Learn ArcGIS

Guided lessons based on real-world problems

Interpolate Temperatures Using the Geostatistical Wizard

By Witold Frączek and Heather Smith

learn.arcgis.com/

380 New York Street
Redlands, California 92373 – 8100 USA

Interpolate Temperatures Using the Geostatistical Wizard

Time: 30 minutes

Overview

Many natural phenomena occur continuously across a landscape, for example, the distribution of nutrients in soil, dissolved oxygen in seawater, or rainfall. It is unrealistic to take measurements of these phenomena everywhere; you can only collect measurements at a set of sample locations. Using geostatistics, these sample measurements can be used to predict values at other, unmonitored locations. Many geostatistical methods also allow you to measure the uncertainty of your predictions.

Geostatistics is used widely for mapping natural phenomena in many fields, including meteorology, oceanography, geology, forestry, and the soil sciences. It is also applied to unnatural phenomena such as pollution. You have likely seen and used many maps that were created with the help of geostatistics.

Note: The Geostatistical Analyst license is required to complete this lesson.

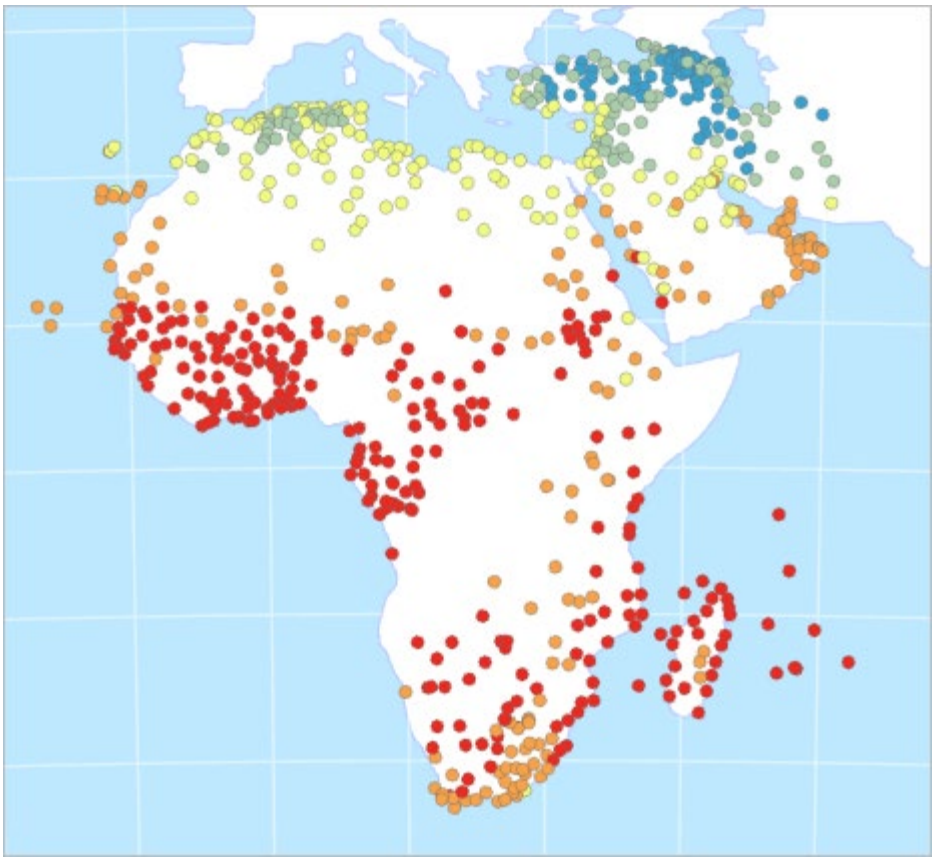
In this lesson you will learn to do the following:

- Create a continuous temperature map of Africa and the Middle East from a dataset of sample points.
- Analyze the distribution of the sample data with histograms.
- Create interpolated surfaces using the inverse distance weighting and kriging methods.
- Compare the accuracy of different surfaces using cross-validation.
- Map the standard error of prediction for a chosen surface.

Create histograms of data distribution

1. Download the InterpolateTemperatures project package.
2. Locate the downloaded file on your computer. Double-click **InterpolateTemperatures.ppkx** to open it in ArcGIS Pro.

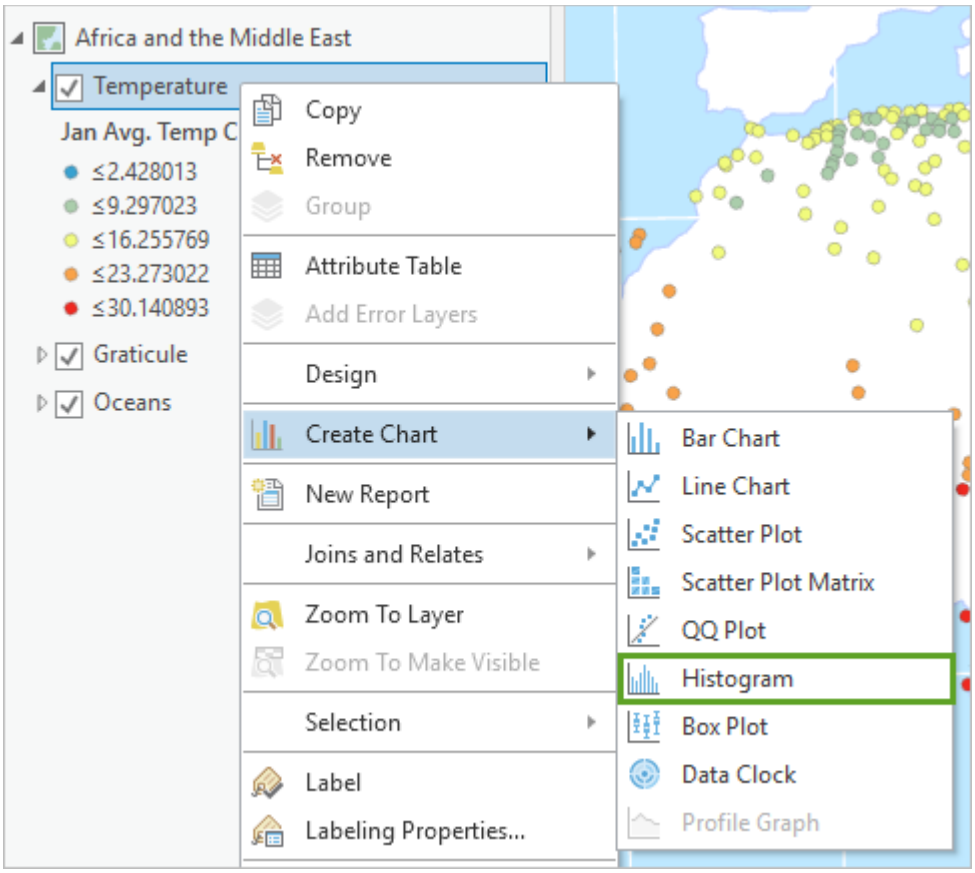
   Note: If you don't have ArcGIS Pro or an ArcGIS account, you can sign up for an ArcGIS free trial.



The points on the map represent temperature samples. Each point stores average temperature values for each month. You will examine the data distribution of some of these fields to determine which to use for interpolation.

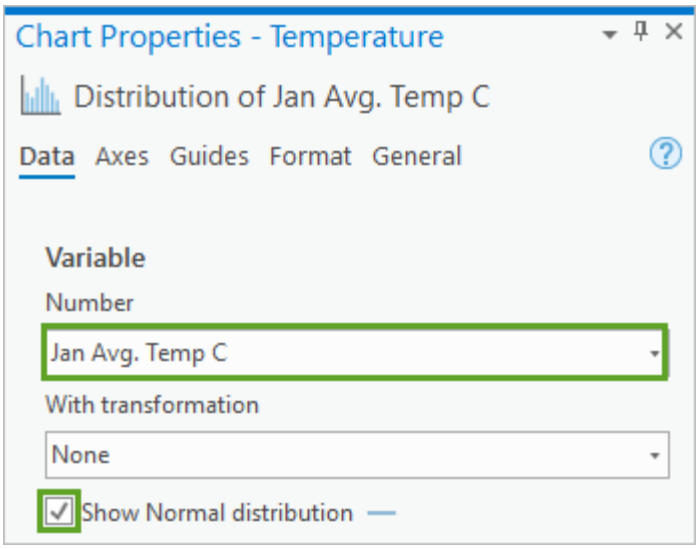
Note: You can find the full dataset on the Living Atlas: [World Historical Climate – Monthly Averages for GHCND Stations for 1984 - 2010.]

3. In the **Contents** pane, right-click **Temperature**. Point to **Create Chart** and choose **Histogram**.

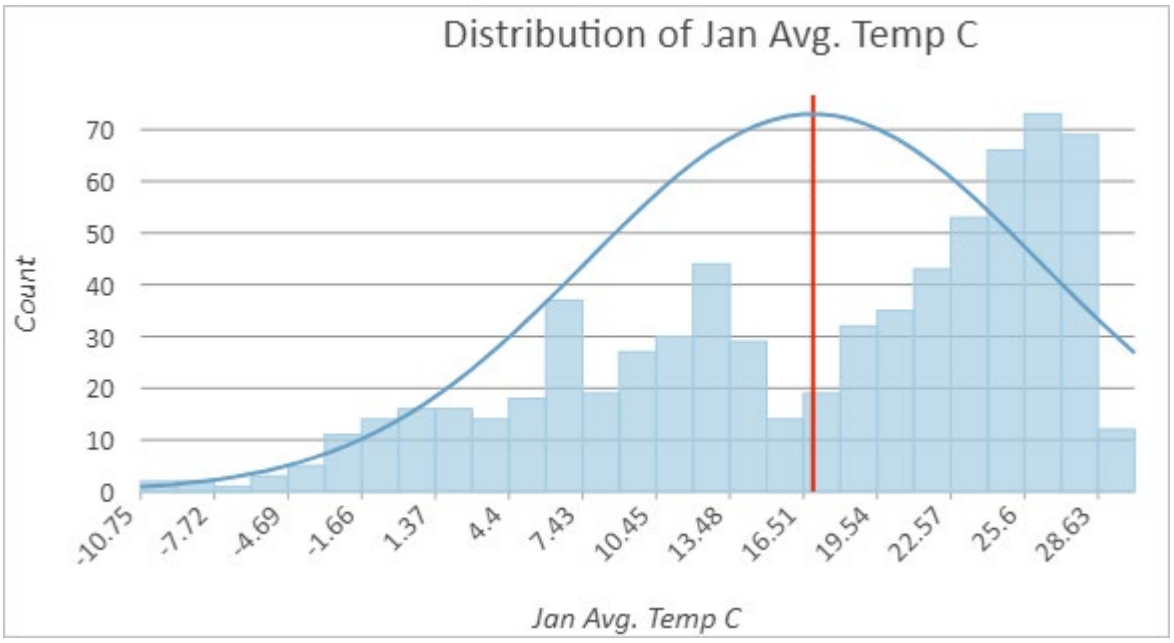


Both the **Chart Properties** pane and an empty chart view appear.

4. In the **Chart Properties** pane, change **Number** to **Jan Avg. Temp C** (short for January Average Temperature in Celsius) and check the box for **Show Normal distribution**.

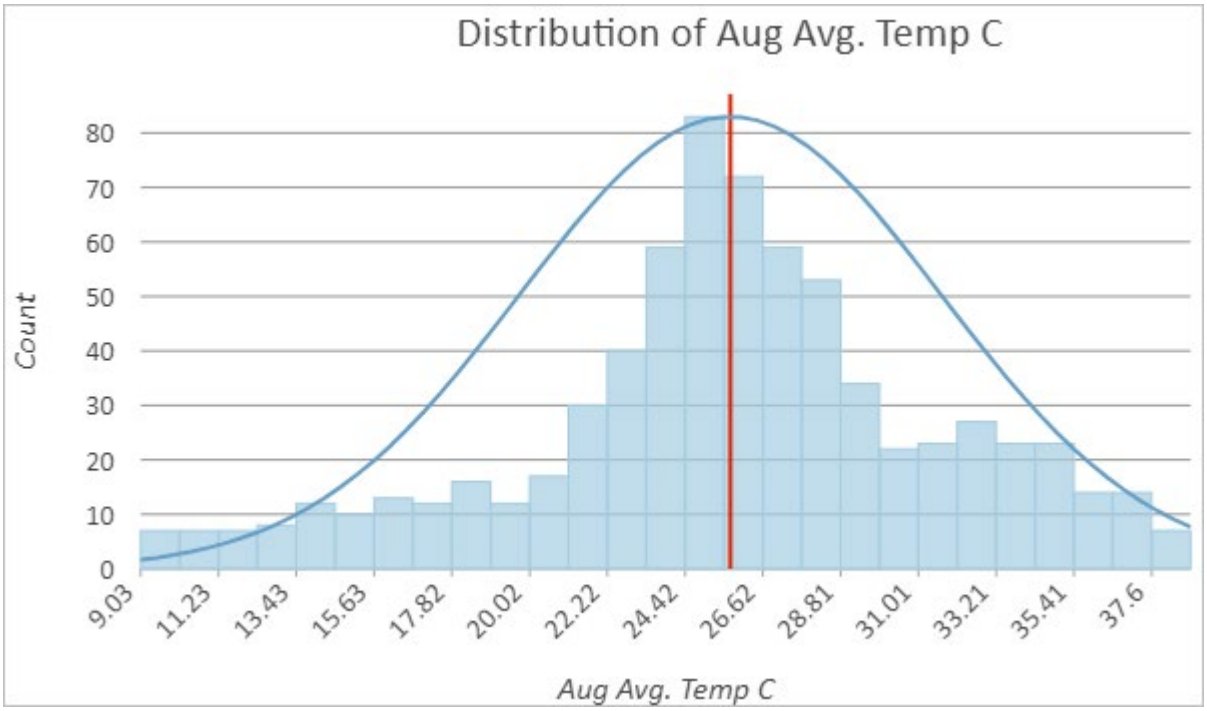


The chart view updates to show a histogram representing the maximum temperature values from the point data. You can see that the values range from -10.2 to 30.1° Celsius. The values shown on the axis may vary, depending on the width of the pane.



The curved blue line represents the normal distribution of the chart. Data with a normal distribution has a bell-shaped curve. You can see that the distribution of average temperatures in January is not normal, but rather it is skewed to the right.

5. In the **Chart Properties** pane, change **Number** to **Aug Avg. Temp C**. The histogram updates to the new field.



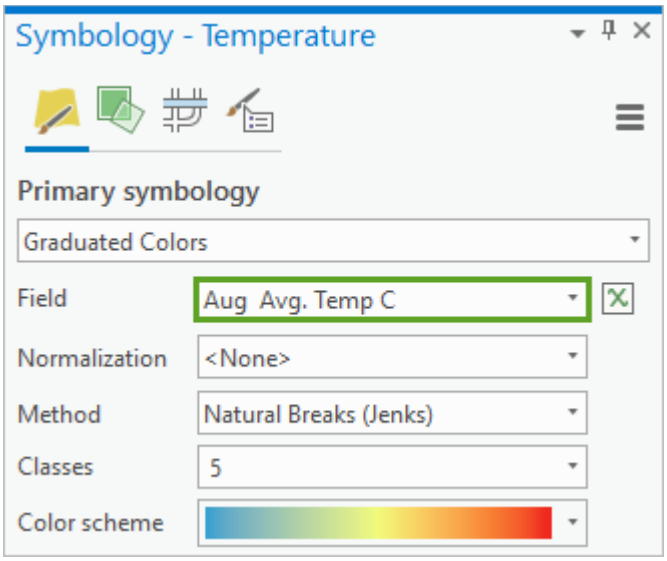
Temperatures in August have more of a normal distribution. Interpolation methods are most effective when the data is close to a normal (bell-shaped) distribution, and some geostatistical methods require that the data be normally distributed. For this reason, you will use **Aug Avg. Temp C** for the rest of this lesson.

Note: If your data does not follow a bell curve, you can apply a transformation to make it closer to a normal distribution. Read about this process at Box-Cox, arcsine, and log transformations.

6. Close the chart view.

7. On the **Contents** pane, right-click **Temperature** and choose **Symbology**.

8. The **Symbology** pane appears. Change **Field** to **Aug Avg. Temp C**.



The map updates to show temperatures for August.

Create geostatistical surfaces using inverse distance weighting

Next, you will create surfaces of predicted temperature values for all of Africa and the Middle East using the sample data. You are able to do this because geostatistics makes the assumption that things that are closer together are more alike than things that are far apart. Therefore any unknown location is probably going to have a similar value to the known locations nearest to it.

The Geostatistical Wizard in ArcGIS Pro offers many different interpolation methods for creating predicted surfaces. Usually you will not know which one to use until you have tried several and compared their results. The first method you will try is inverse distance weighting, also sometimes called IDW.

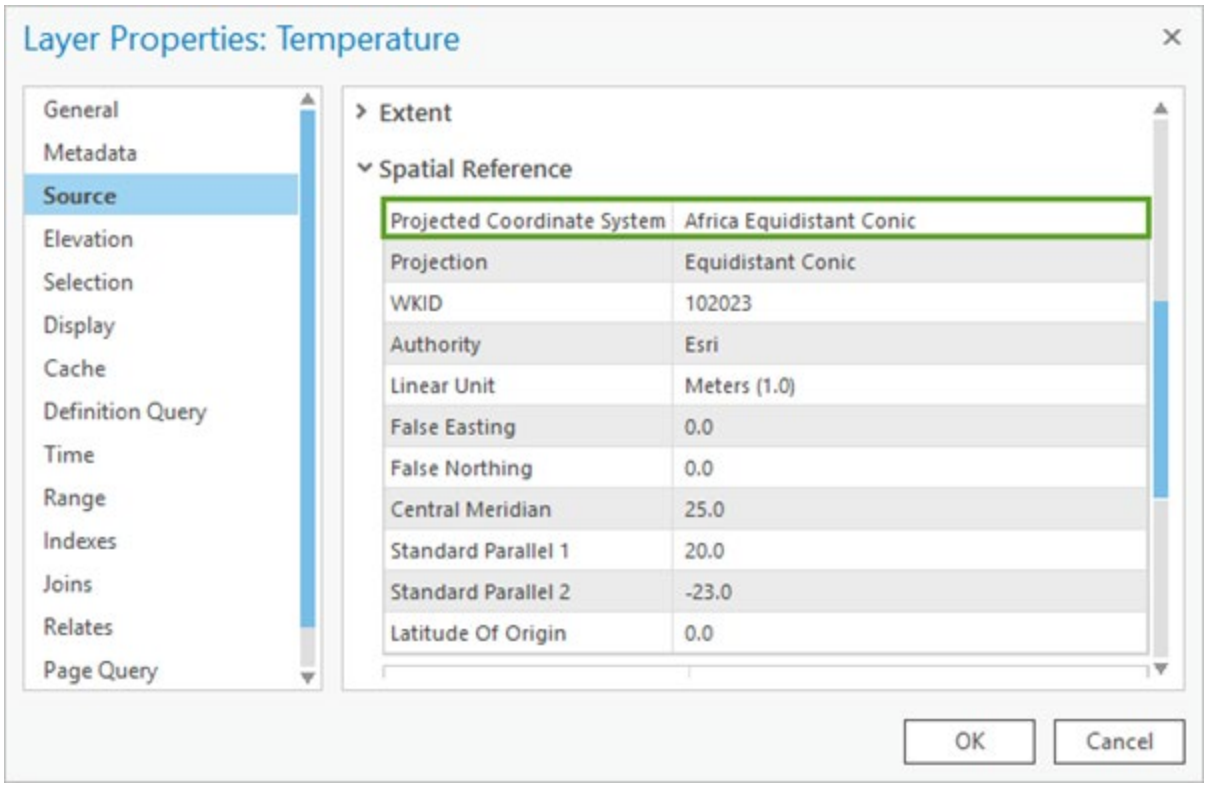
IDW is an exact method. This means that the resulting surface will not vary from the sample values. It is also one of the simpler methods to execute. You can read more about IDW at [How inverse distance weighted interpolation works](#).

1. In the **Contents** pane, right-click **Temperature** and choose **Properties**.

2. Click the **Source** tab.

3. Scroll down and click **Spatial Reference** to expand that section.

   The table should begin with **Projected Coordinate System**.

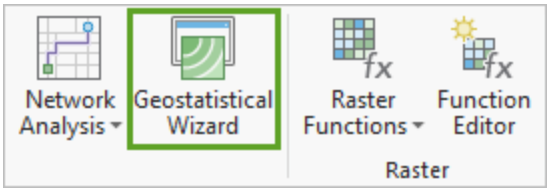


Geostatistics relies on distance measurements. To minimize the distortion of these distances, your input data must use a projected (rather than geographic) coordinate system. You can give it one using the **Project** geoprocessing tool.

This data uses an Equidistant Conic projection centered on Africa. There is no projection that can perfectly preserve all distances on your map, but equidistant projections will do a better job of this than others. The choice of projection is more important when mapping a large area, such as a continent.

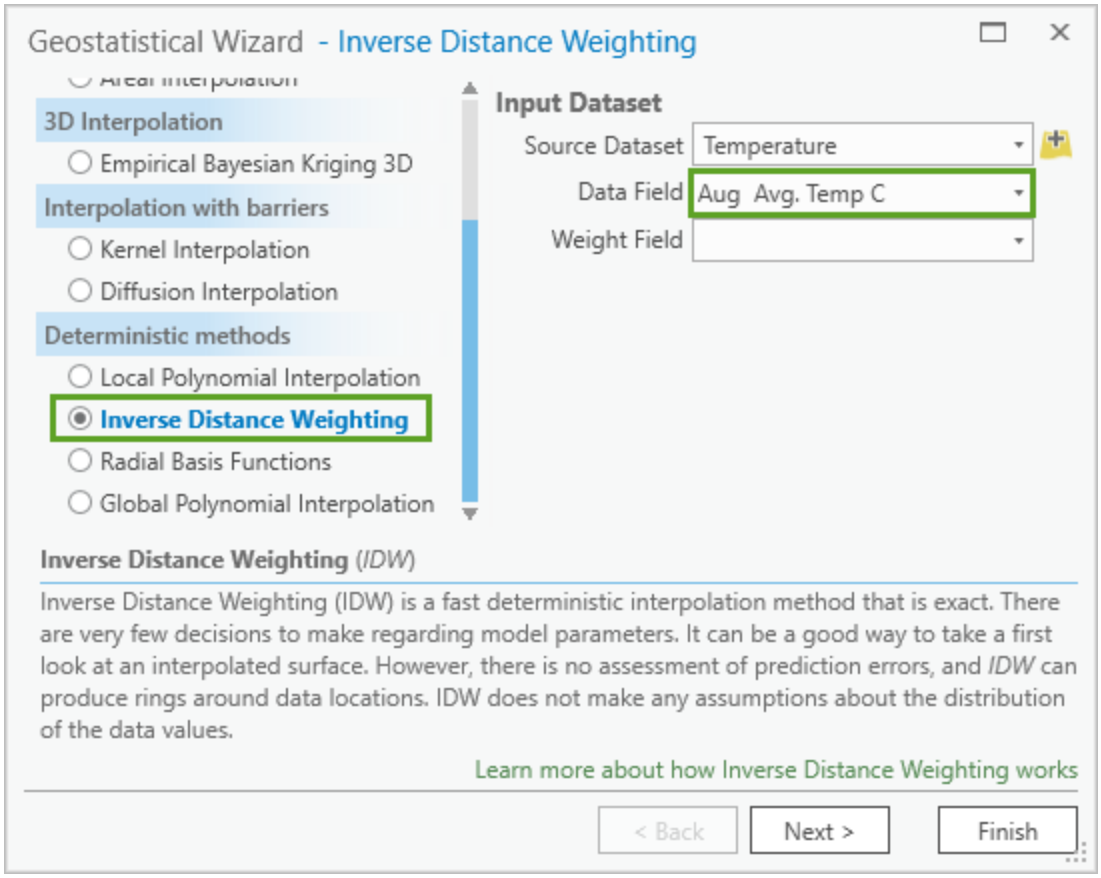
4. Click **Cancel** to close the **Layer Properties** window.

5. On the ribbon, on the **Analysis** tab, in the **Tools** group, click **Geostatistical Wizard**.

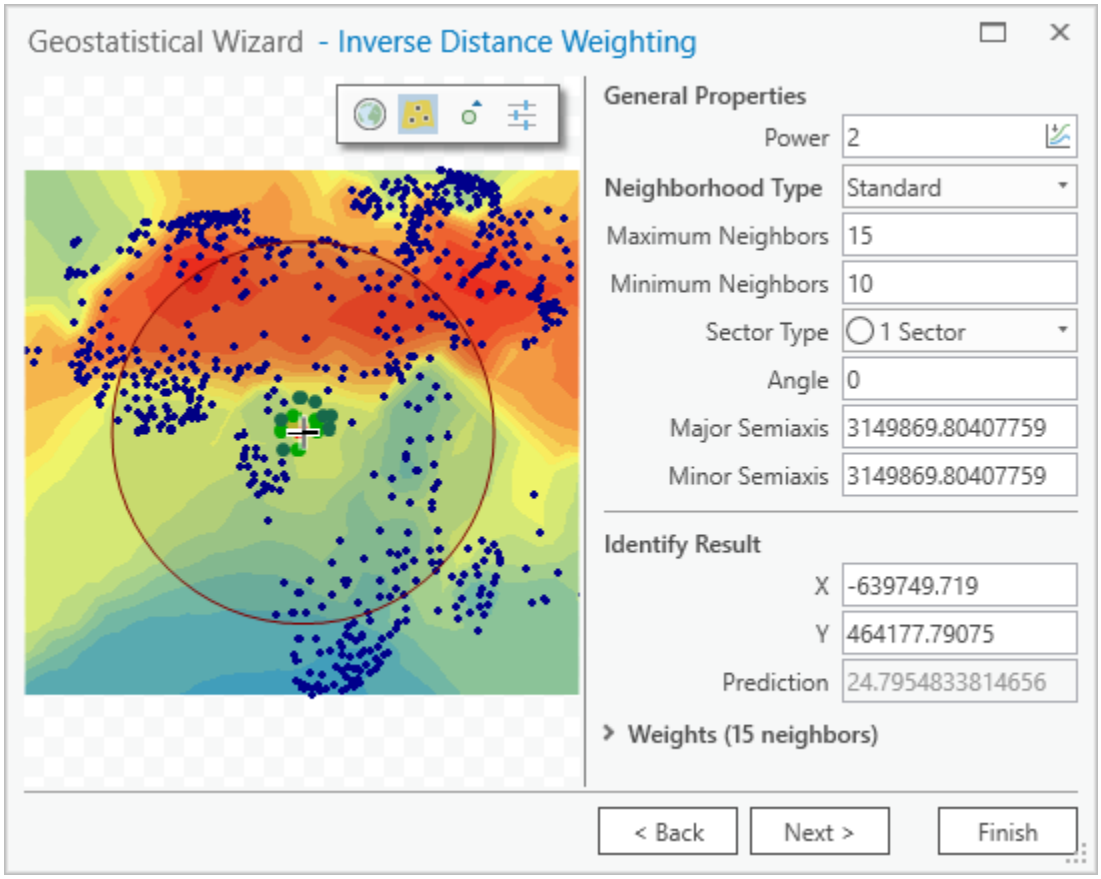


6. Under **Deterministic methods**, select **Inverse Distance Weighting**. You may need to scroll down to find this option.

7. For **Data Field**, choose **Aug Av. Temp C**.



8. Click **Next**.

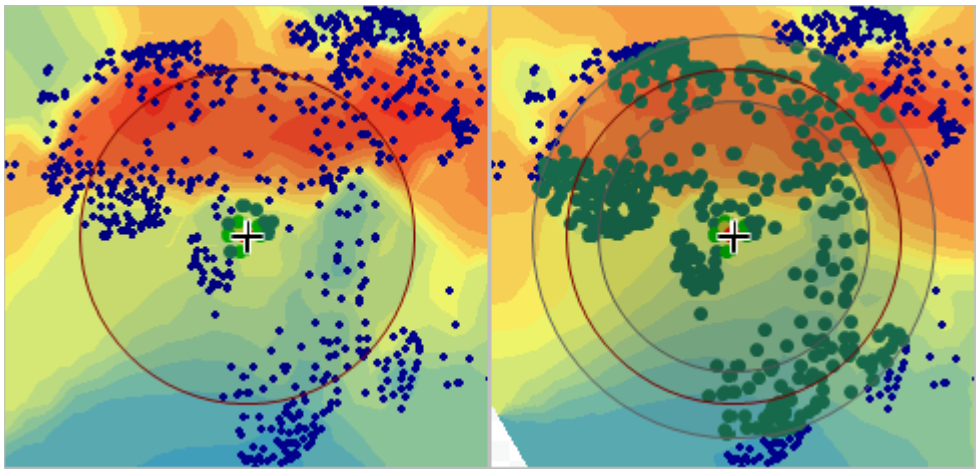


On this page you can interactively change the parameters of the IDW method and see how the model responds in the preview map. The **Identify Result** section tells you the predicted value for any location.

9. In the Geostatistical Wizard, click some different parts of the preview map to see the predicted temperature for that area in the **Identify Result** section.

10. Change **Neighborhood Type** to **Smooth**. The smooth option will generally make the prediction surface smoother and less jagged.

    The preview map updates. When **Neighborhood Type** is **Standard,** there is only one circle on the preview map. When it is **Smooth**, there are three concentric circles.



The circles on the preview map represent the search neighborhood. To predict a new value, only the sample points that are nearby—within the search neighborhood—are considered. You can read more about this process, including the smooth neighborhood type, at Search neighborhoods.

11. Verify that **Smoothing Factor** is set to **0.2**.

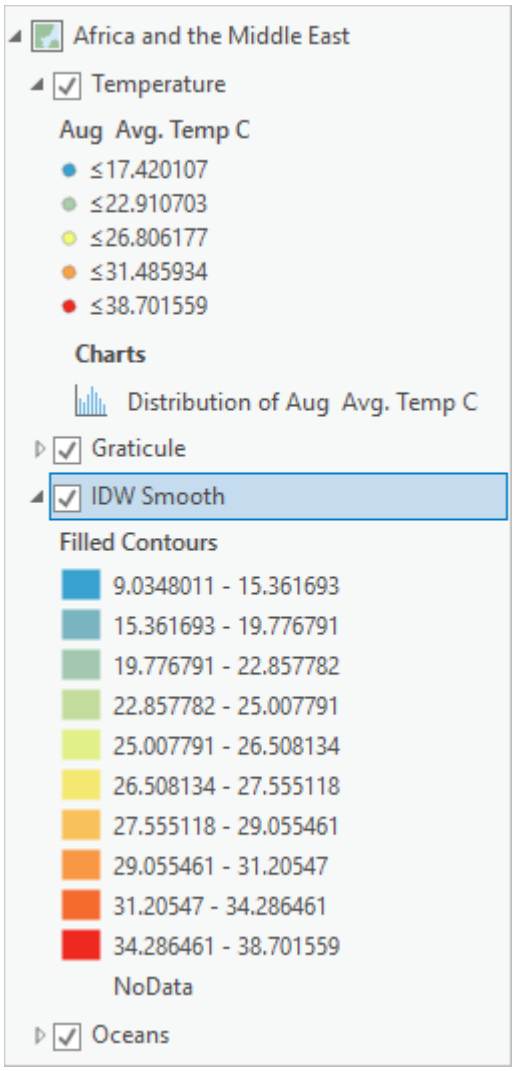
12. Click **Finish**.

13. On the **Method Report**, click **OK**.

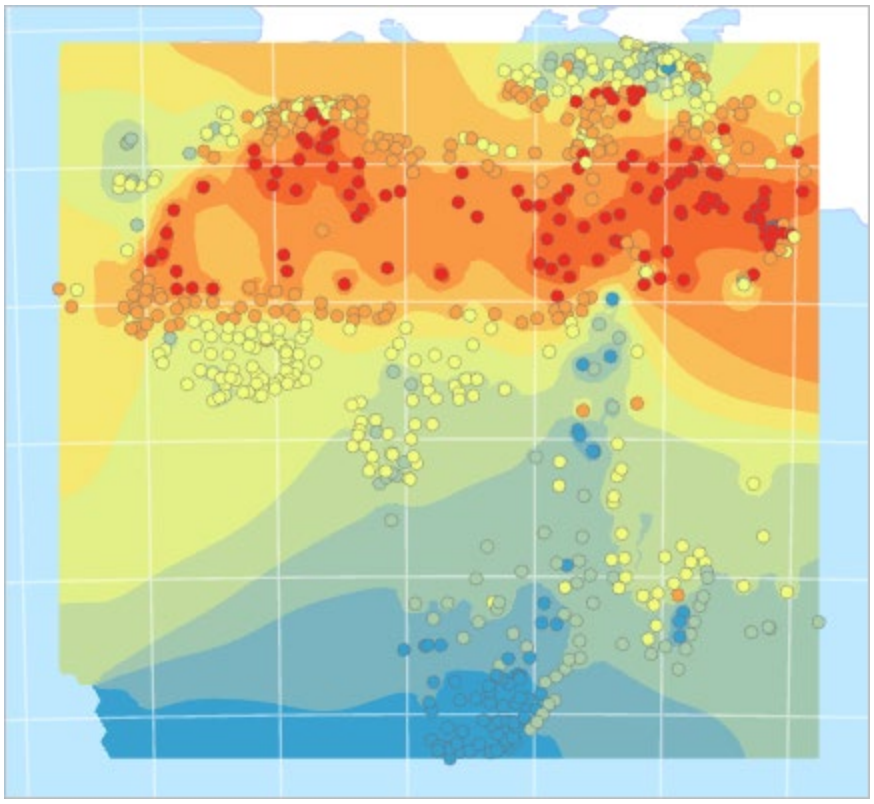
    A new layer is added to the map, representing a surface of maximum temperature for the Africa region.

14. In the **Contents** pane, select **Inverse Distance Weighting** and press F2 on the keyboard to make the name editable. Rename the layer `IDW Smooth`.

15. Drag **IDW Smooth** above **Oceans** and expand it.



The map now shows temperature predictions for places that had no temperature data.



Next you will create a slightly different surface using the same data and the same method.

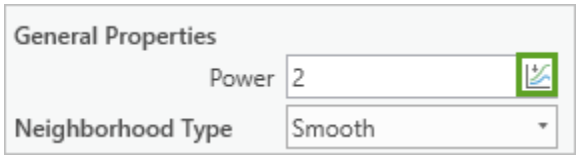
16. Open the Geostatistical Wizard.

    Hint: On the ribbon, on the **Analysis** tab, click **Geostatistical Wizard**.

17. Confirm that the selected method is **Inverse Distance Weighting** and the selected **Data Field** is **Aug Avg. Temp C**. Click **Next**.

18. For **Neighborhood Type**, choose **Smooth**.

19. Click the **Optimize** button in the **Power** parameter.



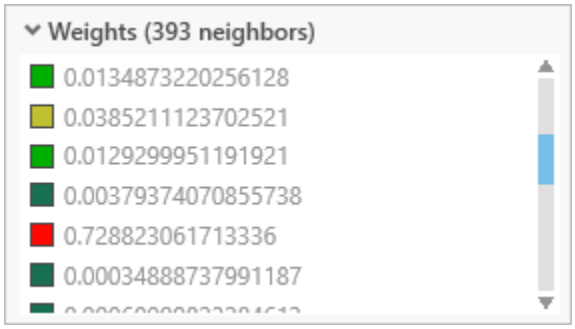
The **Power** value changes to 3.1076.

Not all of the points in the search neighborhood are considered equal. Those that are nearer to the location being predicted are given more weight in the calculation.

If **Power** is 0, all points in the neighborhood are weighted equally. The higher the power, the more rapidly the weights decrease with distance. A higher power of 3.1 results in a surface that appears more localized and less general, since points that are farther away have less of an influence.

20. Expand **Weights** and scroll through the list to find weights of different colors.

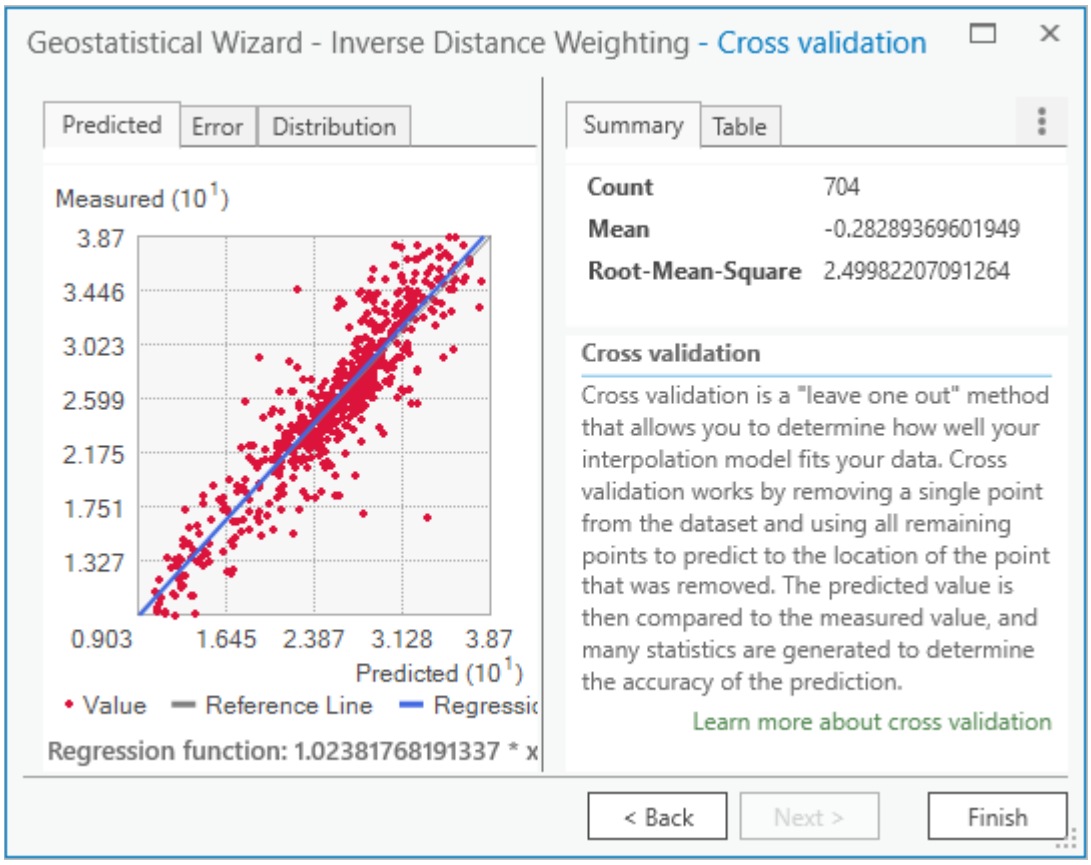
    This list represents all of the points within your search radius and includes the weights assigned to them.



    Click some of the values in the list to see the points selected on the preview map. Red points will exert more influence over the prediction than green ones.

21. Collapse **Weights** and click **Next**.

The **Cross validation** window provides information about how reliable your interpolation will be.



The information on this page allows you to assess the accuracy of the prediction surface. It does this by removing a single point from the dataset and using all remaining points to predict the value of the removed point.

The scatterplot compares predicted values (on the x-axis) to measured values (on the y-axis) and is considered best when the thin gray line coincides with the thick blue line.

The **Mean** value tells you if the model is skewed toward predicting values that are too high or too low. It is best when it is closest to 0.

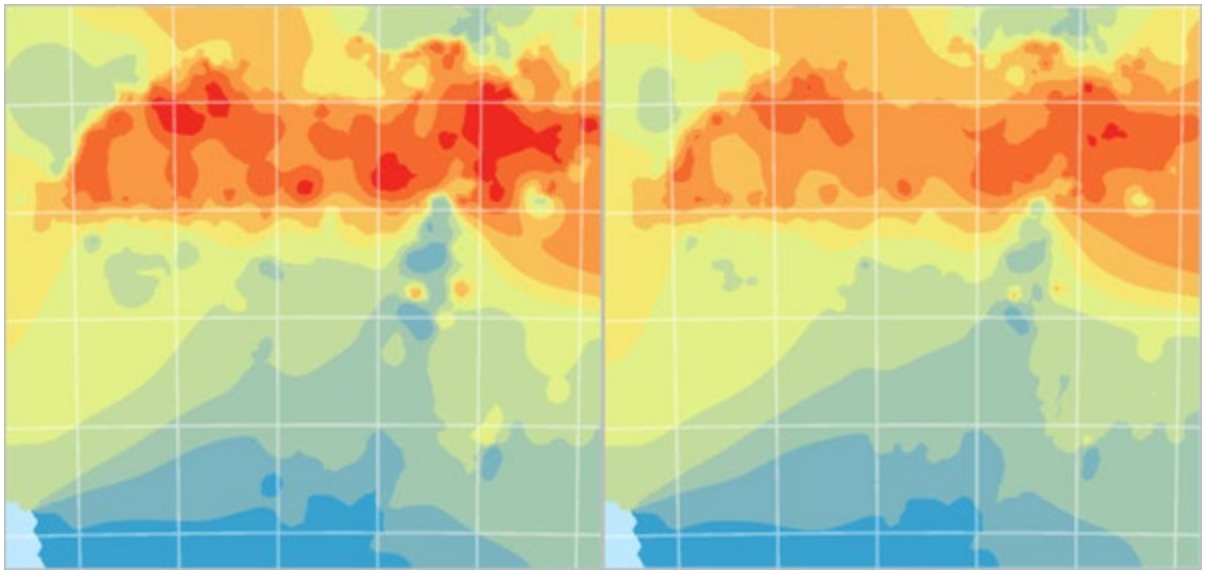
The **Root-Mean-Square** value is almost 2.5. This indicates that on average, the predicted temperature values differed from the measured values by about 2.5° Celsius.

22. Click **Finish**, and on the **Method Report** window, click **OK**.

23. A new layer is added to the map. Rename it `IDW Smooth Optimized`.

24. In the **Contents** pane, turn off the **Temperature** point layer.

25. Uncheck and check **IDW Smooth Optimized** to compare it with **IDW Smooth**.



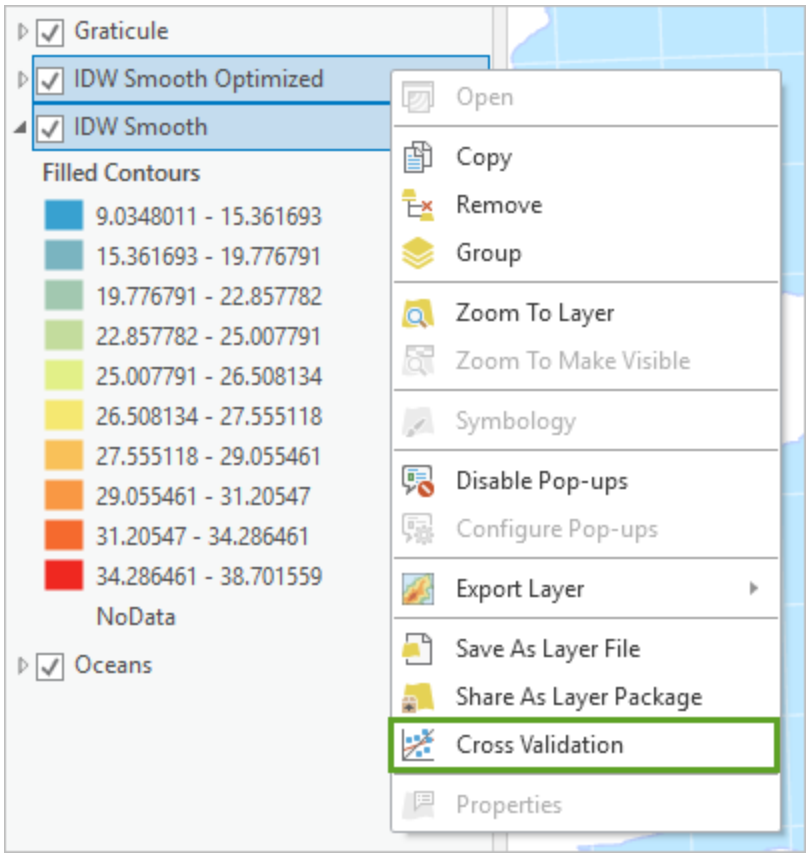
IDW Smooth Optimized (left) compared to IDW Smooth (right)

The two layers are similar, but the newer layer has more red. Which one is better? You can compare the accuracy of the two layers to help you decide.

26. In the **Contents** pane, select both I**DW Smooth** and **IDW Smooth Optimized**.

    Hint: To select more than one layer, press the Shift key on your keyboard while selecting layers.

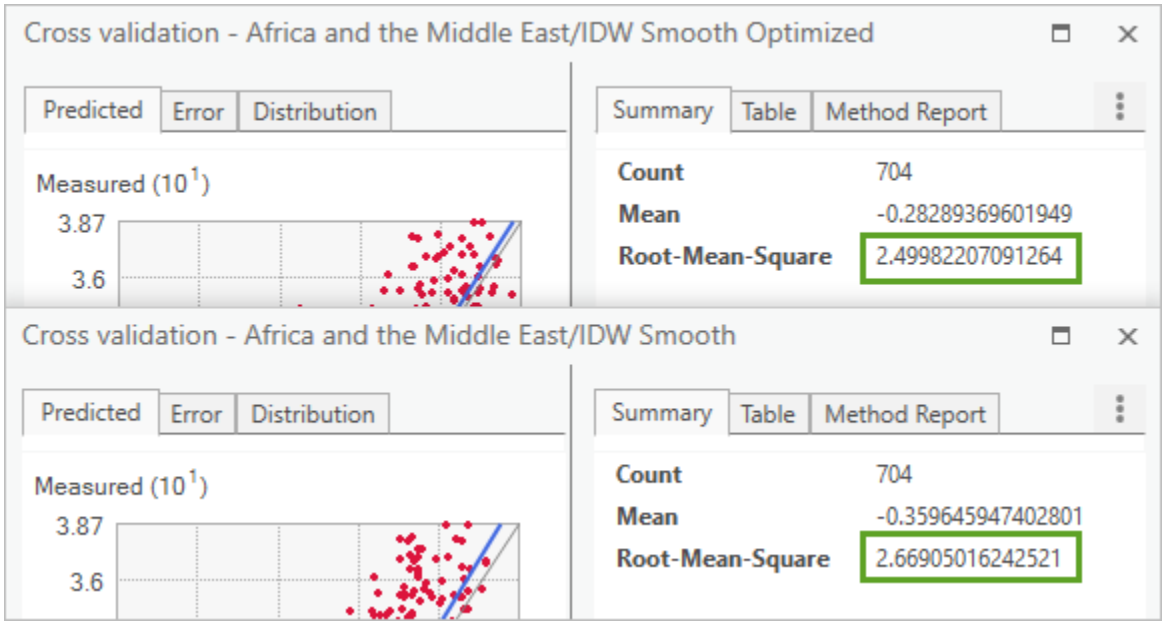
27. Right-click and choose **Cross Validation**.



28. Two **Cross validation** windows appear. One of them is blocking the other from view. Move it aside so you can see both at once.

    These are the same **Cross validation** windows that were shown in the Geostatistical Wizard. You already reviewed one of them, but the results are sometimes more useful when you can compare them between multiple prediction surfaces.

    The **Summary** tab reports numerical errors for each surface. The closer the **Root-Mean-Square** value is to 0, the more accurate the created surface is.

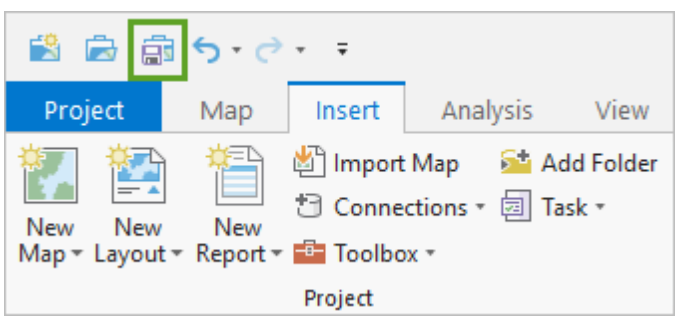


IDW Smooth Optimized has the smaller error value and so can be considered the more reliable prediction surface.

29. Close both **Cross validation** windows.

30. In the **Contents** pane, select only **IDW Smooth**. Right-click this layer and choose **Remove**.

31. On the toolbar at the top corner of the ribbon, click the **Save** button.



Inverse distance weighting is considered an easy and fast interpolation method. It is good for getting an initial picture of the phenomenon you are mapping, and sometimes you may need to use it because it will follow measured values exactly. But it can also produce a ring effect around islands in your data. Next you will try kriging to see if you can get more accurate results.

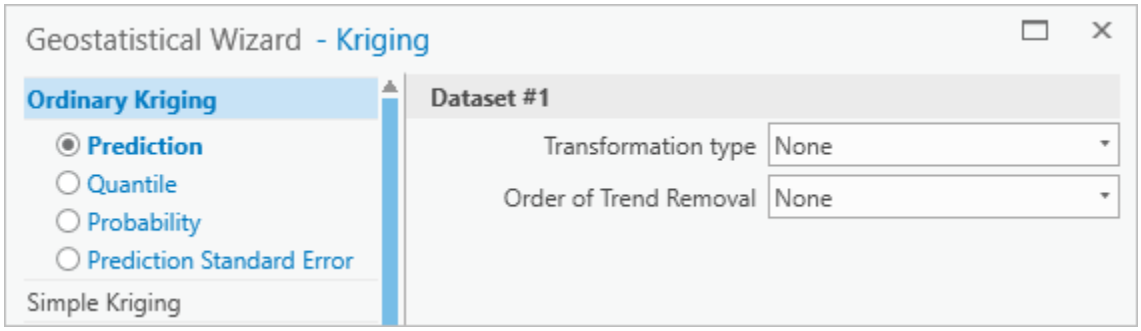
Create geostatistical surfaces using kriging

Kriging is a very flexible geostatistical method. This means that you can adapt it in many ways to suit your data, but it also means that there are many more choices that must be made.

1. Open the **Geostatistical Wizard**.

2. Under **Geostatistical methods**, select **Kriging / CoKriging** and click **Next**.

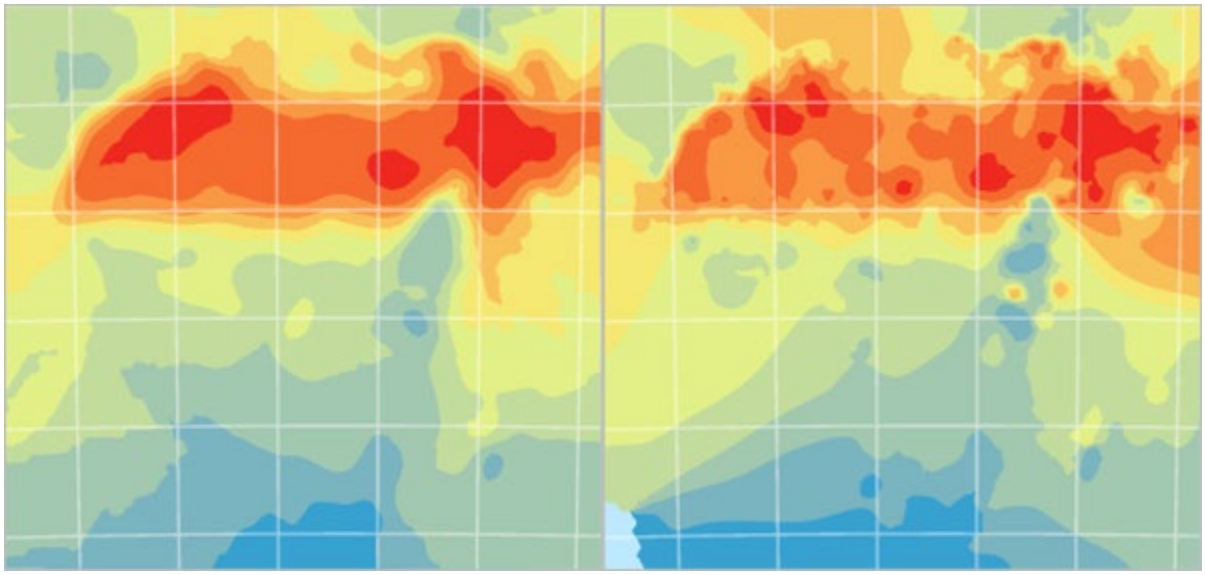
3. Under **Ordinary Kriging**, choose **Prediction** to create a surface of predicted values similar to the ones you created earlier using IDW.



4. For now, you will create a surface with the default parameters for ordinary kriging. Click **Finish** and click **OK**.

5. A new layer is added to the map. Rename it to `Kriging Default`.

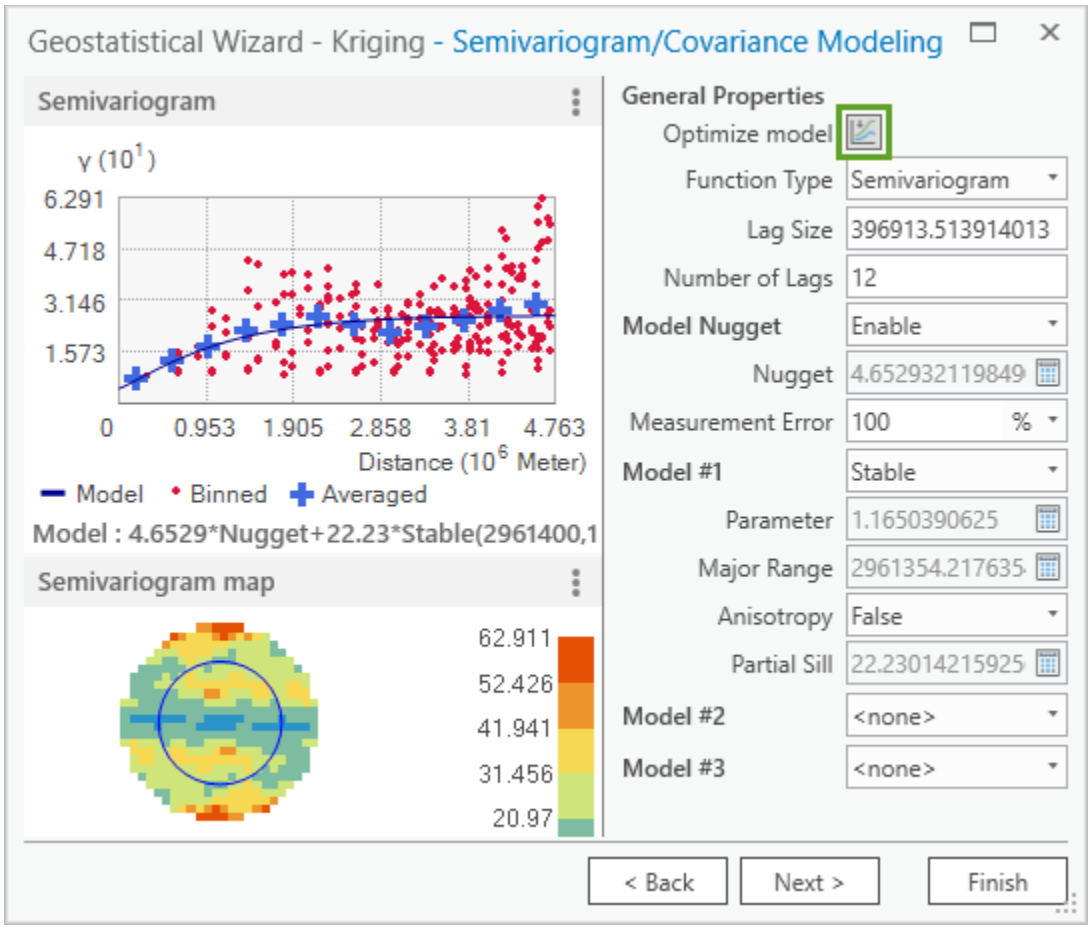
6. Compare **Kriging Default** to **IDW Smooth Optimized**.



Kriging Default (left) compared to IDW Smooth Optimized (right)

The new layer is much more general in its pattern. Next, you'll change some of the parameters to try to create a better geostatistical surface.

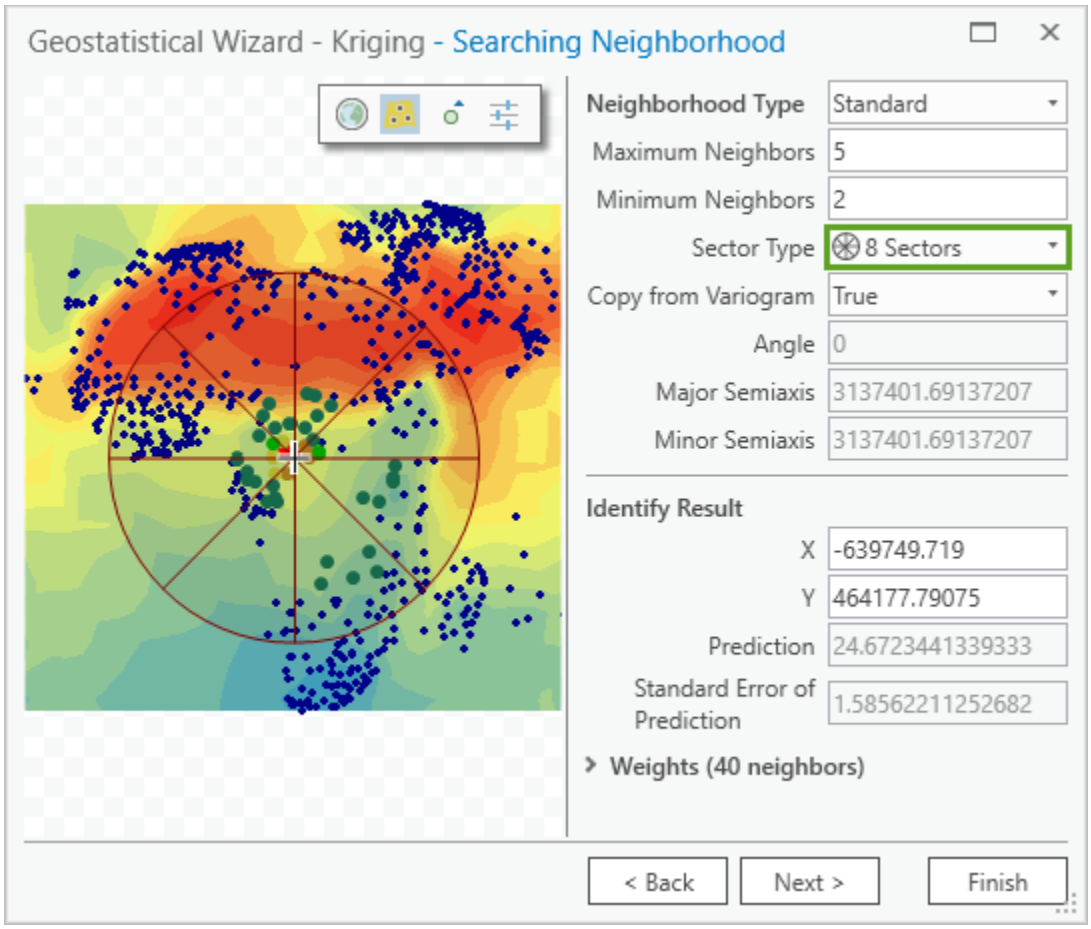
7. Open the **Geostatistical Wizard**.
8. Confirm that the selected method is **Kriging / CoKriging** and click **Next**.
9. Under **Ordinary Kriging**, select **Prediction** and click **Next**.
10. On the **Semivariogram/Covariance Modeling** page, click the **Optimize model** button.



The optimize button will find the parameters that result in the smallest prediction errors. Notice that the **Semivariogram** graph and some of the parameters have changed. In this case the change is minimal.

11. Click **Next**.

12. On the **Searching Neighborhood** page, change **Sector Type** to **8 Sectors**.



Increasing the number of sectors ensures that neighbors are searched for in all directions, and a large cluster of nearby points in only one direction will not have all of the influence over the predicted value.

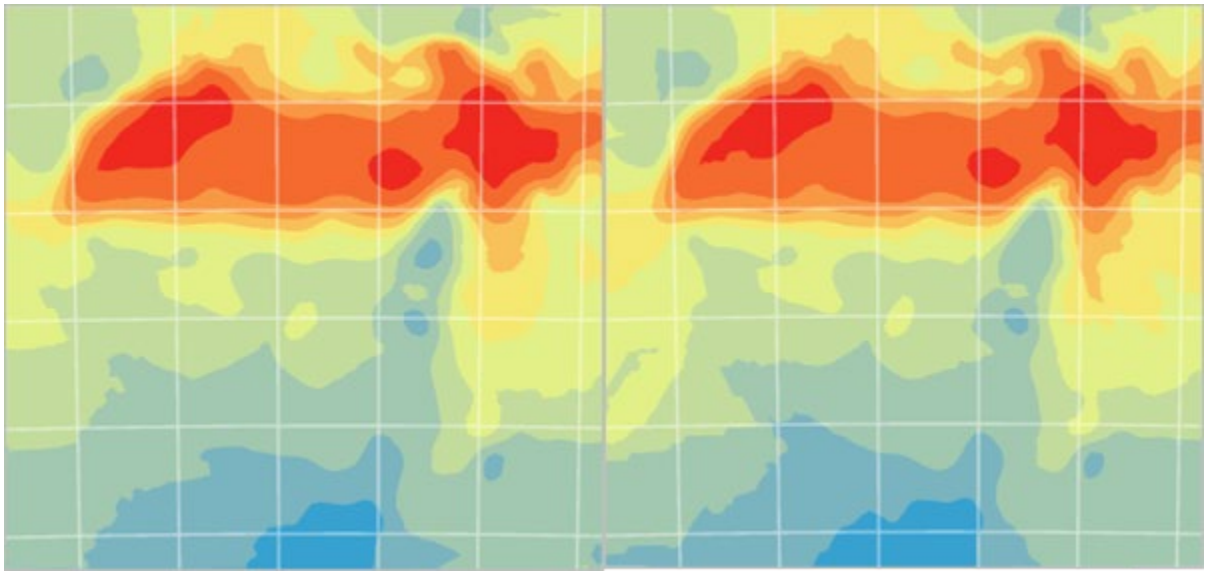
13. Click **Next** and review the results on the **Cross validation** window. Note that kriging provides you with many more values than inverse distance weighting.

Count	704
Mean	-0.0236115793560031
Root-Mean-Square	2.28300339706507
Mean Standardized	-0.00288787516931152
Root-Mean-Square Standardized	0.840726823721056
Average Standard Error	2.77497937323438

14. Click **Finish** and click **OK**.

15. Another layer is added to the map. Rename it to `Kriging Modified`.

16. Compare **Kriging Modified** to **Kriging Default**.



Kriging Modified (left) compared to Kriging Default (right)

They are very similar.

17. In the **Contents** pane, select **Kriging Default** and **Kriging Modified**. Right-click and choose **Cross Validation**.

18. Arrange the windows so you can see both at once. Analyze the values on the **Summary** tab.

Numbers closer to zero indicate better accuracy. The exception is **Root-Mean-Square Standardized**. In this case, values closer to 1 are desired.

	Kriging Default	Kriging Modified
Mean	-0.013	-0.024
Root-Mean-Square	2.294	2.283
Mean Standardized	-0.001	-0.003
Root-Mean-Square Standardized	0.854	0.841
Average Standard Error	2.740	2.775

It is not immediately obvious from these values which surface is better. **Kriging Default** has better values for every category except **Root-Mean-Square**. However, this does not necessarily mean it is better.

If any of these values are too far off, you should eliminate that layer. But in this scenario, both layers show good cross-validation results, so you can use **Root-Mean-Square** as the tie breaker value. It is also desirable that the **Root-Mean-Square** and **Average Standard Error** values be close to one another. If there is a large difference between these values it may indicate that the prediction is unstable.

The **Cross validation** report indicates that **Kriging Modified** is slightly more reliable than **Kriging Default**.

19. Open the **Cross validation** window for **IDW Smooth Optimized**.

    This surface has a **Root-Mean-Square** value of 2.5. It is less reliable than either of the kriging surfaces.

20. Close all three **Cross validation** windows.

21. Remove **IDW Smooth Optimized** and **Kriging Default** from the map.

22. Save the project.

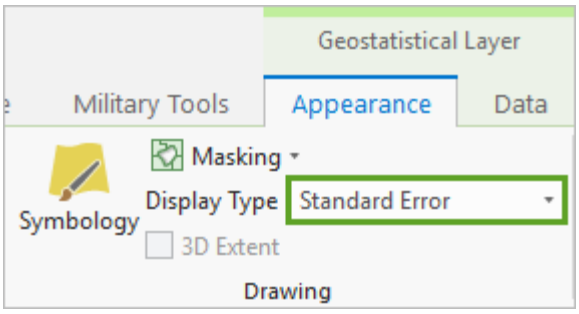
Kriging is a more advanced method than IDW and requires you to make more decisions. But this allows you to experiment with the parameters until you find those that are a good fit for your data and phenomenon. Kriging also gives you more tools to assess the accuracy of your results, such as a map of the standard error estimates, which you will create next.

Map the standard error estimates

You have now made four different surfaces of temperature covering Africa and the Middle East. Each was interpolated from the same data, but each showed a different surface. Clearly these predictions are useful, but they cannot be taken as fact. Some parts of the surface (where there are many data points) can be considered more accurate and reliable than others (where the data is scarce). It is useful to map these degrees of uncertainty to aid decision makers.

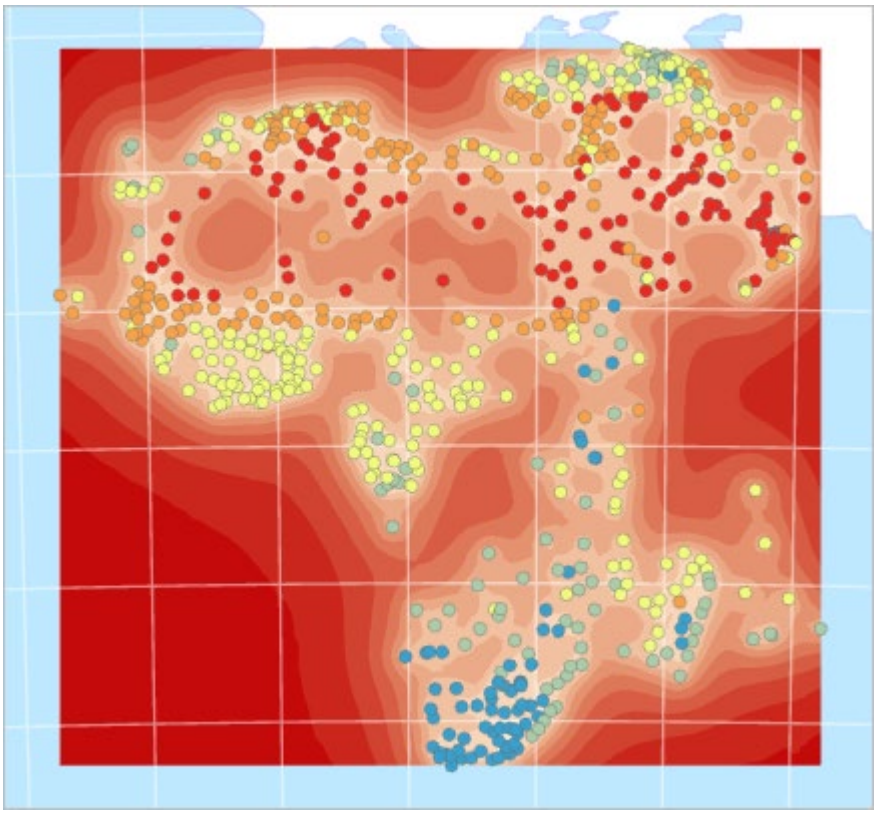
1. In the **Contents** pane, select **Kriging Modified**.

2. On the ribbon, on the **Appearance** tab, change **Display Type** to **Standard Error.**



The map changes to become mostly red.

3. On the **Contents** pane, turn on **Temperature**.

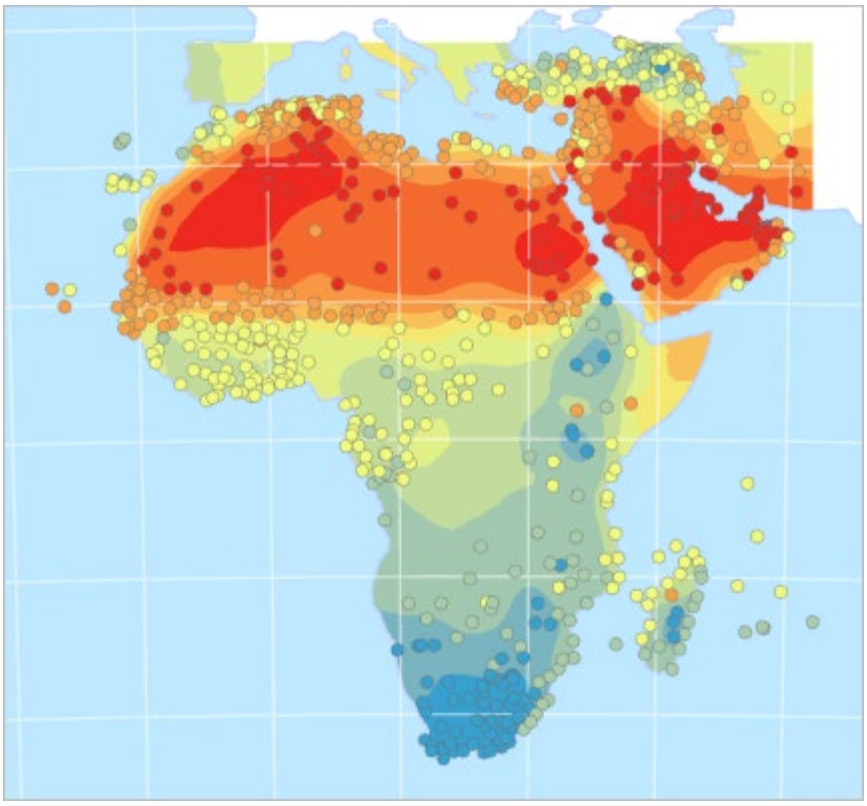


Standard errors are measures of uncertainty for the predicted values. The dark red areas on the map have larger standard error values and therefore lower certainty in the predicted values. Lighter areas are those where you can place more trust in the results. This map suggests that the results have the greatest standard error in the ocean. This makes sense, because there were no sample measurement points in the ocean (although there were some on small islands).

4. For the **Kriging Modified** layer, change the **Display Type** back to **Prediction**.

For this map, you are only interested in predicting land temperatures, so the ocean can be masked out.

5. In the **Contents** pane, drag the **Oceans** layer above **Kriging Modified**.



6. Save the project.

Summary

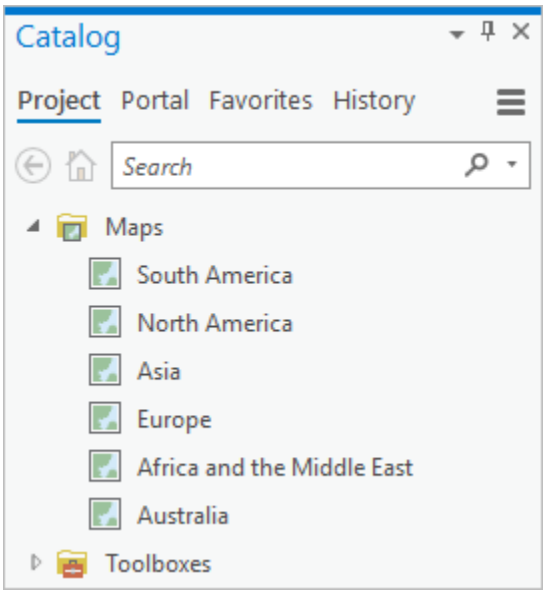
Geostatistics can help you map many phenomena as continuous surfaces even though you only have discrete point data. This can be incredibly useful for visualizing patterns and performing analysis. You may not have a weather station in your study area, but a set of weather stations in a wider region can provide the data you need to understand and predict temperatures everywhere.

The Geostatistical Wizard offers many interpolation methods, and each one has parameters that can be tweaked to produce different results. Why? Depending on the phenomenon you are mapping, and the data you have available, one model may give you more reliable results than another. If you are going to make decisions based on an interpolated surface, finding the most accurate model is critical.

You can compare the cross-validation results to determine which method is working best for your data. Once a surface has been created, some parts of it will offer more accurate predictions than others. You can visualize the surface by its standard prediction error to understand where the prediction is most reliable.

The four maps you made were all derived from the same input data, but they looked different from one another. Now that you know how maps with interpolated surfaces are made, do you trust them more or less? Geostatistical models can be tweaked to create more accurate results. On the other hand, the map maker might have an agenda that they want to promote, and they may tweak the geostatistical parameters to emphasize a trend.

This project contains five more maps—one for each of the other continents. You can find them in the **Catalog** pane, on the **Project** tab, in the **Maps** folder.



For an extra challenge, work through this lesson again using one of these maps. For Africa and the Middle East, you found that **Aug Avg. Temp C** was the best field, and **Kriging Modified** was the best surface. For another continent, you may find that different parameters yield better results.