Exercise 2: Predict spatial pattern changes

How can I print an exercise to PDF format?

Technical note

Depending on your computer's hardware, running this prediction model may take 20 minutes to 1 hour. For this reason, you may choose to not run the model during the exercise and instead use the result files that are provided. See the exercise for further instructions.

Software requirements

- ArcGIS Online
- ArcGIS Pro 3.2
- ArcGIS Spatial Analyst extension

Introduction

In many parts of Africa, food and income security are tied to the agricultural production of maize. However, as a result of climate change and its related impacts, food security for many vulnerable populations is diminishing.

In the previous exercise, you prepared a training dataset for the Presence-Only Prediction (MaxEnt) tool, which will be used to model current and future suitability using baseline variables against the SSP3 7.0 climate scenario for 2050. After a training dataset of presence points and explanatory variables is created, the model is ready to be trained. And, once trained, the model can be run using projected explanatory variables to make a prediction of future conditions.

Scenario

Imagine the following scenario: You are continuing your work as a GIS analyst for a global nonprofit organization in Africa. As part of the agricultural ministers' efforts to protect farmers and safeguard food security, you want to use your training dataset for the Presence-Only Prediction (MaxEnt) tool to create an output that models future maize suitability.

The training dataset has already been created. Therefore, in this exercise, your task is to create prediction points, train the model, and, finally, run the model to predict maize suitability in Africa.

Note: The exercises in this course include View Result links. Click these links to confirm that your results match what is expected.

Estimated completion time in minutes: 90 to 120 minutes

Expand all steps | Collapse all steps

- Step 1: Create prediction points

A training dataset is used to teach a model to perform a task, such as prediction. To train a model, you must first run the tool using the training dataset and baseline explanatory variables. To create a presence-only prediction model, you must then run the tool a second time using prediction points and future explanatory variables.

In this step, you will create the prediction points for your model.

a. If necessary, open your MaizePredictionModeling project.

- Hint

  Start ArcGIS Pro and, under Recent Projects, click MaizePredictionModeling.

Note: If you did not download the project, go to the previous exercise, *Prepare data for a prediction model*, and follow the instructions in the *Download the exercise data files* step to download and open the MaizePredictionModeling Pro project.

b. In the Contents pane, turn off the PresencePoints layer.

You will use the GFSAD_Maize layer to create the prediction points. However, this layer covers the entire globe, so you will first clip it to the study area of Africa.

c. In the Geoprocessing pane, search for and open the **Extract By Mask** (Spatial Analyst Tools) tool.

d For Input Raster, choose GFSAD_Maize.

e For Input Raster Or Feature Mask Data, choose World Continents.

f For Output Raster, type **GFSAD_Maize_Clip**.

g Leave all the other default parameters and click Run.

*Step 1g***: Create prediction points.*
The colors on your map may differ from the image.

You have created a study area for your prediction points by clipping the GFSAD_Maize layer to the continent of Africa.

Now that you have identified your study area, you can create the prediction points. You will use the GFSAD_Maize_Clip layer to create a point for each pixel, or cell, of the raster layer. Each cell in the GFSAD layer measures 1 square kilometer. Africa measures approximately 31 million square kilometers; therefore, by converting your raster layer to points, you will end up with approximately 31 million points. This conversion may take some time to process in ArcGIS Pro depending on your computer's processing power.

h In the Geoprocessing pane, search for and open the **Raster To Point** (Conversion Tools) tool.

i For Input Raster, choose GFSAD_Maize_Clip.

j For Field, confirm that Value is selected.

k For Output Point Features, type **PredictionPoints**.

l At the bottom of the map, click the Pause Drawing button ‖.

*Step 1l***: Create prediction points.*

By pausing the map display, you can run the tool but not draw the 31 million points on the map. You only need the points as input for the model—you do not need to visualize them on the map.

Note: Running the Raster To Point tool can take up to 10 minutes, depending on your computer's processing power. **If you do NOT want to run the Raster To Point tool, you can use the PredictionPoints layer found in the Results.gdb file of your project.**

m If you would like to run the Raster To Point tool, click Run.

n After the tool finishes running, in the Contents pane, turn off the PredictionPoints layer.

o At the bottom of the map, click the Pause Drawing button to resume the map display.

You have created the prediction points for your presence-only prediction model.

p Save your project.

- **Step 2: Add bioclimate data to prediction points**

After creating prediction points for the presence-only prediction model, explanatory variables need to be added to each point. The training dataset only uses baseline variables as explanatory variables to establish the baseline presence for the model. The dataset used for prediction uses both baseline and projected variables so that the model can predict future suitability.

In this step, you will finalize your prediction points by adding both the baseline bioclimate variables and the projected bioclimate variables from the climate scenario for 2050. To create the training dataset, you will use the same workflow that you used to extract the baseline bioclimate variable values to your presence points.

a In the Geoprocessing pane, search for and open the **Extract Multi Values To Points** (Spatial Analyst Tools) tool.

b For Input Point Features, choose PredictionPoints.

c For Input Rasters, click the Browse button .

d On the left, under Project, click Folders, and then double-click the MaizePredictionModeling folder.

e Double-click the BioclimateBaselineVariables folder.

f Click BioclimateBaseline01.crf, press Shift on your keyboard, and then click BioclimateBaseline19.crf.

g Click OK.

You have added all 19 baseline bioclimate variables.

Next, you will add the bioclimate variables for 2050.

h In the Geoprocessing pane, scroll to the bottom of the Input Rasters list.

i In the empty row, click the Browse button, as indicated in the following graphic.

j In the Input Rasters dialog box, go to the BioclimateProjectionVariables2050 folder.

- Hint

Click Folders, double-click MaizePredictionModeling, and then double-click BioclimateProjectionVariables2050.

k Click BioclimateProjections01_2050.crf.

l Press Shift on your keyboard and click BioclimateProjections19_2050.crf

m Click OK.

All the baseline and projected bioclimate variables are added to the geoprocessing pane. By adding all the bioclimate variables to the prediction points, you have a complete prediction points layer to model future maize suitability using the Presence-only Prediction (MaxEnt) tool.

Note: Running the Extract Multi Values To Points tool can take up to 20 minutes, depending on your computer's processing power. **If you do NOT want to run the Extract Multi Values To Points tool, you can use the PredictionPoints layer found in the Results.gdb of your project. However, if you added the PredictionPoints layer from the Results.gdb in the previous step, do NOT add it again in this step—you already have all the bioclimate variables required to run your model.**

n If you would like to run the Extract Multi Values To Points tool, click Run.

The PredictionPoints layer now contains all the necessary bioclimate variables and is ready for your prediction model.

o Turn off the GFSAD_Maize_Clip layer and save your project.

- Step 3: Train the prediction model

Before performing the prediction, the model must be trained using the baseline variables.

In this step, you will train the model using the training dataset, which includes the maize presence points with the baseline bioclimate variables.

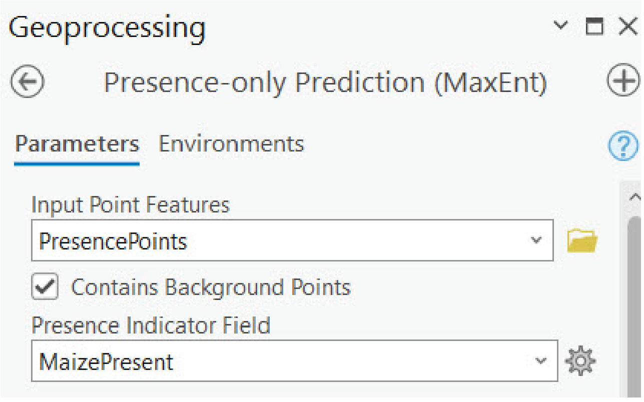
a In the Geoprocessing pane, search for and open the **Presence-only Prediction (MaxEnt)** (Spatial Statistics Tools) tool.

The input point features are the known areas of maize presence in the training dataset that you created for the study area. Because you already prepared your data by reclassifying values to identify where maize is present, you can enable the option that indicates that the training point dataset contains background points.

b For Input Point Features, choose PresencePoints.

c Check the box for Contains Background Points.

d For Presence Indicator Field, choose MaizePresent.

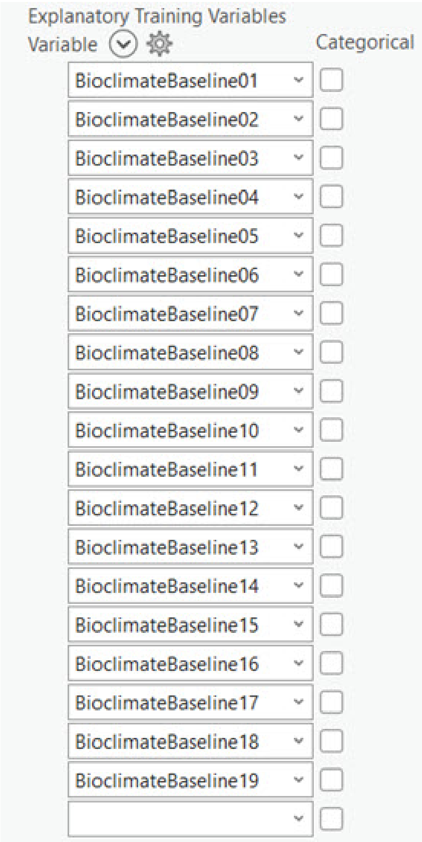


*Step 3d***: Train the prediction model.*

To train the model, you will use the 19 baseline bioclimate variables as the explanatory training variables.

e Under Explanatory Training Variables, click the Add Many button.

f Check the box for each of the 19 BioclimateBaseline variables listed and click Add.



*Step 3f***: Train the prediction model.*

You have added the baseline bioclimate variables to train the model.

You will now set the explanatory variable expansions and spatial thinning parameters.

Explanatory Variable Expansions specifies the basis function to be used to transform the explanatory variables for the model. Basis functions incorporate more intricate relationship forms into the model. If multiple basis functions are selected, the tool will run through all of them and use the variable that creates the best prediction.

Spatial thinning reduces sampling bias in your model when points are clustered. Highly clustered points can cause sampling bias by artificially inflating model accuracy scores.

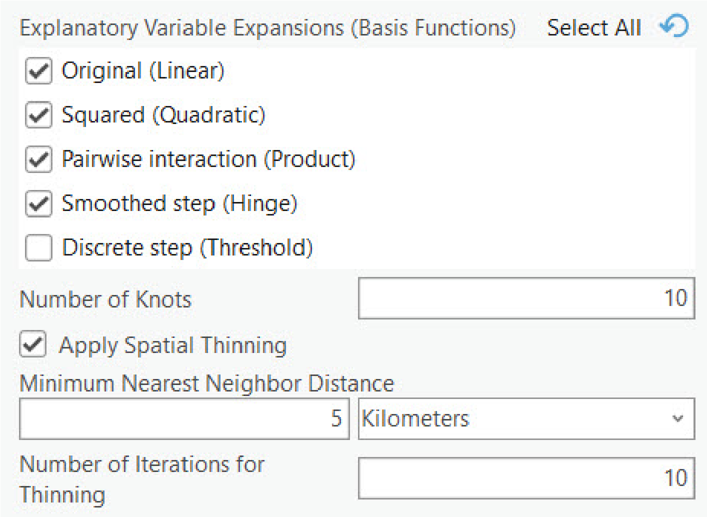
g Under Explanatory Variable Expansions (Basis Functions), ensure that Original (Linear) is checked, and then check the box for the following additional options:

- Squared (Quadratic)
- Pairwise Interaction (Product)
- Smoothed Step (Hinge)

h For Apply Spatial Thinning, check the box.

i For Minimum Nearest Neighbor Distance, in the first field, type **5**.

j In the second field, click the down arrow and choose Kilometers.

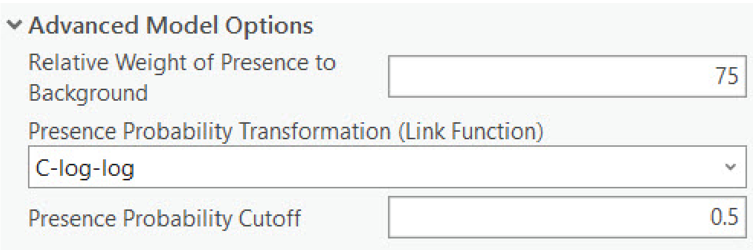


*Step 3j***: Train the prediction model.*

k Expand Advanced Model Options.

The Relative Weight Of Presence To Background parameter should be a value between 1 and 100, with 100 being the default. A higher value indicates that presence points are the primary source of information for training the model. A lower value indicates that background points also contribute valuable information. The proportion of background points to presence points can significantly affect prediction results. Creating a prediction model is an iterative process that requires you to adjust parameters, such as the proportion of background points to presence points, until you have a model that is suitable for your research.

l For Relative Weight Of Presence To Background, delete 100 and type **75**.

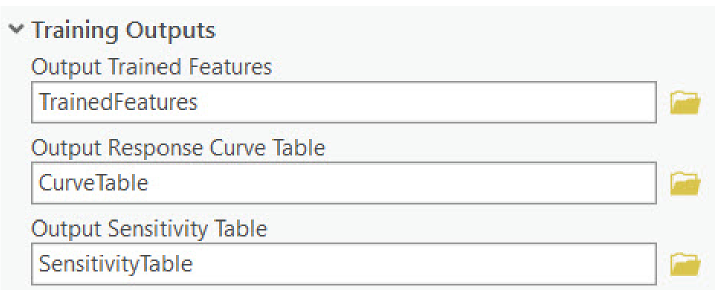


*Step 3l***: Train the prediction model.*

The Training Outputs section allow you to create outputs that can be used to review the diagnostics and reliability of the prediction model. You will want to review these diagnostics, so you will name the training outputs.

m Expand Training Outputs and set the following parameters:

- For Output Trained Features, type **TrainedFeatures**.
- For Output Response Curve Table, type **CurveTable**.
- For Output Sensitivity Table, type **SensitivityTable**.



*Step 3m***: Train the prediction model.*

n Expand Prediction Options.

The Prediction Options section allows you to create the output layers that show the baseline and future suitability for maize based on your trained model. The output prediction feature is a point feature layer that will be added to the map after the tool runs and will show baseline maize suitability because you are using the baseline bioclimate variables as the explanatory variables.

o For Input Prediction Features, choose PredictionPoints.

p For Output Prediction Features, type **BaselineMaizeSuitability**.

q Under Match Explanatory Variables, confirm that the Bioclimate Baseline Prediction and Training variables match.

*Step 3q***: Train the prediction model.*

Note: Training the model can take up to 30 minutes, depending on your computer's processing power. **If you do NOT want to train the model, read through the next exercise step, *Run the prediction model*, and follow the instructions to add the results to your project.**

r If you would like to train the model, click the Pause Drawing button .

s Click Run.

t After the model finishes running, in the Contents pane, turn off the BaselineMaizeSuitability layer.

The model has run successfully, but there is a warning. In the Geoprocessing pane, you can click the View Details option to opens the tool's Message window. The Message window provides information about the warning and includes the diagnostic tables for the model. Reviewing your model's diagnostics gives you a better idea of how your prediction model is performing and how you might fine tune it.

For more information on model diagnostics for the Presence-only Prediction (MaxEnt) tool, see ArcGIS Pro Help: How Presence-only Prediction (MaxEnt) works.

u Click the Pause Drawing button to resume the map display.

You have trained your prediction model using the baseline bioclimate variables.

v Save your project.

- Step 4: Run the prediction model

After the prediction model has been trained, the tool is rerun using the projected explanatory variables to create a prediction output.

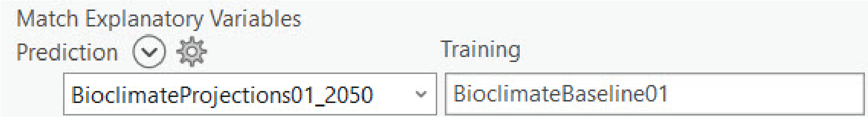
In this step, you will rerun the model using the projected bioclimate variables from the SSP3 7.0 climate scenario for 2050. The prediction output will model maize suitability for 2050.

a In the Prediction Options section, for Output Prediction Features, delete BaselineMaizeSuitability and type **FutureMaizeSuitability**.

You will update the prediction explanatory variable to use the 2050 bioclimate projection variables.

b For Match Explanatory Variables, under Prediction, click the down arrow for BioclimateBaseline01.

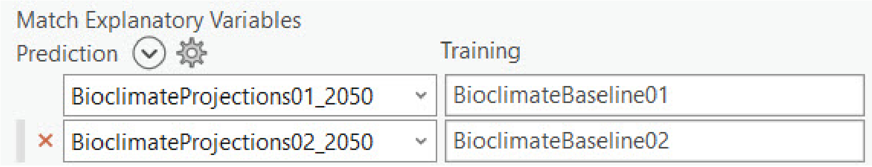
c Scroll down the list of variables to find and select BioclimateProjections01_2050.



*Step 4c***: Run the prediction model.*

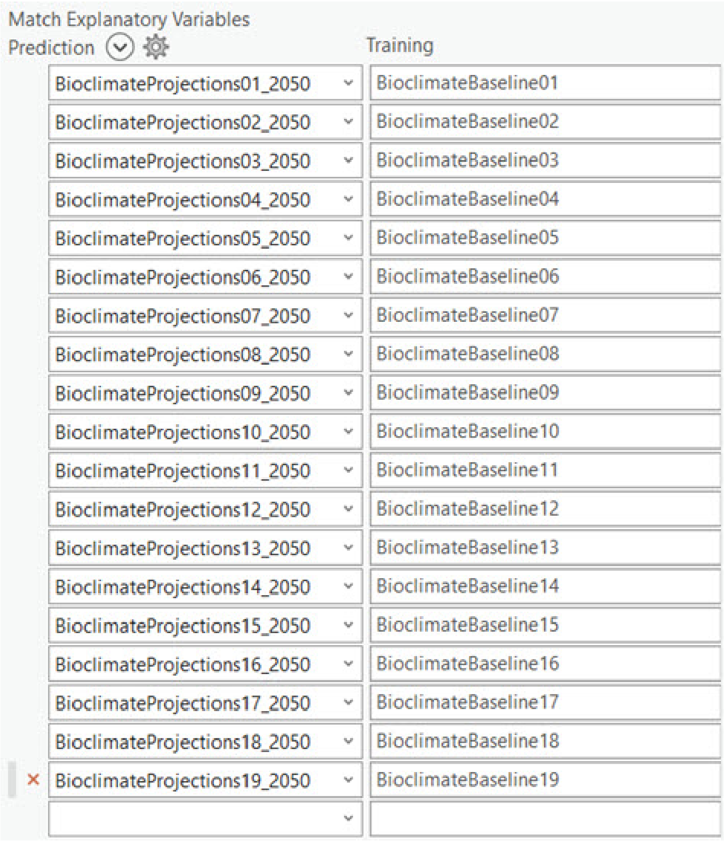
You will complete this step for all 19 variables to match the prediction variables with the baseline variables that you used to train the prediction model.

d Under Prediction, click the down arrow for BioclimateBaseline02 and select BioclimateProjections02_2050.



*Step 4d***: Run the prediction model.*

e Change the Prediction variable for the remaining bioclimate projection variables for 2050.



*Step 4e***: Run the prediction model.*

All the variables for 2050 that you will use to predict future maize suitability are matched with the baseline bioclimate variables that were used to train the model. You are ready to run to tool.

Note: Running the model can take up to 30 minutes, depending on your computer's processing power. **If you decided NOT to train the model in the previous step and will NOT run the prediction model in this step, follow these instructions to add the results to your map:**

- In the Catalog pane, expand Databases.
- Expand Results.gdb.
- Right-click FutureMaizeSuitability_Raster and choose Add To Current Map.
- Right-click BaselineMaizeSuitability_Raster and choose Add To Current Map.

f If you would like to run the model, click the Pause Drawing button, and then click Run.

g After the model finishes running, in the Contents pane, turn off the FutureMaizeSuitability layer.

h Click the Pause Drawing button to resume the map display.

You have created a prediction output for future maize suitability that can help agricultural ministers plan for future climate change and safeguard farmers and food security.

i Save your project.

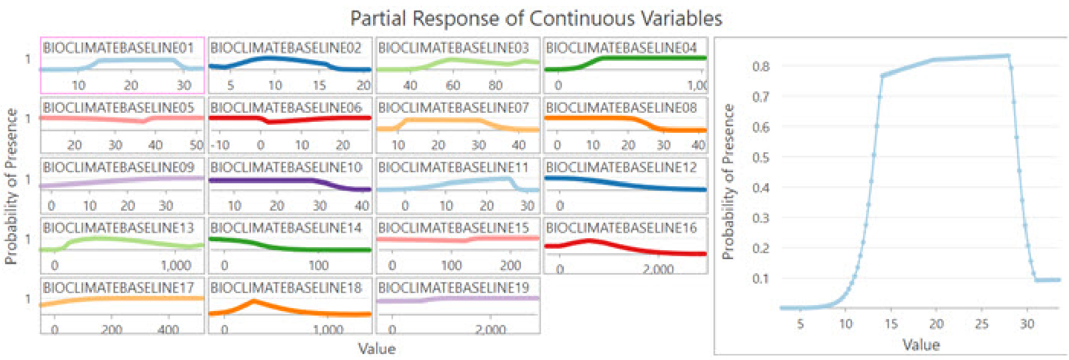
- Step 5: Review model diagnostics

When you run a model using the Presence-Only Prediction (MaxEnt) tool, a series of charts and tables are generated that display the performance diagnostics of your model. It is important to review these performance diagnostics to get a sense of how accurately your model is able to predict.

In this step, you will access two of these charts to review the performance diagnostics of your model.

Note: **If you chose NOT to run the model, read through the following steps and use the graphics provided (Partial Response Of Continuous Variables and Classification Result Percentages) to examine your model's performance diagnostics.**

a At the bottom of the Contents pane, under Standalone Tables, open Partial Response Of Continuous Variables.

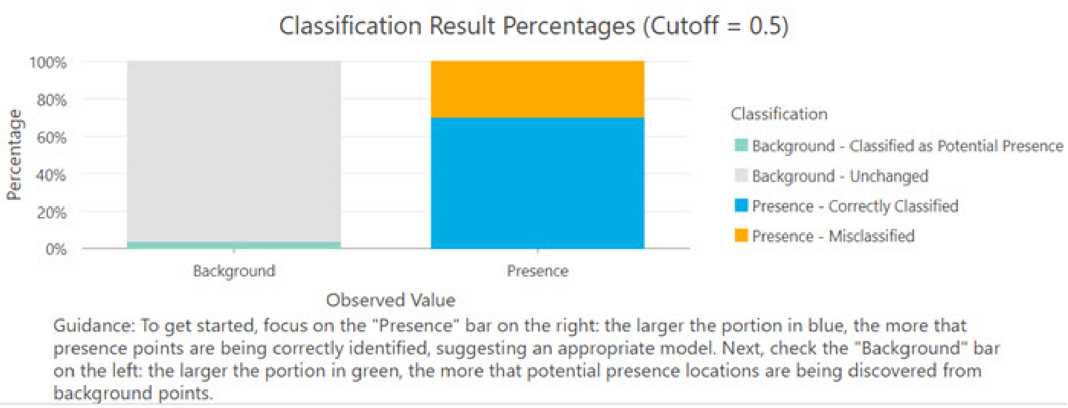


The Partial Response Of Continuous Variables chart summarizes and visualizes the impact of each bioclimate variable on the probability of presence of maize. You can click individual charts to explore the effect of a specific bioclimate variable in more detail.

The variable currently highlighted is baseline bioclimate variable 01, which represents annual mean temperature. The presence probability of maize will be low if the annual mean temperature is too cold or too hot. The presence probability of maize increases quickly from 0°C to approximately 15°C, stays stable and slightly increases from approximately 15°C to 30°C, and then declines rapidly as annual mean temperatures increase above 30°C. These charts are helpful for identifying suitable and unsuitable bioclimates for maize based on a single variable.

You will now review a second chart.

b In the Contents pane, under Trained Feature, double-click Classification Result Percentages (Cutoff = 0.5) to open the chart.

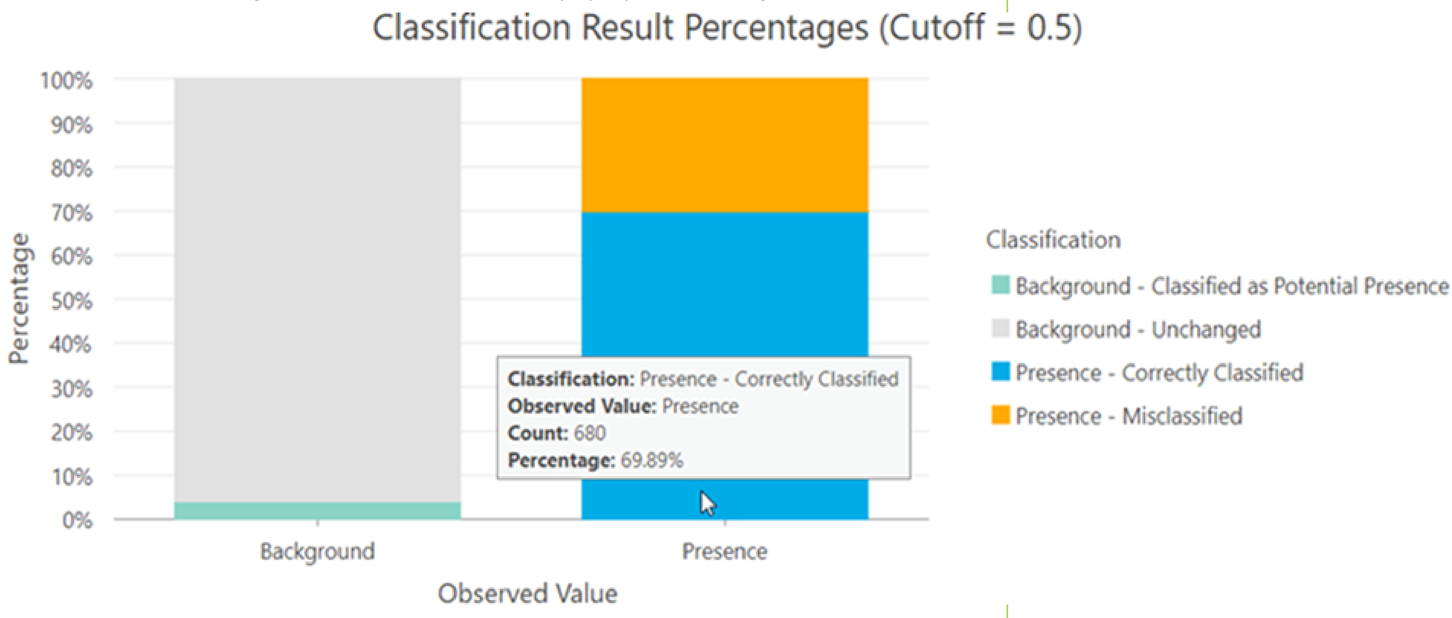


The Classification Result Percentages chart compares the observed and predicted classifications. You can use this chart as a training assessment to check how your prediction model performs when recognizing presence. The prediction model's accuracy is measured as the percentage of presence points that were correctly classified by the model.

c Point to the Presence - Correctly Classified classification portion of the bar chart, shown in blue, and use the pop-up to answer the following question.

What percentage of the presence points was correctly classified?

(Hint: Click the following View Result link to view the pop-up, if necessary.)



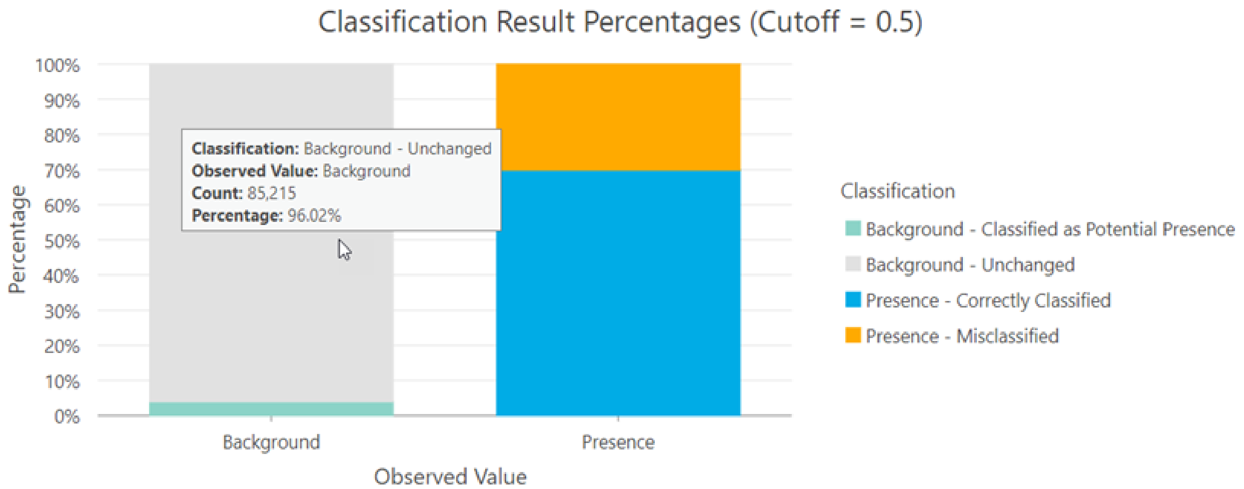
*Step 5c***: Review model diagnostics.*

- Answer
About 70 percent of known presence locations were correctly classified by the model.

For this purposes of this exercise, 70 percent is a reasonable performance number for the model. Creating predictions models is an iterative process; you can change the parameters of the model based of the model's diagnostics or add new data sources to help enhance the performance of your model. How accurate you need your model to be is based on the spatial analysis question that you are trying to answer.

d Now point to the Background- Unchanged classification of the bar chart, shown in gray.

(Hint: Click the following View Result link to view the pop-up, if necessary.)



Guidance: To get started, focus on the "Presence" bar on the right: the larger the portion in blue, the more that presence points are being correctly identified, suggesting an appropriate model. Next, check the "Background" bar on the left: the larger the portion in green, the more that potential presence locations are being discovered from background points.

*Step 5d***: Review model diagnostics.*

The Background - Unchanged classification of the bar chart confirms that a lack of presence was observed and that, according to the prediction model, no presence should be observed in those areas. If the Background - Unchanged classification and the Presence - Correctly Classified classification are the largest portions of your chart, then your prediction model is performing well.

e Close both charts and save your project.

In this step, you used two charts to review the performance diagnostics of your model

- **Step 6: Convert point features to raster**

Reviewing the performance diagnostics for a model allows you to confirm that the model is performing as intended and that the results are valid for use. After reviewing the performance diagnostics for your maize suitability model, you have decided that the results are valid.

Because the baseline and future maize suitability layers are both point feature layers containing 31 million points, you want to convert them to raster layers, a format that will be easier to visualize and share. In this step, you will use the Point To Raster tool to convert both of the presence-only prediction output layers from point feature layers to rasters.

a In the Geoprocessing pane, search for and open the **Point To Raster** (Conversion Tools) tool.

Note: **If you chose NOT to run the model and instead added the results to your map, read through the instructions in this step but do NOT run the Point To Raster tool; you already have the final raster layers that you need to compare the results in the next step.**

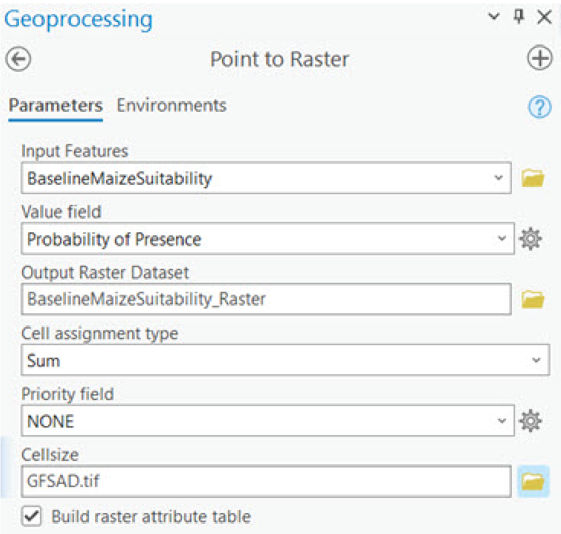
b Set or confirm the following parameters:

- Input Features: BaselineMaizeSuitability
- Value Field: Probability Of Presence
- Output Raster Dataset: **BaselineMaizeSuitability_Raster**
- Cell Assignment Type: Sum
- Priority Field: NONE

c For Cellsize, click the Browse button .

d Under Project, double-click Folders.

e Double-click the GFSAD folder, select GFSAD.tif, and click OK.



*Step 6e***: Convert point features to raster.*

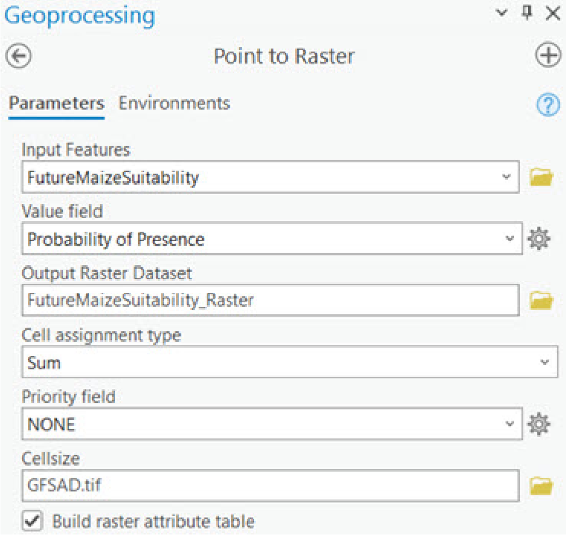
Value Field indicates which field in the attribute table will be used to assign values to the output raster layer. Probability Of Presence classifies the probability of the presence of maize for each cell in the layer. Cell Assignment Type lets you choose the best method of determining how the cell value is assigned when more than one feature falls within the cell. For Cellsize, you used the GFSAD layer to indicate that the cell size for the new raster is 1 square kilometer.

f Click Run.

You will now convert the FutureMaizeSuitability layer.

g In the Geoprocessing pane, for Input Features, click the down arrow and select FutureMaizeSuitability.

h For Output Raster Dataset replace BaselineMaizeSuitability_Raster by typing **FutureMaizeSuitability_Raster**.



*Step 6h***: Convert point features to raster.*

It is important to leave all the other parameters the same because you are comparing the baseline and future layers.

i Click Run.

j After the tool finishes running, in the Contents pane, turn off the TrainedFeatures layer.

k Save your project.

You have converted the baseline and future maize suitability layers from point feature layers to raster layers to facilitate visualizing and sharing the results with the agricultural ministers in Africa.

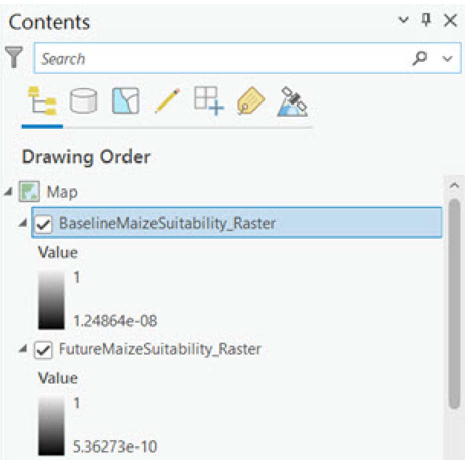
- Step 7: Compare the results

The goal of using a presence-only prediction model is to compare the change between baseline and projected suitability. However, when you converted your prediction output layers from point feature to rasters layers in the previous step, the layers were added to your map using symbology that is difficult to interpret.

In this step, you will update the color scheme of the symbology to make it easier to compare the results and visualize the change in maize suitability in Africa. In preparation, you will first organize the layers in the Contents pane.

a In the Contents pane, click and drag the FutureMaizeSuitability_Raster layer to the top of the drawing order.

b Click and drag the BaselineMaizeSuitability_Raster layer above the FutureMaizeSuitability_Raster layer.

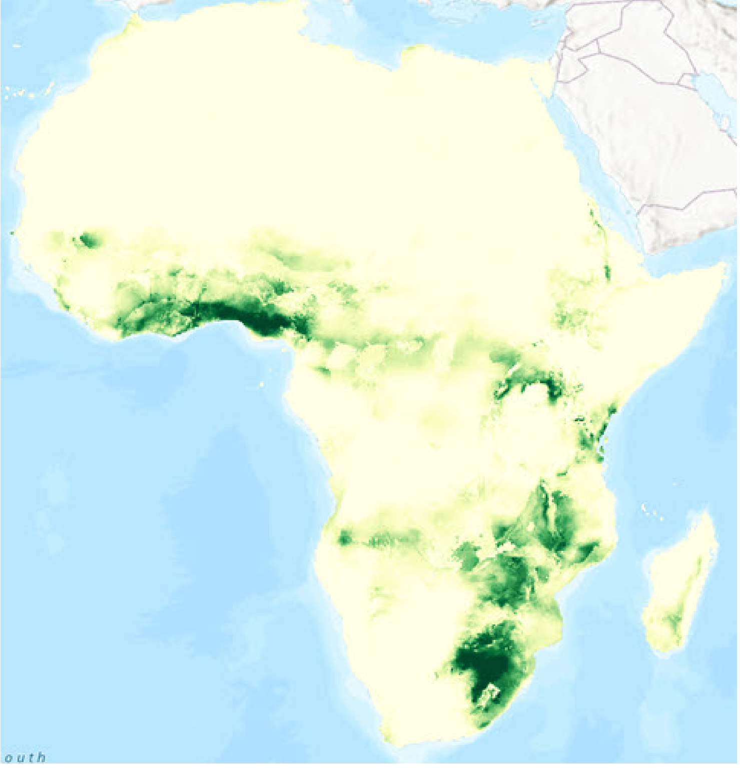


*Step 7b***: Compare the results.*

c Right-click the BaselineMaizeSuitability_Raster layer and choose Symbology.

d In the Symbology pane, for Color Scheme, click the down arrow.

e Check Show Names, and then select the Yellow-Green (Continuous) color scheme.



*Step 7e***: Compare the results.*

The yellow areas on the map represent the baseline, or current, areas that are not suitable for growing maize. The green areas represents the areas that are suitable for growing maize, with the darkest green areas being the most suitable.

To compare the baseline results with the future results, you will use the same color scheme for the FutureMaizeSuitability_Raster layer.

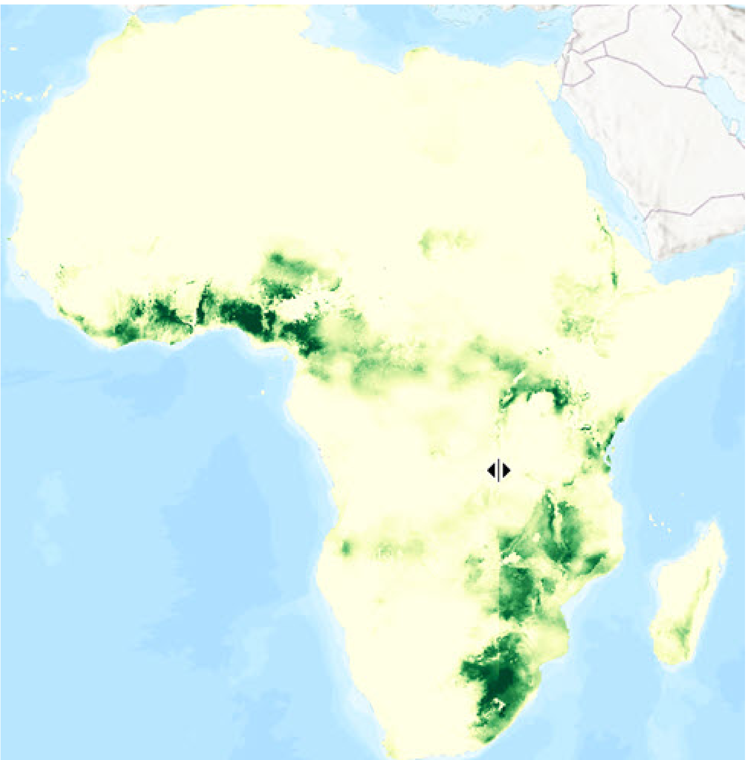
f Apply the Yellow-Green (Continuous) color scheme to the FutureMaizeSuitability_Raster layer.

With both the baseline and future suitability layers symbolized using the same color scheme, you can now use the Swipe tool to visually compare change in suitable maize locations in Africa.

g In the Contents pane, select the BaselineMaizeSuitability_Raster layer.

h On the ribbon, click the Raster Layer tab and, in the Compare group, click Swipe.

i Click anywhere on right side of the map, and then drag your mouse from right to left to see the changes.



*Step 7i***: Compare the results.*

Comparing the baseline and future results shows that, for the SSP3 7.0 climate scenario, suitable land area for maize will shrink significantly and shift north and east across the continent by 2050. The loss of even small areas of crop land can put the food security of a country at risk. Prediction modeling allows forecasting and planning for such changes to minimize the impact of climate change, especially for vulnerable populations that depend on smallholder agriculture as their main source of income.

j Save your project.

You have explored the results of your prediction model by comparing baseline and future suitability for growing maize in Africa.

- Step 8: Share your results to ArcGIS Online

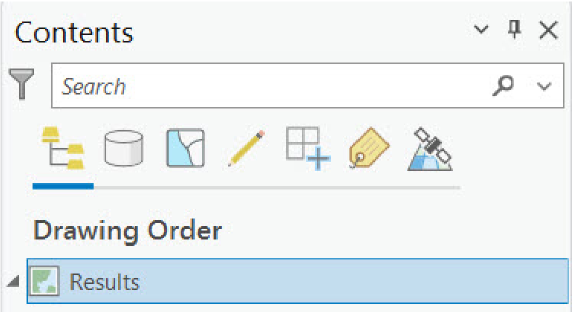
You want to share your results with agricultural ministers in Africa so that they can better plan for future climate change and ensure food security. Sharing your results from ArcGIS Pro to ArcGIS Online will allow you to share your map as a link that anyone can access on the web.

In this step, you will share your results as a web map to ArcGIS Online. You will begin by creating a new map in your project and adding only the layers that you need to share with the agricultural ministers.

a In the Catalog pane, right-click Maps and choose New Map.

To keep your project organized, you will rename the new map view.

b In the Contents pane, click Map, type **Results**, and press Enter.



*Step 8b***: Share your results to ArcGIS Online.*

You will now copy and paste the two maize suitability result layers from the Prediction Modeling map to the Results map.

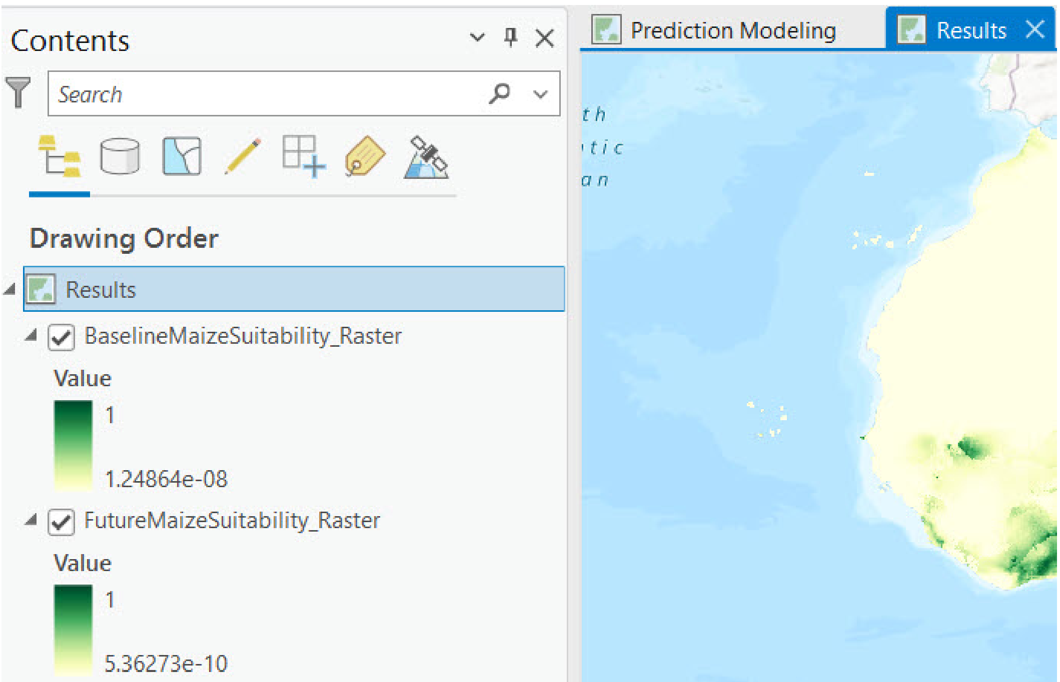
c At the top of the map, click the Prediction Modeling map tab.

d In the Contents pane, right-click FutureMaizeSuitability_Raster and choose Copy.

e Click the Results map tab.

f In the Contents pane, right-click Results and choose Paste.

g Using the same process, copy and paste the BaselineMaizeSuitability_Raster to the Results map.



*Step 8g***: Share your results to ArcGIS Online.*

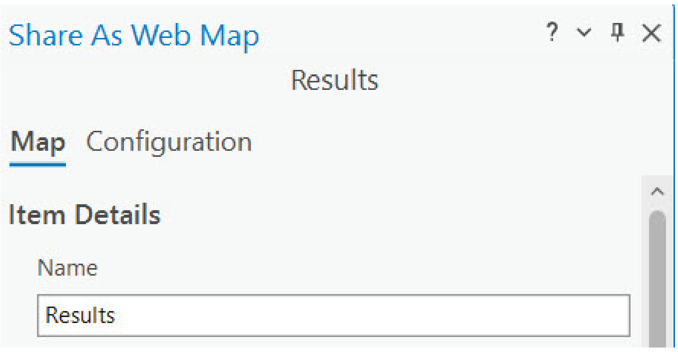
You are now ready to share your results to ArcGIS Online.

h Save your project.

i On the ribbon, click the Share tab.

j In the Share As group, click Web Map.

k In the Share As Web Map pane, confirm that you are sharing the Results map.



*Step 8k***: Share your results to ArcGIS Online.*

You will fill out the item details for the Results map before sharing it to ArcGIS Online.

l In the Share As Web Map pane, enter the following details:

- Name: **Prediction modeling in Africa <your first and last name>**
- Summary: **A web map showing baseline and future areas in Africa that are suitable to grow maize based on a prediction model created using the Presence-Only Prediction (MaxEnt) tool in ArcGIS Pro.**
- Tags: **Africa, Maize suitability, Climate change**

m For Share With, check Everyone.

You have filled out the item details and are now ready to analyze your map for sharing.

n Click Analyze.

o Resolve any errors found for your map.

- Hint

Double-click the the 00230 error message to expand the message, and then double-click the first error listed. Now double-click the second 00230 error message listed. Click Analyze again to also clear the 00079 error.

p Click Share.

Note: Because of the size of the imagery layers, it might take a couple of minutes to share your map.

You have successfully shared your Results map as a web map to ArcGIS Online. You can now share this web map with agricultural ministers or embed it in a web application.

q Save your project and exit ArcGIS Pro.

In this exercise, you trained a prediction model using baseline bioclimate variables to identify current maize suitability in Africa. Then, you used your trained model to create a prediction model for future maize suitability based on the SSP3 7.0 climate scenario for 2050. Finally, you created a web map by sharing your map from ArcGIS Pro to ArcGIS Online.