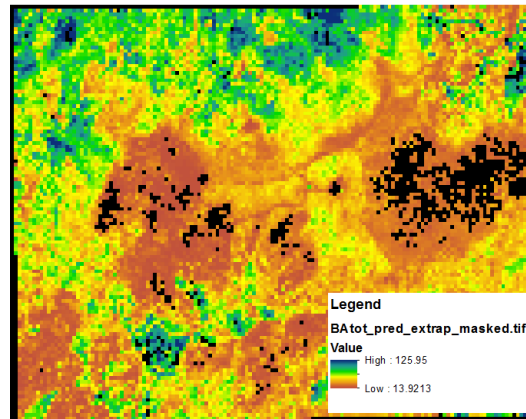# EXERCISE 4b
Modeling in ArcMap

Introduction

Regression models are used to characterize relationships between field measurements and lidar metrics at the plot level in order to generate spatial predictions. Building these regression models does require some understanding of statistics, including the assumptions made and how to evaluate models and interpret model results. The workflow presented in this exercise is built on a larger workflow where model equations are created using R. However, this exercise doesn't include any programming, and all of the work done in R is provided to you. This exercise simply starts you off with some lidar first order metrics (gridmetrics) and shows you how to apply a simple modeling equation in ArcGIS.

Many predictive models take the form of relatively simple mathematical functions or equations that can be applied to raster data using band math (e.g., using ArcMap's Raster Calculator). More complex modes or decision trees are less easily applied using standard GIS software and it may be desirable to apply models directly using R. In the main part of this exercise, we will use simple map algebra in ArcMap to generate spatial predictions from the linear model created in the previous exercise.

It is important to remember the wise old saying:

> *"all models are wrong but some models are useful…"*
> - George Box.

Objectives

- Apply linear regression model to lidar derivatives in ArcMap

Required Data

- **Tile_layout.shp**
- **1st_cover_above_mean_25METERS.asc**

- **all_1st_cover_above_mean_25METERS.asc**
- **elev_P40_2plus_25METERS.asc**
- **elev_P70_2plus_25METERS.asc**

Prerequisites

- It is recommended that you are somewhat proficient using ArcMap however it is not necessary for this exercise to have an extensive knowledge of ArcMap.

Table of Contents

Part 1: Review and Prepare the Spatial Predictor Layers

A. Load the AOI

1. If it is not already open, start ArcMap by clicking on the Start button and navigating to **All Programs, ArcGIS** and then open **ArcMap.**

2. Open a **blank map**.

3. Add the example study area extent:

   i. Use the **Add Data** button to add the **tile_layout.shp** from the **Vector** data folder (**…\Track2_VegetationDerivatives\Vector**). This is a subset of the tile layout files showing the footprint of the lidar tiles used to create the gridmetrics you'll load in the next step.

4. Load the lidar predictor variables included in the linear model. Use the Add Data button to add four of the ascii gridmetrics from the **Lidar** folder (**…\Track2_VegetationDerivatives\Lidar**)
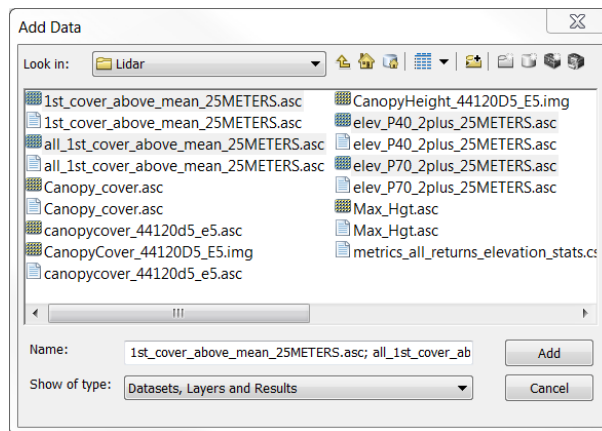

   i. **1ˢᵗ_cover_above_mean_25METERS.asc**

   ii. **all_1ˢᵗ_cover_above_mean_25METERS.asc**

   iii. **elev_P40_2plus_25METERS.asc**

   iv. **elev_P70_2plus_25METERS.asc**

   v. You may get an "Unknown spatial reference" warning. Click **OK**. You'll address this in the next section.

B. Define Projection

> *The lidar grid metrics were generated in a lidar processing software called FUSION. But ArcMap doesn't recognize FUSION projection definitions, so it is a good idea to redefine these prior to your analysis. Below will take you through this process.*

1. Open the **Define Projection** tool in batch mode

   i. Open ArcToolbox and navigate to **Data Management, Projection, Transformations, and** then open the **Define Projection** tool**.**

ii. Right click on the tool and select **Batch…** as shown in the graphic below.
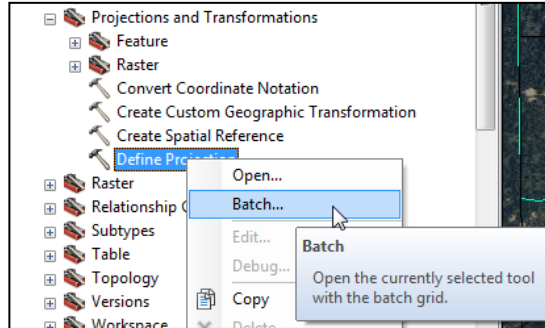
Note: you can use a batch operation on most of the tools available in the Arc Toolbox. This can be very useful when you need to perform a certain operation on several different files.

2. Load the Input Datasets

i. Double click in the **Input Dataset or Feature Class** field and select the first lidar gridmetrics from the dropdown (e.g., elev_P70_2plus_25METERS.asc), and click **OK.** The Coordinate System column will display "Unknown" if the selected file does not have a defined projection.

ii. Use the **+** icon to the right to add another row

iii. Double click the second **Input Dataset or Feature Class** field and specify the second lidar gridmetric variable

iv. Repeat steps ii and iii twice more to add all of the lidar grid metrics.
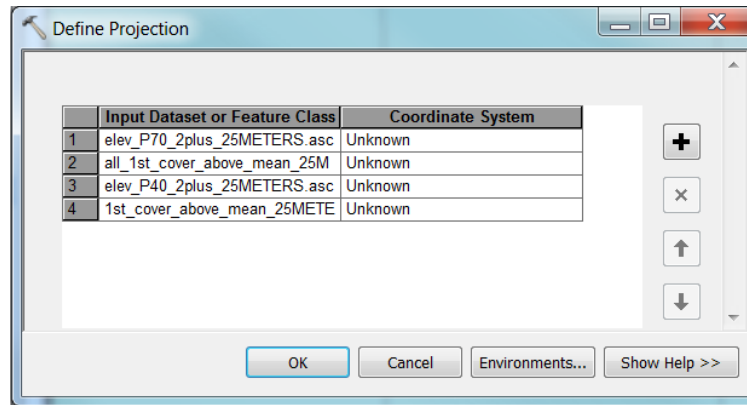
3. Import the **Coordinate System** from an ancillary layer

i. Double click on the first **Unknown** in the **Coordinate System** field.

ii. In the Define Projection: 1 window that opens, click the **Coordinate System** button

iii. In the spatial reference properties window that opens, you can click the **+** sign next to Layers to open the available spatial reference properties from the layers you have loaded. The "Unknown" properties come from the current lidar gridmetrics. To adjust the

gridmetrics to the spatial reference from the tile_layout shapefile, click on
NAS_1983_UTM_Zone_12N.

iv. Click **OK** in the spatial reference properties window, then **OK** in the Define Projection
window.

v. Repeat steps i-iv for all four gridmetrics, or copy and paste the coordinate system for all 4
rows.

4. Click **OK** to run the tool

C. Inspect The Lidar Metric Layers

1. Note areas where height and density are high and low

2. Note the presence and location of **No Data** values (hint: change the display in the raster
properties)

The output products from FUSION's Lidar Tool Kit (LTK) are seamless GIS layers, each representing a single lidar canopy metric which we refer collectively to as gridmetrics. Often, it is preferable to do some preprocessing on these gridmetrics layers to facilitate their incorporation into your GIS. Preprocessing routines may differ according to your project and goals.

Part 2: Open the Raster Calculator Tool

A. Open the Raster Calculator Tool

1. Make sure that the Spatial Analyst extension is turned on.

i. From the main menu, select **Customize** and choose **Extensions**. Make sure that the box
next to **Spatial Analyst** is checked.

2. Expand the **ArcToolbox** and navigate to and open the **Raster Calculator** tool (**ArcToolbox,
Spatial Analyst, Map Algebra, Raster Calculator**)

B. Input the Model Prediction Equation

1. Copy the below equation:

```
Model Equation formatted for Raster Calculator:

10.7485969 + 0.97001955 * "elev_P70_2plus_25METERS.asc" + 0.1248454  *
"elev_P70_2plus_25METERS.asc" + 0.0413567 * ("elev_P70_2plus_25METERS.asc" *
"all 1st cover above mean 25METERS.asc")
```

2. Verify the correctness of the syntax edits prior to running Raster Calculator

    i. Variable names must correspond *exactly* to the variable names of the gridmetrics layers. Variable names must be in quotes, and the interaction term (height x density) must be enclosed in parenthesis to ensure proper order of operations in the calculation.

3. Save the output calculation as a raster in the **outputs** folder.

    i. Navigate to this folder **Track2_VegetationDerivatives\Outputs\LidarOutputs** and specify the name as **BAtot_pred.tif.**

C. Run the Model to Create the Basal Area Prediction

1. Click **OK** to execute the expression and create the prediction.

   i. If you encounter an error, refer back to the expression in the text editor and look for syntax errors.

> **Tip:** Text editors such as **Notepad++** or **Textpad** have automatic syntax checking which can help you troubleshoot syntax errors.

2. The output prediction layer should automatically be added to your ArcMap project. If not, add it before continuing to the next step.

Part 3: Mask Out Extrapolation and Non-Forest Areas

A. Visually Inspect the Prediction of Basal Area

1. Explore this basal areal prediction layer.

    i. Note the maximum and minimum predicted values and where high and low values occur on the landscape.

    ii. *Optional* — Change the color ramp of the raster display.
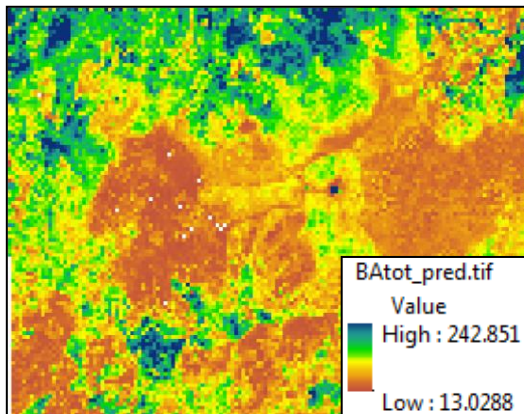
Why does this matter? Recall that predicting beyond the ranges of the original data will result in our models extrapolating which can result in serious prediction errors. While ultimately, it is ideal to be able to create versatile and generalizable models, it is often impractical to sample the full range of variability possible across the landscape. Often, the safest way to deal with this issue is to mask out areas where models extrapolate beyond the range of observation.

B. Get the Ranges of the Observed Plots

1. The code below is an example of the range of observed lidar plots being obtained in the R Console. This is how we get the ranges for each variable in our model.

   i.   This shows the return of the maximum and minimum values in the plot data, shown below.

Minimum observed

Maximum observed value

```
> range(lidarData$P70)

[1]  2.62613 27.47566

> range(lidarData$all_1st_cover_mean)

[1] 1. 551981 80.713702
```

C. Use the Conditional Tool to Create Extrapolation Masks for Each Variable—Identify Values Beyond the Range of the Plot Data and Set Them to Zero

1. In ArcMap, open the **Con** tool in the Spatial Analyst Conditional toolset.

2. Create an extrapolation mask for the first predictor layer (the height metric).

   i. For the Input conditional raster, select the **elev_P70_2plus_25METERS.asc**

   ii. Use the following expression: **"VALUE" > 2.62613 And "VALUE" < 27.47566**

     (a) The value ranges are acquired from rstudio, seen in section B.

   iii. Set the Input true raster or constant value to **1** and the Input false raster or constant value to **0.**

   iv. Save the output mask in the **…Track2_VegetationDerivatives\Outputs\LidarOutputs** folder and name it **elev970_extrap_mask.tif.**
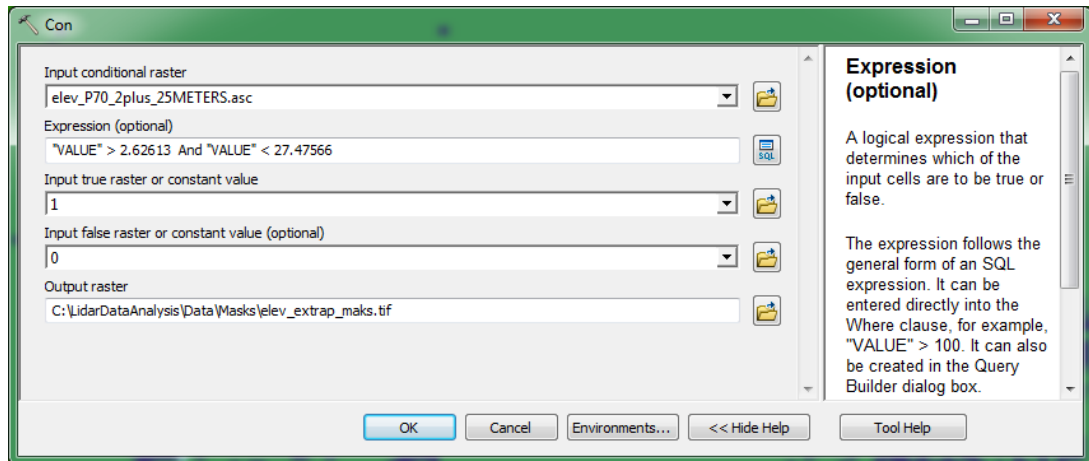
3. Create an extrapolation mask for the second predictor layer (cover metric). Open the **Con** tool in ArcMap.

    i. For the Input conditional raster, select the **all_1st_cover_above_mean_25METERS.asc**

    ii. Use the following expression: **"VALUE" > 1. 551981 And "VALUE" < 80.713702**

       (a) Again, the value ranges were acquired from rstudio, seen in section B.

    iii. Set the Input true raster or constant value to **1** and the Input false raster or constant value to **0.**

    iv. Save the output as **all1stcover_extrap_mask.tif.**

## D. Combine the Extrapolation Layers to Create a Model Extrapolation Mask

1. Multiply these two masks you just created together using the **Times** tool, which you can find by opening **Spatial Analyst** and then the **Math** toolbox.

    i. Navigate to and open the **Times** tool.

    ii. In the **Input raster or constant value 1** field, specify the first extrapolation mask layer created in section C step 2 above (**elev970_extrap_mask.tif**).

    iii. In the **Input raster or constant value 2** field, specify the second extrapolation mask layer created in section C step 3 above (**all1stcover_extrap_mask.tif**).

    iv. In the **Output raster** field, navigate to the **outputs** folder and name the output raster **combined_extrap_mask.tif**

    v. Click **OK** to run the tool.

> *In the resulting layer, pixels where predictor variable values are beyond the range of the data will have a value of **0**. Pixels where **both** variables are within the range of observations will have a value **1**.*

E. Apply the Combined Extrapolation Mask to the Basal Area Prediction Using the Set Null Tool

1. Open the **Set Null** tool by navigating to Spatial Analyst, then Conditional.

    i. We will use this tool to set pixel values in extrapolation areas to the NoData value.

2. In the **Input conditional raster** field, select the combined extrapolation mask created in section D above.

    i. Choose **combined_extrap_mask.tif** from the dropdown.

3. In the **Expression** field, specify the values in the mask that should be set to NoData (i.e., masked out of the prediction layer).

    i. Enter the expression "VALUE" = 0.

4. Specify the prediction layer in the **Input false raster or constant value** field

    i. Select **BAtot_pred.tif** from the dropdown.

    ii. Where the pixel values in the mask are *not equal to 0,* the values from this layer will be used in the output.

5. In the **Output raster** field, navigate to the outputs directory and name the output with the additional description of what the layer represents.

    i. Use the name **BAtot_pred_extrap_masked.tif**
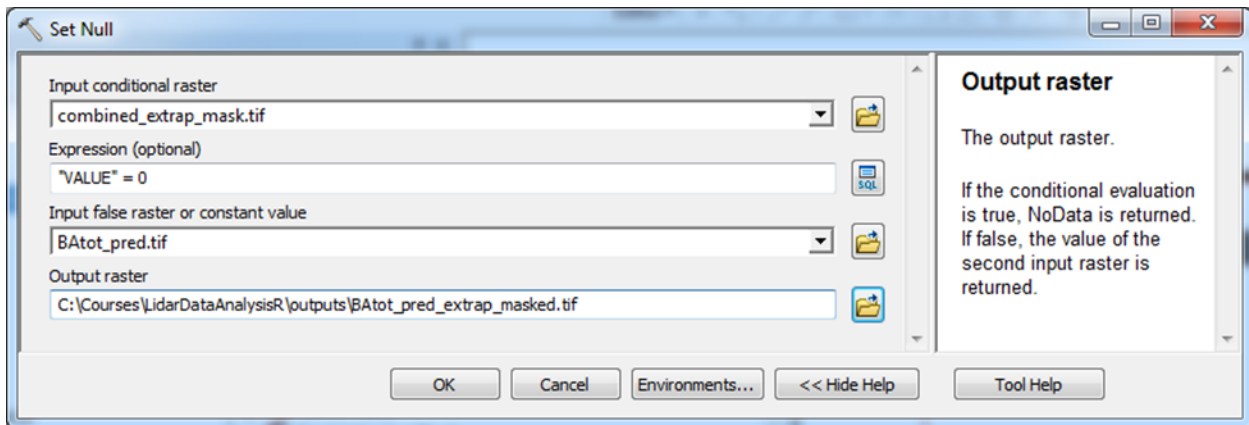
6. Click **OK** to run the tool to create the masked prediction layer.

    i. The result should be added automatically, if not, add the result layer to your ArcMap project.

F. Examine the Masked Prediction Layer and Compare it to the Unmasked Layer

*In the masked layer, prediction values where the lidar predictor values were beyond the range of the observation data are now coded as **NoData** and will not be included in the raster layer statistics.*

1. Look at the properties of each prediction layer to compare the maximum and minimum values predicted for basal area.

    i. Right click on the layer to open the raster properties, select the **Source** tab and scroll down to view the layer statistics.

    ii. Read the text box below for a discussion of this.

2. Change the display to view the NoData values in black (shown in the graphic below)

    i. Right click on the layer to open the **Properties...** and under the **Symbology** tab, change the **Display NoData as...** to black.

> **Note:** *Before we applied our mask to eliminate areas where lidar metric values are beyond the range of the observed data, predicted total basal area values ranged from 13.0 to 242.9 kg/hectare. After we identified and masked out these areas, the values ranged from 13.9 to 126.0 kg/hectare. It's possible that these large predicted total basal areas are reasonable; however given that our linear model is extrapolating beyond the range of observations, we have no way to estimate the accuracy in these areas.*
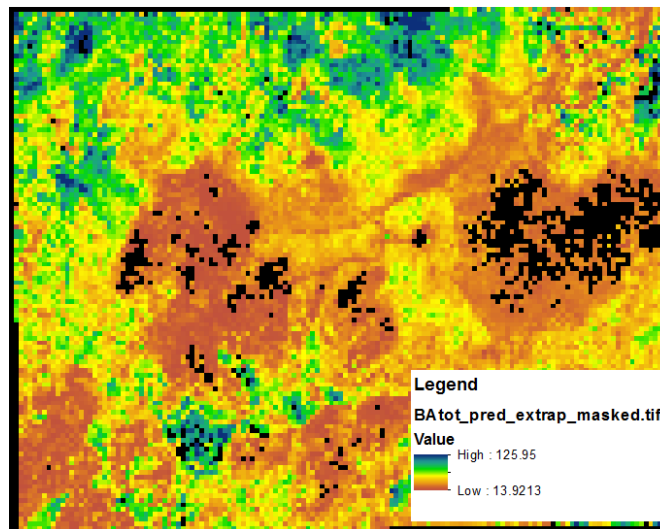

G. *Optional*—Clip the Prediction to a Forest Area Polygon Shapefile

1. Load the forest polygon shapefile into your ArcMap project.

    i. Use the **Add Data** button to add the **forest.shp** polygon layer from the folder **…Track2_VegetationDerivatives\Vector**

    ii. Optional—change the display so that you view prediction values inside and outside the forest polygon.

    iii. Note that many of the extrapolation areas are outside of the forest area.

2. Open the **Clip** tool for rasters.

    i. This tool is located in the Data Management group in the ArcToolbox (Data Management Tools>Raster>Raster Processing>Clip).

3. Select the **BAtot_pred_extrap_masked.tif** layer as the **Input Raster.**

4. For the **Output Extent** select the **forest.shp**

5. Select the option **Use Input Features for Clipping Geometry (optional).**

6. Specify the name of the **Output Raster Dataset as the BAtot_pred_forestMask.tif**

    i. Save this to the course data folder in the outputs folder.

7. Click **OK** to run the tool

> **Note:** *Care should also be taken not to apply models to other forest or vegetation types as this represents another type of extrapolation. In other words, if plot data represents a single forest type or species, your model should only be applied in the areas represented by your sample (White et al 2013).*

Masking out non-forest areas in the inventory area could be accomplished using an ancillary vegetation type dataset as demonstrated here. Reliable data is often not available and it is usually preferable to use the lidar metrics themselves to generate a forest/non forest mask. For example a forest mask could be created by applying thresholds to lidar metrics such that forest pixels are assumed to have greater than 2% canopy cover and a minimum height value of 2 meters.
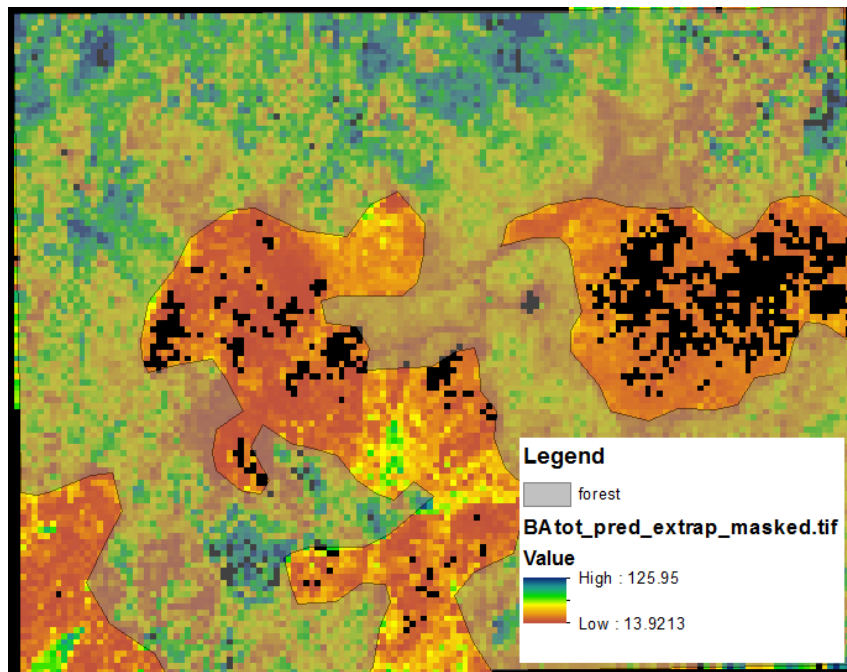
Congratulations! You have completed this exercise and used map algebra in ArcMap to apply a linear model equation across the landscape to predict total basal area from lidar predictor gridmetrics. This process represents the final step in the lidar modeling workflow. Remember that the steps leading up to this exercise required some work in R.

The purpose of this exercise has been to introduce you to using familiar tools in ArcGIS to work with and generate spatial data from models. If you are interested in this process or plan to use modeling methods other than multiple linear regression, we recommend that you learn to use R and Rstudio. GTAC offers additional training on the subject.