute β - β -coefficients of all predictors. In the case of Ridge Regression, we used the optimal value of λ , selected by the minimum cross-validation error. After training the model in transformed variables, we extracted the absolute values of the coefficients at each predictor. These absolute values were normalised in the same way as in OLS by dividing by the sum of all absolute coefficients and expressing the result as a percentage. For the Generalised Additive Model (GAM), the significance of the predictor was assessed through the F-statistic of each smooth term. The F values were converted to absolute values and expressed as a percentage of the sum of all F-statistics to ensure comparability with linear approaches. The most flexible approach, Random Forest, allowed us to measure permutational importance: We measured the increase in the mean squared

error (IncMSE) after randomly permuting its values for each predictor. The resulting raw IncMSE values were also normalised to a percentage of their sum. For each model, the final table shows each predictor's specific (%) contribution, facilitating a direct comparison of the importance of different algorithms.

Results

The findings of this study demonstrate that, across all models examined, including ordinary linear regression (OLS), Ridge Regression (Ridge), Generalised Additive Models (GAM) and Random Forest (RF), the water index of soil reaction (pH) is the most significant factor in explaining the spatial variability of corn area (Table 1). The respective shares in the OLS contribution, Ridge, GAM, and RF are 17.2%, 9.4%, 12.8%, and 5.8%, respectively. The second most significant factor in each model is distinct. In the simple linear model (OLS), the proportion of sand in the soil (12.9%) is substantial. In Ridge and GAM, the second most crucial factor is the climatic indicators of temperature (after adjustment for the general trend), and in RF, it is again the sand content in the soil (4.7%). Silt particles and organic carbon in the soil characterise the third position. The third position in the models is as follows: silt in the OLS and the Ridge models (11.8% and 3.5%), organic carbon in the GAM model (10.5%), and the RF model demonstrated a mixed effect of silt (3.7%) and organic carbon (4.7%). Linear models (e.g. OLS and Ridge) permit the inclusion of soil characteristics and residual climate components. At the same time, more flexible approaches (e.g. GAM and RF) primarily focus on basic soil properties (e.g. pH, sand, silt, and organic carbon). This finding suggests that the spatial distribution of maize crops can be most effectively explained by integrating physical and chemical soil properties, with climatic factors playing a comparatively less significant role.

Table 1
Specific contribution of predictors to explaining the variability of the share of area under maize according to different regression models (OLS, Ridge, GAM, RF)

Variables	Variable code	Contribution of variables by regression models (%)			
		linear model (OLS)	ridge regression	GAM	random forest
Water index of soil reaction (pH)	phh2o	17.2	9.4	12.8	5.8
Average daily temperature amplitude	bio2	6.3	2.4	1.0	5.2
Organic carbon content in the soil	soc	–	6.8	10.5	4.7
Residual component of the maximum temperature of the warmest month	bio5_resid	–	2.2	1.1	4.4
Annual rainfall	bio12	5.6	7.5	6.5	4.4
Average temperature of the coldest quarter	bio11	14.1	13.9	18.3	4.3
Share of sand in the soil	sand	–	1.3	6.5	4.1
Residual component of the average temperature of the wettest quarter	bio8_resid	11.8	3.5	4.3	3.8
Residual isothermal component	bio3_resid	–	0.7	1.2	3.8
Residual component of the average temperature of the warmest quarter	bio10_resid	–	0.1	0.8	3.7
The share of silt in the soil	silt	–	2.9	4.5	3.7
Residual component of the minimum temperature of the coldest month	bio6_resid	1.8	1.4	3.2	3.7
Residual component of the annual temperature range	bio7_resid	–	1.3	0.4	3.7
Residual component of precipitation in the wettest quarter	bio16_resid	–	5.7	3.7	3.6
Residual component of precipitation in the coldest quarter	bio19_resid	–	4.4	1.4	3.4
Residual component of temperature seasonality	bio4_resid	–	2.1	1	3.4
Average temperature of the driest quarter	bio9	4.8	0.6	1.6	3.4
Bulk density of the soil	bdod	12.9	4.3	1.2	3.2
Volume fraction of coarse impurities in the soil	cfvo	–	8.6	4.4	3.2
Residual component of precipitation in the driest quarter	bio17_resid	7.2	1.5	0.7	3.1
Residual component of precipitation in the wettest month	bio13_resid	–	7.5	2.9	2.8
The proportion of clay in the soil	clay	–	0.5	1.3	2.8
Residual component of precipitation in the driest month	bio14_resid	12.4	5.6	1.4	2.7
Density of organic carbon in the soil	ocd	3.2	1.3	0.9	2.7
Residual component of mean annual temperature	bio1_resid	–	0.4	3.2	2.7
Total nitrogen content in the soil	nitrogen	2.7	0.4	0.4	2.5
Residual component of precipitation of the warmest quarter	bio18_resid	–	2.8	4.1	2.5
Residual component of precipitation seasonality	bio15_resid	–	0.8	0.8	2.4

The application of simple linear methods showed that OLS explains only 60 per cent of the variance ($R^2 = 0.60$) with RMSE = 47.40 and MAE = 30.59, while Ridge Regression slightly improves the performance ($R^2 = 0.64$; RMSE = 44.99; MAE = 28.12) (Table 2). Generalised Additive Model captured non-linear dependencies, with $R^2 = 0.79$, RMSE = 32.86, and MAE = 19.19. The best results were obtained using Random Forest with $R^2 = 0.96$, RMSE = 14.77, and

MAE = 6.73. In the plots of observed versus predicted values, the points for OLS are widely scattered around the 1:1 line, while Ridge Regression shows only a slight compaction. The GAM plot exhibits a significant narrowing of the spread, and in the case of Random Forest, almost all points lie close to the perfect line (Fig. 1). Quantitative metrics and graphical diagnostics indicate the superiority of Random Forest as a model for this dataset.

Table 2Comparison of regression model performance in terms of explained variation (R^2), RMSE, and MAE

Regression models	Explained variation (R^2)	Root mean square error (RMSE)	Mean absolute error (MAE)
Linear model (OLS)	0.60	47.40	30.59
Ridge Regression	0.64	44.99	28.12
GAM	0.79	32.86	19.19
Random Forest	0.96	14.77	6.73

The high cation-exchange capacity of soil, ranging 12 to 16 cmol(+)/kg, has impacted crop productivity significantly. However, lower or higher values of this capacity can reduce soil suitability for crop growth by limiting nutrient exchange. As the volumetric proportion of coarse fragments (particles > 2 mm) rises, the area suitable for sowing declines sharply due to restricted root development. The optimal clay content, ranging 15 to 30 per cent, has been shown to facilitate root activity. However, values exceeding 40 per cent have been observed to compromise drainage and aeration, impeding growth. An increase in total nitrogen in the fine-earth fraction results in a nearly linear rise in suitability, thereby underscoring the pivotal role of nitrogen nutrition. The optimal pH in a water extract is 6.0–7.5: Lower values markedly reduce suitability due to acidity, and higher values do so because of limited micronutrient availability in alkaline conditions. Furthermore, the texture of the fine-earth fraction has been demonstrated to play a critical role. An approximately 30–40% sand content is conducive to adequate drainage and aeration, while exces-

sively high levels can impede moisture retention. A 20–35% silt content is essential for maintaining a balanced water–air regime; however, extreme values can compromise productivity. It has been demonstrated that increasing the organic carbon content of soil up to approximately 30 g/kg enhances its suitability. Nonetheless, the trend decreases after this point, indicating soil saturation with organic matter and improved structure.

Slight deviations from the average annual temperature (moderate cooling or warming) generally increase the favourability of maize cultivation. By contrast, excessive warming reduces yields (Fig. 3). Differences between day and night temperatures of about seven to ten degrees are optimal, while too small and too large fluctuations hurt plants. A ratio of daily fluctuations to the annual temperature range in the average values is characterised by the best results. At the same time, excessive or insufficient isothermicity decreases the productivity. A pronounced seasonal variation between winter and summer temperatures contributes to better crop growth.

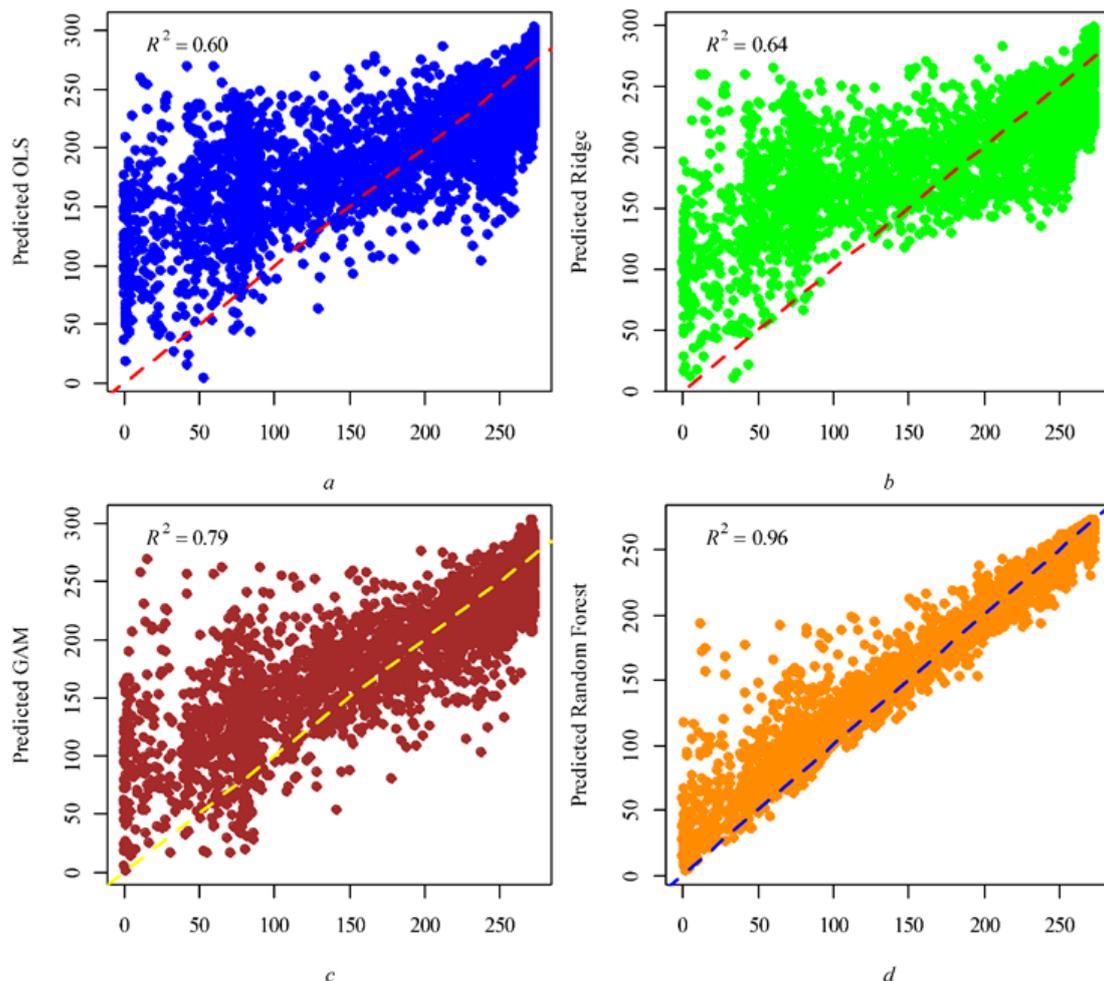


Fig. 1. Correlation between observed (abscissa) and predicted (ordinate) regression models of maize area (in ha per pixel of $0.05^\circ \approx 5.55$ km on the equator) for the models: a is OLS, b is Ridge Regression, c is GAM, and d is Random Forest

The maximum temperature of the warmest month should remain within moderate limits. Excessive dryness in this period reduces yields, while too mild a heat does not provide sufficient soil warming. Similarly, the minimum temperatures of the coldest month should not fall below the critical level to avoid the risk of frost. A moderate an-

nual amplitude range of temperatures favours root development and metabolism, while too great contrasts between winter and summer create stress for plants. Average temperatures in the wettest quarter should be moderately high to maintain sufficient moisture, and in the driest quarter should be low enough to avoid drought, but not so low

as to cause drought. Generally, the most favourable conditions for maize are moderate temperatures with distinct seasons, balanced daily and annual fluctuations and no temperature extremes.

Total annual precipitation appears to be the key climatic factor: Between about 300 and 600 mm, the favourability of maize cultivation increases linearly, and then the increase slows down and becomes almost constant (Fig. 4). The amount of precipitation in the wettest month is indicative of an optimum of approximately 100–120 mm; at lower values, plants suffer from a lack of moisture, and too high values lead to the risk of waterlogging. In the driest month, the favourability index increases with precipitation levels of up to approximately 40 mm, thus emphasising the significance of maintaining a minimum water supply during periods of extreme drought.

The seasonality of precipitation, expressed as a coefficient of variation, shows that a moderate level of irregularity (around 40–60%)

provides the best conditions: An overly even distribution of precipitation is unlikely to offer moisture reserves for the entire season, while excessive irregularity creates stress from the alternation of prolonged droughts and heavy rains (Fig. 5). Similar response models for precipitation in the wettest quarter and the driest quarter confirm that both maximum and minimum seasonal precipitation are essential, and residuals from the trend of these indicators within moderate fluctuations are most favourable. All four climate scenarios show a similar U-shaped trend in the spatial variability of the favourable conditions for maize cultivation. Thus, in the first future interval (2021–2040), the minimum suitability values decrease compared with the historical background. In the middle interval (2041–2060), they reach their minimum. In the long term (2061–2080), there is a partial recovery of the baseline conditions.

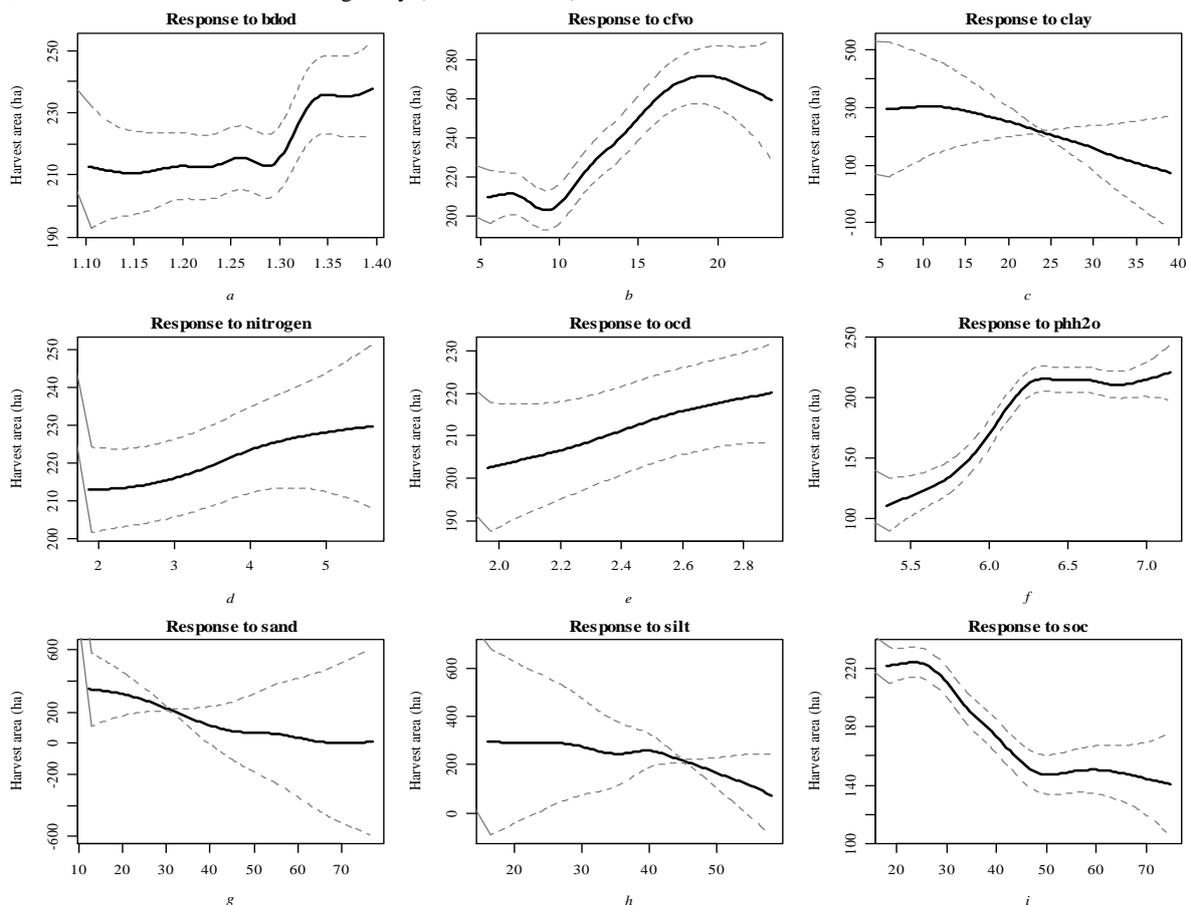


Fig. 2. Response models of the share of maize area to soil factors (partial-dependence plots of the GAM model): *a* (bdod) is the bulk density of the fine soil fraction (kg/dm^3); *b* (cfvo) is the volume fraction of coarse debris ($> 2 \text{ mm}$) in the fine soil fraction (%); *c* (clay) is the volume fraction of clay ($< 0.002 \text{ mm}$) in the fine soil fraction (%); *d* nitrogen is the total nitrogen content (g/kg); *e* (ocd) is the organic carbon density (kg/m^3); *f* (phh2o) is the soil pH in the water extract; *g* (sand) is the proportion of sand ($> 0.05 \text{ mm}$) in the fine soil fraction (%); *h* (silt) is the proportion of dust ($0.002\text{--}0.050 \text{ mm}$) in the fine soil fraction (%); *i* (soc) is the organic carbon content in the fine soil fraction (g/kg)

In the SSP1-2.6 scenario, the minimum level of the area share is about 21% in 2021–2040 (Fig. 5a), while in 2041–2060 it decreases to about 14% (Fig. 5b), after which it returns to about 42% in 2061–2080 (Fig. 5c). The maximum ($\approx 273 \%$) remains almost unchanged due to a stable pool of regions with the most favourable agroclimatic conditions. Under the moderate scenario SSP2-4.5, the minimum suitability is initially higher ($\approx 28\%$, Fig. 5d), decreases to $\sim 20\%$ in the middle of the century (Fig. 5e), and then slightly increases to $\sim 24\%$ (Fig. 5f). Scenarios SSP3-7.0 and SSP5-8.5 show similar patterns, but with deeper minimum values in the middle interval ($\approx 20\%$ for SSP3-7.0, $\approx 20\%$ for SSP5-8.5) and a greater recovery at the end of the century ($\approx 37\%$ and $\approx 41\%$, respectively; Fig. 5g–i). Spatially, the most stable and consistently favourable areas are concentrated in the northern part of the territory, where climatic conditions remain relatively moderate. In the central and southern zones, there are pronounced pockets of declining suitability in the middle of the century, especially

under more severe climate scenarios. By the end of the modelling period, most of these "weak" pockets show a recovery in their potential due to adaptive changes in agricultural practices and a slight mitigation of extreme conditions due to long-term climate cycles. In all scenarios, the increased intensity of global change leads to a short-term decline in the area of land suitable for maize in the medium term. Still, the long-term forecast indicates a potential recovery in the favourability due to compensatory climate processes and changes in precipitation and temperature patterns.

All four climate scenarios demonstrate similar spatial dynamics of deviations of maize cultivation suitability from the historical base (Fig. 6). Under the low-emission scenario SSP1-2.6, the initial period of 2021–2040 is characterised by a moderate decrease in the suitability index (minimum ≈ -53), the medium-term period of 2041–2060 shows the most profound decline (minimum ≈ -58), and the long-term period of 2061–2080 still shows a partial recovery of the condi-

tions (minimum $\approx -58 \rightarrow -52$) with an unchanged maximum ($\approx +170...+180$) (Fig. 6a–c). In the moderate scenario SSP2-4.5, there is a more profound initial decline (minimum ≈ -64 in 2021–2040), a

slightly smaller decrease in 2041–2060 (minimum ≈ -59), and stabilisation at a higher level in 2061–2080 (minimum $\approx -60...-59$ with a maximum of $\approx +172...+188$) (Fig. 6d–f).

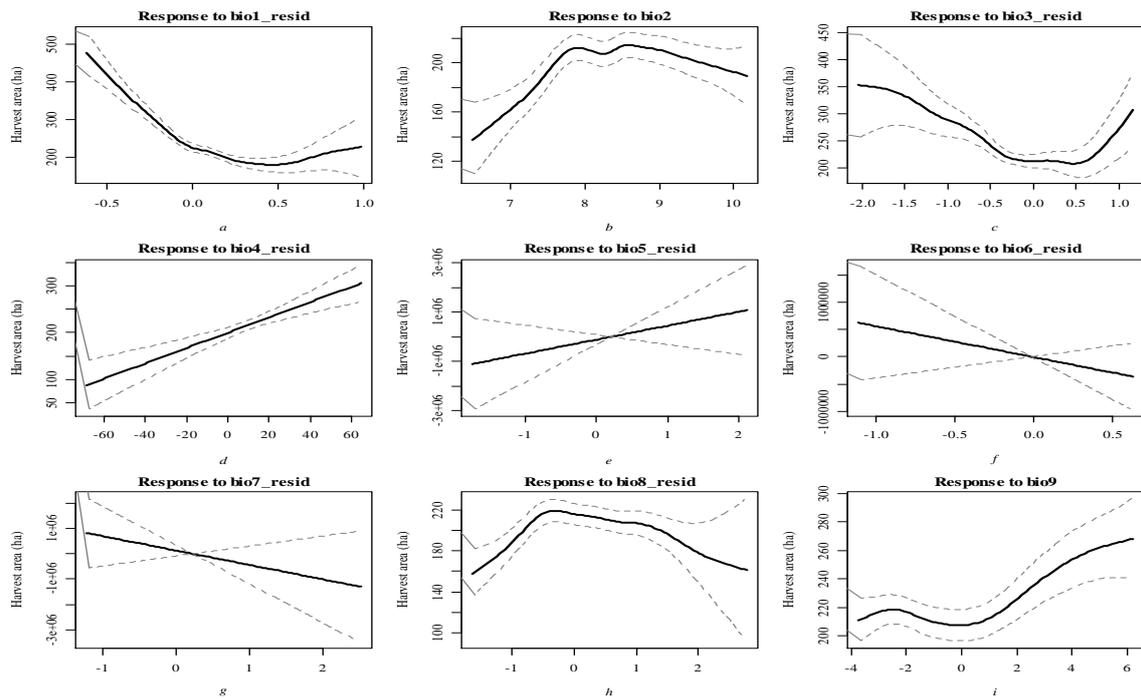


Fig. 3. Models of response of the share of maize area to bioclimatic variables (partial-dependence plots of the GAM model): *a* (bio1_resid) is the residual of the mean annual temperature from the linear trend ($^{\circ}\text{C}$); *b* (bio2_resid) is the residual of the mean daily temperature range from the trend ($^{\circ}\text{C}$); *c* (bio3_resid) is the residual of the isothermicity (ratio of daily to annual temperature range $\times 100$) from the trend; *d* (bio4_resid) is the seasonality of temperature (standard deviation $\times 100$) (-); *e* (bio5_resid) is the residual of the maximum temperature of the warmest month from the trend ($^{\circ}\text{C}$); *f* (bio6_resid) is the residual of the minimum temperature of the coldest month from the trend ($^{\circ}\text{C}$); *g* (bio7_resid) is the annual amplitude range of temperatures ($^{\circ}\text{C}$); *h* (bio8) is the average temperature of the wettest quarter ($^{\circ}\text{C}$); *i* (bio9) is the average temperature of the driest quarter ($^{\circ}\text{C}$)

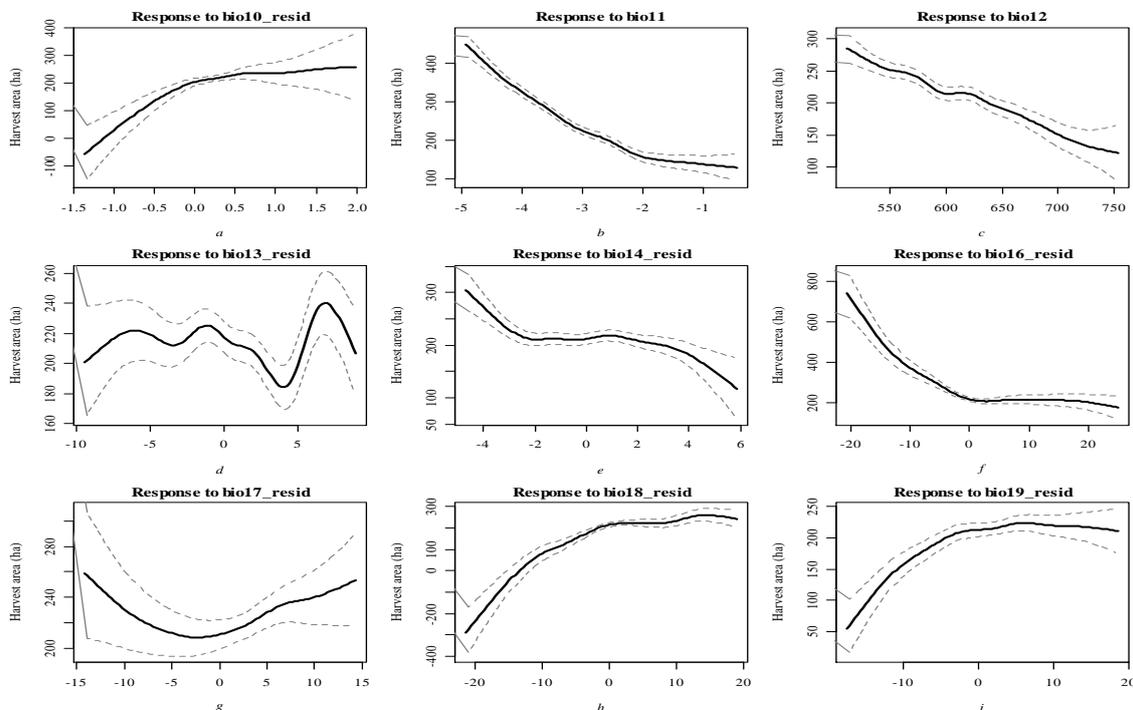


Fig. 4. Models of response of the share of corn area to additional bioclimatic variables (partial-dependence plots of the GAM model): *a* (bio10_resid) is the residual of the average temperature of the warmest quarter from the trend ($^{\circ}\text{C}$); *b* (bio11) is the average temperature of the coldest quarter ($^{\circ}\text{C}$); *c* (bio12) is the annual precipitation (mm); *d* (bio13_resid) is the residual of the precipitation of the wettest month from the trend (mm); *e* (bio14_resid) is the residual of the precipitation of the driest month from the trend (mm); *f* (bio15_resid) is the residual of precipitation seasonality (coefficient of variation $\times 100$) from the trend (-); *g* (bio16_resid) is the residual of precipitation of the wettest quarter from the trend (mm); *h* (bio17_resid) is the residual of rainfall of the driest quarter from the trend (mm); *i* (bio18_resid) is the residual of precipitation of the warmest quarter from the trend (mm)

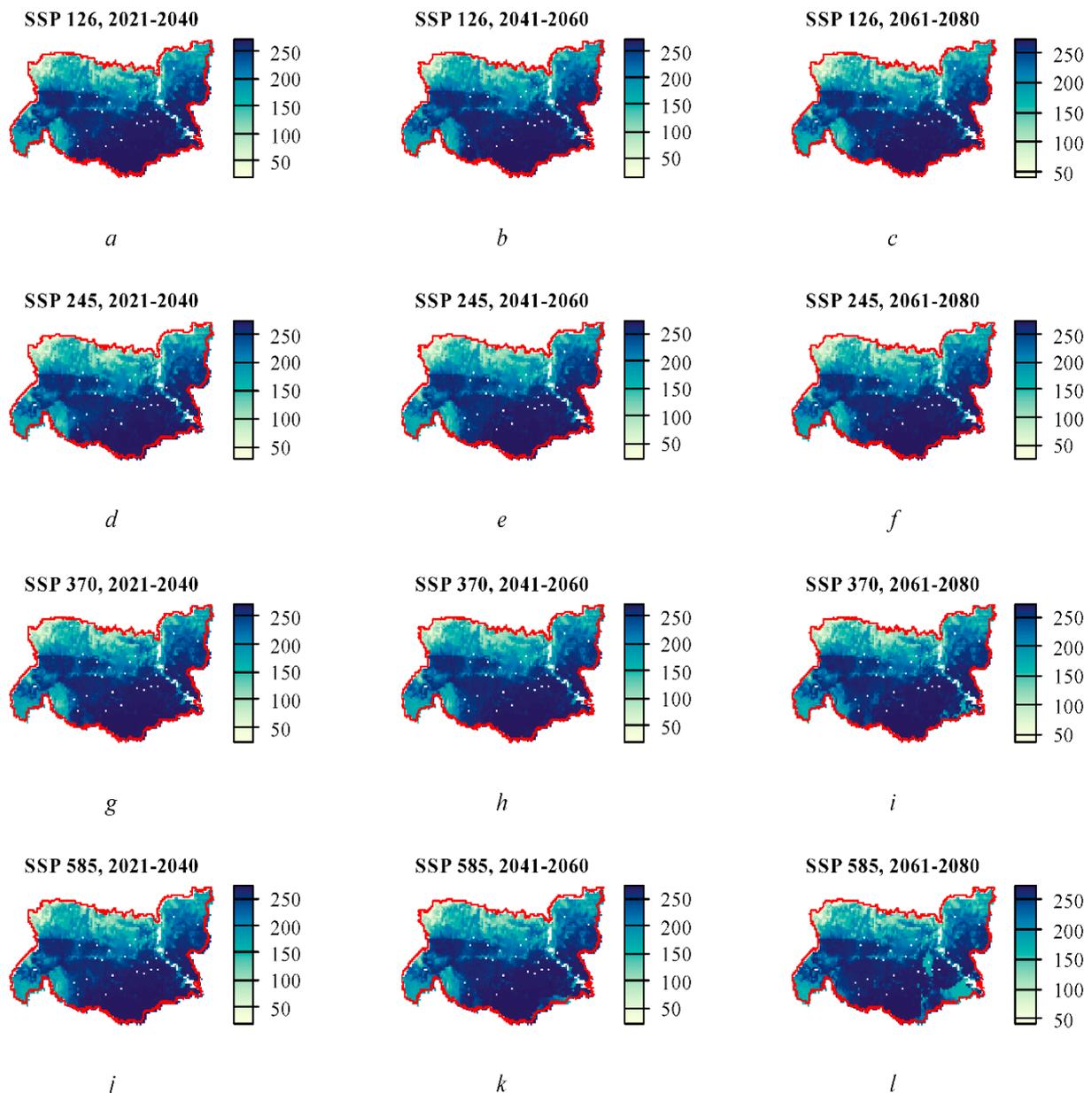


Fig. 5. Projected changes in maize yield under alternative societal development scenarios used for climate change modelling with varying greenhouse gas forcing levels (SSP), 2021–2080: the maps illustrate the spatial variability of projected maize yield (in hectares per pixel) across the study region in Ukraine for three future periods (2021–2040, 2041–2060, 2061–2080) under four shared socio-economic pathways (SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5); these SSP scenarios represent distinct trajectories of societal development combined with differing levels of greenhouse gas emissions, reflecting potential outcomes of climate change impacts on crop production: *a* is the SSP1-2.6 scenario for 2021–2040; *b* is the SSP1-2.6 scenario for 2041–2060; *c* is the SSP1-2.6 scenario for 2061–2080; *d* is the SSP2-4.5 scenario for 2021–2040; *e* is the SSP2-4.5 scenario for 2041–2060; *f* is the SSP2-4.5 scenario for 2061–2080; *g* is the SSP3-7.0 scenario for 2021–2040; *h* is the SSP3-7.0 scenario for 2041–2060; *i* is the SSP3-7.0 scenario for 2061–2080; *j* is the SSP5-8.5 scenario for 2021–2040; *k* is the SSP5-8.5 scenario for 2041–2060; *l* is the SSP5-8.5 scenario for 2061–2080

The SSP3-7.0 and SSP5-8.5 scenarios indicate even more severe temporal fluctuations: The minimum values decrease to ≈ -73 in the medium term (2041–2060), (SSP3-7.0) and ≈ -101 (SSP5-8.5), and a partial recovery to ≈ -116 (SSP3-7.0) and ≈ -113 (SSP5-8.5) with increasing maximums to $\approx +184$... $+209$ is expected at the end of the century (Fig. 6g–l). Spatially, the smallest losses of favourability are mostly recorded in the northern and north-eastern parts of the territory, where the climate remains more moderate. At the same time, the central and southern zones show significant local "dips" in the medium term, especially under more severe scenarios. The partial recovery by the end of the twenty-first century indicates potential compensatory climatic fluctuations and adaptive changes in agricultural practices that could reduce the adverse effects of global warming.

The climate scenarios foresee an increase in the average area under maize compared with the historical baseline (Fig. 7). Under the low-emissions scenario (SSP1-2.6), a significant increase is already

observed in the period 2021–2040 and is maintained at approximately the same level in 2041–2060 and 2061–2080. The moderate scenario (SSP2-4.5) shows similar dynamics: a sharp increase in the first future interval and subsequent stabilisation at this increased level. Under the high emissions scenario (SSP3-7.0), the growth in 2021–2040 is less pronounced, but the area reaches its maximum in 2041–2060, followed by a slight decline in 2061–2080, although still above historical levels.

The most severe scenario (SSP5-8.5) also shows an increase in 2021–2040 compared with the past, followed by a sharp rise until 2060, and a slight decrease at the end of the century. Still, the average values remain significantly higher than historical values (Fig. 7). All scenarios promise a significant improvement in corn yields in the short term, a maximum increase in the medium term, and in the long term the yields are expected to be stable or slightly lower, but still higher than in the past.

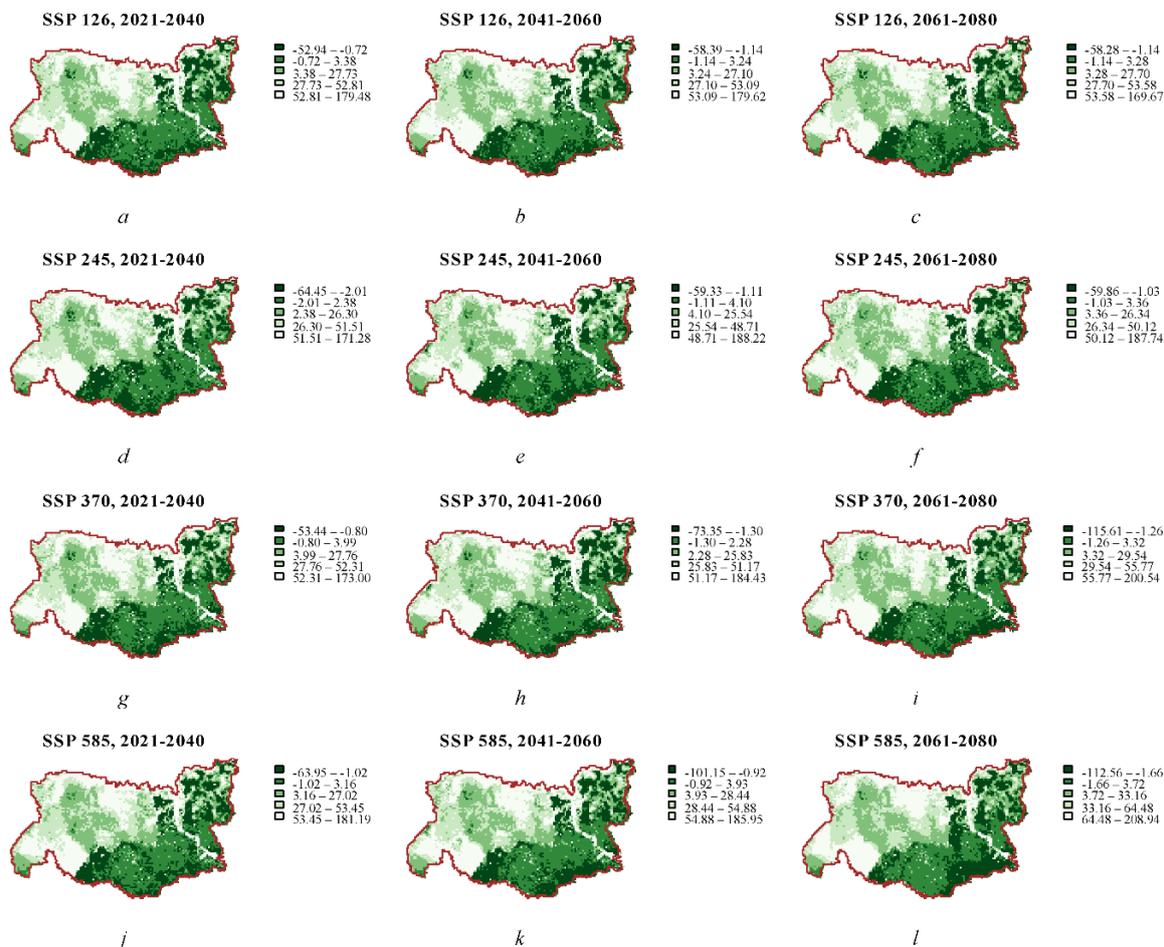


Fig. 6. Projected changes in maize cultivation area relative to the historical baseline (delta), under different SSP scenarios for 2021–2080: values represent the difference between projected and baseline cultivation areas; positive values (lighter shades) indicate an expansion of buckwheat cultivation, while negative values (darker shades) reflect a reduction; the boundaries of the study region are marked in red: *a* is the SSP1-2.6 scenario for 2021–2040; *b* is the SSP1-2.6 scenario for 2041–2060; *c* is the SSP1-2.6 scenario for 2061–2080; *d* is the SSP2-4.5 scenario for 2021–2040; *e* is the SSP2-4.5 scenario for 2041–2060; *f* is the SSP2-4.5 scenario for 2061–2080; *g* is the SSP3-7.0 scenario for 2021–2040; *h* is the SSP3-7.0 scenario for 2041–2060; *i* is the SSP3-7.0 scenario for 2061–2080; *j* is the SSP5-8.5 scenario for 2021–2040; *k* is the SSP5-8.5 scenario for 2041–2060; *l* is the SSP5-8.5 scenario for 2061–2080

Discussion

The study revealed a clear link between the physical and chemical properties of the soil and the level of favourable conditions for maize cultivation. The soil reaction in the range from slightly acidic to slightly dense ($\text{pH} \approx 6.0\text{--}7.5$) creates optimal conditions for mineral nutrition: In this range, micro- and macroelements are in a form accessible to the root system, while beyond this range, either the fixation of nutrients into insoluble compounds or their leaching is observed. These findings are consistent with Pan et al. (2020), confirming sharp increase in the productivity as pH rises to a slightly acidic range; however, the novelty of our work lies in demonstrating the limiting effect of excessively alkaline conditions ($\text{pH} > 7.5$) due to reduced micronutrient availability. The soil particle size distribution, i.e., the balance of sand and silt, determines the water-air regime, which is critical for root aeration and moisture retention (Yu et al., 2024). A 30–40% share of sand and a 20–35% share of silt provide sufficient drainage and water retention, while too light sandy soils dry out and excessively clayey soils stagnate water and limit gas exchange. Organic carbon performs several functions, including structuring soil aggregates and retaining moisture and nutrients in the humus pool. Thus, an increase in organic carbon to approximately 30 g/kg would significantly increase the stability of the soil environment and promote optimal development of the corn root system. Climatic factors have been identified as a primary modifier of these fundamental soil conditions.

Moderate daily temperature amplitudes (approximately 7–10 °C difference between day and night) promote a coordinated develop-

ment of the vegetation phases, while uniformity or excessive fluctuations cause stress responses. The distribution of precipitation throughout the season is crucial for maintaining moisture in the driest months. Insufficient precipitation leads to drought, and excessive water pressure in the soil can cause oxygen deficiency for the roots. Heat and humidity interact with soil characteristics (Chukwudi et al., 2021). High soil organic matter mitigates the adverse effects of short-term droughts, while optimal granulometry contributes to a more even moisture distribution (Kane et al., 2021). The favourable conditions for maize in the study region are primarily determined by a harmonious combination of neutral soil pH, balanced texture, and high organic carbon content, which provide efficient nutrient supply and water-air management. At the same time, climatic parameters only add to local variations, enhancing or mitigating the impact of the underlying soil framework.

The different explanatory power of the models is due to the interactions among factors and responses they can capture and their sensitivity to complex interactions typical of the Polissia and Forest-Steppe regions of Ukraine. Classical linear regression (OLS) assumes that each predictor acts strictly linearly and independently of the others (Roustaei, 2024). In the context of the data under consideration, this limitation precludes the ability to accurately capture peak effects (e.g., optimal pH or texture interval) or interactions between factors. Consequently, OLS provides the least explanation of variance and exhibits elevated error rates. Ridge Regression imposes a penalty on significant coefficients. This penalty is intended to prevent overfitting with many predictors, especially when these predictors are correlated (e.g., different soil texture indicators) (Pavlou et al., 2016). This

slightly improves the stability of the estimates and produces a better average R^2 , but does not solve the problem of non-linear soil-climate interactions. The Generalised Additive Model (GAM) allows each predictor to have its own smooth response function (spline) (Lai et al.,

2024). In this manner, the GAM reflects that maize favourability initially increases with pH, stabilises and decreases, or that the optimum soil moisture lies within a narrow range.

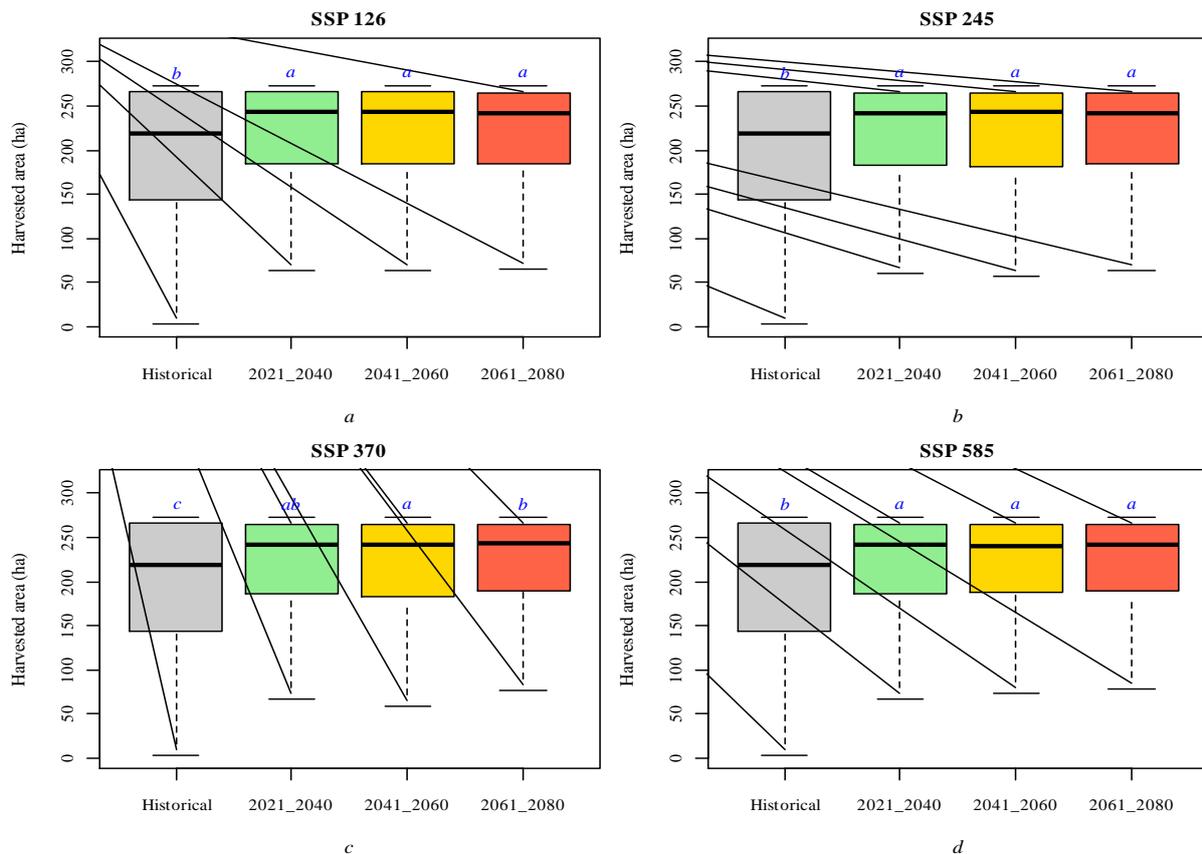


Fig. 7. Boxplots of harvested area (ha) under four SSP scenarios (SSP126, SSP245, SSP370, SSP585) across different periods: historical baseline, 2021–2040, 2041–2060, and 2061–2080; different letters above the boxes indicate statistically significant differences between periods within each SSP according to Dunn's test (with Bonferroni correction, $P < 0.05$); the central line indicates the median; the box represents the interquartile range; whiskers extend to 1.5 times the interquartile range: *a* is the SSP1-2.6; *b* is the SSP2-4.5; *c* is the SSP3-7.0; *d* is the SSP5-8.5

The ability of GAM to capture peak and plateau trends enhances its explanatory power in comparison with linear methods. Random Forest is an ensemble of many decision trees that automatically captures non-linear effects and complex interactions between soil and climate factors without the need to first specify a response shape (Breiman, 2001). This is especially important in the context of the heterogeneous combination of grey forest and glacial soils of Polissia with the black soils of the Forest-Steppe and variable climate: Random Forest can "split" the predictor space into local regions, each with its specific response, and then aggregate the predictions. That is why it demonstrates significantly higher R^2 and lower errors. The simplicity of the model (OLS) limits its ability to reproduce complex ecological processes in the Polissia and Forest-Steppe, the regularisation (Ridge) only reduces noise, and the flexibility of GAM and especially the ability of Random Forest to model non-linearities and interactions makes them the most suitable for describing real conditions of maize cultivation in these regions.

Instead of rigid linear relationships, Generalised Additive Model (GAM) employs smooth splines that automatically adjust to the actual shape of the response. This encompasses optimal plateaus within a neutral pH range, peak moisture values, and U-shaped dependencies on grain size distribution. This makes it possible to accurately reproduce non-linear effects, when the favourability initially increases with a specific value and then drops when it exceeds its optimal range. Separate partial relationship graphs for each factor clearly show what pH or organic carbon level is best for corn and what are stressful, without complex mathematical interpretations. Built-in regularisation mechanisms of GAM automatically prevent overfitting by adjusting the degree of smoothing to the volume and structure of the data,

which is especially important in the heterogeneous conditions of Ukraine's Polissia and Forest-Steppe regions. Generalised Additive Model successfully combines ease of reading results with a high ability to capture complex, locally variable soil-climate interactions, making it a promising tool for modelling crop favourability.

The analysis revealed that several factors show a steadily increasing influence on the favourable conditions for maize cultivation, with no signs of saturation or reversal. This is total nitrogen content in the fine soil fraction. With each increase in available nitrogen concentration, crop productivity steadily increases, as nitrogen is a key element for growth. Annual precipitation in the range of approximately 300 to 600 mm also has a monotonically positive effect: As the amount of moisture in the soil increases, the total moisture supply to plants increases, which ensures a more complete use of the agronomic potential. An apparent seasonal variation in temperature (high seasonality) linearly increases the favourability of cultivation, as pronounced temperature contrasts between winter and summer contribute to the timely completion of corn life stages and optimal distribution of vegetation resources. These factors create the most stable and predictable conditions for corn cultivation in Ukraine's Polissia and Forest-Steppe regions. The only factor that had a monotonically negative response across the entire range was the volume fraction of coarse debris (> 2 mm) in the fine soil fraction: With an increase in the percentage of these coarse particles, the predicted share of the area under maize consistently decreases because the development of the root system becomes hindered due to limited access to nutrients and moisture.

The bell-shaped (unimodal) dependencies of maize cultivation favourability indicate the presence of precise optimal intervals for numerous soil and climatic parameters, beyond which there is a sig-

nificant deterioration in crop productivity. The soil reaction demonstrates the optimal range at approximately 6.0–7.5. In conditions exhibiting slight acidity, the process of nutrient uptake is impeded. Conversely, in alkaline environments, issues with phosphorus fixation emerge, accompanied by deficiencies in micronutrients. A similar unimodal response is observed for soil texture: A moderate proportion of sand (approximately 30–40%) and dustiness (20–35%) creates an optimal balance between drainage and water retention, while excessively light or, conversely, excessively heavy (clay) soils lead to moisture and aeration stress. The cation exchange capacity attains optimal efficiency within the 12–16 resin(+) range per kilogram, as a diminished capacity impedes nutritional intake. By contrast, an excess capacity may signify an abundance of sodium or calcium, thereby reducing the availability of other ions. A parallel between precipitation levels and plant development can be drawn, with the wettest month being the most significant in determining outcomes. A range of 100–120 mm of water supply is optimal for plant development. By contrast, less precipitation can lead to moisture deficiency, and excessive rainfall may result in waterlogging, which can hinder plant development. The maximum temperature of the warmest month also has its own optimum: Moderate heat promotes photosynthesis and growth, while extreme heat leads to water stress. These unimodal dependencies indicate a universal biological principle, whereby each factor exhibits an optimum within which a plant can optimise its potential. In the context of Ukraine's Polissia and Forest-Steppe agroecosystems, it is insufficient to merely increase one factor (e.g., apply more fertiliser or expect more rainfall); it is imperative to maintain each factor within its optimal range. Only under such conditions can a synergistic effect be achieved, whereby the root system develops under ideal water and nutrient conditions, and the climatic component enhances yields without causing stress reactions.

Understanding the nature of maize's non-linear responses to individual soil and climate factors provides a foundation for explaining how conditions will change under global warming. Unimodal (bell-shaped) relationships, such as pH or soil texture, mean that even small shifts away from the optimum will reduce suitability. Therefore, in areas where climate change, occurring through changes in precipitation or temperature, will lead to accelerated leaching of calcium or organic matter, shifts of pH or soil texture will cause a marked decline in yields. Monotonically positive trends (e.g., in total nitrogen or temperature seasonality) explain why in some spatial regions where precipitation or overall seasonal variability increases, we see an increase in favourability despite other stresses. In other words, if an area is projected to receive more moisture or clearer seasons, its relative suitability will increase due to these monotonic relationships. Factors exhibiting strongly negative trends (e.g., the volume fraction of coarse debris) assist in elucidating why certain agricultural landscapes will not become more suitable, even under favourable climate conditions: They are constrained by static soil structure impervious to global change. The amalgamation of these three types of responses (unimodal, monotonous, and negative) directly constitutes the spatial mosaic of increasing and decreasing favourability in the 2021–2080 scenarios. Regions that combine optimal soil conditions with the strengthening of monotonously beneficial climate factors will improve their suitability. Conversely, areas that diverge from the optimum of unimodal factors or experience no additional positive climate impulses will suffer the deepest losses. Providing a tool for localised forecasting and adaptation measures is paramount in maintaining the equilibrium of internal soil optimality and maximising beneficial climate trends.

The practical significance of this work lies in the fact that a clear understanding of maize response patterns to specific soil and climatic factors opens up new opportunities for targeted management of agroecosystems. The established optimal pH ranges, particle size distribution, and organic carbon content allow agronomists to adjust soil composition by liming, adding organic fertilisers, or mulching exactly where they are needed rather than applying standard recipes to all fields in the same way. Knowing that productivity increases linearly with annual rainfall or seasonal temperature differences allows farmers to adjust sowing dates, hybrid selection, and irrigation methods

promptly and in line with weather forecasts and expected global warming scenarios. The resulting models and suitability maps can form the basis of regional recommendations for Ukraine's Polissia and Forest-Steppe regions, helping to identify priority areas for investment in modernising irrigation systems or liming acidic soils. Finally, integrating knowledge on monotonic, unimodal, and negative feedbacks of factors creates a tool for risk management. By identifying drought or waterlogging hotspots in advance, farmers can implement soil protection technologies and select more resilient varieties, increasing the overall resistance of agroecosystems to a changing climate. The results of this work deepen the scientific understanding of the soil and climate determinants of yields and offer specific practical measures for efficient and adaptive management.

Further research should focus on integrating maize genotypic traits with known soil and climatic response patterns. Analysing different hybrids under the same conditions will help to identify which varieties best realise their potential within the optimal ranges of soil pH, texture, and moisture. In addition, it will be helpful to extend the research timeframe by monitoring changes in soil and climatic factors over the long term, focusing on extreme weather events (heat, drought, floods) and their cumulative impact on yields. Modern remote sensing methods (satellite vegetation indices, unmanned aerial surveys) will allow for real-time monitoring of spatial variability of conditions and verification of forecast models. Modelling should be expanded to be multidisciplinary, combining agrophysical, biochemical, and socio-economic aspects, particularly to study the cost of soil improvement measures and assess their effectiveness in different regions. Finally, an important area is the development of adaptive decision support systems for farmers based on interactive maps and mobile applications that will offer optimal agronomic solutions in real time, considering the current state of the fields and weather forecasts. This will allow for the most efficient use of available resources and a flexible response to the effects of global climate change.

Conclusion

The favourable conditions for maize cultivation in Polissia and Forest-Steppe of Ukraine are determined by the physical and chemical properties of the soil (pH in the neutral–alkaline range of 6.0–7.5, optimal sand and silt shares of 30–40% and 20–35%, respectively, and organic carbon content of about 30 g/kg). Climate factors play a supporting role, correcting local deviations due to monotonically positive trends in precipitation and temperature seasonality and unimodal responses to daily temperature fluctuations and peak precipitation in wet months. The predicted climate scenarios indicate a medium-term reduction in areas with high suitability (2041–2060) and further partial recovery by the end of the century (2061–2080) due to compensatory climatic processes and agrotechnical adaptations.

The generalised additive model has proven effective in reproducing non-linear environmental patterns and can serve as a basis for creating practical recommendations to maintain soil optima (pH control, humus enrichment, texture regulation) as a key element of agroecosystems' adaptation to global warming. The Random Forest ensemble approach showed the best results for reproducing these complex ecological patterns due to its ability to automatically account for non-linear relationships and interactions between soil and climate variables. This regression model provides the highest explained variance and lowest forecast errors.

References

- Agnolucci, P., & De Lipsis, V. (2019). Long-run trend in agricultural yield and climatic factors in Europe. *Climatic Change*, 159(3), 385–405.
- Benitez-Alfonso, Y., Soanes, B. K., Zimba, S., Sinanaj, B., German, L., Sharma, V., Bohra, A., Kolesnikova, A., Dunn, J. A., Martin, A. C., Khashi u Rahman, M., Saati-Santamaría, Z., García-Frале, P., Ferreira, E. A., Frazaõ, L. A., Cowling, W. A., Siddique, K. H. M., Pandey, M. K., Farooq, M., ... Foyer, C. H. (2023). Enhancing climate change resilience in agricultural crops. *Current Biology*, 33(23), R1246–R1261.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

- Charalampopoulos, I., & Droulia, F. (2024). A pathway towards climate services for the agricultural sector. *Climate*, 12(2), 18.
- Chukwudi, U. P., Kuttu, F. R., & Mavengahama, S. (2021). Influence of heat stress, variations in soil type, and soil amendment on the growth of three drought-tolerant maize varieties. *Agronomy*, 11(8), 1485.
- Drobitko, A., Kachanova, T., Markova, N., & Malkina, V. (2024). Modern cultivation technologies in improvement of corn quality. *Ukrainian Black Sea Region Agrarian Science*, 28(1), 19–28.
- Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. M. (2022). Global maize production, consumption and trade: Trends and R&D implications. *Food Security*, 14(5), 1295–1319.
- Feng, X., Tian, H., Cong, J., & Zhao, C. (2023). A method review of the climate change impact on crop yield. *Frontiers in Forests and Global Change*, 6, 1198186.
- Hultgren, A., Carleton, T., Delgado, M., Gergel, D. R., Greenstone, M., Houser, T., Hsiang, S., Jina, A., Kopp, R. E., Malevich, S. B., McCusker, K. E., Mayer, T., Nath, I., Rising, J., Rode, A., & Yuan, J. (2025). Impacts of climate change on global agriculture accounting for adaptation. *Nature*, 642(8068), 644–652.
- Kane, D. A., Bradford, M. A., Fuller, E., Oldfield, E. E., & Wood, S. A. (2021). Soil organic matter protects US maize yields and lowers crop insurance payouts under drought. *Environmental Research Letters*, 16(4), 044018.
- Kapustian, M. V., Muzafarov, N. M., Chernobay, L. M., Kolomatska, V. P., Yegorova, N. Y., Kuzmishina, N. V., & Kuzmishina, N. V. (2021). Yield level and stability in corn hybrids of different ripeness groups. *Plant Breeding and Seed Production*, 120, 16–23.
- Kayode Ayinde, O. O. A., & Nwosu, U. I. (2021). Solving multicollinearity problem in linear regression model: The review suggests new idea of partitioning and extraction of the explanatory variables. *Journal of Mathematics and Statistics Studies*, 2(1), 12–20.
- Koshelev, O., Koshelev, V., Fedushko, M., & Zhukov, O. (2021). Annual course of temperature and precipitation as proximal predictors of birds' responses to climatic changes on the species and community level. *Folia Oecologica*, 48(2), 118–135.
- Lai, J., Tang, J., Li, T., Zhang, A., & Mao, L. (2024). Evaluating the relative influence of predictors in Generalized Additive Models using the gam.hp R package. *Plant Diversity*, 46(4), 542–546.
- Landis, D. A., Gardiner, M. M., van der Werf, W., & Swinton, S. M. (2008). Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes. *Proceedings of the National Academy of Sciences*, 105(51), 20552–20557.
- Lee, C.-C., Zeng, M., & Luo, K. (2024). How does climate change affect food security? Evidence from China. *Environmental Impact Assessment Review*, 104, 107324.
- Masereka, E. M., Ochieng, G. M., & Snyman, J. (2019). Frequency analysis of agricultural drought of maize in Sabie River catchment in South Africa. *Jambá Journal of Disaster Risk Studies*, 11(1), a549.
- Mykhailiuk, T., Lisovets, O., & Tutova, H. (2023). Steppe vegetation islands in the gully landscape system: Hemeroby, naturalness and phytoindication of ecological regimes. *Regulatory Mechanisms in Biosystems*, 14(4), 581–594.
- Nykytiuk, Y., & Kravchenko, O. (2024). Regulatory mechanisms in agroecosystems: A retrospective and forecast of spatial and temporal dynamics of precipitation as a factor of crop yield. *Regulatory Mechanisms in Biosystems*, 15(4), 688–695.
- Nykytiuk, Y., Kravchenko, O., Pitsil, A., Bambura, V., & Seredniak, D. (2025). Global climate change may reduce the anti-erosion regulatory capacity of vegetation cover in Ukraine's Polissya and Forest-Steppe regions. *Regulatory Mechanisms in Biosystems*, 33(1), e25004.
- Olesen, J. E., Trnka, M., Kersebaum, K. C., Skjelvåg, A. O., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J., & Micale, F. (2011). Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy*, 34(2), 96–112.
- Pan, X., Baquy, M. A.-A., Guan, P., Yan, J., Wang, R., Xu, R., & Xie, L. (2020). Effect of soil acidification on the growth and nitrogen use efficiency of maize in Ultisols. *Journal of Soils and Sediments*, 20(3), 1435–1445.
- Pavlou, M., Ambler, G., Seaman, S., De Iorio, M., & Omar, R. Z. (2016). Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine*, 35(7), 1159–1177.
- Pichura, V., Potravka, L., Domaratskiy, Y., & Petrovas, S. (2023). Spatiotemporal patterns and vegetation forecasting of sunflower hybrids in soil and climatic conditions of the Ukrainian Steppe zone. *Ukrainian Black Sea Region Agrarian Science*, 27(3), 31–45.
- Rehman, A., Hua, J., Pinzon, S., Mirabela, F. I. M., Loredana, C., & Milin, I. A. (2024). Fertilizer use and agricultural practices in the paradox of maize crop production and environmental sustainability. *Heliyon*, 10(15), e34743.
- Roustaei, N. (2024). Application and interpretation of linear-regression analysis. *Medical Hypothesis Discovery and Innovation in Ophthalmology*, 13(3), 151–159.
- Shiferaw, B., Prasanna, B. M., Hellin, J., & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, 3(3), 307–327.
- Stefanovska, T., Pidlisnyuk, V., Skwiercz, A., Newton, R. A., Zhukov, O., Ust'ak, S., Szczech, M., & Kowalska, B. (2025). Field scale evaluation of the influence of organic amendments on *Miscanthus* × *giganteus* biomass production and soil bioindicators in the marginal degraded land. *Industrial Crops and Products*, 232, 121281.
- Tang, F. H. M., Nguyen, T. H., Conchedda, G., Casse, L., Tubiello, F. N., & Maggi, F. (2024). CROPGRIDS: A global geo-referenced dataset of 173 crops. *Scientific Data*, 11(1), 413.
- Tanumihardjo, S. A., McCulley, L., Roh, R., Lopez-Ridaura, S., Palacios-Rojas, N., & Gunaratna, N. S. (2020). Maize agro-food systems to ensure food and nutrition security in reference to the sustainable development goals. *Global Food Security*, 25, 100327.
- Tutova, H., Ruchiy, V., Khrystov, O., Lisovets, O., Kunakh, O., & Zhukov, O. (2025). Influence of morphology and functional properties of floodplain water bodies on species diversity of macrophyte communities. *Regulatory Mechanisms in Biosystems*, 33(1), e25012.
- Yakovenko, V., & Zhukov, O. (2021). Zoogenic structure aggregation in steppe and forest soils. In: Dmytruk, Y., & Dent, D. (Eds.). *Soils under stress*. Springer International Publishing, Cham. Pp. 111–127.
- Yu, Z., Wang, H., Yu, D., Yin, N., & Zhang, J. (2024). The effect of aeration and irrigation on the improvement of soil environment and yield in dry-land maize. *Frontiers in Plant Science*, 15, 1464624.
- Yuan, X., Li, S., Chen, J., Yu, H., Yang, T., Wang, C., Huang, S., Chen, H., & Ao, X. (2024). Impacts of global climate change on agricultural production: A comprehensive review. *Agronomy*, 14(7), 1360.
- Zhang, Z., Wei, J., Li, J., Jia, Y., Wang, W., Li, J., Lei, Z., & Gao, M. (2022). The impact of climate change on maize production: Empirical findings and implications for sustainable agricultural development. *Frontiers in Environmental Science*, 10, 954940.
- Zhukov, O., Yorkina, N., Budakova, V., & Kunakh, O. (2022). Terrain and tree stand effect on the spatial variation of the soil penetration resistance in an urban park. *International Journal of Environmental Studies*, 79(3), 485–501.
- Zymarioieva, A., Bondarev, D., Kunakh, O., Svenning, J.-C., & Zhukov, O. (2024). Young-of-the-year fish as bioindicators of eutrophication and temperature regime of water bodies. *Environmental Monitoring and Assessment*, 196, 161.
- Zymarioieva, A., Zhukov, O., Fedoniuk, T., & Svenning, J.-C. (2022). Strong decline in breeding-bird community abundance throughout habitats in the Azov Region (Southeastern Ukraine) linked to land-use intensification and climate. *Diversity*, 14(12), 1028.