Rebecca Risch

EARS 77 - Environmental Applications of GIS

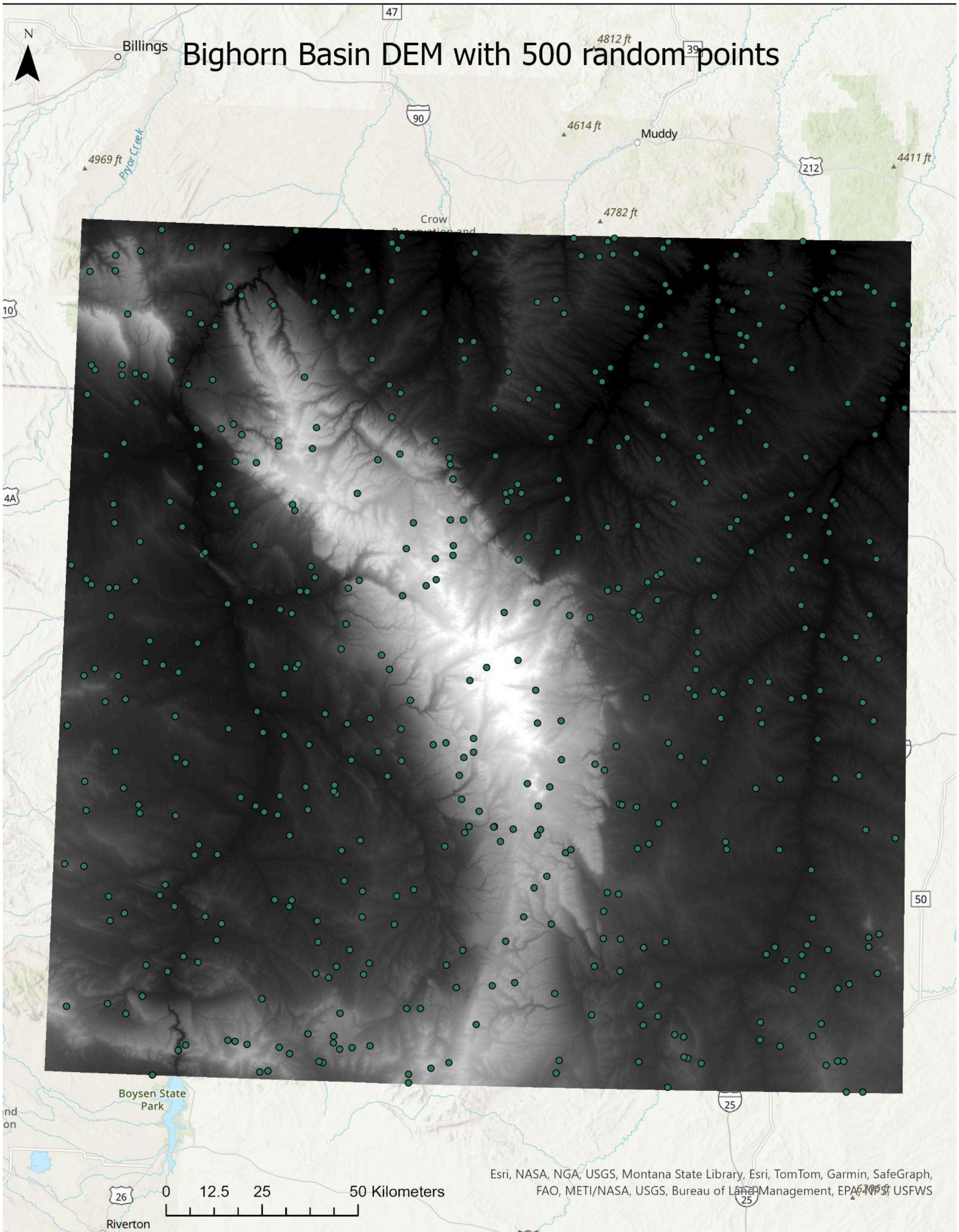
Professor Jonathan Chipman

23 January 2024

Lab 2 Report: Spatial Interpolation Algorithms and Kriging

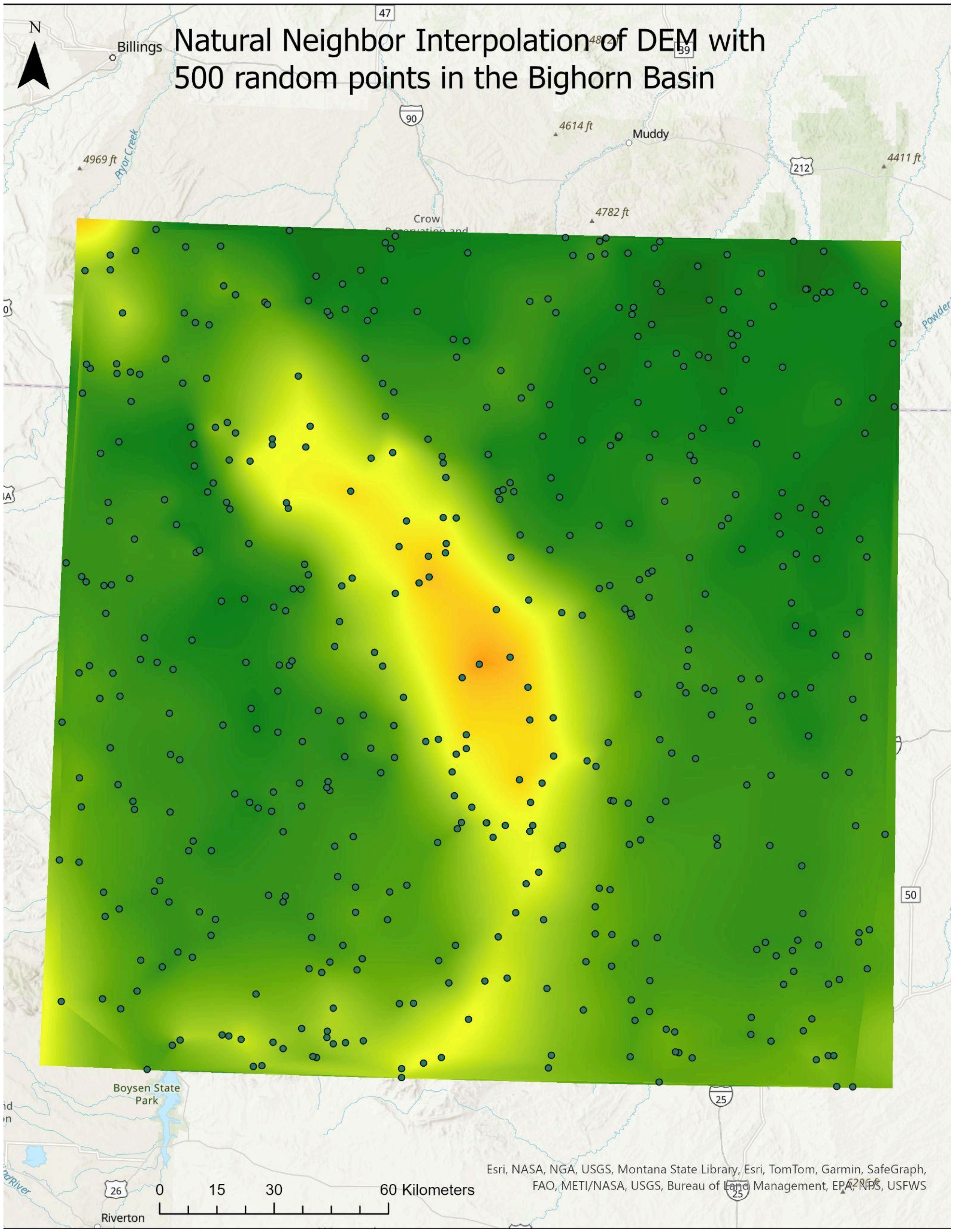
1. For parts A and B, I used the Bighorn Basin DEM.
    a. Rows: 2283
    b. Columns: 2259
    c. Cell size: 100 x 100
    d. Min: 934m
    e. Max: 4010m
    f. Mean: 1578m
    g. UTM Zone: 13N

2.



Bighorn Basin DEM with 500 random points
N
Billings
Muddy
Crow
Boysen State Park
Riverton
4812 ft
4614 ft
4969 ft
4411 ft
4782 ft
Pryor Creek
0 12.5 25 50 Kilometers
Esri, NASA, NGA, USGS, Montana State Library, Esri, TomTom, Garmin, SafeGraph, FAO, METI/NASA, USGS, Bureau of Land Management, EPA, NPS, USFWS

3. After changing the symbology of the error raster layers of each interpolation method to a red-white-blue stretch, I was able to get a better sense of the accuracy of each interpolation method in comparison to the DEM. Darker patches of color represented areas with high error, while light patches (white) represented areas with low error. Inverse Distance Weighting (IDW) to the power of 3 appears to have less error than IDW to the power of 1. Nearest Neighbor interpolation appears to have a bit more error than IDW 3 and produces a coarser error raster. Splining produced quite a bit more error than IDW 3. Natural Neighbor interpolation appears to be the winner however, it produced the least amount of error of the group- less than IDW 3 which is the runner up. The Natural Neighbor interpolation raster is shown below, along with the original 500 random points selected from the DEM:


Natural Neighbor Interpolation of DEM with
500 random points in the Bighorn Basin
N
Billings
Muddy
Crow
4969 ft
4614 ft
4782 ft
4411 ft
Pryor Creek
Boysen State
Park
Riverton
0 15 30 60 Kilometers
Esri, NASA, NGA, USGS, Montana State Library, Esri, TomTom, Garmin, SafeGraph,
FAO, METI/NASA, USGS, Bureau of Land Management, EPA, NPS, USFWS
a.

4.

Algorithm	Min	Max	Mean	Std Deviation
Nearest Neighbor	–1474	1080	-2.1	161
IDW (1st order)	-1085	616	-15.6	150
IDW (3rd order)	-1083	666	-6.8	131
Spline	-1326	1663	3.0	172
Natural Neighbor	-953	626	-3.5	119
Kriging	-970	620	-2.4	112.9

5. Kriging is an interpolation method based on spatial autocorrelation, where the model predicts values of unmeasured points based on the values of the neighboring, measured points, and weighting on distance. We use interpolation to create higher resolution data: interpolation estimates values of data points of which we currently have no value measurement, based on and within the range of the available, measured data points (or the ground truth). Spatial autocorrelation is analyzed with the semivariance metric, and analysts will then create a semivariogram, a graph which depicts variability between the value of data points as a function of distance. In this case, we use lag distance. To find lag distances between points, we take a center point and then group the surrounding points based on distance from the center point. Then we pair each point with the center point, comparing the measured values, and then comparing that difference in values to the other difference in values of the same lag distance group. We plot semivariance (represented by gamma, calculated with the following formula (see 5a)) against lag distance (represented by h). We then fit a model to the trend of the semivariogram to explain semivariance vs

lag distance with a more simple/straightforward formula. There are a few options in models to choose from, including Stable, Circular, Spherical, Exponential, Gaussian, etc. For this lab, I chose the Spherical model as I felt it best fit the data points and general trend of the semivariogram, especially the left side of the graph. Semivariogram models have a few components: notably range, sill, and nugget. Range is the distance between the y axis (where h = 0) and the relative point in the graph where the curve flattens out. The semivariance value (gamma) that the range reaches right as it flattens out is called the sill. The nugget is the y-intercept of the graph, the value of semivariance where lag distance equals zero . The nugget isn't necessarily equal to zero because if you continuously remeasure the center point, you might get different results, meaning its semivariance could be non-zero. For this lab, I kept all defaults (for lag size, amounts of lag groups, etc) in ArcGIS Pro except for the model type, for which I used the spherical model.

$$\gamma_h = \frac{\sum_a \sum_b (z_a - z_b)^2}{2n}$$

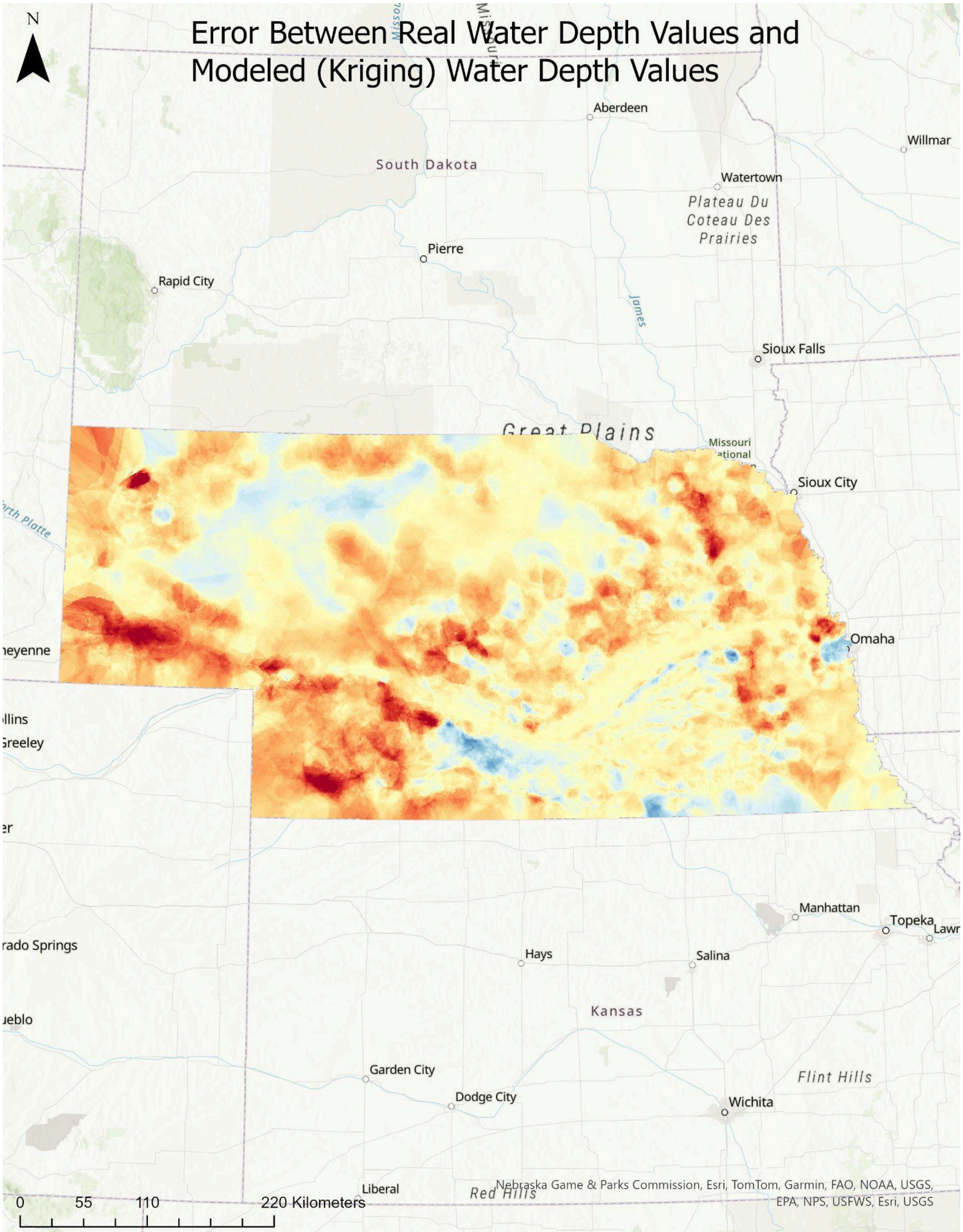
a. **Where**
h = lag distance (within specified tolerance)
z_a and z_b = values of two points separated by lag h
n = number of pairs that are separated by lag h

Small lag distances should have low semivariance
Larger lag distances should have more semivariance

6. See Python notebook

7. From the subtraction of the two Kriging models, the water depth decreased by an average of 6.60m between the two time periods of our data. For the 2020 time period, between the real and modeled sets of water depths, my mean error was 0.60m, my mean absolute error was 7.99m, and my root mean square error was 12.59m. The error between the real water depth values measured at specific locations and the modeled water depth represents the difference in depth values, or how close the model predicted to the actual ground truth data. The mean absolute error (calculated by taking the average of the absolute values of the errors) is less sensitive to outliers, and causes positive and negative errors to cancel out. The RMSE (root mean squared error, calculated by square rooting the average of each error value squared) will be more sensitive to outliers, and best represents the magnitude and variability of errors. These error metrics allow us to assess the accuracy of kriging (lower metrics being more accurate), which is used here as the interpolation method. If you threw a dart randomly at a map of Nebraska and drilled a well at its location, you wouldn't expect these error statistics to apply very well to that well, because even if the well was within the range of our known data, we could interpolate that point, but our interpolation wouldn't be perfect because it's a point that might not necessarily be spatially autocorrelated with the rest of the dataset. Kriging is based on spatial autocorrelation and so adding a data point would change the autocorrelation relationship on which our model is based. Also, the model is based on the broader data set as a whole, and how the water depth changes based on location, and specific wells might have other specific geological, hydrological, and anthropogenic-caused factors influencing the water depth at that location that we wouldn't be able to predict for nor understand yet.

8.



N
Error Between Real Water Depth Values and
Modeled (Kriging) Water Depth Values
Aberdeen
Willmar
South Dakota
Watertown
Plateau Du
Coteau Des
Prairies
Pierre
Rapid City
James
Sioux Falls
Great Plains
Sioux City
Omaha
Manhattan
Topeka
Hays
Salina
Kansas
Flint Hills
Garden City
Dodge City
Wichita
Liberal
Red Hills
Nebraska Game & Parks Commission, Esri, TomTom, Garmin, FAO, NOAA, USGS,
EPA, NPS, USFWS, Esri, USGS
0 55 110 220 Kilometers