Learn ArcGIS

Guided lessons based on real-world problems

Find Fish in the Bering Sea

By Witold Frączek and Heather Smith

learn.arcgis.com/

Time: 30 Minutes

Overview

Alaska pollock (also known as walleye pollock) is an important species for commercial fishing. It is found in greatest numbers in the Bering Sea, a biodiverse region supporting various fisheries. Pollock is currently not an overfished species, but science-based management and monitoring is necessary to ensure that these fish stocks can remain healthy and productive.

Read more about Alaska pollock at the [NOAA Fisheries species page](#).

The fishing industry does not want to waste time fishing in empty seas. At the same time, they don't want to overestimate the pollock population when they fish in a particularly productive area. In order to fish both efficiently and sustainably, it is important to understand where shoals are most likely to be located. Because it's known that Alaska pollock prefer temperatures ranging between 2 and 3 degrees Celsius, you can apply geostatistics to temperature data to predict where to find them in the Bering Sea.

This exercise will allow you to familiarize yourself with some of the interpolation methods available in ArcGIS Geostatistical Analyst, and a process for finding a best fit model for your data. These methods will allow you to predict the ocean temperature continuously across your study area using discrete temperature measurements.

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

In this lesson you will learn to:

- Map ocean bottom temperatures in the Bering Sea using the Geostatistical Wizard.

- Build prediction surfaces using polynomial, inverse distance weighting, and kriging interpolation methods.

- Compare cross-validation errors for different surfaces to determine which is the most accurate.

Explore the data

1. Download the [BeringSea project package](#).

2. Locate the downloaded file on your computer. Double-click **BeringSea.ppkx** to open it in ArcGIS Pro.

    Note: If you do not have ArcGIS Pro or an ArcGIS account, you can sign up for an ArcGIS free trial.
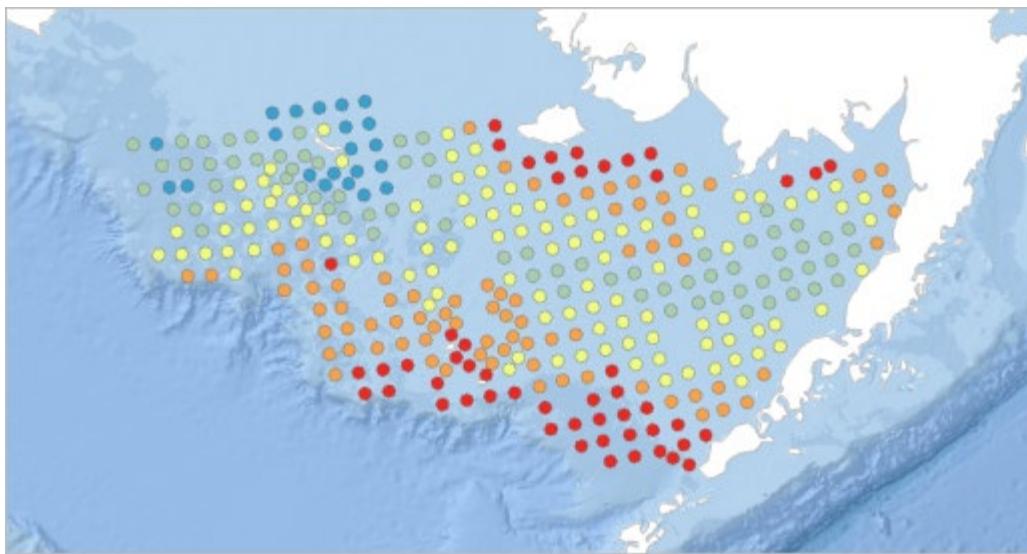

The 328 points in the **BottomTemp** layer represent water temperatures from the bottom of the eastern part of the Bering Sea. The data was created by NOAA Fisheries, Alaska Fisheries Science Center, Resource Assessment and Conservation Engineering Division. The blue points represent the coldest water, while red points represent the warmest. These temperatures were recorded in the summer.

The temperature data points all lie on the continental shelf. This part of the Bering Sea is shallow, with a depth not exceeding 150 meters. In this area, you can find schools of Alaska pollock, but it is easier to find the fish if you can predict the temperatures in different parts of the sea.

Predict ocean temperature using global polynomial interpolation

First, you will try the global polynomial interpolation method. This method is useful for visualizing the general trend of your data and getting an initial picture of its pattern.

1. On the ribbon, on the **Analysis** tab, in the Tools group, click **Geostatistical Wizard**.
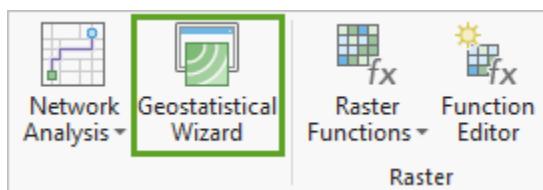
2. Under **Deterministic methods**, select **Global Polynomial Interpolation**. You may need to scroll down to find this option.

3. For **Source Dataset**, choose **BottomTemp**.

4. For **Data Field**, choose **BOTTOMTEMP**.
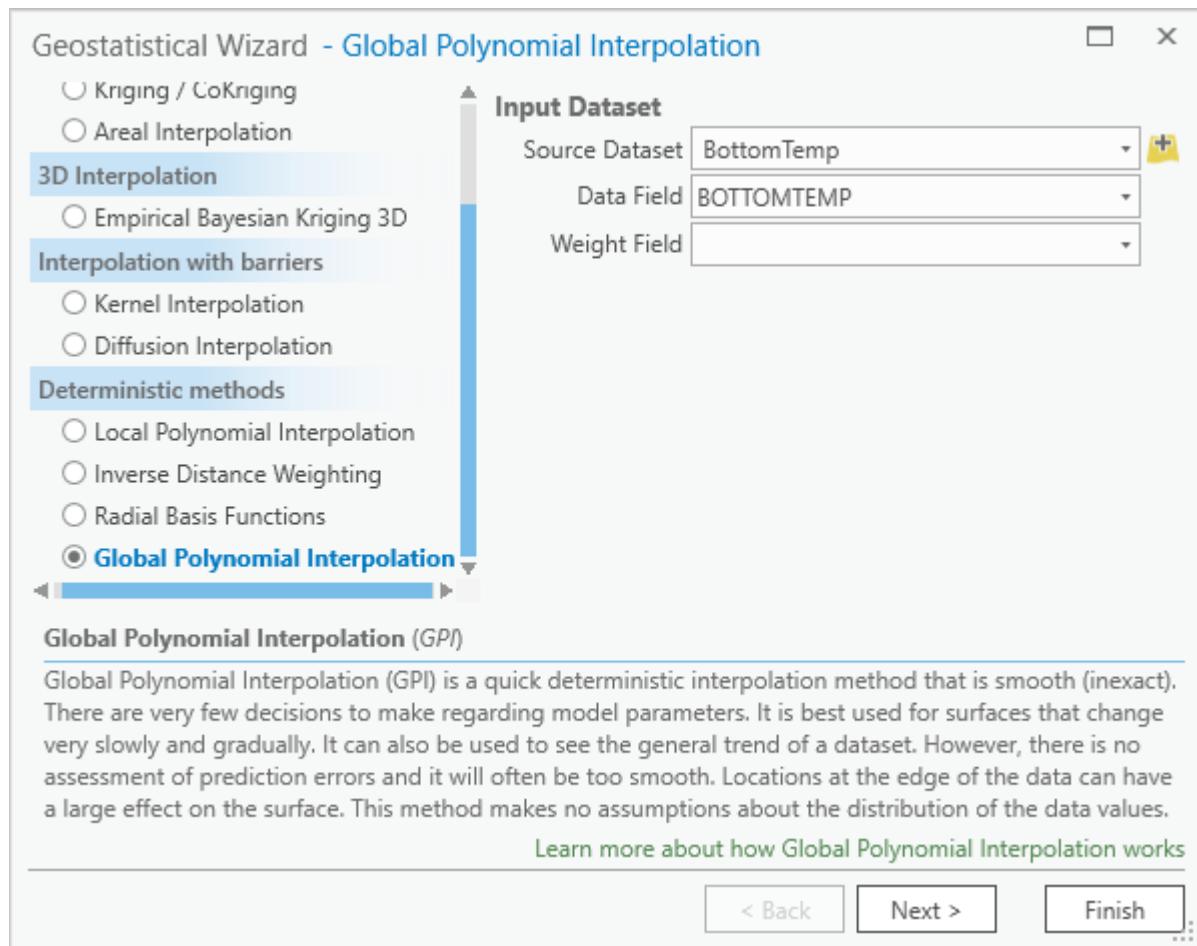
5. Click **Next**.

6. Under **General Properties**, change the **Order of polynomial** setting to different values and observe the effect it has on the preview map.
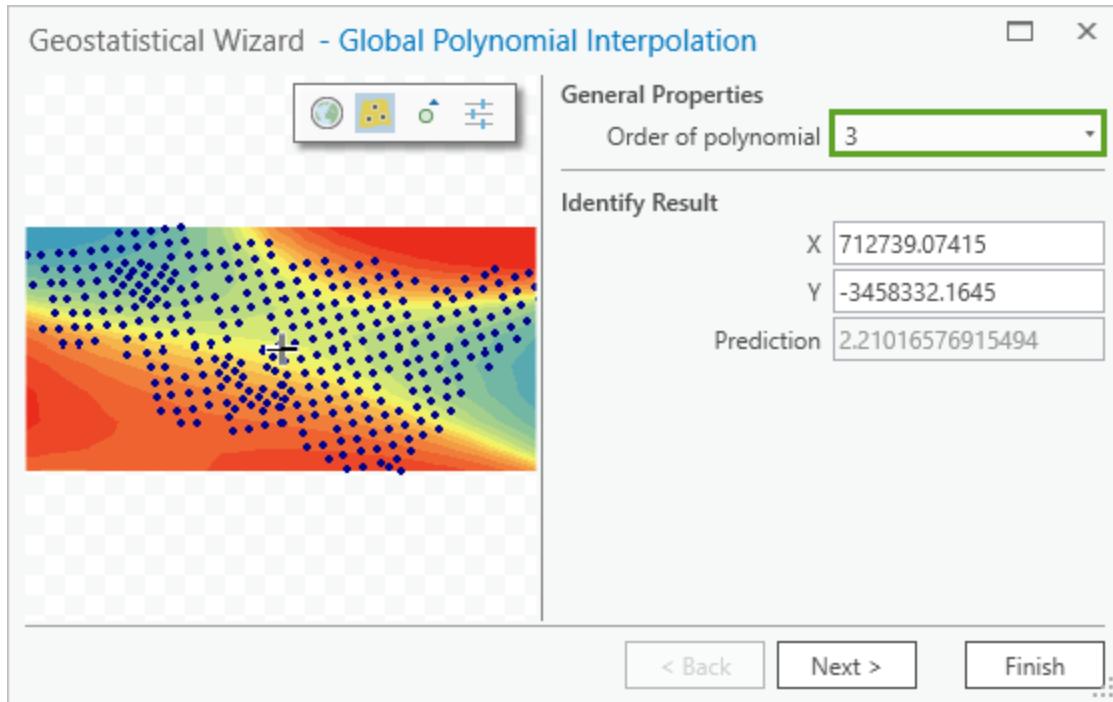
An order of polynomial of 1 is too general. But the higher numbers produce strange results.

This method fits a very smooth surface through the points. The order of polynomial is how many times this smooth surface will bend. Higher orders produce strange results because the surface is forced to bend too many times.

7. For **Order of polynomial**, choose **3** and click **Finish**.

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

A new layer is added to the map, representing predicted bottom temperatures for the Eastern Bering Sea.
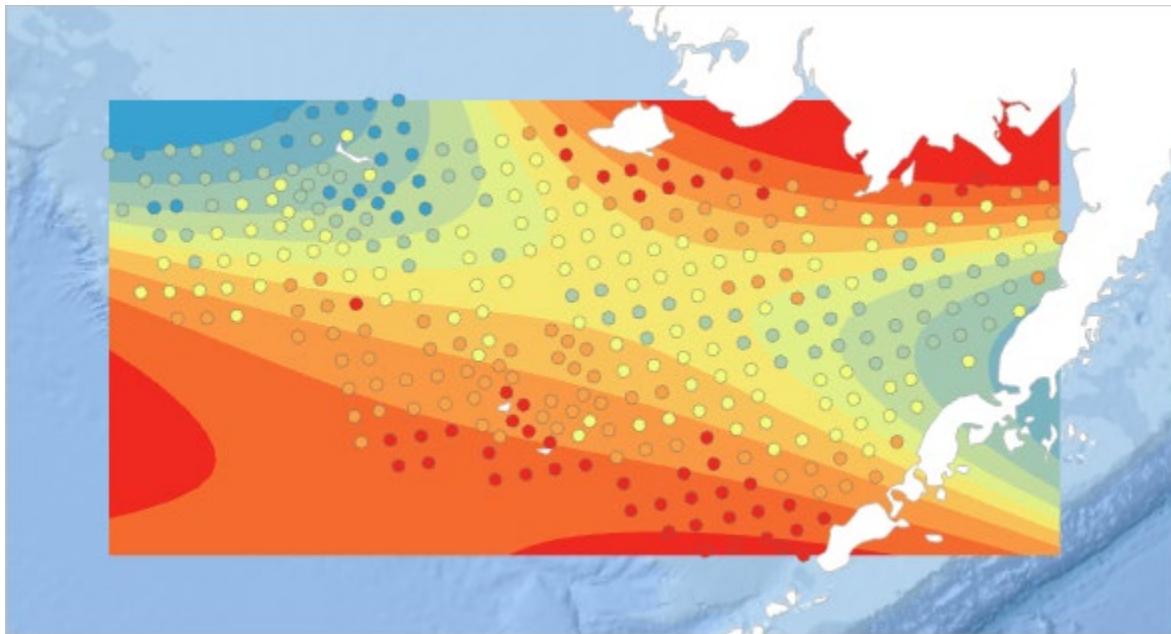
You have not yet quantitatively assessed the accuracy of this surface, but visually, this result does not seem to reflect the natural phenomenon very closely. Next you will try the local polynomial interpolation method to see if you can create a better surface with it.

Predict ocean temperature using local polynomial interpolation

Local polynomial interpolation is similar to global polynomial interpolation, but it fits multiple planes over the surface instead of one.

1. Open the **Geostatistical Wizard**, and under **Deterministic methods**, select **Local Polynomial Interpolation**.
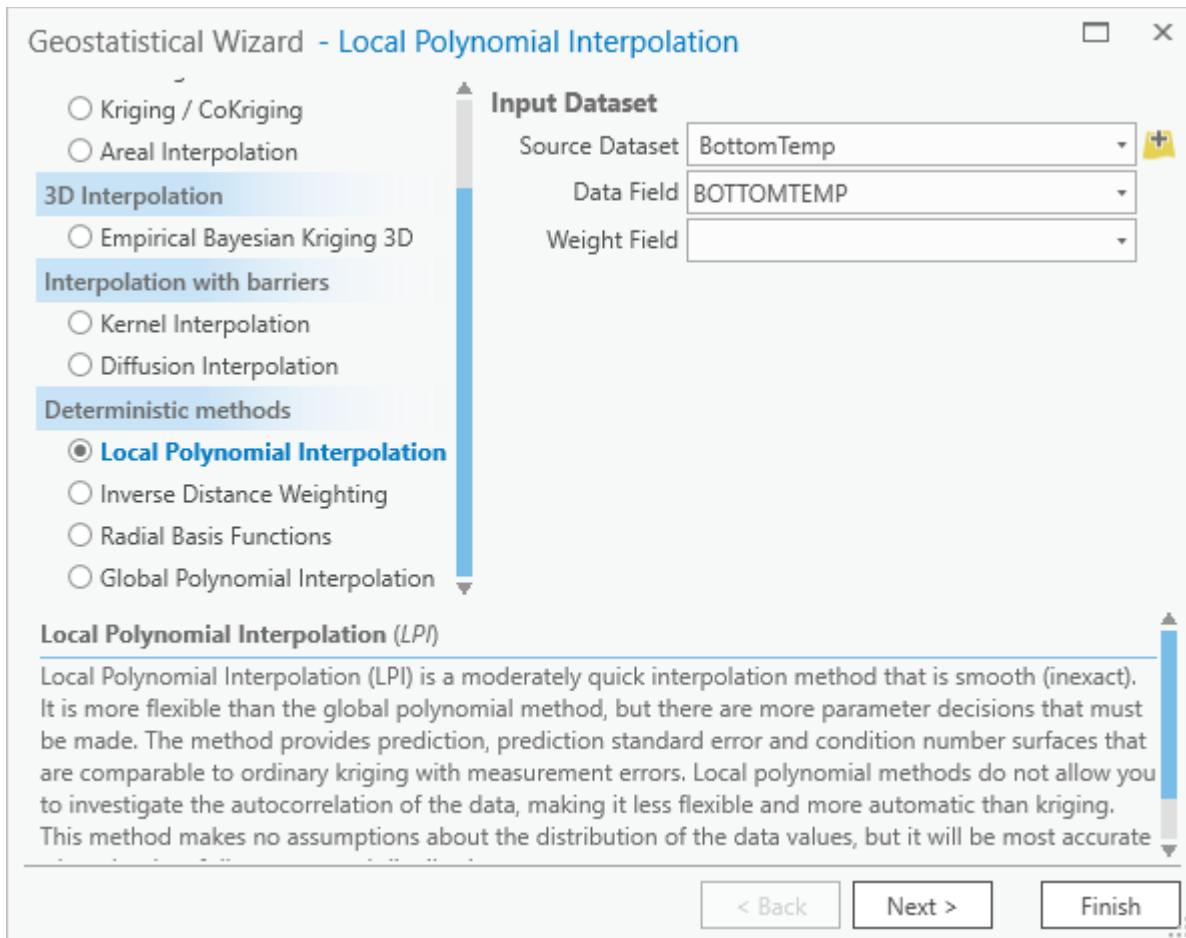
2. Click **Next**.

3. For **Order of polynomial**, select **2** and observe the change in the preview map.
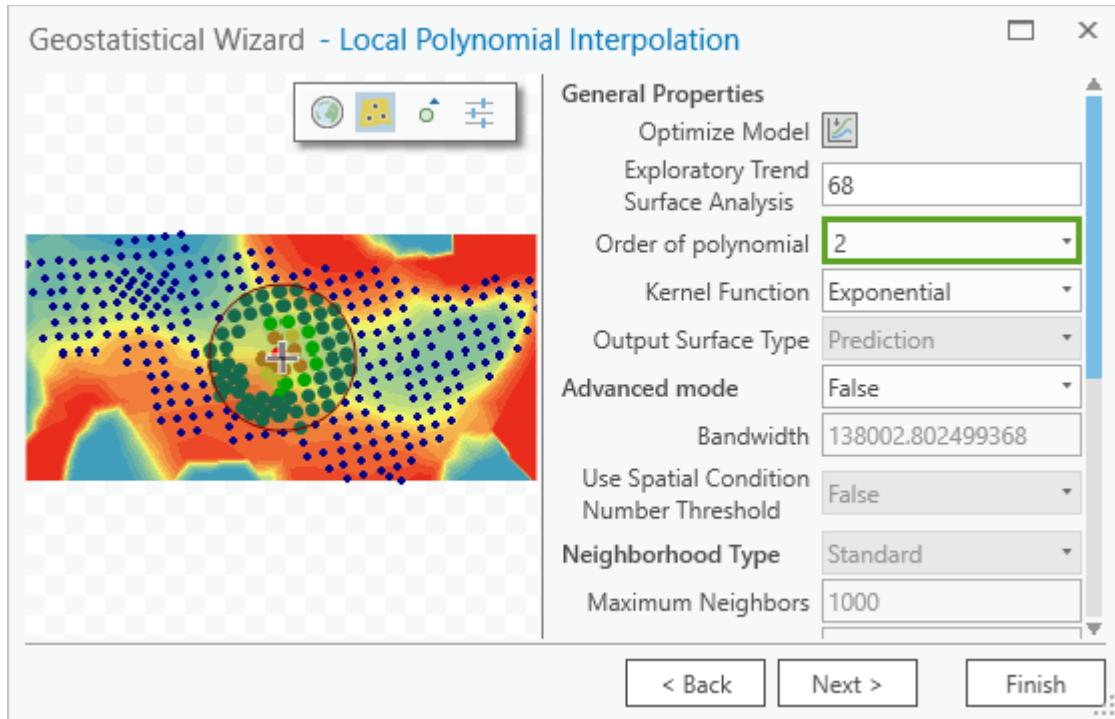
4. For **Exploratory Trend Surface Analysis**, experiment with different values and observe the effect that they have on the preview map.

If this value is set to 0, the resulting surface will be identical to global polynomial interpolation. A smaller value will result in a more general interpolation. A larger value will result in a more local one.
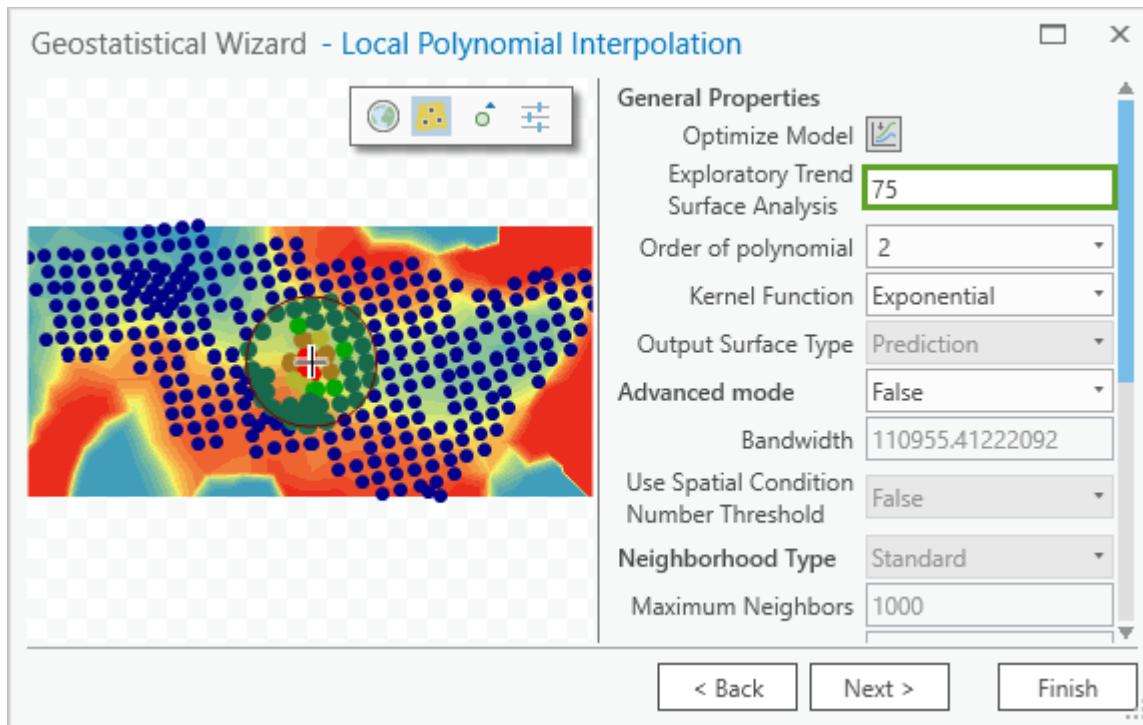
5. Set **Exploratory Trend Surface Analysis** to *75*.

6. Click **Finish** and click **OK**.

   A new interpolated surface is added to the map.
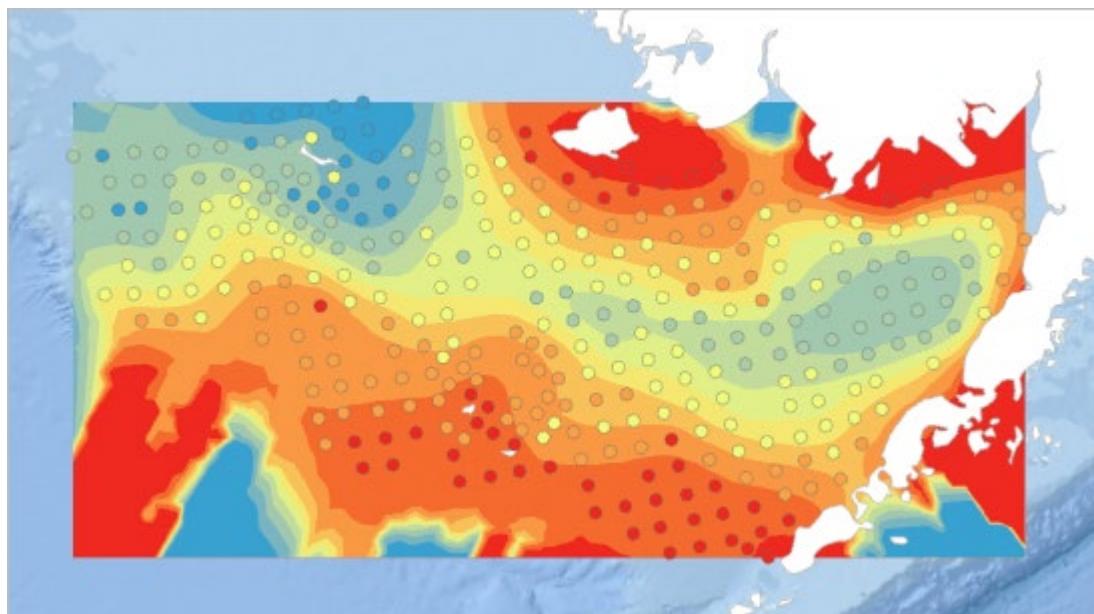
7. In the **Contents** pane, check and uncheck **Local Polynomial Interpolation** to compare it with **Global Polynomial Interpolation**. The newer layer is less simplistic.

8. On the toolbar above the ribbon, click the **Save** button.
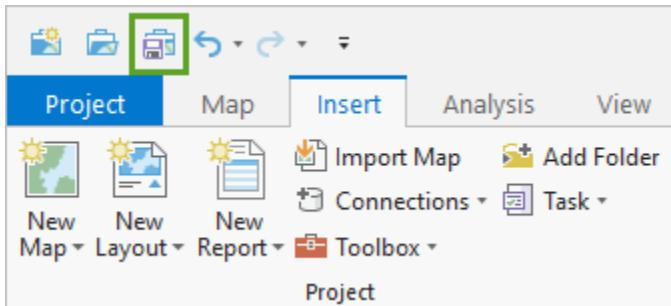
Predict ocean temperature using inverse distance weighting

Inverse distance weighting (IDW) is an interpolation method that is considered fast and automatic. It is a good choice if you need the predicted surface to follow the input points exactly.

1. Open the **Geostatistical Wizard**, and under **Deterministic methods**, select **Inverse Distance Weighting**.

2. Click **Next**.

3. Change **Neighborhood Type** to **Smooth**.

   The smooth option will generally make the prediction surface smoother and less jagged.

4. Click **Next**.

   The **Cross validation** window helps you decide if your model is accurate enough to rely on for predictions. It does this by removing a single point from the dataset and using all remaining points to predict the value of the removed point.
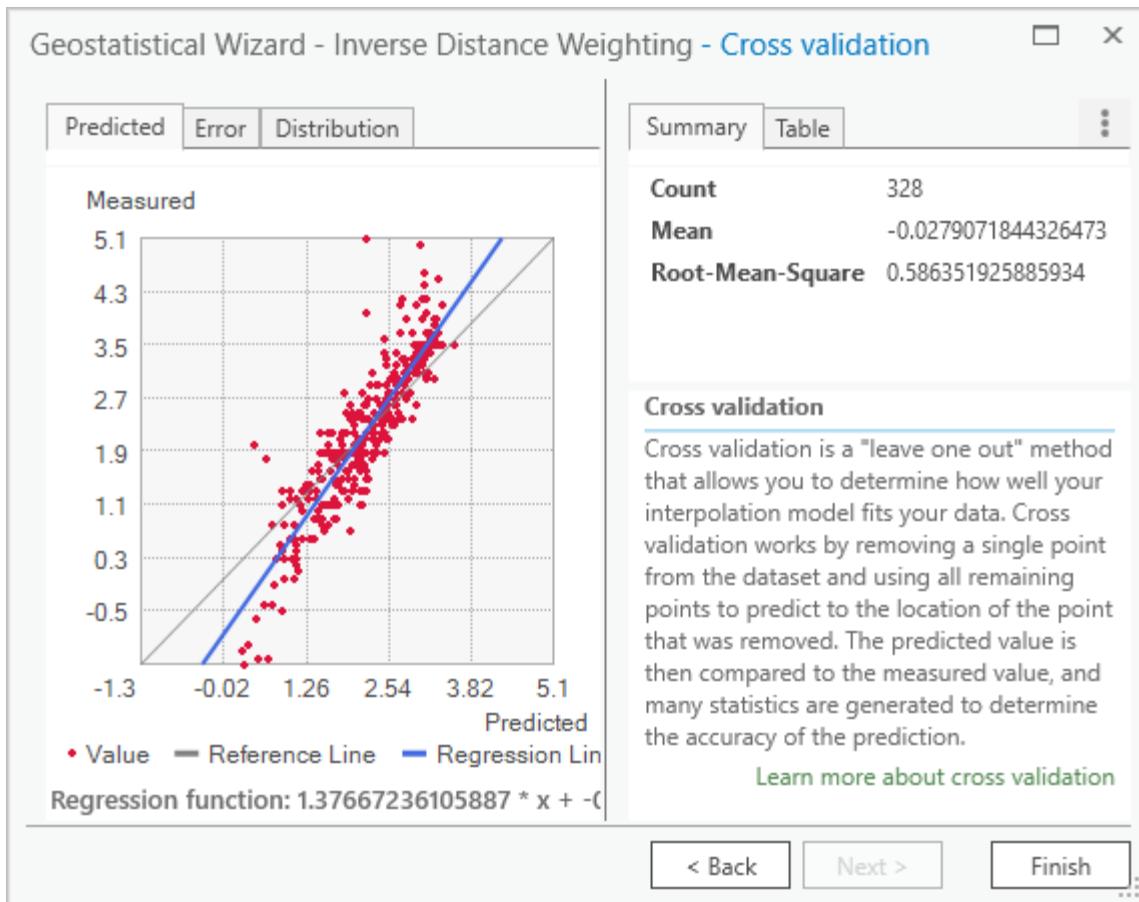
Mean and **Root-Mean-Square** should be close to zero. On the scatterplot, the gray reference line should be close to the blue regression line.

The scatterplot in this case might be improved. You can make some changes to the model to try to get better cross-validation results.

5. Click **Back**.

6. Next to **Power**, click the **Optimize** button.
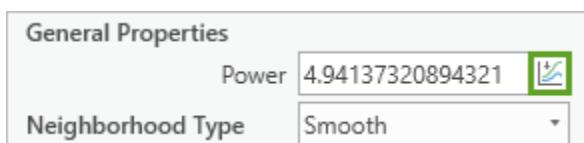
This will try to find a value that will minimize the root mean square (RMS) error.

7. Click **Next**.
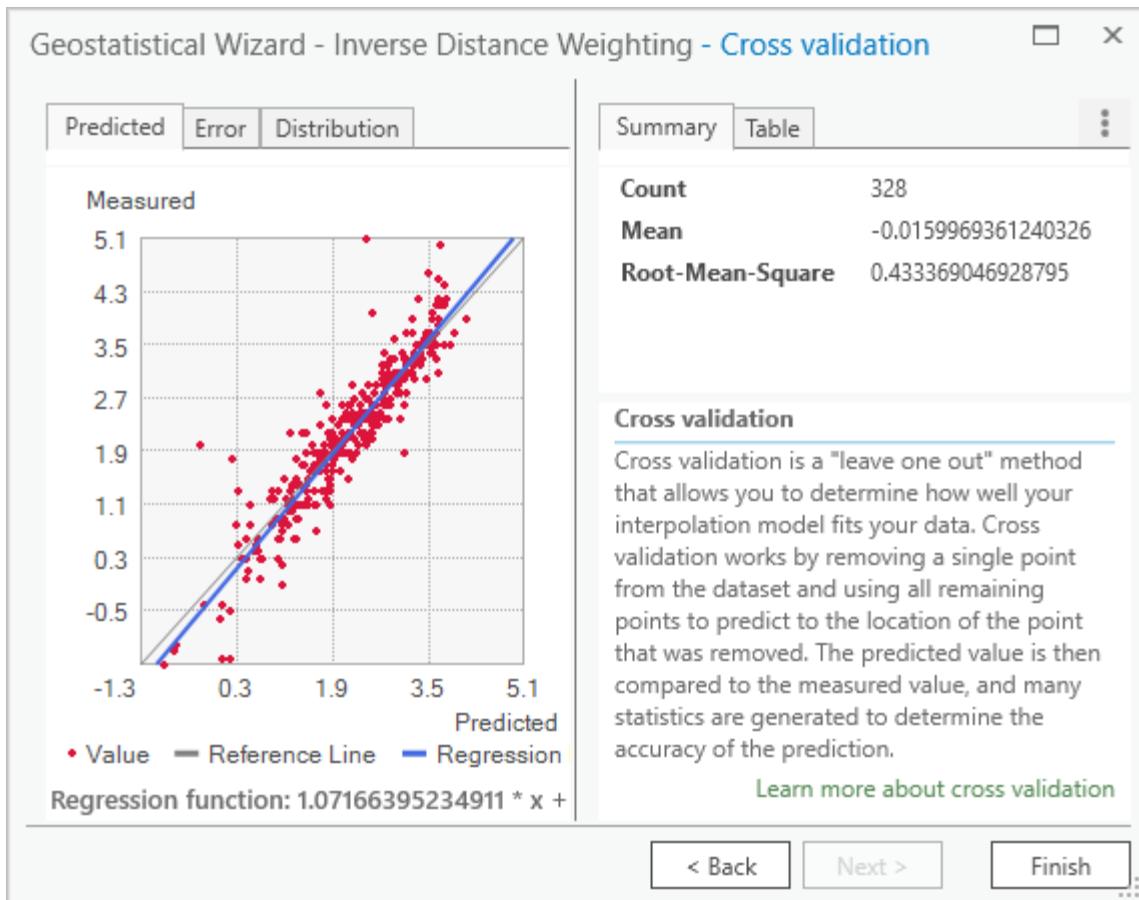
The **Cross validation** window shows you better results this time. This model gives more accurate predictions of ocean temperature than the previous one.

8. Click **Finish** and click **OK**.
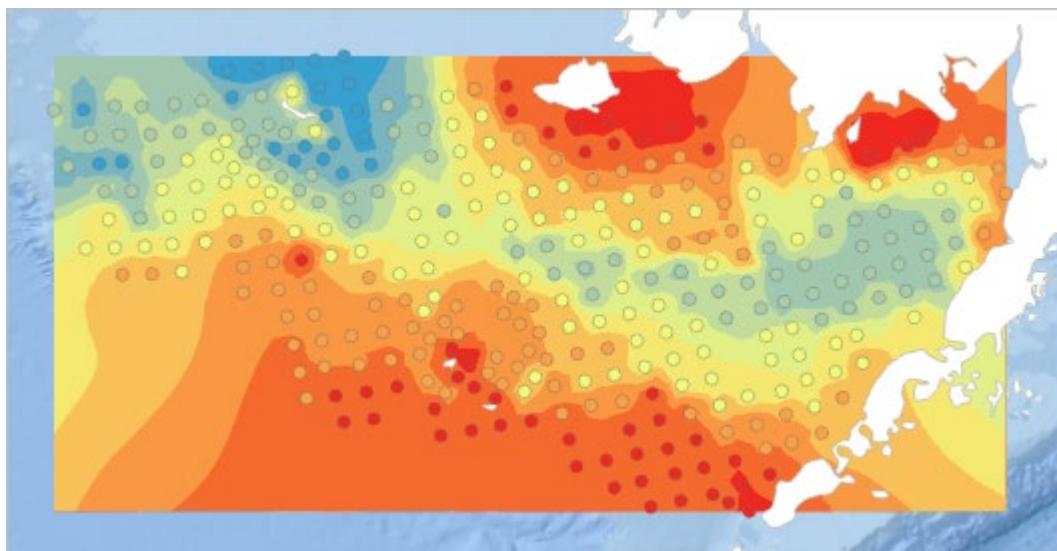
9. Save the project.

Optionally generate a new surface by entering your own input parameters for **Power**, **Neighborhood Type**, and **Smoothing Factor**. Which values create a better surface showing the distribution of temperature? The more accurate you can make your surface, the more useful it will be as a tool for understanding the distribution of pollock.

Predict ocean temperature using kriging

Kriging is a more sophisticated interpolation method than polynomial or IDW. But kriging works on some assumptions that require that your input data have a normal distribution. First you will create a histogram to determine if the distribution of bottom temperatures in the Bering Sea is normal.

1. In the **Contents** pane, right-click **BottomTemp**. Point to **Create Chart** and choose **Histogram**.
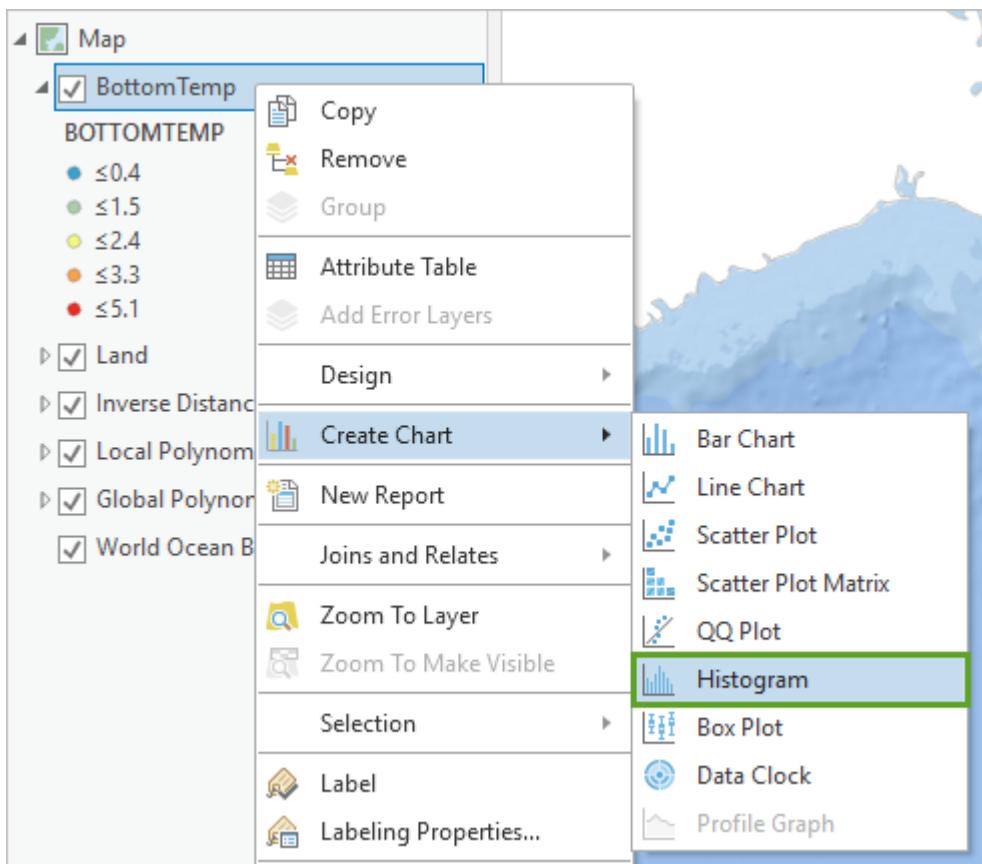
An empty chart view appears below the map, along with a **Chart Properties** pane.

2. In the **Chart Properties** pane, change **Number** to **BOTTOMTEMP** and check the **Show Normal distribution** and **Median** boxes.
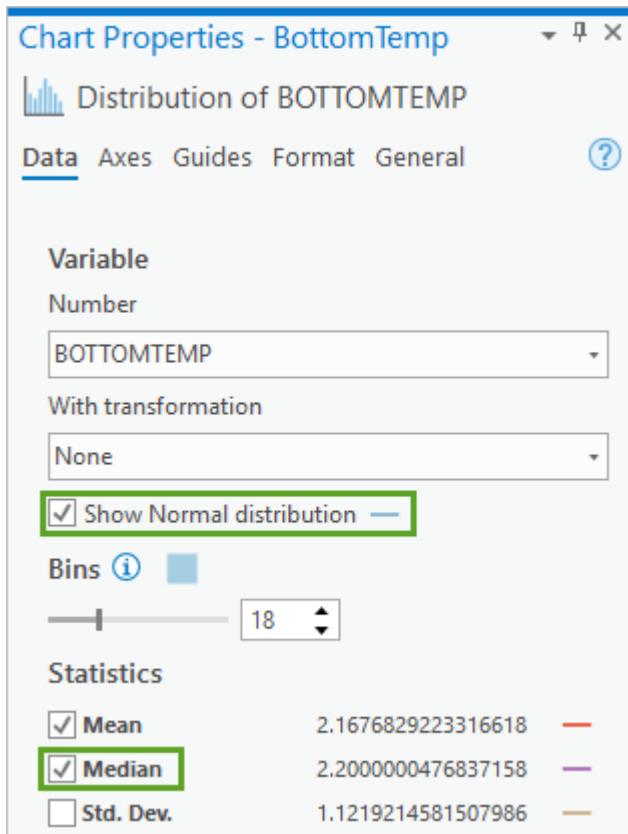
The chart view updates to show a histogram representing the temperature values of the point data.
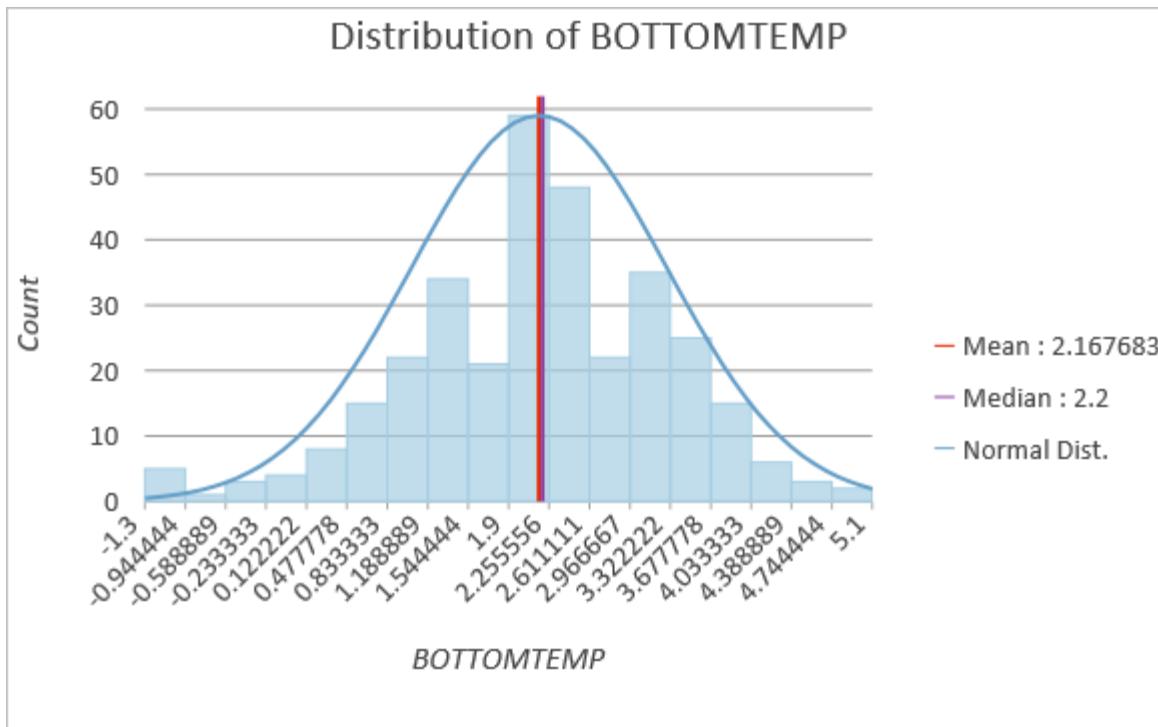
The histogram shows a normal distribution, also called a Gaussian distribution, which has a bell-shaped curve. This indicates that you can get reasonable results from the kriging geostatistical methods. The red and purple lines represent the mean and median, and they are very close together, which is another good indication that the distribution is normal.

You can learn more about normal distribution at this blog post from Insights for ArcGIS.

3. Close the chart view.

4. Open the **Geostatistical Wizard**, and under **Geostatistical methods**, select **Kriging / CoKriging**.

5. Click **Next**.

6. Under **Ordinary Kriging**, select **Prediction.**

7. For **Order of Trend Removal**, select **Second**.
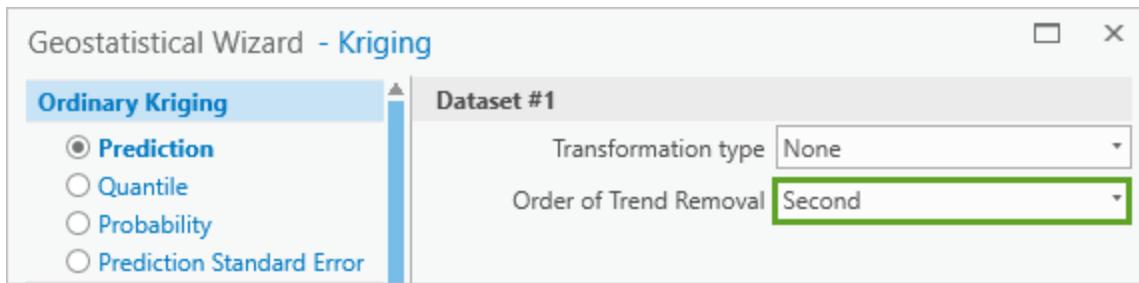
Some data sets will have a trend—for example, wind direction, or elevation increasing as you climb a mountain range. More often it is difficult to detect trends. The Bering Sea data might have a trend running parallel to the edge of the continental shelf.

When you apply trend removal, the model will do an initial interpolation to find broad trends and remove them, and then do the interpolation again on the detrended data.

8. Click **Next**.

9. For **Exploratory Trend Surface Analysis**, type *25* and click **Next**.

10. In the **Semivariogram/Covariance Modeling** window, click the **Optimize model** button.
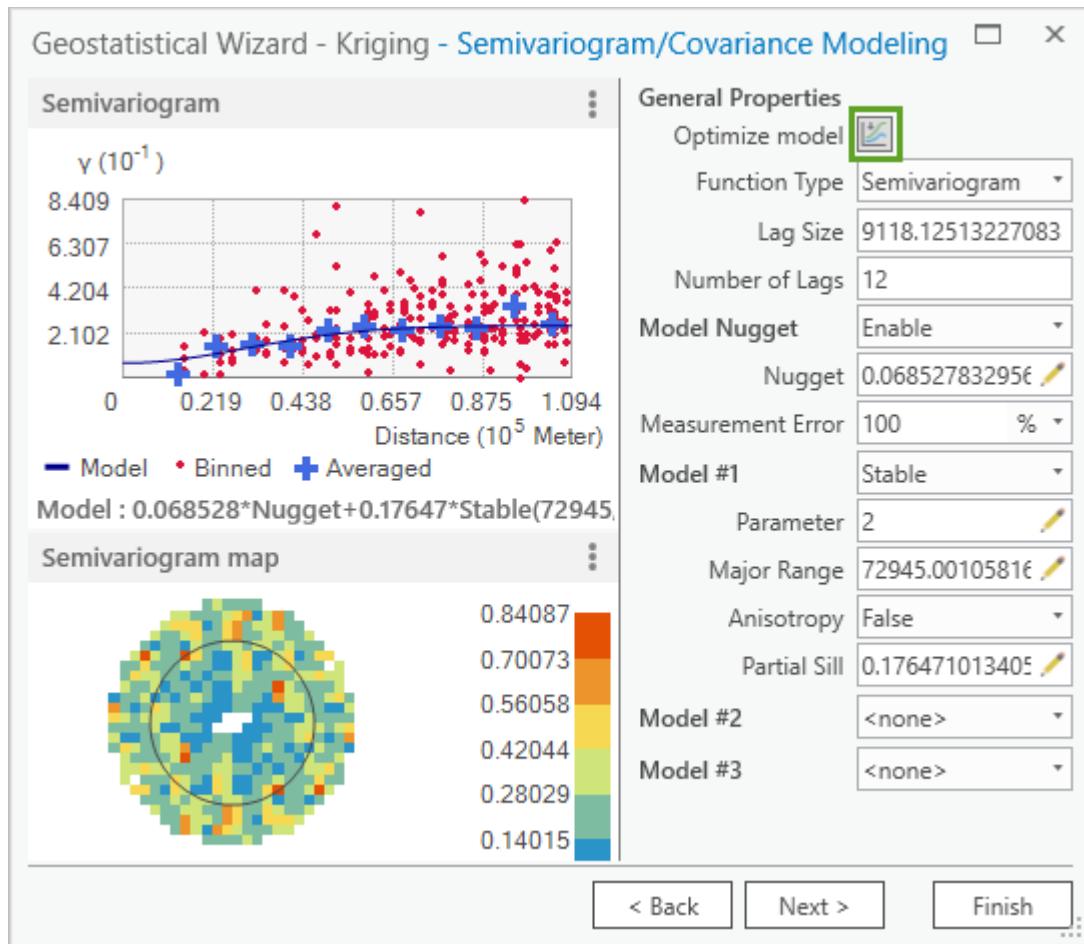
The semivariogram is very important to the kriging process and can help you determine if your model is good enough.

11. Click the options button next to **Semivariogram** and turn off **Show binned values**.
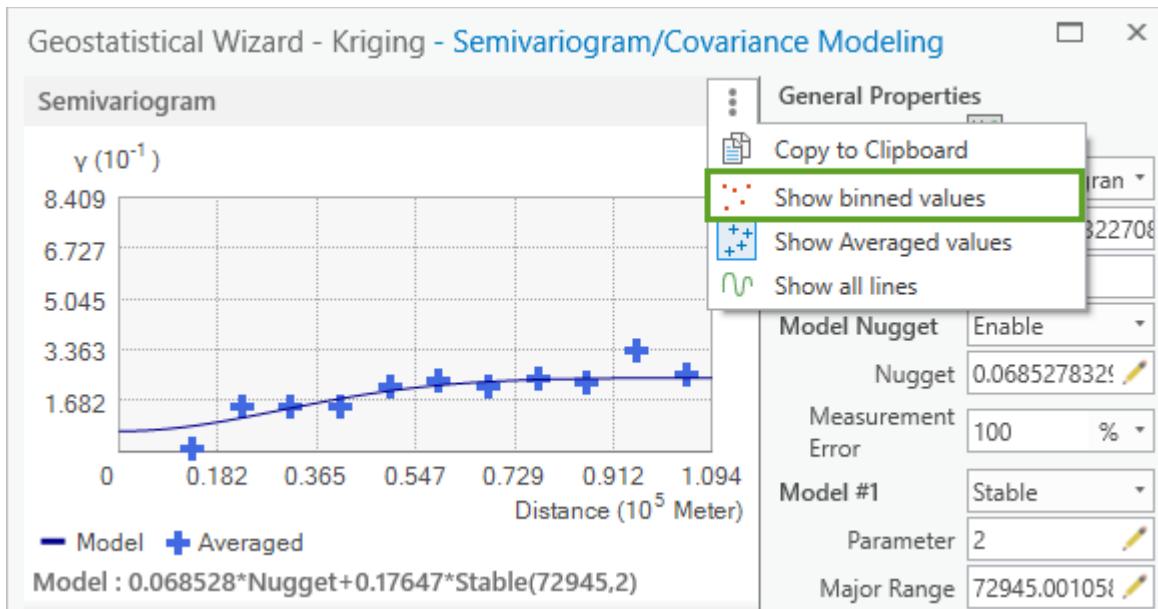
The semivariogram examines pairs of points in your data and compares the distance between them (on the x-axis) to their similarity (on the y-axis). Because things that are closer together are assumed to be more similar than things that are farther apart, a typical semivariogram will start near to zero on the y-axis, rise, and then level off. If the blue **Model** line on your graph differs from this shape, or if it does not follow the blue **Averaged** points, you should go back and change some of your kriging parameters.

12. Click **Next** twice to observe the values on the **Cross validation** page.

   All values should be close to zero except for **Root-Mean-Square Standardized**, which should be close to 1.

13. Click **Finish** and click **OK**.

14. Compare the Kriging surface to the others by checking and unchecking their boxes in the **Contents** pane.
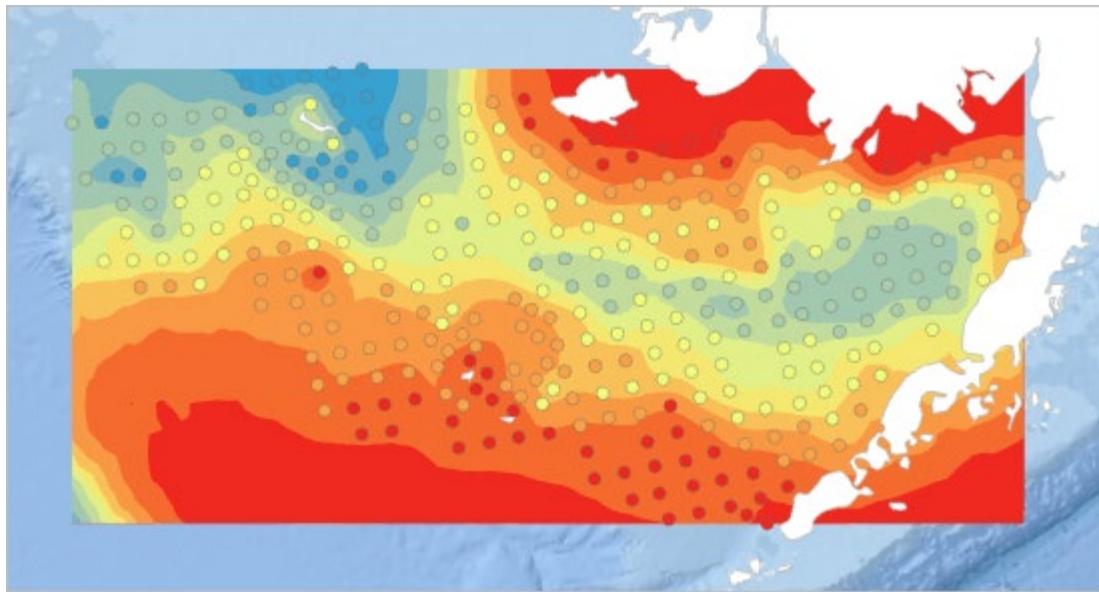
The Kriging layer is more general and less precise.

15. Save the project.

While polynomial and IDW are purely mathematical methods of interpolation, kriging uses the statistical properties of the sample points. Often, this creates a more general surface, which does not match the input data exactly, but which tends to produce more reliable predictions.

Choose a prediction surface

You have now created four prediction surfaces using four different methods: global polynomial interpolation, local polynomial interpolation, inverse distance weighting, and kriging. Which one is the best? You can use cross-validation to decide.

1. In the **Contents** pane, select all four prediction surfaces. Right-click and choose **Cross Validation**.

    Hint: To select more than one layer, press the **Shift** key on your keyboard while selecting layers.
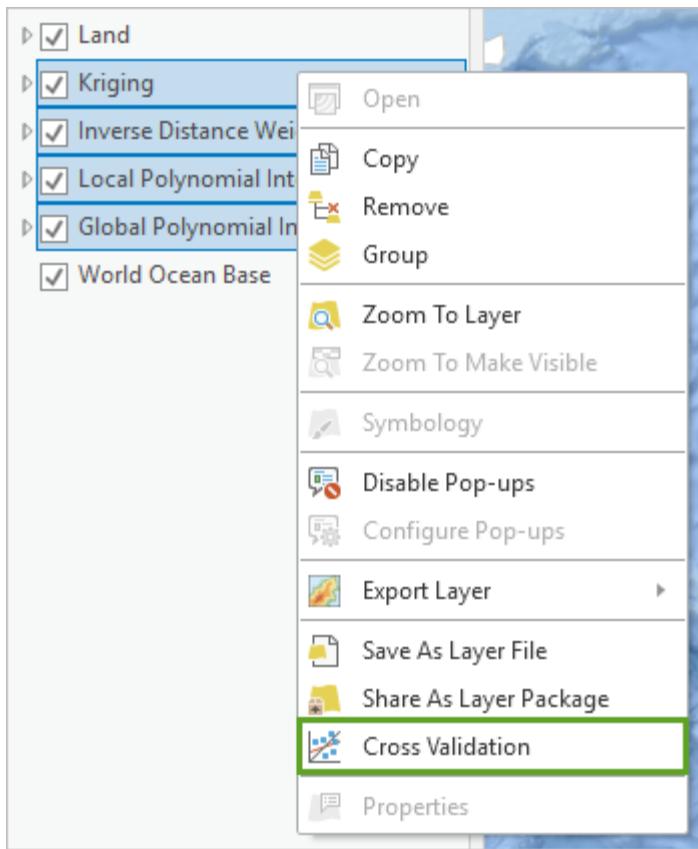
2. Four windows appear, all on top of one another. Rearrange them so you can compare them.

   Cross-validation is a report on the accuracy of a surface. Comparing the reports of different surfaces can help you decide which one to use.

3. Compare the scatterplots for the four layers. The closer the thin gray line is to the thicker blue line, the better the match of predicted values.

Cross validation - Map/Inverse Distance Weighting

Predicted | Error | Distribution

Regression function: 1.07166395234911 * x + -0.18202586875245{
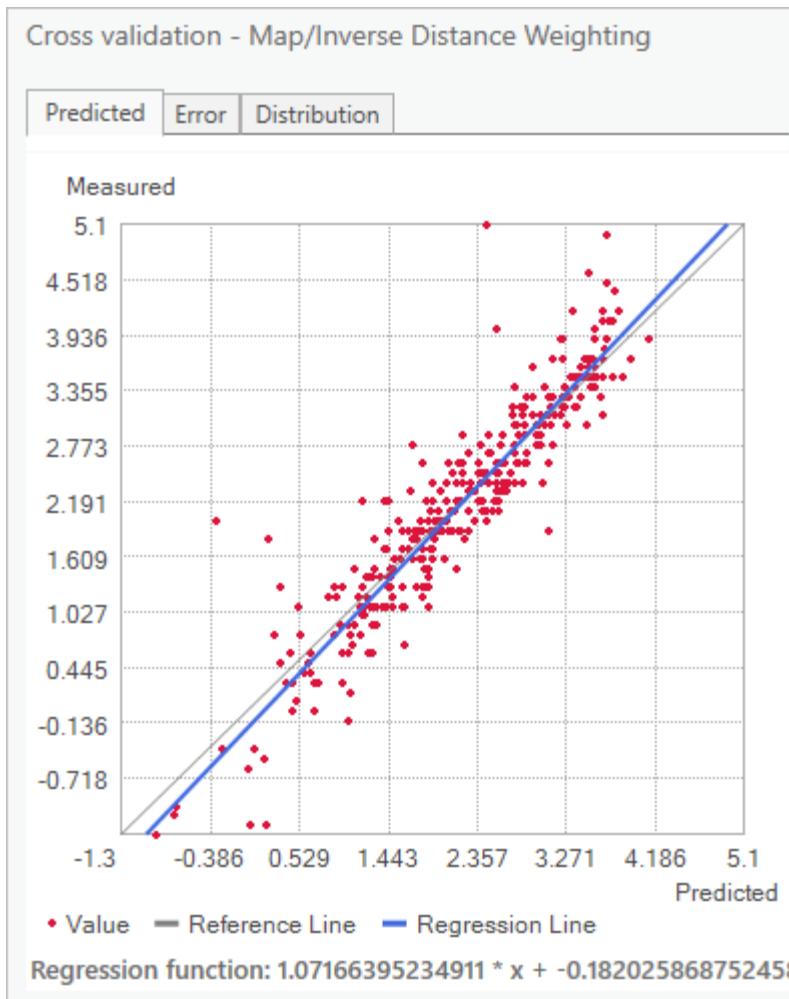
In this case, Kriging has the best scatterplot.

The **Root-Mean-Square** (RMS) value is calculated for all interpolation methods. If the other values are close to zero and do not cause concern, this one can be useful to determine which of your results is best.

4. Compare the **Root-Mean-Square** error value for each of the layers.

Model	Root-Mean-Square
Global Polynomial Interpolation	0.672
Local Polynomial Interpolation	0.417
Inverse Distance Weighting	0.433

Kriging	0.387

The ideal value for RMS is zero, but how close it should be to zero depends on the unit. The unit is the same as the input data—in this case, degrees Celsius. All of the models above produced good RMS error values, but kriging produced the best one.

5. Close all of the **Cross validation** windows.

6. In the **Contents** pane, make sure only **Inverse Distance Weighting** is selected. Right-click this layer and choose **Remove**.

7. Also remove **Local Polynomial Interpolation** and **Global Polynomial Interpolation**.

8. In the **Contents** pane, expand the **Kriging** layer.
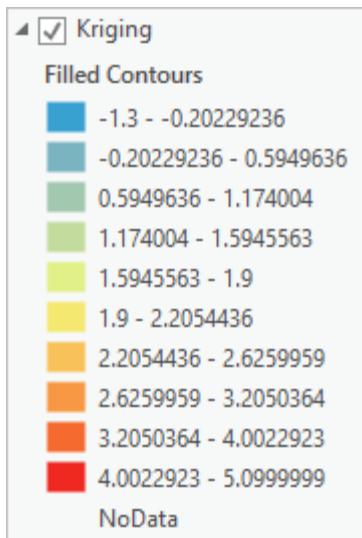
The legend shows temperature values with up to 8 decimal places. This suggests a precision that is not true. Earlier you saw that the RMS error value was almost 0.4 degrees Celsius. You can change the labels of the legend to be more general.

9. In the **Contents** pane, right-click **Kriging** and choose **Symbology**.

The **Symbology** pane appears.

10. In the **Class Breaks** table, click in one of the **Label** cells and press **F2** on the keyboard to make it editable.

Symbol	Upper value	Label
	≤ -0.202292364216872	-1.3 - -0.20229236
	≤ 0.594963601122714	-0.20229236 - 0.5949636
	≤ 1.17400403152556	0.5949636 - 1.174004
	≤ 1.59455631914413	1.174004 - 1.5945563
	≤ 1.89999997615814	1.5945563 - 1.9
	≤ 2.20544363317215	1.9 - 2.2054436
	≤ 2.62599592079072	2.2 - 2.6259959
	≤ 3.20503635119357	2.6 - 3.2
	≤ 4.00229231653316	3.2 - 4.0
	≤ 5.09999990463257	4.0 - 5.1

11. Edit each of the labels so they are rounded to one decimal place.

The legend is now easier to read.

◢ ☑ Kriging	
Filled Contours	
	-1.3 - -0.2
	-0.2 - 0.6
	0.6- 1.2
	1.2 - 1.6
	1.6 - 1.9
	1.9 - 2.2
	2.2 - 2.6
	2.6 - 3.2
	3.2 - 4.0
	4.0 - 5.1
	NoData
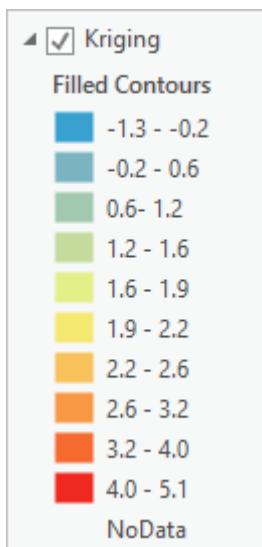
Alaska pollock prefer temperatures between 2 and 3 degrees Celsius. Based on this legend, those parts of the map with yellow, light orange, or medium orange colors are the most likely habitats for this kind of fish.

12. Save and close the project.

Summary

There are many different interpolation methods available to you in the Geostatistical Wizard. How can you know which one to use? It may be true that kriging is usually better than IDW, but it may not be better for your particular dataset. It is worthwhile to try many methods and experiment with their parameters to find the best fit.

You may have specific requirements for your model. If that is the case, you can refer to Classification trees of the interpolation methods offered in Geostatistical Analyst.

A more accurate knowledge of fish distribution is only one of the pieces necessary for improving fishing practices, alongside observation of fish behavior and modification of fishing equipment to prevent it from catching incidental species. Together, these and other studies help plan for a more sustainable catch.

To learn more about the interpolation methods you used in this lesson, see the following:

How global polynomial interpolation works

How local polynomial interpolation works

How inverse distance weighted interpolation works

Kriging in Geostatistical Analyst