



Jun 02, 2020

Protocol for predicting specialist herbivore distribution from host distribution data

Joshua Fuller¹¹University of North Carolina at Chapel Hill

1 Works for me dx.doi.org/10.17504/protocols.io.9gvh3w6

Joshua Fuller

ABSTRACT

A necessary, but difficult, part of ecology and conservations is to try to create order and meaning from chaos. An example of such a task is the need to predict where an organism may exist at any given point in a landscape. In the past, this was done by systematically surveying the entire landscape, a lengthy and expensive endeavor. With the current pace of habitat loss and climate change, new methods were created to efficiently determine the distribution of a species using computers and new statistical algorithms. These new methods still require people in the field taking GPS points of the target species. A way to ease the workload on researchers in the field is to use data generated through citizen science projects such as iNaturalist. Citizen scientists can rapidly generate a tremendous amount of species GPS points by just exploring parks and backyards. That data can be directly utilized by researchers to create a species distribution model (SDM). The goal of this protocol is to take that a step further by generating an SDM for a plant, *Tsuga canadensis*, and then using that model to determine the distribution of an invasive specialist, *Adelges tsugae*, on the plant. By the end of this protocol, you should be able to create an SDM, generate a map using GIS software, and critically evaluate the results of the model to determine any conclusions that can be drawn from the model.

GUIDELINES

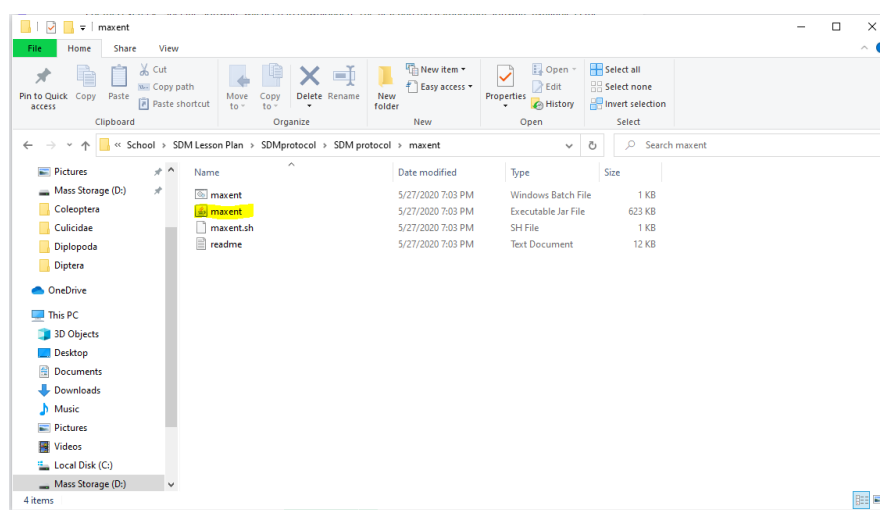
While a basic background in GIS is not required, it is helpful. This procedure is developed primarily for PC users and specific steps may be different for IOS users.

BEFORE STARTING

READ THE ABSTRACT. Make sure to download and unzip the zip file titled "SDMprotocol.zip"

Setting UP Software

- 1 For this exercise, specific software will need to be downloaded. The first and most important software available is the statistical package called MaxEnt. You can find MaxEnt in the SDM Protocol folder in a Folder titled MaxEnt. This program should be ready to use and require no installation steps. You can read more about it here: https://biodiversityinformatics.amnh.org/open_source/maxent/

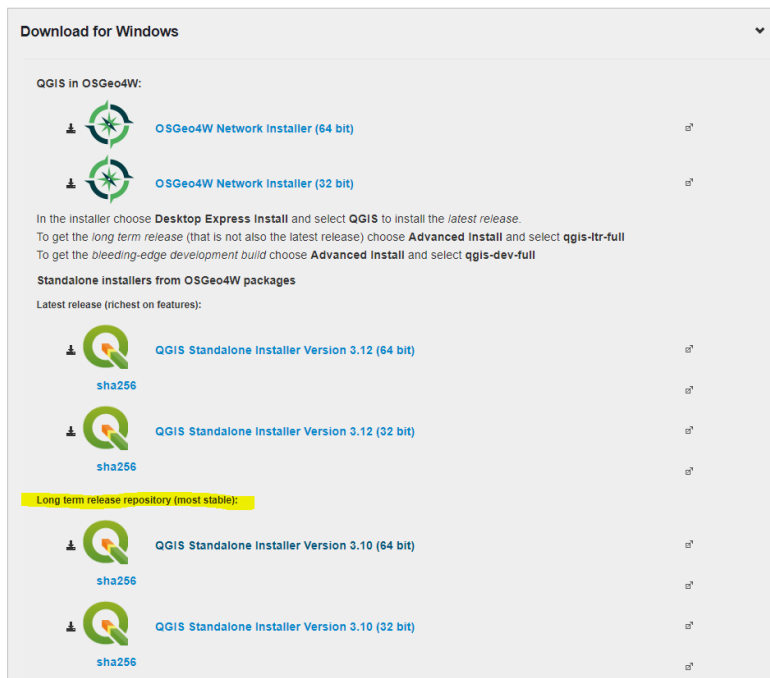


The MaxEnt program

MaxEnt requires Java to run, so make sure you have the latest version installed:

<https://www.java.com/en/download/win10.jsp>

- 2 Next, we will download the GIS software that will enable us to easily visualize the map. ArcGIS is one software that is widely used, but for the purposes of this guide, we will be using the free open-source program called QGIS. The download can be found at this link: <https://qgis.org/en/site/forusers/download.html>. At the download screen, download the long term release repository, which is QGIS version 3.10 at the time this was written.

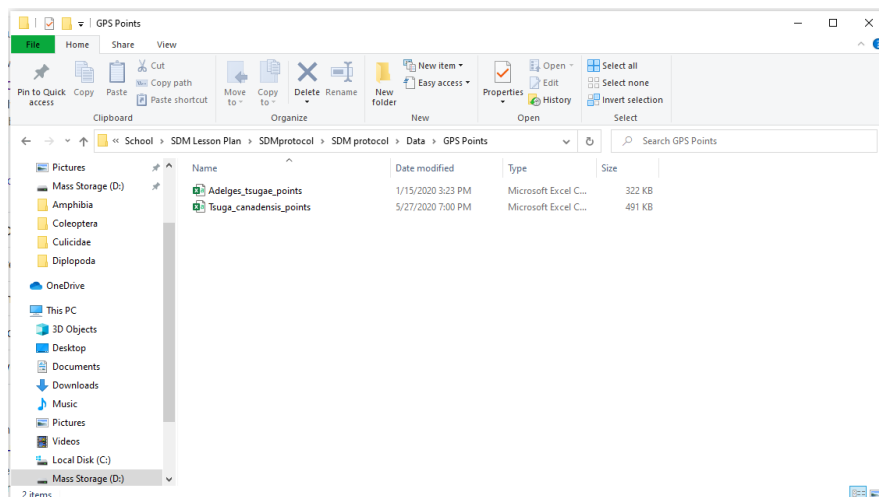


Download the "Long term release repository (most stable)"

The link you choose will depend on what operating system you have, but make sure it is the long term stable release version. Once downloaded, install the program and follow the prompts in the setup launcher while choosing the default setup options.

Understanding the Data I: Point Data

- 3 Let's review the data we will be using to complete this project. There are two types of data we will use to create a Species Distribution Model: point and raster data. Point data is data that gives the exact location of an object on a surface and has no shape or volume. We will be using GPS locations of plants, *Tsuga canadensis*, and insects, *Adelges tsugae*, for our point data. The GPS points can be located in the data folder.



The two GPS point files that we will be using

- 4 The point data is a simple table listing the name of the species and the latitude and longitude for each occurrence of the species. Open one of the point data files to view the data.

	A	B	C	D
1	species	decimalLatitude	decimalLongitude	
2	Adelges tsugae	36.643943	-81.741199	
3	Adelges tsugae	42.69432	-77.60707	
4	Adelges tsugae	42.68121	-77.60306	
5	Adelges tsugae	42.69409	-77.60716	
6	Adelges tsugae	42.69631	-77.508438	
7	Adelges tsugae	40.320422	-79.087348	
8	Adelges tsugae	40.17567	-76.803789	
9	Adelges tsugae	42.555533	-70.843942	
10	Adelges tsugae	41.913536	-72.436997	
11	Adelges tsugae	34.684776	-83.15374	

The point data for *Adelges tsugae*

This point data was obtained freely from a citizen science program called iNaturalist. iNaturalist allows everyday people to take pictures and GPS points of organisms and upload them to a central database. Experts and even advanced AI can then correctly ID the organisms. This data can then be downloaded from the website for any organism a researcher is interested in.

- 5 What information can be determined from this data?

What are some ways that point data could be collected?

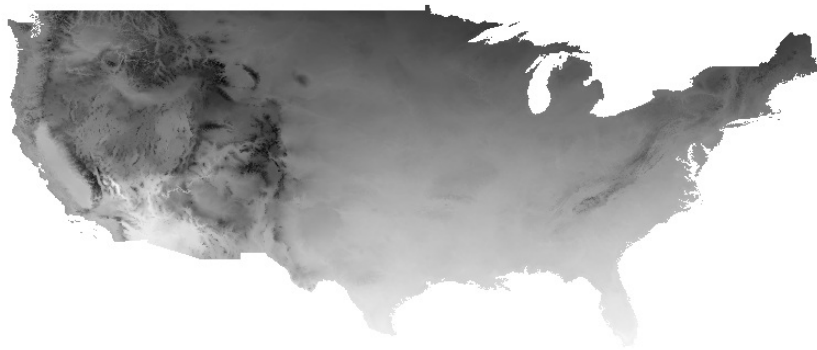
Would collecting point data for plants differ from insects?

How could citizen science help with this data collection?

What are some potential drawbacks of using citizen science to collect point data?

Understanding the Data: II Raster Data

- 6 The next type of data we will use is raster data that will be an approximation of the climatic factors at any given area of the contiguous U.S. Raster data is data organized into a grid layout with each grid being the same size but with a unique numerical value associated with it. Raster data is best visualized by looking at a computer screen. Raster grids are analogous to the pixels that make up a computer screen. Each screen is made of millions of pixels and each pixel has a specific numerical value that we see as color. When the pixels are combined together they create an image that we see on the screen. Our raster data will be a map of the Contiguous United States that has been divided into 800m x 800m grids. Numerical values are then added to each individual grid based on the environmental conditions such as average temperature, maximum precipitation, or average diurnal range. These grids come together to give us an image of the variable climatic factors across an entire landscape.

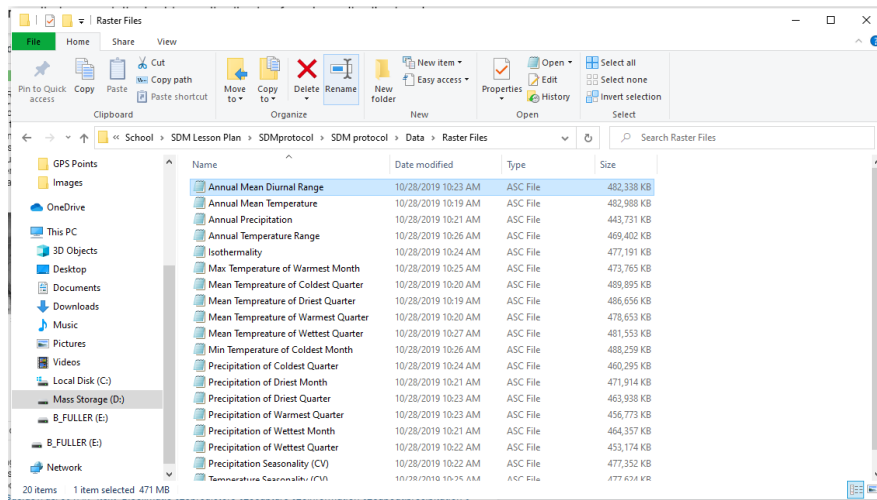


Visualization of annual mean temperature using raster data with darker colors as colder temperatures in Celcius

- 7 We will be using raster files of 20 different climatic factors created by the USGS to model the distribution of our chosen species. Together, these files allow ecologists in the U.S. to model bioclimatic conditions for various purposes. the link to Information about the creation of these files, how the data were collected can be found here, and to download the files can be found here:

<https://pubs.usgs.gov/ds/691/#:~:text=Bioclimatic%20predictors%20capture%20information%20about,precipitation%20of%20the%20wettest%20and>

You will be able to find these files in the data folder. **Do not open the files at this time.**



Raster files of the bioclimatic variables

8 What information can be determined from this data?

What are some ways that climatic data could be collected?

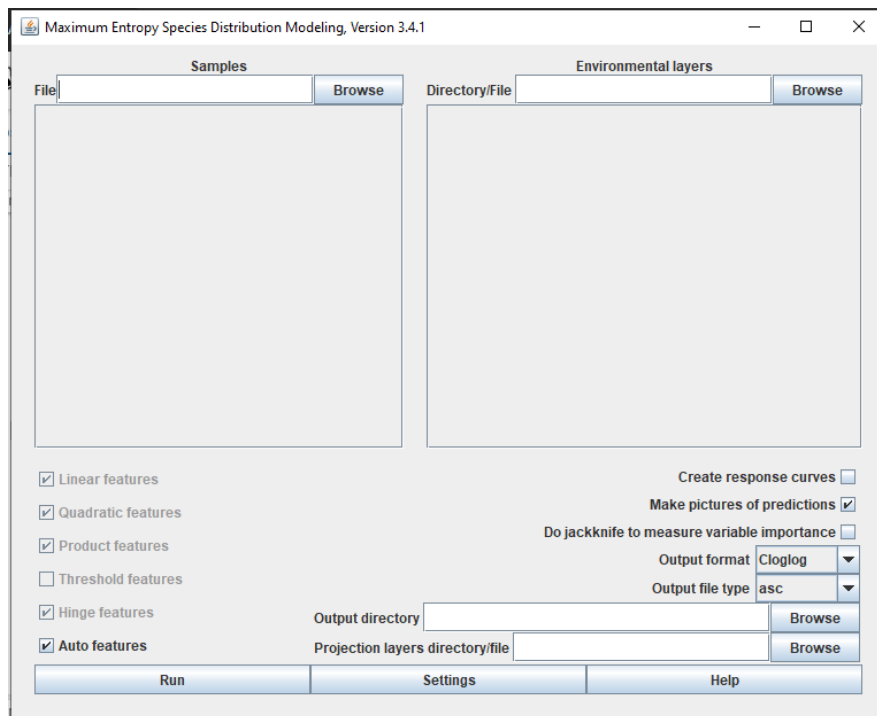
Would collect climatic data for plants differ from insects?

How could citizen science help with this data collection?

What are some potential drawbacks of using citizen science to collect climatic data?

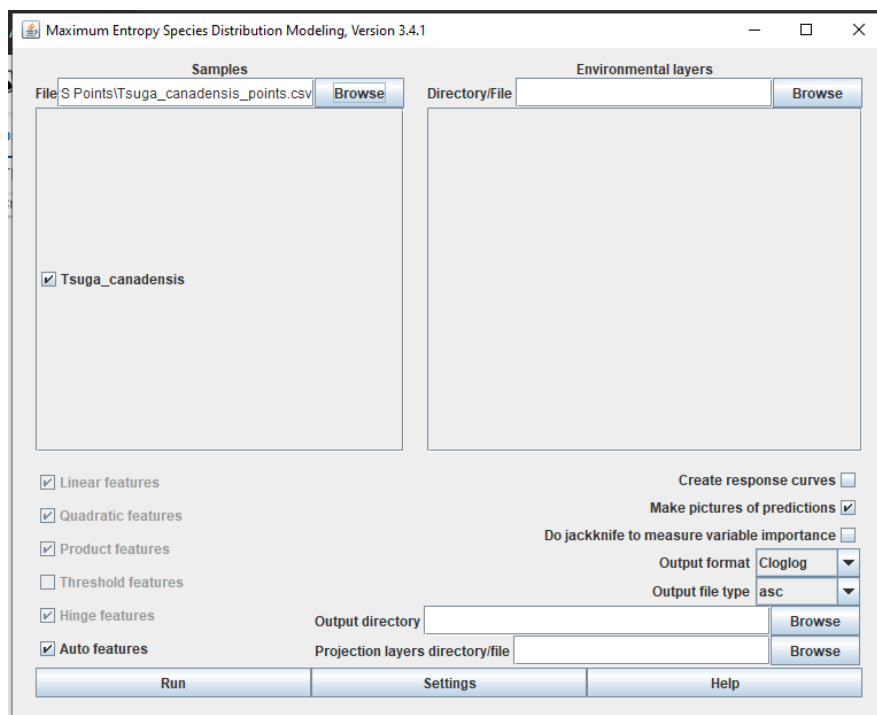
Creating Plant SDM

9 Now that we understand our data, we can move on to using the data to create SDM's. Our first step will be to open MaxEnt. You will see this:



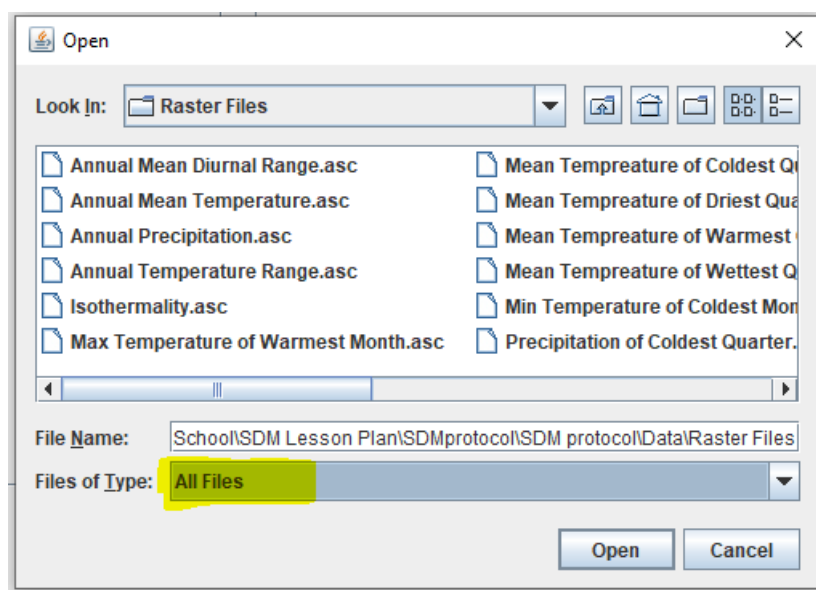
MaxEnt Home Screen

- 10 Under the "Samples" category, click the browse button. Go to the GPS points folder in the data folder and open the file called "Tsuga_canadensis_points.csv". MaxEnt should load the points and create a label with a checkbox called "Tsuga_canadensis".



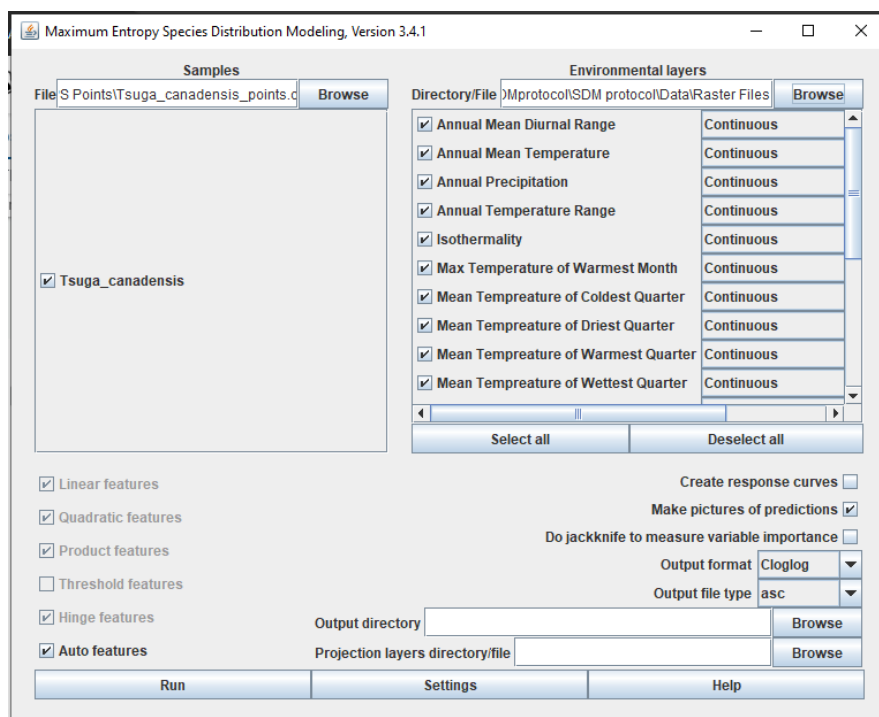
Tsuga_canadensis_points.csv loaded into MaxEnt

- 11 Next, click the browse button under the "Environmental layers" category. We will now locate the bioclimatic layers under the raster data folder. The raster folder will appear empty unless you have "All Files" selected for types of files.



Make sure "All Files" is selected

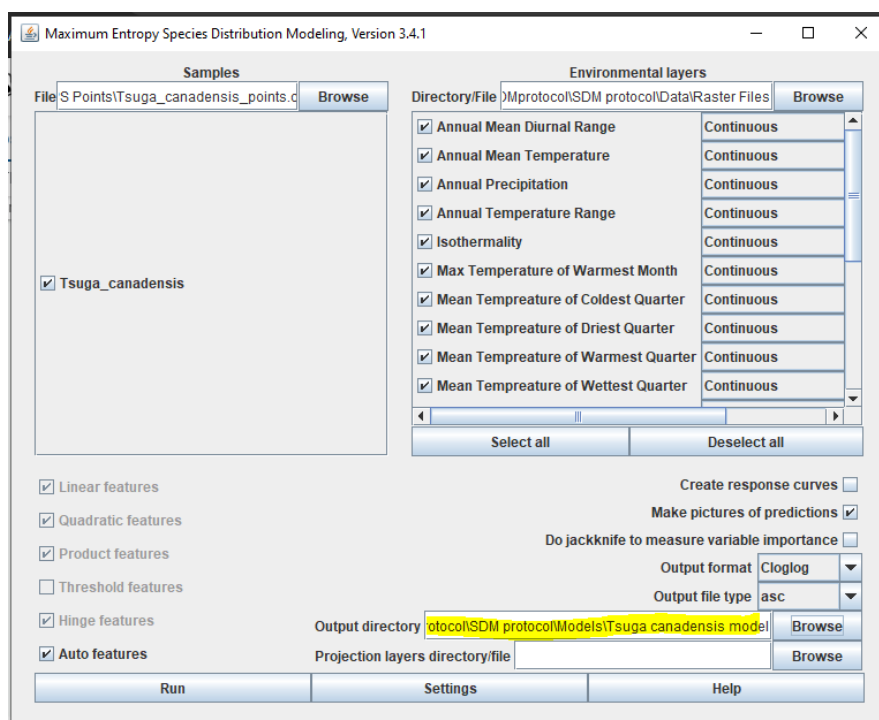
Open any of the bioclimatic files. MaxEnt should automatically import all of the bioclimatic files from the raster folder.



The MaxEnt home screen after importing both point and raster data

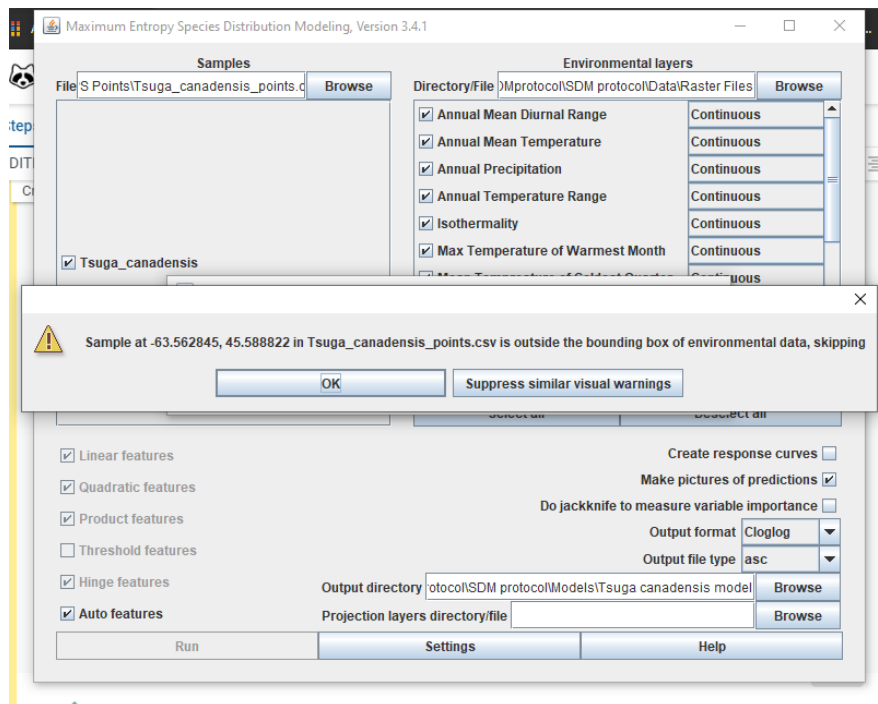
Make sure all of the bioclimatic files have "continuous" selected next to the name of the file.

- 12 Normally, we would check the options under environmental layers called "Create response curves" and "Do jackknife to measure variable importance", but these are time-consuming calculations and will skip them. I will provide the outputs to the two previously mentioned calculations later on in this guide. Leave all of the other options set to the default settings, and click browse next to the "Output directory" input. Under the "Models" folder, open the "Tsuga canadensis model" folder.



Output directory with the "Tsuga canadensis model" folder selected

- 13 Click Run. You will immediately be given a warning.



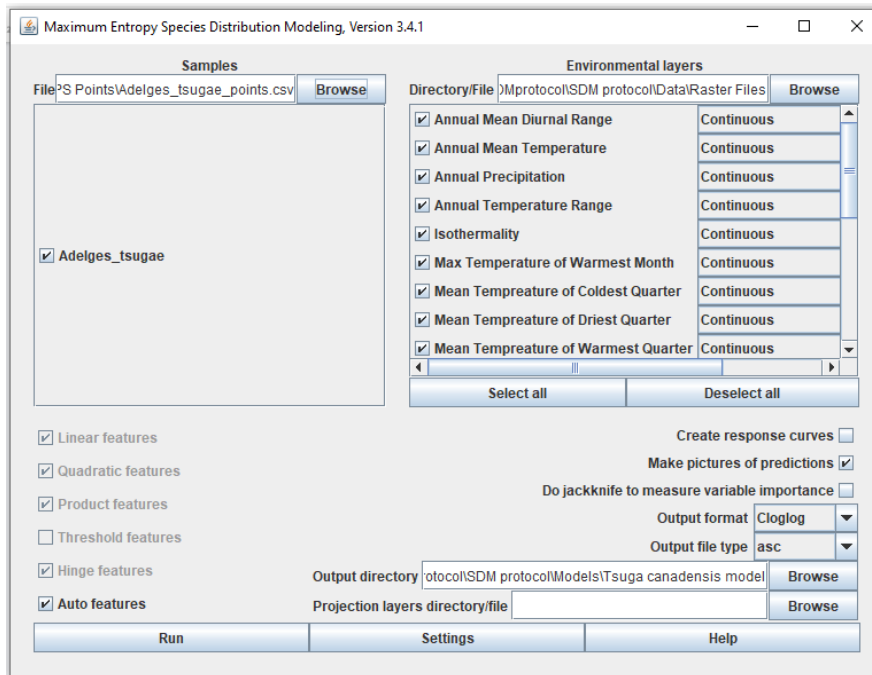
Outside bounding box warning

The GPS points we are using include those for Canada and Alaska which are outside the range of our environmental files. This is ok and will not affect our model. Click the "Suppress similar visual warnings" button now and any other time it may arise. MaxEnt will now create an SDM for *Tsuga canadensis* and store it in the "Tsuga canadensis Model" folder. This process can take 10+ mins (it's creating a model for the entire contiguous US, so 10+ mins is incredibly fast), so use this time to open QGIS and familiarize yourself with its layout.

- 14 Once MaxEnt is finished running, leave it open as we will immediately create two models for *Adelges tsugae*.

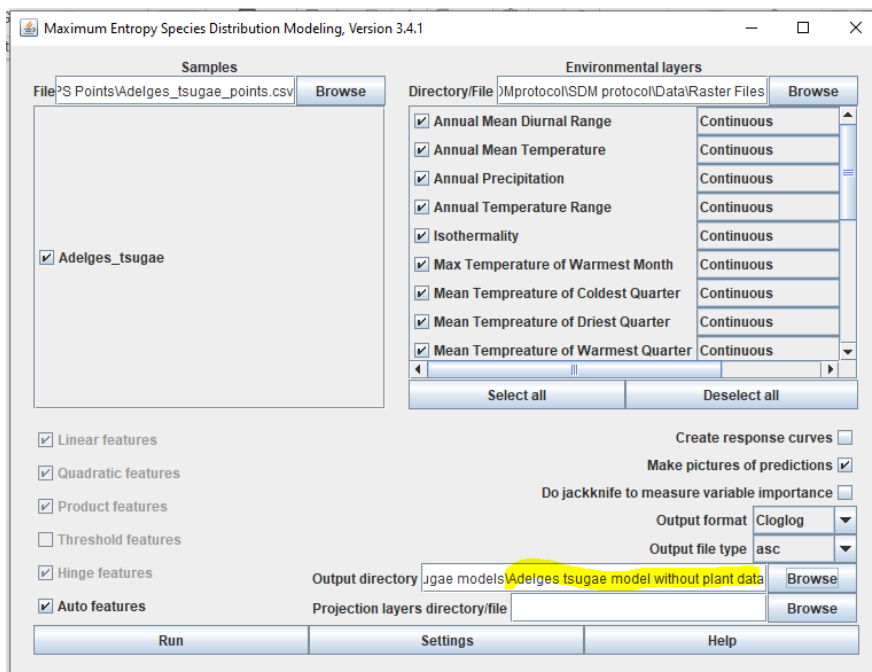
Creating Insect SDM

- 15 Creating the *Adelges tsugae* models will be fairly straightforward. We will be creating one model without and one model with the *Tsuga canadensis* output we created in the previous section. This will allow us to compare how the insect model changes when we include information about the distribution of the host plant.
- 16 Click the browse button under the "Samples" category. Find the point data for the insects named "*Adelges tsugae*_points.csv", and open this file. MaxEnt will import the point data the same as before.



MaxEnt home screen with the Adelges tsugae point data imported

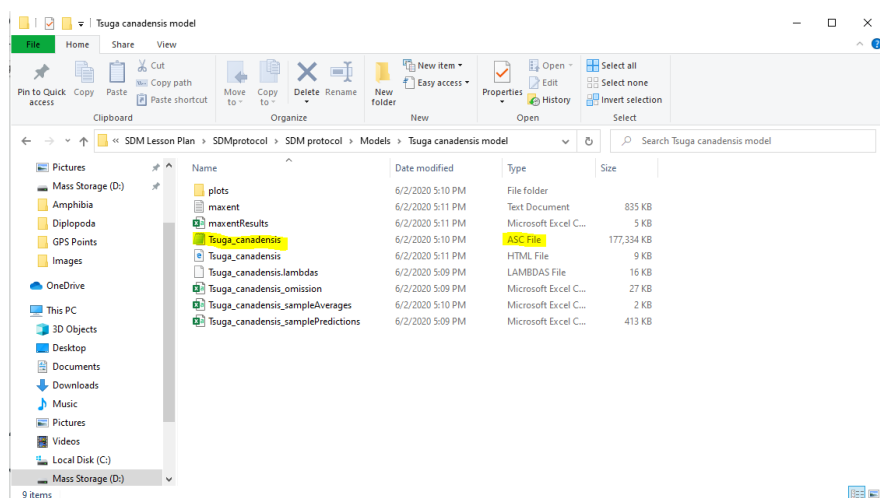
- 17 Now, we need to change the output directory. Click browse next to the output directory, find the folder called "Adelges tsugae models" and open the folder named "Adelges tsugae model without plant data".



MaxEnt home screen with the output directory changed to the "Adelges tsugae model without plant data" folder

- 18 Click the run button and suppress any warnings that pop up. This second run should take less time as MaxEnt still has temporary bioclimatic files cached from the previous run. After the run has finished, proceed to the next step.

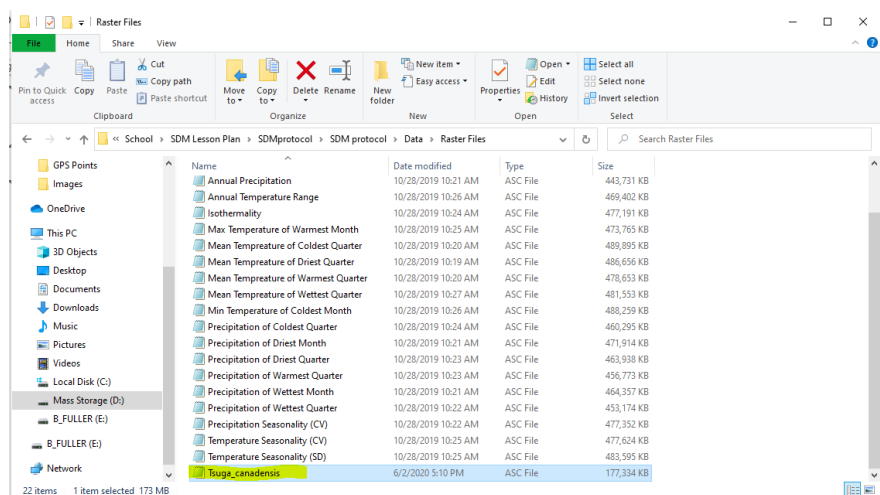
- 19 Now, we need to create a model based on the host plant data. The first step will be to transfer the raster file containing plant SDM to the raster folder containing the bioclimatic layers using the file explorer on your computer. This can be most easily accomplished by opening the "Tsuga canadensis model" folder. Finding and copying the ASC file called "Tsuga_canadensis".



The ASC file called "Tsuga_canadensis" that needs to be copied.

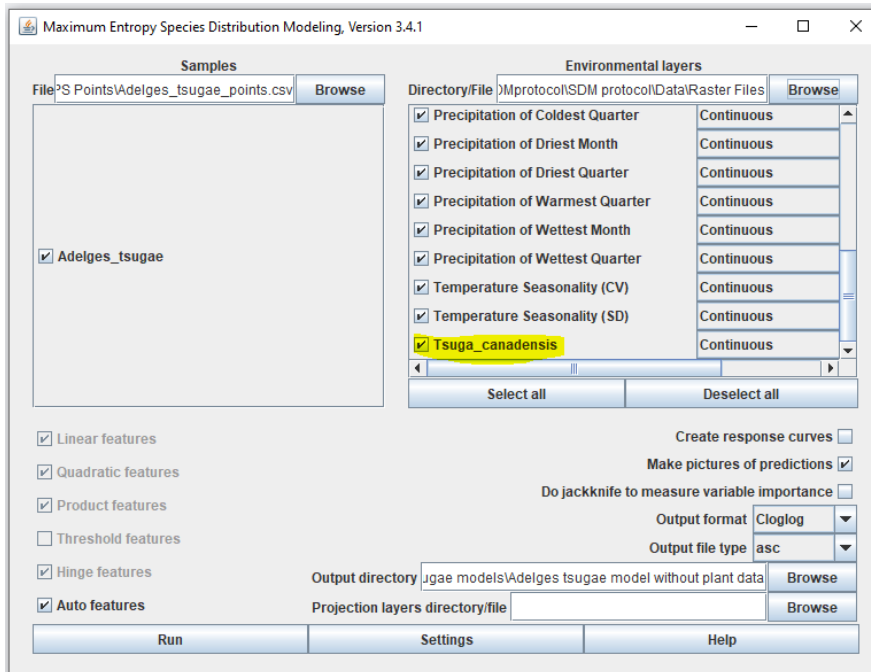
The ASC file called "Tsuga_canadensis" is the raster file that contains the SDM for the host plant.

- 20 Next, we need to paste the ASC file into the raster folder containing the bioclimatic layers.



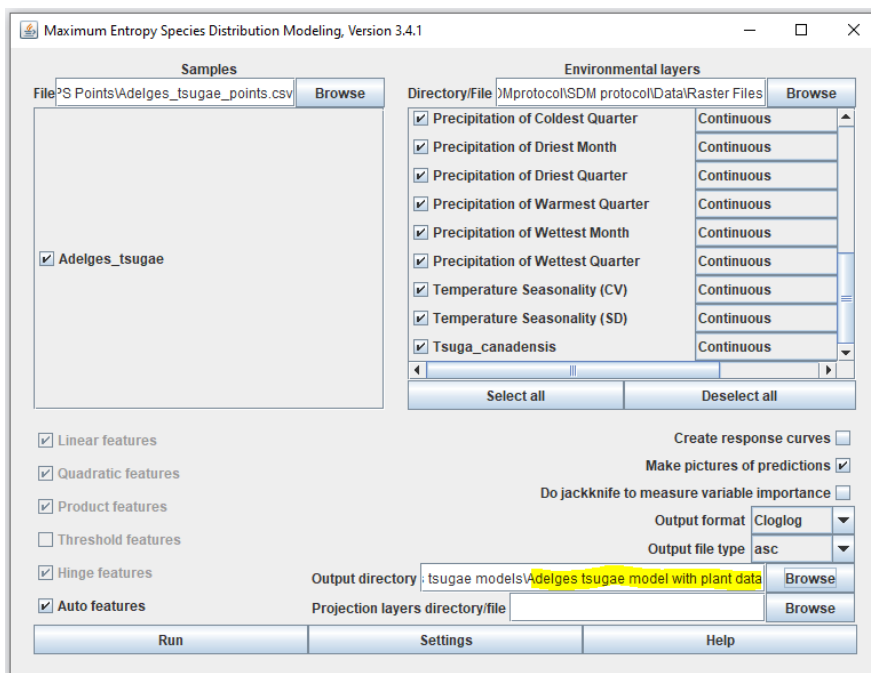
The ASC file called "Tsuga_canadensis" pasted into the raster folder

- 21 The "Environmental layers" category on MaxEnt now needs to be updated to include the plant host model, so click the browse button for this category. Go to the "Raster Files" folder, like last time, make sure "All Files" is selected next to "type of file", and open any of the bioclimatic files. The host plant model should now be located at the bottom of the "Environmental layers" category.



MaxEnt home screen with the environmental layers updated to include the plant model

- 22 Change the Output directory once again. This is the same process as before, but this time change the output directory to the folder named "Adelges tsugae model with plant data".

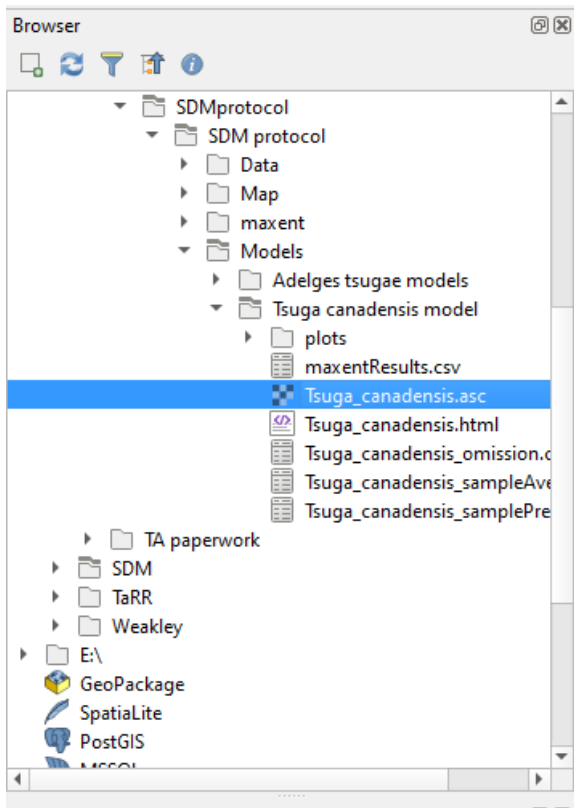


MaxEnt home screen with the output directory changed to the "Adelges tsugae model with plant data" folder

- 23 Click run. After MaxEnt has run the model move to the next section.

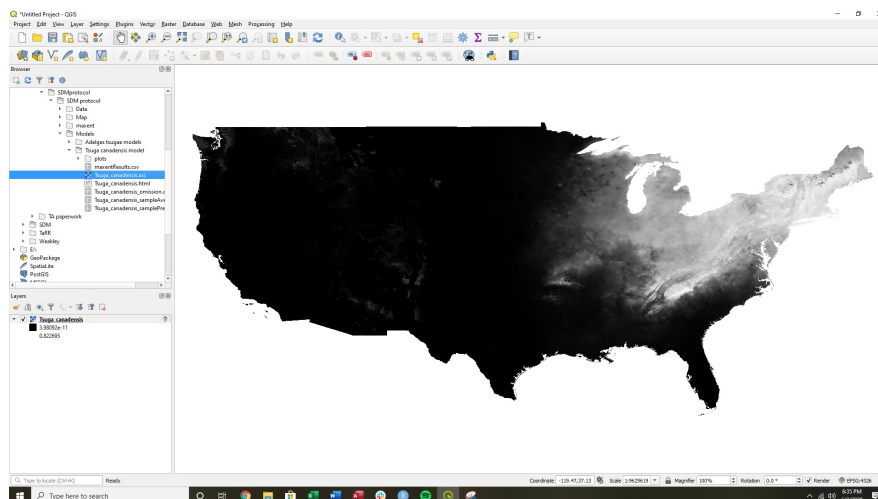
Visualizing the Models

- 24 Lets see how our models turned out. Open QGIS and start a new GIS project. On the left, you will see a folder browser. Search through the browser to find the folder you saved the plant and insect models too. Lets start with the plant model, so double click on "Tsuga_canadensis" with the checkered image next to it.



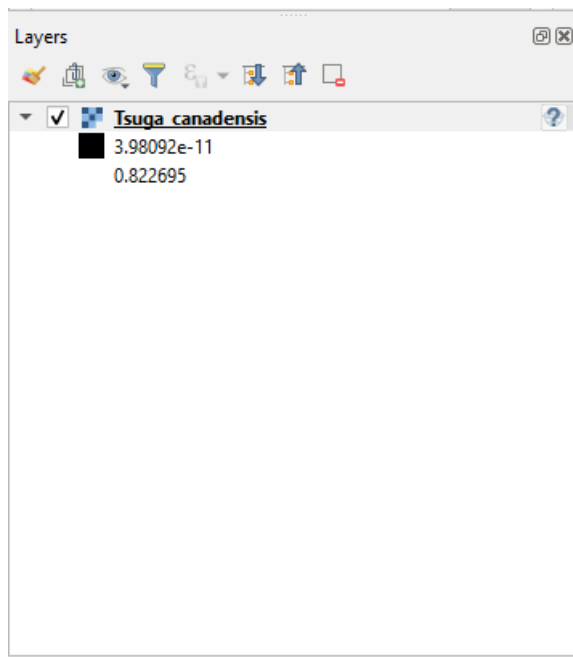
The Tsuga canadensis ASC file

This will load the raster file with the model into QGIS and will look like this.



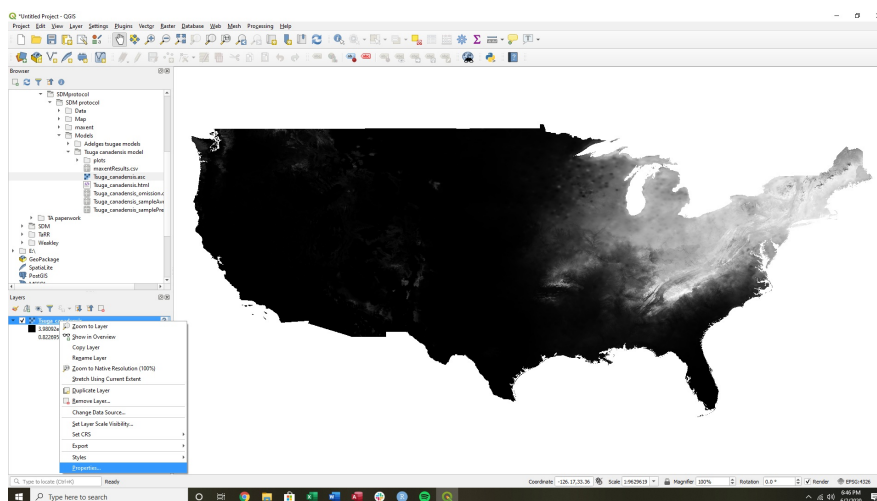
Raw Tsuga canadensis model.

The MaxEnt models are based on probabilities, so the whiter the area the higher the probability the species will be found at that location. The values on the bottom left tell us what the max and min probabilities are.



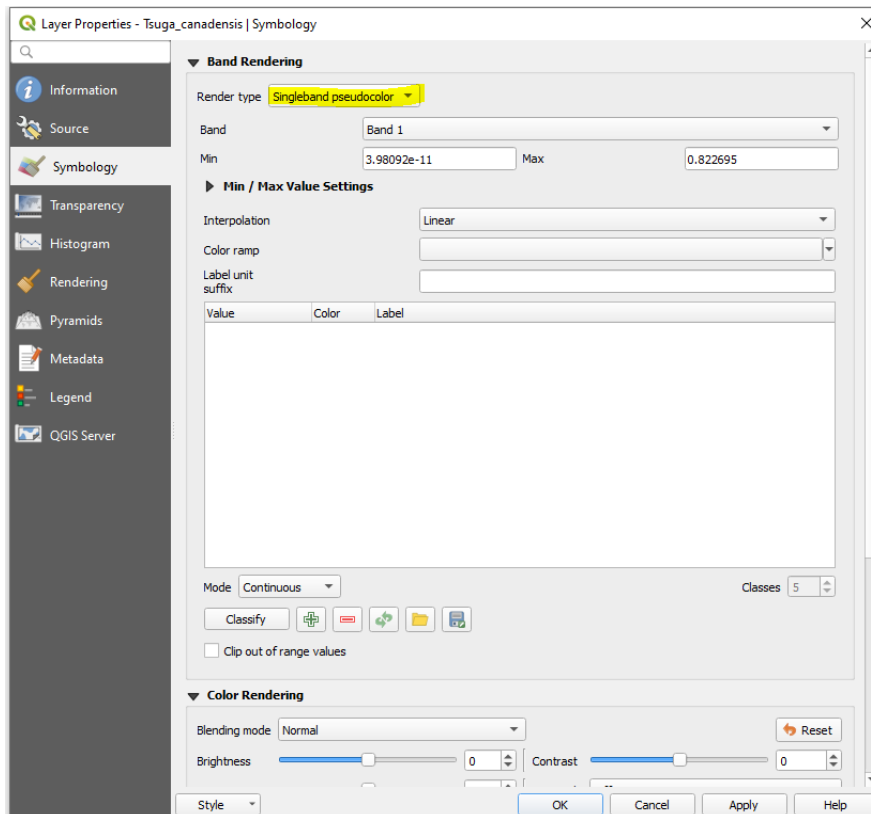
Max and Min probabilities for *Tsuga canadensis*

- 25 While the black and white map tells us what we need to know, there are better ways to visualize the data to more easily see gradients in the distribution, so let design the map to be easier to decipher. Right-click on *Tsuga canadensis* under the layers section and click on properties.



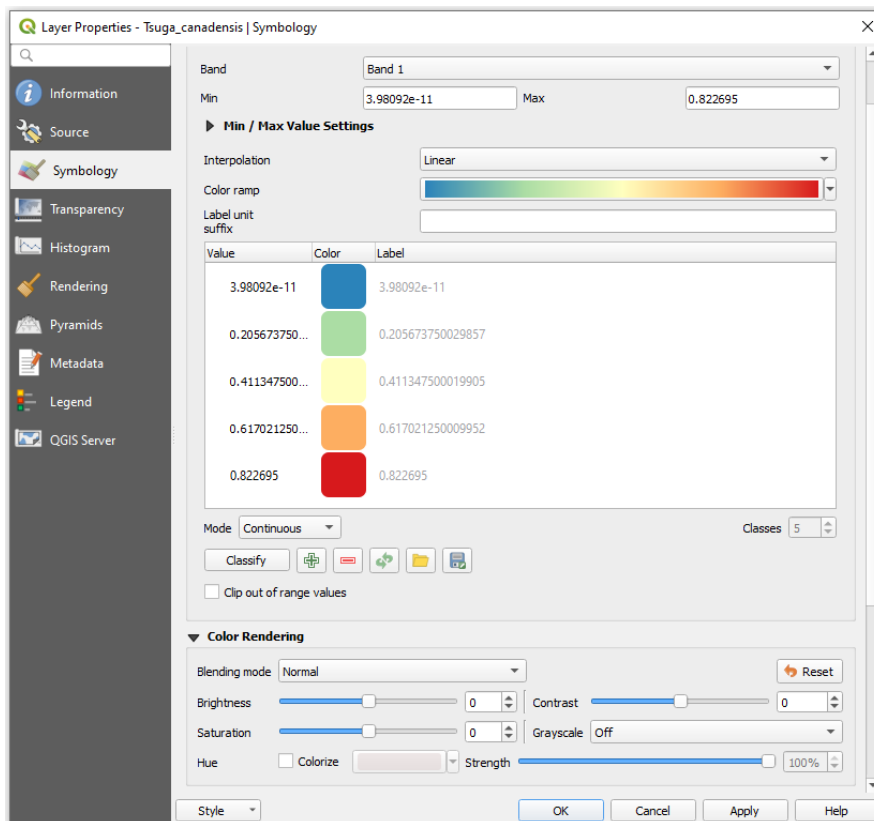
Click on properties

- 26 Open the "Symbology" tab, and change the render type to "Singleband pseudocolor".



Render type should be "Singleband pseudocolor"

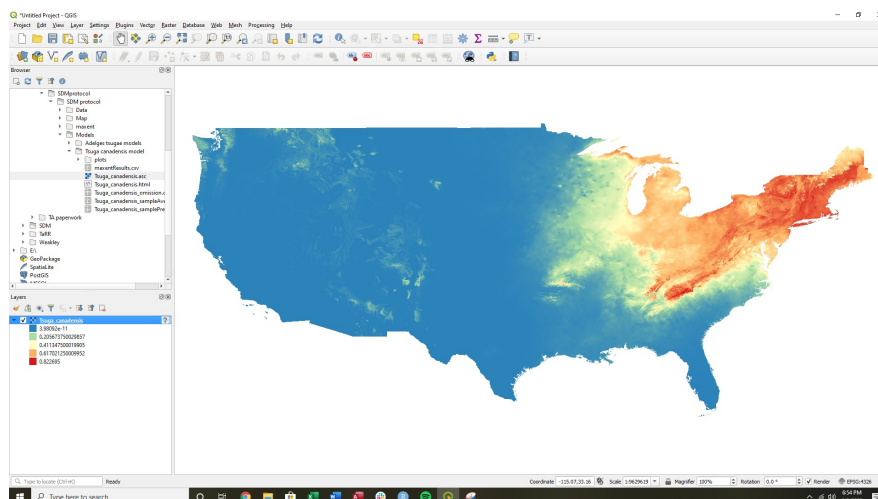
- 27 From the "Color ramp" drop-down menu, choose the "spectral" color palette. Open the same drop-down menu again, and this time click "Invert Color Ramp". Your settings should look like this now.



Symbology menu with correct color settings

28

Click ok to save the settings and close the properties menu. Your map should look like this now.



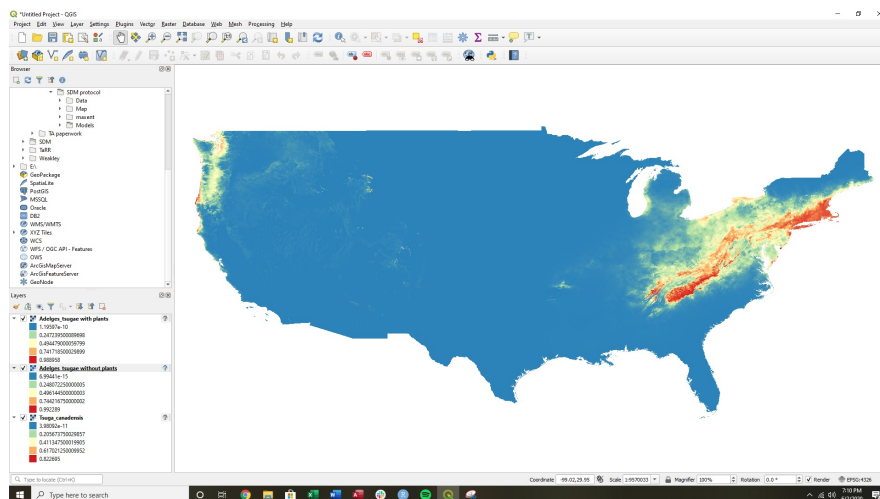
The correct color scheme for the *Tsuga canadensis* map

The areas that are hotspots for the species distribution are now clearly visible.

29

Repeat steps 24-28 for both *Adelges tsugae* models (with and without plants). Make sure to rename the insect models in the layer section to "with plants" or "without plants" depending on the model as they will both have the same name.

30 Once all three maps are loaded and transformed, then your QGIS project should look like this.



Your QGIS project should look like this

You've created three species distribution models, congratulations! To switch between the three maps, you can either deselect the ones you don't want to see or move the one you want to see to the top of the "layer" section list. I prefer the former option.

31 Is there a major difference between the insect model with and without plants? why or why not?

Adelges tsugae is an invasive species whose range is expanding. From the maps you have created, what areas of the lower 48 do you predict the insect could expand to? Hint: compare the insect models to the plant model.

What aspects of the citizen science data could be influencing how the models are created and how are they influencing it?

If you were a research scientist in an ecological field, what questions would you answer with and how would you use these species distribution models?