Agrometeorology

https://doi.org/10.1590/1678-992X-2023-0119

ISSN 1678-992X

Scientia Agricola

Research article

Differences between ecological niche models when predicting the potential distribution of soybean

Lijuan Gong[1*], Xiufen Li[1], Dan Liu[1], Lixia Jiang[1], Lanqi Jiang[1], Yuguang Li[2]

[1]Heilongjiang Province Institute of Meteorological Science – Innovation and Opening Laboratory of Regional Eco-Meteorology in Northeast – China Meteorological Administration, DianTan, 71 – 150030 – Harbin – China.

[2]Heilongjiang Eco-meteorology Center, DianTan, 71 – 150030 – Harbin – China.

*Corresponding author <footprint05@126.com>

Edited by: Fábio Ricardo Marin

Received November 16, 2022
Accepted June 21, 2023

ABSTRACT: Against the background of a sharp decline in soybean planting, rising imports, and natural disasters in China, finding appropriate distribution areas for soybean has become a matter of urgency so that soybean planting policies can be formulated and soybean food security ensured. Among the numerous ecological niche models, the most suitable one for predicting potential distribution areas of soybean in the frigid region must be identified. Based on 65 soybean occurrence points and nine environmental variables, three ecological niche models, MaxEnt, BIOCLIM, and DOMAIN, were applied to the prediction of potential distribution areas for soybean. According to the analytical comparison, the three models predicted the potential distribution of soybean and, specifically, MaxEnt, stood out above the other two models as regards predicting the soybean distribution (Receiver Operating Characteristic curve, AUC = 0.916, Kappa = 0.685). The potential distribution areas (from low suitability to high suitability) predicted by MaxEnt was the largest and accounted for 59.5 % of the total area. The potential suitable distribution area of soybean was mainly concentrated in relatively flat terrain. The Sanjiang Plain and the Northeast Plain accounted for 9.4 % of the total area in the frigid region and are highly suitability for soybean. At the same time, annual mean temperature, elevation and Apr solar radiation were the key determinants affecting soybeans' habitat. On the whole, the selection of ecological niche models and the prediction of soybean potential distribution can provide an essential reference for soybean planting and planning. Moreover, it would be a reliable example for the subsequent related research on soybean habitats.

Keywords: AUC, model evaluation, environmental variables, suitability

Introduction

Soybean [*Glycine max* (L.) *Merr.*], originating in China, is cultivated worldwide and planted in the wild all over the country. From the perspective of China, soybean production depends on agricultural climate resources, and is mainly rainfed. Spatial distribution and yield fluctuation of soybean are all deeply affected by climate conditions. Compared with rice and maize, the economic benefits of soybean less remarkable than the first two crops due to a relatively low yield. The willingness of farmers to plant soybean is not high, resulting in small growing areas and yield, from a peak of 959 × 10^6 km^2 and 1740 × 10^9 kg in 2004 to 6.8 × 10^6 km^2 and 12.15 × 10^9 kg in 2014 (Dong et al., 2017). In recent years, domestic demand for soybean, which is not only a nitrogen fixation crop but also the primary source of protein, feed, and edible oil, has continued to grow unabated (Yao et al., 2020). Against such a background, it is imperative to expand the cultivation of oil crops such as soybeans.

The effects of climate change on crops are reflected in changes in their planting system, suitable distribution range, and productivity; even the centroids of geographic distribution move along with climate change (Duan and Zhou, 2011). Researchers found that dominant vegetation types are not only the adaptation to the regional climate patterns but also a reflection of climate (Roy et al., 2006). Species distribution models (SDMs), including BIOCLIM, DOMAIN, MaxEnt, based on species distribution information and environmental variables have emerged one after another, and are widely used in the simulation of vegetation distribution. Previously, reports on potential distribution of plant species had primarily focused on endangered species (Yi et al., 2016; Yi et al., 2017), invasive species (Núria et al., 2012; West et al., 2016), and special plants (Qin et al., 2020). A few years ago, researchers studied the potential distribution of rice and maize, which are widely distributed in China by MaxEnt (Duan and Zhou, 2011; He and Zhou, 2011). However, reports on the potential distribution of soybean with niche models are rare. How to improve the production layout of soybean crops, exploit potential production, realize the stable yield of soybean in China, and achieve the goal of increasing production in the future are urgent problems currently requiring solutions.

The main goal of this study was to compare and analyze the prediction effects of the three models, BIOCLIM, DOMAIN, and MaxEnt, to find an optimal model that can predict the potential distribution of soybean in the frigid region. The potential distribution of soybean was explored to provide a reference for scientific planning of soybean planting and policy making against climate changes in the frigid region.

Materials and Methods

Occurrence data

Information on soybeans was derived from the crop growth and development dataset of the National Meteorological Information Center in China. The points obtained were strictly screened to eliminate duplicate points. Sixty-five points were obtained as occurrence data which were filtered, one record per 5 km^2 cell. The sample years ranged from 1982 to 2008. For the main part, the distribution of soybean was at 38°40′6″and 53°33′29″ N, 115°29′00″ and 135°5′8″ E, altitude 0-2268 m. The coordinate system was WGS84. The longitude and latitude of the points were edited using Excel in the format of CSV and input ArcGIS, which reduced the study area.

From 1971 to 2000, statistics from several provinces' statistical yearbooks were used to compile the information on soybean yield. Three cities (counties), where the occurrence locations were located, provided statistics on soybean production from 1990 to 2000. Soybean yield statistics were gathered for 63 cities and counties in total (two cities and counties with no data were omitted).

Environmental variables

The study's environmental variables relating to temperature and precipitation were from the World Climate Dataset (https://www.worldclim.org/data/index.html). There were nineteen bioclimatic variables (bio_01 – bio_19), twelve monthly solar radiation variables (version 2), and elevation (Fick and Hijmans, 2017). The variables were averaged for the years 1970-2000 with a spatial resolution of 2.5° then resampled into the raster data with a resolution of 5 km.

Pearson's correlation coefficient method was used to test the correlation of the above variables to avoid collinearity between variables, keeping the variables within a coefficient less than 0.80. One of the ecologically significant variables was retained for the two correlated variables with a coefficient in excess of 0.80, resulting in nine environmental variables (Table 1).

Table 1 –The variables retained for this study.

Environmental variable	Description	Unit
bio_01	Annual mean temperature	°C
bio_02	Mean diurnal range	°C
bio_03	Isothermality	%
bio_05	Max temperature of the warmest month	°C
bio_12	Annual precipitation	mm
bio_15	Precipitation seasonality (Coefficient of Variation)	%
elev	Elevation	m
srad_04	Apr solar radiation	KJ m^{-2} d^{-1}
srad_06	June solar radiation	KJ m^{-2} d^{-1}

Ecological niche models

DIVA-GIS software (version 7.5) was downloaded from http://www.diva-gis.org/download, for BIOCLIM and DOMAIN model analysis. MaxEnt model software (version 3.3.3k) was acquired from https://biodiversityinformatics.amnh.org/open_source/maxent/., for maximum entropy model prediction. ArcGIS 10.2 is a comprehensive GIS platform for formatting layers, data conversion and reclassification.

Prediction methods

Ten groups of training sets in the format of .CSV and environmental variables sets in the format of .ASC were input into the MaxEnt model, setting jackknife to measure the importance of the variables. Prediction results were displayed in the ASC format using defaults. Seventy five percent of the soybean occurrence points were set as the training set with a Sample Points tool, and 25 % of the remaining points and ten times the background points were randomly selected as the test set. To compare the results of the three models, ten sets of training data and corresponding test data sets were randomly generated for model prediction and verification.

In the DIVA-GIS, the training set in Shp format was added. The nine environmental layers converted into the format of GRD (Grid) were generated in the stack dataset. Next, the environmental stack dataset was added into the Molding-BIOCLIM and DOMAIN module to predict the potential distribution.

Processing of soybean yield data

Using the five-year moving average method, the soybean yield per unit area (Y, hereinafter referred to as "yield") in each city (county) was divided into trend yield and meteorological yield. $Y = y(t) + y(w)$, where $y(t)$ was the trend yield, reflecting the long-term yield component of agricultural production level, and $y(w)$ the meteorological yield, reflecting the short-term yield component impacted by meteorological factors. The ratio of meteorological yield to trend yield was used to construct relative meteorological yield, which took into account variations in soybean production across various varieties. The relative meteorological yield was comparable across regions and unaffected by time, space, or farming technology changes. The root mean square (Y_{RMS}) of the yield decline rate was the root mean square of all years with negative relative meteorological yield. The multi-year yield decline rate for a city (or county) was measured using the Y_{RMS}. The more vulnerable it was to climate change and the less stable the yield, the higher the Y_{RMS} value. The term "high-stable yield" was defined as $Y_{HS} = (1 - Y_{RMS}) \times Y_{mean}$, where Y_{HS} stood for "soybean high-stable yield" in a city(county), and Y_{mean} "soybean mean annual yield".

Accuracy assessment of models

Receiving operating characteristic (ROC) and Kappa statistics were used to verify the accuracy of predicting the models´ performance. The ROC, a highly recognized diagnostic evaluation index, has been widely applied to evaluating the current predictions of ecological niche models (Peterson et al., 2007). The area under the curve of ROC is the AUC (area under curve) value, which ranges from 0.5 to 1. The AUC = 0.5 means that the model is randomly distributed. The closer the value is to one, the farther it is from a random distribution and the more accurate the model predicts the results (Wang et al., 2007).

The Kappa statistic also assessed the prediction of the models extensively. The range of the values is [–1, 1]. When the kappa statistic is 1, the prediction model is perfect. However, when the value is equal and less than 0, it represents that the prediction of the model is not as effective as a random model (Segurado and Araújo, 2004).

Results

Comparison of potential distribution from different models

Based on nine environmental variables, the potential distribution areas of soybean in the frigid region of China were predicted by BIOCLIM, DOMAIN, and MaxEnt. A map with the most significant AUC, selected from each model's ten sets of prediction maps as the base map, was converted to ASCII format to import into ArcGIS. Next, the raster map was reclassified. Based on the prediction results of the models, the potential distribution of each model was obtained by using the quantile classification method to classify the potential distribution of soybean.

The predictive range and area of potential distribution in the region studied varied across the three models. The BIOCLIM model predicted climate suitability to be unsuitable and marginal for East Four Leagues in Inner Mongolia (FL), the mountain area, and most of the coastal area. The potential distribution of low to high suitability concentrated in the Sanjiang Plain and the Northeast Plain, accounted for 33.8 % of the total area. Finally, the area of high suitability (10.7 %) was in the center of Heilongjiang (HLJ) and Jilin (JL) provinces (Figures 1A and 2A). The reduction of the map by DOMAIN was similar to the prediction by BIOCLIM. The potential distribution from low to high suitability was also primarily located in the two plains. The range predicted by DOMAIN was more extensive than that of BIOCLIM, with the area increasing by 17 %. However, the area of high suitability with fragmental distribution was reduced to 9.4 % (Figures 1B and 2B). The northernmost point of the area of high suitability reached 52.1° N, 4.2° higher than the northernmost point predicted by BIOCLIM. Meanwhile, the southernmost area of high suitability extended to the Liaodong Bay, the range the former BIOCLIM prediction had never reached. Projected potential distribution from low to high suitability by MaxEnt, 59.5 % of the total area, which was significantly bigger than the preceding two algorithms prediction (Figures 1C and 2C). The coherent covered most of the three northeastern provinces and parts of the FL connected with them. The high suitability distribution was in Sanjiang Plain and Northeast Plain, which was larger than the prediction range of the first two models and was more consistent than the prediction range by DOMAIN.

Assessment of prediction in different models

The main hypothesis of BIOCLIM is that a species can survive and colonize in places where the climate matches its current distribution area (Busby, 1991). BIOCLIM first generates a series of bioclimatic variables that are considered of great biological significance and can describe the climatic information of the distribution areas according to a known distribution area or the relevant data of species growth. Then, it uses a percentage algorithm to capture the area where all environmental variables are within the bioclimatic envelope as the potential distribution area. The DOMAIN model predicts the potential distribution of species by calculating the point-to-point Gower distance, and assignes classification values to each location to be selected (Carpenter et al., 1993). Next, the predicting distribution range is determined according to all the points to be selected below and the threshold decided on by users. A maximum entropy model (MaxEnt) simulated the species distribution based on the maximum entropy principle (Phillips et al., 2006). The MaxEnt chooses a distribution with the largest entropy as the optimal distribution from the distributions that meet the conditions. A feature space, the known distribution area of species, is initially determined, and constrained conditions (environmental variables) that restrict species distribution construct a constraint set. Finally, the relationship is built between the feature space and environmental variables. The three models have their advantages and defects, respectively, as determined by different algorithms.

In the study, three models (BIOCLIM, DOMAIN, and MaxEnt) were assessed for predicting the soybean distribution areas using the same environmental variables and soybean occurrence points. The AUC values of the three models ranged from 0.812 to 0.916, indicating the prediction effect of each model was better than that of the random model (AUC = 0.5) (Table 2), where the

Table 2 – The average AUC (area under curve) and Kappa values of the three models.

Model	AUC	Kappa
BIOCLIM	0.812	0.505
DOMAIN	0.910	0.688
MaxEnt	0.916	0.685

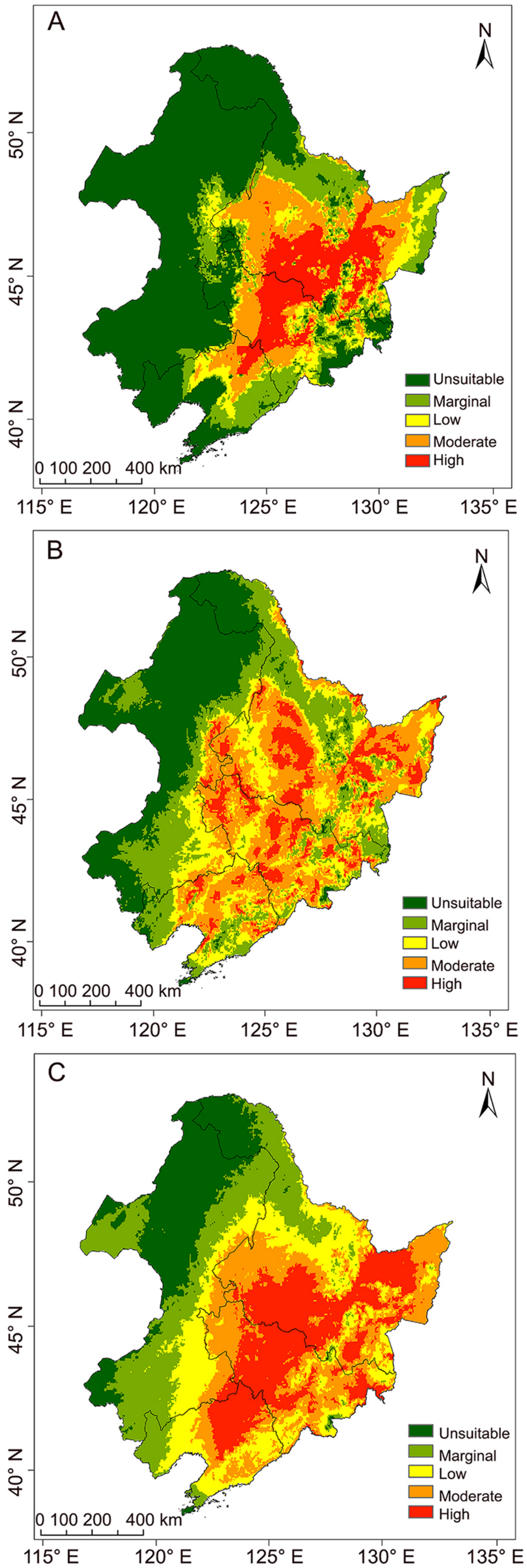

Figure 1 − Potential distribution of soybean based on different niche models: (A) BIOCLIM (B) DOMAIN (C) MaxEnt.

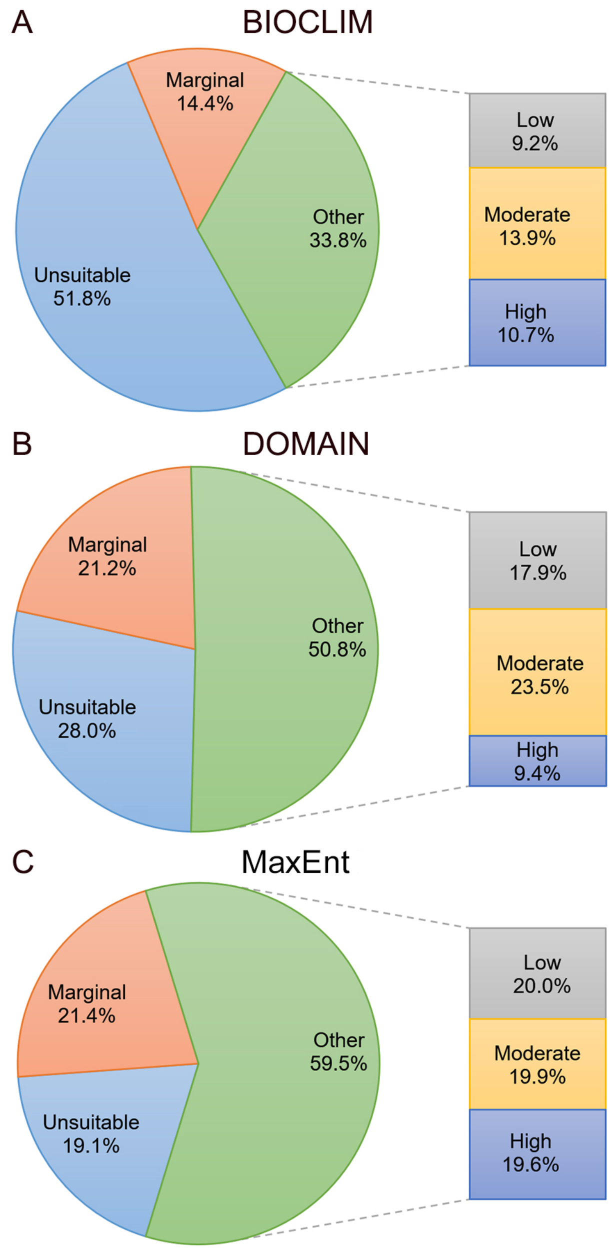

Figure 2 − The proportion of potential suitable area predicted by three models. Other represents total proportion of low, moderate and high suitability. (A) BIOCLIM (B) DOMAIN (C) MaxEnt.

values for DOMAIN and MaxEnt were greater than 0.9. The overall predictive performance of DOMAIN and MaxEnt was better than that of BIOCLIM. Kappa statistics were similar to AUC values. The greater the AUC value, the greater the corresponding Kappa statistics. Moreover, the AUC values and Kappa statistics of DOMAIN and MaxEnt were almost the same, and the difference between the two models was no more than 0.006. In view of the small difference between the two models, one of them could be selected to evaluate the suitability of soybean in frigid regions.

The three indicators of Y, Y_{RMS} and Y_{HS} in various cities and counties were compared to the suitability (after normalization) to support futher the accuracy of the models' predictions (Figure 3A-C). The correlations between suitability based on BIOCLIM and yield indicators, as well as suitability based on MaxEnt and $Y_{RMS'}$ were significant (reaching a significance level of 0.05), except for the weak correlations based on the DOMAIN between suitability and yield indicators. The Y, $Y_{RMS'}$ and Y_{HS} were not substantially connected to the suitability results based on DOMAIN. The suitability results based on MaxEnt were more accurate in conveying the yield characteristics than those of the BIOCLIM. When suitability increased, the yield-related indicators did not exhibit a linear pattern, but the change rate decreased with the suitability increase. This showed that the yield tended to stabilize as suitability increased, consistent with the actual production. Typically, the soybean yield tended to be stable when the suitability rose to a certain level. In addition, the Y_{RMS} (BIOCLIM) tended to fall first and then rise as suitability increased, which was inconsistent with real production. In conclusion, the suitability predicted by MaxEnt was better in the other two models as regards evaluating the habitat of soybean in the frigid region in China.

Environmental variables' response to suitability

Taking the MaxEnt as an example, the environmental variables in different suitable areas were comprehensively analyzed. The predominant environmental factors affecting species distribution were obtained from the contribution rate of MaxEnt´s environmental variables. The main environmental factors affecting the distribution of soybean were as follows: bio_01 (48.8 %), elev (17.9 %), srad_04 (15.7 %), bio_12 (5.8 %) and bio_15 (4.8 %). The total contribution rate of the five environmental variables was 93.0 %. It indicated that bio_01 was the most important environmental factor affecting the distribution of soybean, followed by elev and srad_04. The bio_12 and bio_15 also had important effects on the distribution of soybean, although their contribution rates were lower than those of the elev and srad_04.

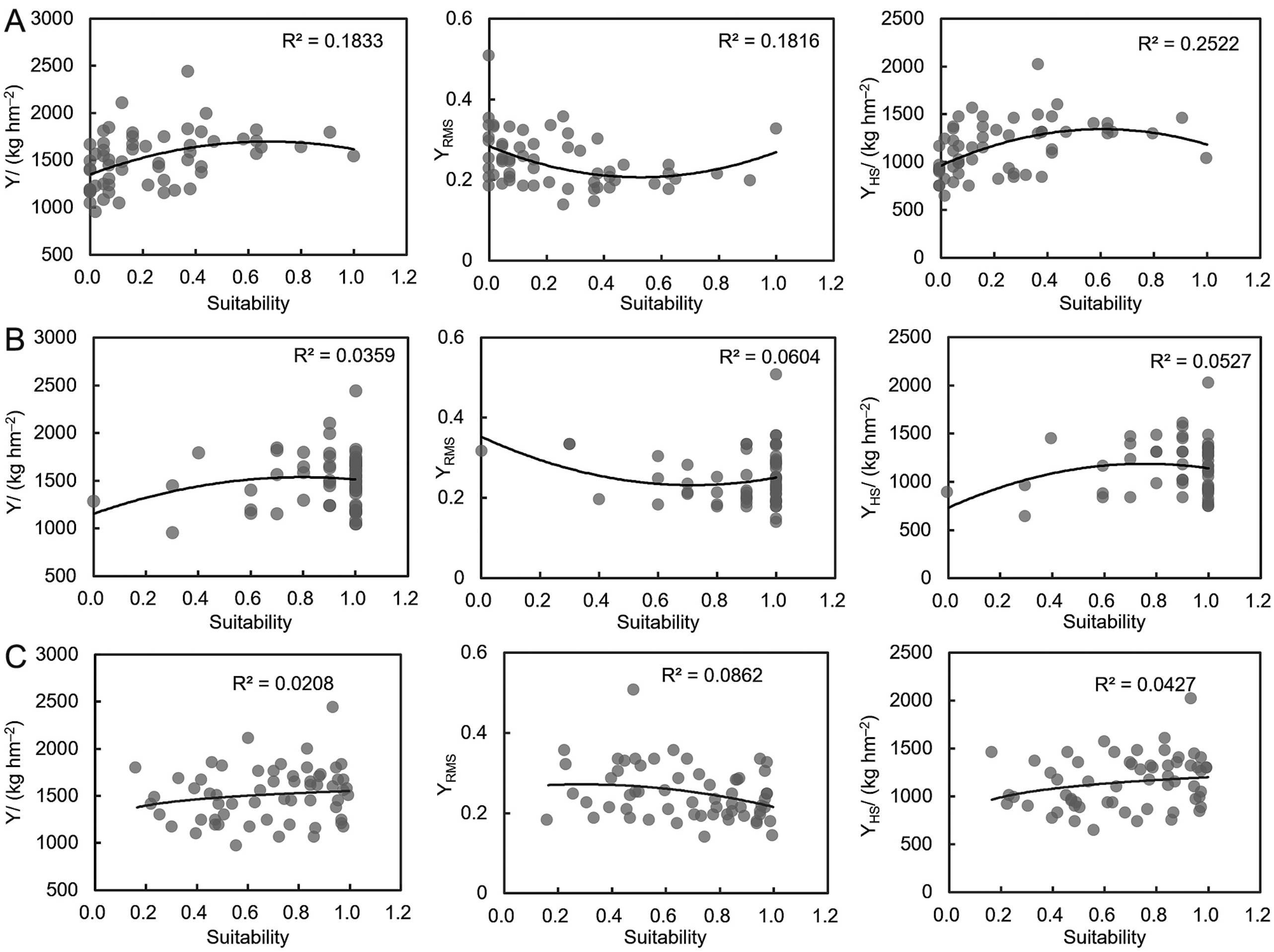

Figure 3 − Correlation between suitability and yield index of soybean in the frigid region in China: (A) BIOCLIM, (B) DOMAIN, (C) MaxEnt.

The spatial range of each suitable grade was extracted from the ArcGIS module spatial analyst tools. The parameters range (minimum and maximum value), mean and standard deviation of nine environmental variables in five suitable grades distribution areas were counted, respectively (Table 3). The results showed that in the potential distribution areas with different suitable grades, the variation ranges of the five dominant environmental variables (bio_01, elev, srad_04, bio_12 and bio_15) affecting the distribution of soybean showed changes in trends as follows: the variation range was gradually narrowed from the unsuitable area to the highly suitable area. The standard deviation presented a "sole peak" curve. However, the five factors showed obvious downward trends from the low to the highly suitable area. It suggested that the main environmental factors affecting the potential distribution of soybean tended to be the same in each highly suitable distribution area. On average, the area most suitable for the soybean to grow was an annual mean temperature of 4.5 °C, an elevation of 181.5 m, Apr solar radiation of 17673.4 KJ m^{-2} d^{-1}, an annual precipitation of 566.2 mm and precipitation seasonality of 105.8 %.

Discussion

In this study, three niche models were applied to simulate the potential distribution of soybean based on the environmental variables in the frigid region. The low to highly suitable areas were concentrated in the plains, which were basically consistent with the soybean production areas in this region (Dong et al., 2017; Gong et al., 2022b). Therefore, as confirmed by the prediction results, the Northeast Plain could be regarded as the main soybean cultivation area. However, there was divergence in the results of the models founded on different principles. The potential distribution areas from low to highly suitable by DOMAIN and MaxEnt exceeded 50 % of the total area of the studied area, above that by BIOCLIM, perhaps on account of the different principles adopted. BIOCLIM was based on the principle of environmental envelope and the MaxEnt on the species distribution reaching the maximum entropy under given environmental conditions (Busby, 1991; Carpenter et al., 1993; Phillips et al., 2006). The DOMAIN predicted that the highly suitable area would span over a wide expanse, from the Bohai bay in the south to the Sino-Russian boundary in the north. A basic principle of DOMAIN was to classify the similarity matrix between points (Carpenter et al., 1993). The results, meanwhile, were greatly affected by the sampling points. Almost all places with occurrence points could be suitable areas. Therefore, in an area with such a large span, there was either a low or highly suitable area due to the existence of occurrence points.

ROC and Kappa statistics are widely used to evaluate the accuracy of the niche model prediction (Koo et al., 2015; Qin et al., 2020). The prediction accuracy of

Table 3 – Statistical analysis of environmental variables in different suitable classes of soybean by MaxEnt.

Variables	Unsuitable		Marginal		Low		Moderate		High	
	Range	Mean ± std.	Range	Mean ± std.	Range	Mean ± std.	Range	Mean ± std.	Range	Mean ± std.
bio_01	− 6.4~11.6	2.7 ± 2.1	− 3.4~11.0	1.8 ± 3.5	− 1.6~10.8	3.6 ± 3.1	0.1~9.6	3.9 ± 2.0	0.4~9.3	4.5 ± 1.6
bio_02	5.6~17.7	14.0 ± 1.2	7.2~15.5	13.3 ± 0.9	7.0~15.5	12.3 ± 1.1	8.4~14.5	11.9 ± 0.9	9.5~14.0	11.8 ± 0.6
bio_03	18.5~29.6	25.4 ± 1.5	19.8~30.5	16.0 ± 2.1	19.2~29.1	24.7 ± 1.8	19.4~27.9	23.7 ± 1.9	20.0~27.2	23.3 ± 1.4
bio_05	17.2~29.1	23.6 ± 1.3	19.3~30.1	26.0 ± 1.7	20.9~30.2	26.3 ± 1.8	20.8~30.1	26.8 ± 1.4	22.9~29.4	27.2 ± 0.9
bio_12	216~979	448.2 ± 74.4	188~897	447.6 ± 104.8	376~1006	584.2 ± 146.4	390~994	573.9 ± 132.4	397~951	566.2 ± 77.4
bio_15	92.7~125.2	116.7 ± 6.6	79.2~129.1	113.3 ± 8.7	74.6~130.75	108.4 ± 10.9	69.7~129.6	105.6 ± 13.2	80.0~125.6	105.8 ± 8.8
elev	12~2268	842.4 ± 250.1	1~1804	585.8 ± 227.4	0~1500	391.1 ± 222.5	1~1151	273.5 ± 197.4	4~741	181.5 ± 98.5
srad_04	16143~19994	17433.6 ± 944.1	16232~19900	18110.0 ± 1219.2	16412~19620	17817.2 ± 1009.0	16605~19199	17688.5 ± 697.1	16808~18852	17673.4 ± 520.2
srad_06	17760~23162	21861.9 ± 739.12	17781~23316	22232.7 ± 741.8	17478~23166	21342.2 ± 1295.3	17511~23164	21125.9 ± 1276.8	17299~22837	21139.2 ± 924.5

DOMAIN and MaxEnt was higher than that of BIOCLIM, both of which could simulate the potential distribution of soybean. In comparison, MaxEnt performed better with the highest AUC value. Although its Kappa value was relatively small, Kappa statistics represented the maximum discrimination value of the series threshold, reflecting the point discrimination. In contrast, the AUC reflected the comprehensive value of the discrimination of the series threshold. The strong discrimination of the model at a certain threshold did not reflect its strength at other thresholds. However, the AUC value was not affected by the thresholds, and it could be better used for a comparative study between different models of the same species (Stoklosa et al., 2015; Wang et al., 2007). Several scholars have compared and analyzed the prediction of 16 niche models on the suitable habitat of many species, which showed that MaxEnt presented a better prediction effect than BIOCLIM and DOMAIN (Elith et al., 2006). MaxEnt also showed a predominant effect in predicting the habitat of other alien species such as *Euplatypus parallelus* in China (Tang et al., 2019). MaxEnt was applied to predict plant distribution, including rice, maize and *Dipterocarpus alatus* (Duan and Zhou, 2011; Kamyo and Asanok, 2020; He and Zhou, 2011). Moreover, the MaxEnt results had a more detailed spatial distribution and a better expression of yield reduction and stability, than the suitability results of BIOCLIM and DOMAIN. In this study, the MaxEnt was selected to predict the habitat and response of environmental variables to the suitability, constituting the final results, showing that the model had a high fitting degree for soybean. However, the results may be different in the case of other species. Each model may perform differently when predicting the suitable distribution areas of other species. Therefore, the optimal model should be selected from multiple niche models when simulating the potential for suitable distribution of a species.

The above research suggested that the potential distribution of soybean was predominantly distributed in the Northeast Plain. The range of suitable distribution was essentially compatible with the findings of previous researchers (Dong et al., 2017), and the potential distribution of soybean was in the temperate and warm temperate zone with flat terrain. The jackknife test, comprising five variables relating to temperature, elevation, radiation, and precipitation were the leading environmental variables that affect soybean distribution. These correlated closely to the characteristics of soybean, which is a short-day and temperature-bias plant with cold tolerance. Compared with other studies, the affecting factors selected were only environmental variables, which were less in number than other research factors, such as accumulative temperature, meteorological factors at different crop growing stages, and growing degree days (Zhang et al., 2019). However, the five environmental variables obtained by Pearson's correlation coefficient method and the jackknife test in the MaxEnt reflected climate, landform, ecosystem, and habitat niche of the study region, which are significant factors that affect the species distribution (García-Aldés et al., 2015; Hannah, 2015). The variables acquired containing the climate and terrain information were the primary factors in determining the biodistribution and played a decisive role on a large scale (Davis et al., 1998).

Soybean is a crop that is greatly affected by human activity. Human-originating components, such as irrigation, variety improvement, cultivation management, market demand, and other management strategies may expand the distribution areas of soybean, which would lead to the predicted distribution area of soybean being larger than the actual one. The plain is an ideal location for soybean cultivation. To boost real yield and production efficiency, the soybean planting pattern in the study region may be adjusted depending on suitability orientation. According to the current distribution of soybean and the variations in the types of suitable areas, the planting area may be adequately expanded into high and moderately suitable regions and progressively reduced or maintained in areas of low suitability. The northernmost potential area of soybean by MaxEnt might stretch to Heihe city at the junction of China and Russia. The city is in fact one of the main producing areas, and more suitable for soybean production with relatively lower requirements for accumulated temperature than those for corn (Yin et al., 2017). The application of the MaxEnt to predict the potential distribution of soybean in future climatic scenarios had been carried out, and the prediction can better express the spatial change information of crops (Gong et al., 2022a). Most China's soybean imports come from countries such as Brazil and America (Cattelan and Dall'Agnol, 2018). However, concerns such as international trade conflicts and the COVID-19 pandemic have significantly influenced on China's edible oil production, processing, and import and export commerce in recent years. Certain risks and hidden perils exist in China's reliable supply of soybeans. In addition to expanding the planting area, optimizing the spatial layout of the soybean industry, coordinating allocation of investment elements, strengthening economic policy support, and other such measures to release the potential of soybean yield per unit area are effective ways to ensure increases in safe and sustainable soybean production. The impact of these factors on soybean production in China still needs to be further discussed in future research.

Authors' Contributions

Conceptualization: Gong LJ, Li XF. **Data curation:** Jiang LQ, Li YG. **Methodology:** Gong LJ, Jiang LX. **Formal analysis:** Gong LJ, Li YG. **Software:** Liu D, Li XF. **Writing-review & editing:** Gong LJ, Li XF, Jiang LQ, Li YG.

References

Busby JR. 1991. BIOCLIM: a bioclimate analysis and prediction system. Plant Protection Quarterly 6: 8-9.

Carpenter G, Gillison AN, Winter J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 2: 667-680.

Cattelan AJ, Dall'Agnol A. 2018. The rapid soybean growth in Brazil. Oilseed and fats Crops and Lipids 25: D102. https://doi.org/10.1051/ocl/2017058

Davis AJ, Lawton JH, Shorrocks B, Jenkinson LS. 1998. Individualistic species responses invalidate simple physiological models of community dynamics under global environmental change. Journal of Animal Ecology 67: 600-612.

Dong F, Liu A, Feng Z, Xue L, Yang X. 2017. Food crop planting structure feature and dynamics changes in traditional soybean producing area-taking the Nenjiang county as an example. Chinese Journal of Agricultural Resources and Regional Planning 38: 79-85 (in Chinese, with abstract in English). https://doi.org/10.7621/cjarrp.1005-9121.20170313

Duan J, Zhou G. 2011. Potential distribution of rice in China and its climate characteristics. Acta Ecologica Sinica 31: 6659-6668 (in Chinese).

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-m spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302-4315. https://doi.org/10.1002/joc.5086

García-Aldés R, Svenning J, Zavala MA, Purves DW, Araújo MB. 2015. Evaluating the combined effects of climate and land-use change on tree species distributions. Journal of Applied Ecology 52: 902-912. https://doi.org/10.1111/1365-2664.12453

Gong L, Li X, Wu S, Jiang L. 2022a. Prediction of potential distribution of soybean in the frigid region in China with MaxEnt modeling. Ecological Informatics 72: 101834. https://doi.org/10.1016/j.ecoinf.2022.101834

Gong L, Liu D, Jiang L, Li X, Lv J. 2022b. Distribution characteristics of climate potential productivity of soybean in frigid region and its response to climate change. Environmental Science and Pollution Research 29: 7452-7464. https://doi.org/10.1007/s11356-021-15879-y

Hannah L. 2015. Modeling species and ecosystem response. p. 237-262. In: Hannah L. ed. Climate change biology. 2ed. Academic Press, New York, NY, USA. https://doi.org/10.1016/B978-0-12-420218-4.00011-1

He Q, Zhou G. 2011. Climatic suitability of potential summer maize planting zones in China. Acta Geographica Sinica 66: 1443-1450 (in Chinese, with abstract in English). https://doi.org/10.11821/xb201111001

Kamyo T, Asanok L. 2020. Modeling habitat suitability of *Dipterocarpus alatus* (Dipterocarpaceae) using MaxEnt along the Chao Phraya River in Central Thailand. Forest Science Technology 16: 1-7. https://doi.org/10.1080/21580103.2019.1687108

Koo KA, Kong WS, Nibbelink NP, Hopkinson CS, Lee JH. 2015. Potential effects of climate change on the distribution of cold-tolerant evergreen broadleaved woody plants in the Korean Peninsula. Plos One 10: e0134043. https://doi.org/10.1371/journal.pone.0134043

Núria G, Thuiller W, Pino J, Vilà M. 2012. Potential distribution range of invasive plant species in Spain. NeoBiota 12: 25-40. https://doi.org/10.3897/neobiota.12.2341

Peterson AT, Papeş M, Eaton M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30: 550-560. https://doi.org/10.1111/j.2007.0906-7590.05102.x

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Qin A, Jin K, Batsaikhan M, Nyamjav J, Li G, Li J, et al. 2020. Predicting the current and future suitable habitats of the main dietary plants of the Gobi Bear using MaxEnt modeling. Global Ecology and Conservation 22: e01032. https://doi.org/10.1016/j.gecco.2020.e01032

Roy PS, Joshi PK, Singh S, Agarwal S, Yadav D, Jegannathan C. 2006. Biome mapping in India using vegetation type map derived using temporal satellite data and environmental parameters. Ecological Modelling 197: 148-158. https://doi.org/10.1016/j.ecolmodel.2006.02.045

Segurado P, Araújo MB. 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31: 1555-1568. https://doi.org/10.1111/j.1365-2699.2004.01076.x

Stoklosa J, Daly C, Foster SD, Ashcroft MB, Warton DI. 2015. A climate of uncertainty: accounting for error in climate variables for species distribution models. Methods in Ecology and Evolution 6: 412-423. https://doi.org/10.1111/2041-210X.12217

Tang J, Li J, Lu H, Lu F, Lu B. 2019. Potential distribution of an invasive pest, *Euplatypus parallelus*, in China as predicted by Maxent. Pest Management Science 75: 1630-1637. https://doi.org/10.1002/ps.5280

Wang Y, Xie B, Wan F, Xiao Q, Dai L. 2007. Application of ROC curve analysis in evaluating the performance of alien species' potential distribution models. Biodiversity Science 15: 365-372. https://doi.org/10.1360/biodiv.060280

West AM, Kumar S, Brown CS, Stohlgren TJ, Bromberg J. 2016. Field validation of an invasive species Maxent model. Ecological Informatics 36: 126-134. https://doi.org/10.1016/j.ecoinf.2016.11.001

Yao H, Zuo X, Zuo D, Lin H, Huang X, Zang C. 2020. Study on soybean potential productivity and food security in China under the influence of COVID-19 outbreak. Geography and Sustainability 1: 163-171. https://doi.org/10.1016/j.geosus.2020.06.002

Yi Y, Cheng X, Yang Z, Zhang S. 2016. Maxent modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China. Ecological Engineering 92: 260-269. https://doi.org/10.1016/j.ecoleng.2016.04.010

Yi Y, Zhou Y, Cai Y, Yang W, Li Z, Zhao X. 2017. The influence of climate change on an endangered riparian plant species: the root of riparian Homonoia. Ecological Indicators 92: 40-50. https://doi.org/10.1016/j.ecolind.2017.05.004

Yin R, Feng X, Zhang Z. 2017. Changes of soybean planting area in northeast China and the Huang-Huai region in 2017 and its production outlook. Agriculture Outlook 7: 42-47

Zhang Q, Zhang D, Wu M, Guo J, Sun C, Xie C. 2017. Predicting the global areas for potential distribution of *Gastrodia elata* based on ecological niche models. Chinese Journal of Plant Ecology 41: 770-778. https://doi.org/10.17521/cjpe.2016.0380 (in Chinese, with abstract in English).

Zhang K, Zhang Y, Tao J. 2019. Predicting the potential distribution of *Paeonia veitchii* (Paeoniaceae) in China by incorporating climate change into a Maxent Model. Forests 10: 190. https://doi.org/10.3390/f10020190