

## DIVA-GIS --- Exercise 2

### Modeling the range of wild peanuts

Before starting this exercise, you should have gone through the tutorial and at least glanced through the manual. It is also a good idea to first do Exercise 1.

This exercise focuses on the use of 'ecological niche modeling' to predict the distribution of species. The wild relatives of *Arachis hypogaea*, the cultivated peanut, are used as an example.

#### 1) Introduction

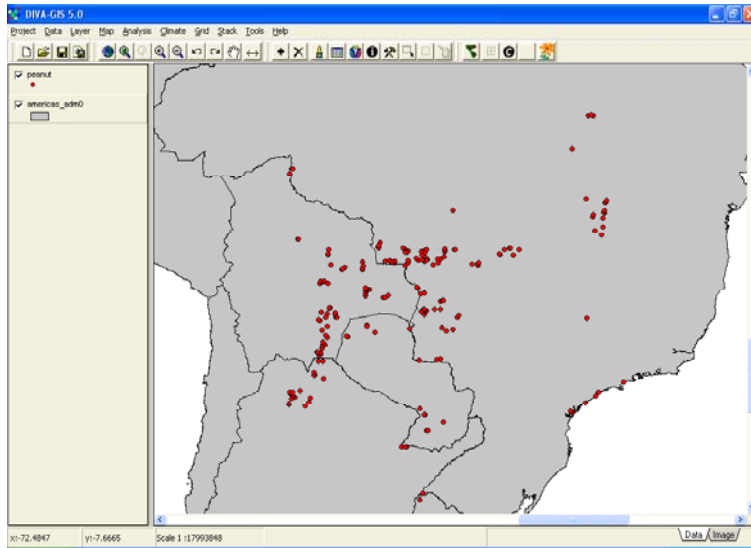
The centre of origin of the cultivated peanut is thought to lie in the area of northern Argentina and southern Bolivia (Stalker and Simpson, 1995), but various questions remain unanswered regarding its evolutionary history and domestication. The closest known relatives of the peanut are 26 wild species of the genus *Arachis*, section *Arachis* that occur in south-central South America (Bolivia, Argentina, Brazil, Paraguay and Uruguay). Some of the wild peanut species have been useful in breeding improved peanut varieties (e.g., Simpson and Starr, 2001). Further reading, and basis of this exercise:

Jarvis, A., L. Guarino, D. Williams, K. Williams, I. Vargas, and G. Hyman, 2002. The use of GIS in the Spatial Analysis of Wild Peanut Distributions and the Implications for Plant Genetic Resources Conservation. *Plant Genetic Resources Newsletter* 131:29-35.

Jarvis, A., M.E. Ferguson, D.E. Williams, L. Guarino, P.G. Jones, H. T. Stalker, J, F. M. Valls, R, N. Pittman, C, E. Simpson and P. Bramel, 2003. *Biogeography of Wild Arachis: Assessing Conservation Status and Setting Future Priorities*. *Crop Science* 43: 1100-1108.

#### 2) Data preparation

- a. Make a new folder, and unzip the data for this exercise into that folder.
- b. Make a shapefile called "peanuts.shp" of the points in "arachis\_accessions.dbf" and save it in the new folder to the new directory. This file has 397 localities representing collecting sites. Data obtained from SINGER (<http://singer.cgiar.org>) and GRIN (<http://www.ars-grin.gov/npgs>) and herbarium specimens from herbaria in Argentina and Paraguay (Krapovickas and Gregory, 1994).
- c. Add the shapefile to the map; also add the Americas\_adm0 file (download from the Data page on the DIVA-GIS.ORG website).



- d. If you do not have them, now download the global climate files at 10 minute resolution for current conditions. Check if you have climate data files by clicking on *Climate/Point* and the clicking on any part of the map (where there is land). If you cannot get windows like the two shown below (and get the error message "no climate databases found"), you must probably download the climate data from <http://www.diva-gis.org> OR you need to change the folder where DIVA-GIS looks for the climate data (under *Tools/Options/Climate*) OR you are using a map that is not in geographic coordinates. The default climate data that you can download from the DIVA-GIS website are not projected and only work with data in geographic coordinates (it is possible to import your own climate data that may be projected).

Climate by point

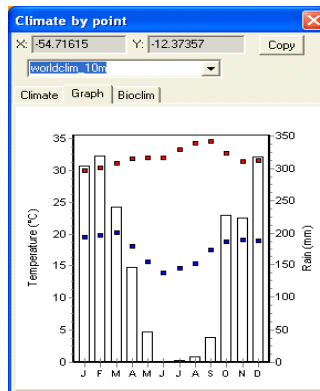
X: -54.71615 Y: -12.37357 Copy

worldclim\_10m

Climate | Graph | Bioclim

Altitude: 357

|      | Min. Temp | Max Temp | Precipitation |
|------|-----------|----------|---------------|
| Jan  | 13.5      | 30       | 304           |
| Feb  | 19.8      | 30.3     | 320           |
| Mar  | 20.3      | 31.1     | 241           |
| Apr  | 18.1      | 31.8     | 145           |
| May  | 15.6      | 31.9     | 47            |
| Jun  | 13.8      | 32       | 1             |
| Jul  | 14.5      | 33.2     | 3             |
| Aug  | 15.2      | 34.2     | 9             |
| Sep  | 17.5      | 34.5     | 38            |
| Oct  | 18.8      | 32.7     | 228           |
| Nov  | 19.1      | 31.4     | 223           |
| Dec  | 19        | 31.5     | 318           |
| Year | 17.6      | 32.1     | 1877          |



Climate by point

X: -54.71615 Y: -12.37357 Copy

worldclim\_10m

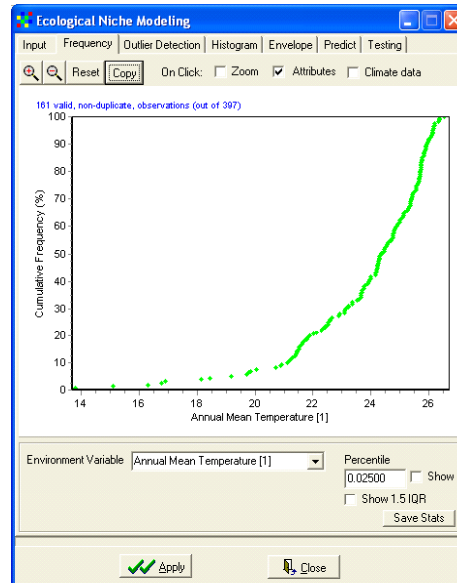
Climate | Graph | Bioclim

|  |      |
|--|------|
| Annual Mean Temperature [1]              | 24.8 |
| Mean Monthly Temperature Range [2]       | 14.5 |
| Isothermality [2/7] (* 100) [3]          | 69.8 |
| Temperature Seasonality (STD * 100) [4]  | 91.8 |
| Max Temperature of Warmest Month [5]     | 34.5 |
| Min Temperature of Coldest Month [6]     | 13.8 |
| Temperature Annual Range (5-6) [7]       | 20.7 |
| Mean Temperature of Wettest Quarter [8]  | 25.0 |
| Mean Temperature of Driest Quarter [9]   | 23.8 |
| Mean Temperature of Warmest Quarter [10] | 25.7 |
| Mean Temperature of Coldest Quarter [11] | 23.5 |
| Annual Precipitation [12]                | 1877 |
| Precipitation of Wettest Month [13]      | 320  |
| Precipitation of Driest Month [14]       | 1    |
| Precipitation Seasonality (CV) [15]      | 83.3 |
| Precipitation of Wettest Quarter [16]    | 942  |
| Precipitation of Driest Quarter [17]     | 13   |
| Precipitation of Warmest Quarter [18]    | 489  |
| Precipitation of Coldest Quarter [19]    | 51   |

## Distribution Modeling

### 3) Frequency Distributions.

- Now we are going to explore the climate data associated with the peanut accessions that we have imported. Make the “peanuts” layer the active layer.
- Use *Climate/Ecological Niche Models*
- Go to the second tab (*Frequency*) and press *Apply*

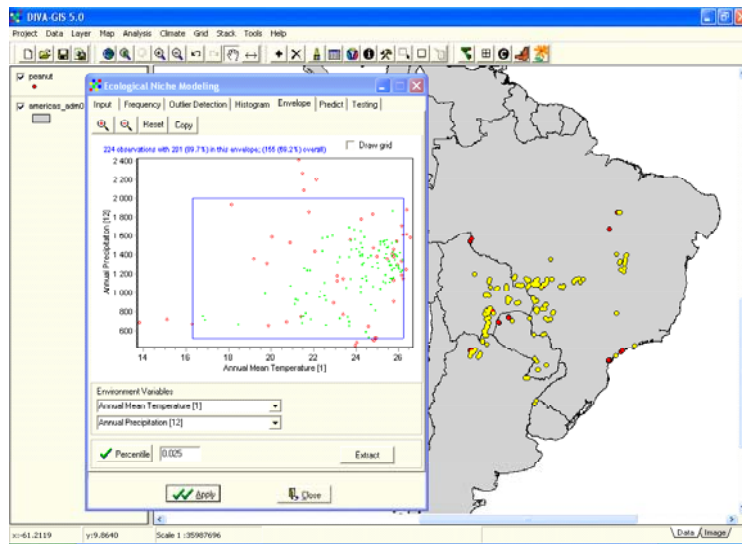


- Note that there are 161 non-duplicate observations. Points that are on exactly the same location are included only once. Points that are at a different location but in the same cell of the climate grid are also included once, in this case (this is an option that can be changed on the first tab; after which there are 224 unique observations).
- Explore this tab by changing the climate variables. This is helpful to get an idea about the general characteristics of the distribution of these points in ecological space (as opposed to geographical space). It may also reveal ecological outliers, perhaps caused by incorrect coordinates. Check the *Show percentile* and *Show IQR* (inter quartile range) boxes.
- If you click on a point in the graph different things can happen, depending on the state of checkboxes above the graph. Check this out.

### 4) Environmental Envelope

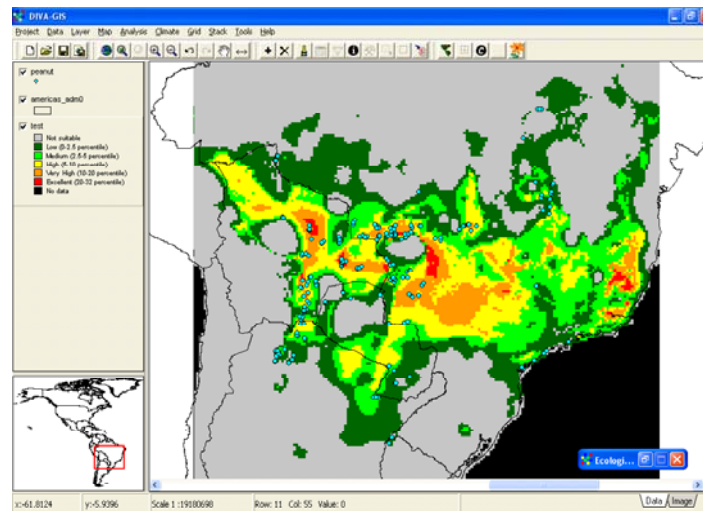
- Go to the *Envelope* tab. Here you can look at the distribution of the points in two dimensions of the ecological space.
- Define an “Envelope” using the *Percentile* button and adjacent text box. The points inside the multidimensional (of all climatic variables) envelope are colored green; the other ones are colored red. Of course there are never green points outside the two-dimensional envelope that can be seen on the graph.
  - The percentiles are calculated for each variable individually, and then combined. The percentage of the points that are inside a multidimensional envelope at a certain percentile will be different for each data set.

- The points that are inside the (two-dimensional) envelope on the graph are now colored yellow on the map. In this way we have established a link between the distributions in ecological and geographic space.



## 5) Modeled distributions

- Go to the *Predict* tab.
- First enter the area for which you want to make a prediction. Use these values: MinX = -71; MaxX = -39; MinY = -33; MaxY = -4
- Select the *Bioclim* output, choose an output filename, and press Apply. The result should be like in the image below. Note that this result suggests that there is a large area in south central Brazil where the climate appears to be suitable for peanuts but where no peanuts were collected (or even occur?).

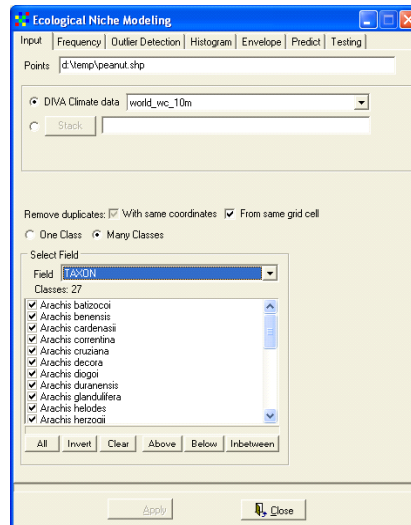


- Now make distribution models with different output types (use "Bioclim True/False" and "Bioclim Most Limiting Factor"). Note that they are all representations of the same result. The True/False surface has two possible values:

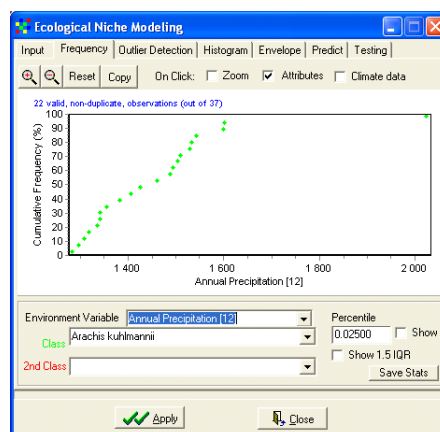
0 (False) or 1 (True), and that the result is dependent on the percentile cut-off you choose (explore this cut-off in the graph of the *Envelope* tab). The “Most Limiting Factor” outputs the variable with the lowest score in each grid cell for which there is prediction.

## 6) Single Species Models

- a. So far, we have treated the data as representatives of one single taxon, the genus *Arachis*. Now let’s go to the species level. Go to the *Options* tab and select “Many Classes” and select the “Taxon” field. Go over the list of taxa to assure that there are no errors in the list of taxa. A typical problem is that the same species occurs more than once because of spelling mistakes.

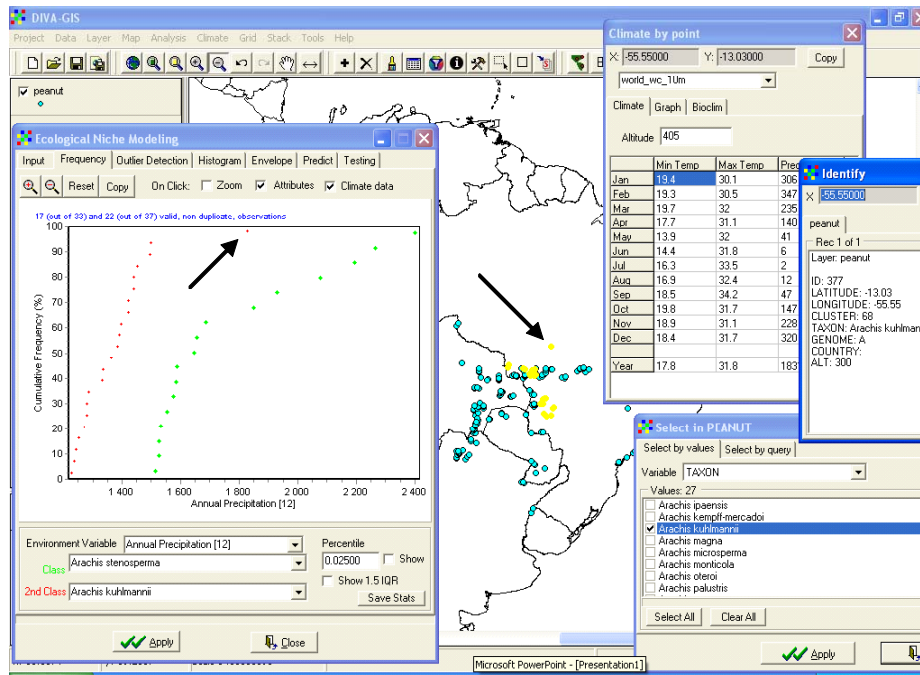


- b. Now go to the *Frequency* tab and select one or two classes (species in our case). In the example below we compare the annual rainfall for two species. *A. kuhlmannii* is clearly distributed in drier areas than *A. stenosperma*. Note that there is one huge outlier in the rainfall distribution of *A. kuhlmannii*.



- c. Click on that outlier on the graph, to find out what record it belongs to, what its associated climate data are, and where it is located on the map (use the

checkboxes above the graph). For example, first select the species, using *Layer/Select Records* so that the records for that species are colored yellow on the map. Then click on the outlier on the 'frequency' graph, and see to which point on the map it corresponds. It is clearly a geographical outlier as well. No members of the species (or any other *Arachis* species) have been observed in this area. In this case it would be prudent to check the coordinates against the locality description. However, that information is not provided here so there is not much we can do to assess the whether this record is valid or not.



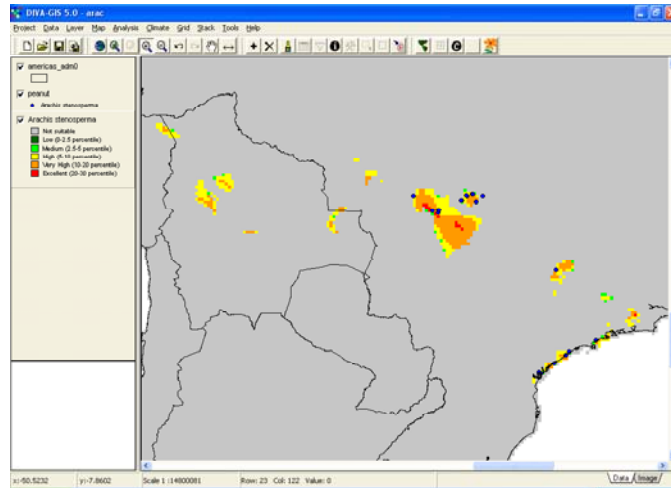
- d. Go to the *Predict* tab and make modeled predictions (Bioclim) for the same area as before (set the coordinates again as before, or make the previously modeled distribution the active layer and press the "read from layer" button to get the coordinates). In this case, however, use the *batch* option. This will create a prediction for each species. Save the output file in a new folder. Note that the output file is a "stack". A stack is a collection of 1 or more gridfiles of the same origin, extent, and resolution.
- e. Open some of the gridfiles that you have just made and compare the modeled range map with the observed points.

There are a number of ways to look at the points of one species at a time. One option is to select the records for the species in question. The default color for the selected points is yellow, which, in this case is also a color on the grid. You can change the color of the selected records under *Tools/Options*. You can also save the selected records to a new shapefile. Try this.

However, the easiest way to show a single species is by using *Layer/Filter*. Make the peanuts layer the active layer; Select the field "Taxon" and the value "*A. stenosperma*".

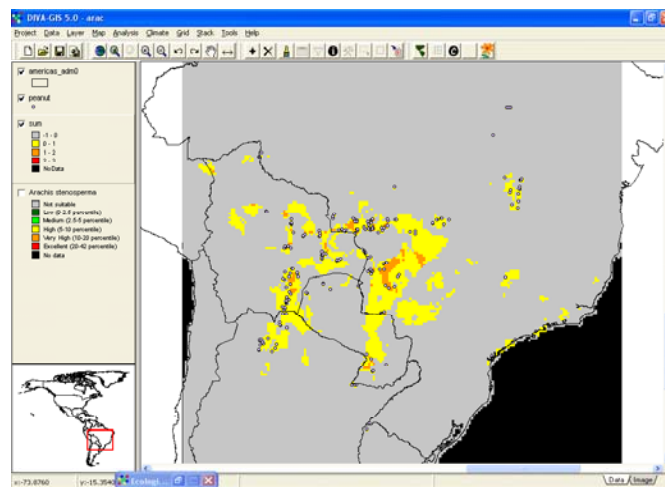
*A. stenosperma* has a disjunct distribution (a number of populations along the coast

and another group more inland). There is one record right in the middle of the two distributions. This record again warrants verifying. If it is correct it would make an interesting specimen for study. A disjunct distribution can be difficult to model. It may very well be that the groups are in fact different ecotypes with a quite different climatic adaptation. In this case, the predicted range has all the inland localities on its margins, and shows a large area of apparent suitability from where there are no records of this or another *Arachis* species. Also note some small isolated areas far away from the main area of the observed records and predicted range. These areas are probably irrelevant because they are very small, and far away from any locality where the species has been observed.

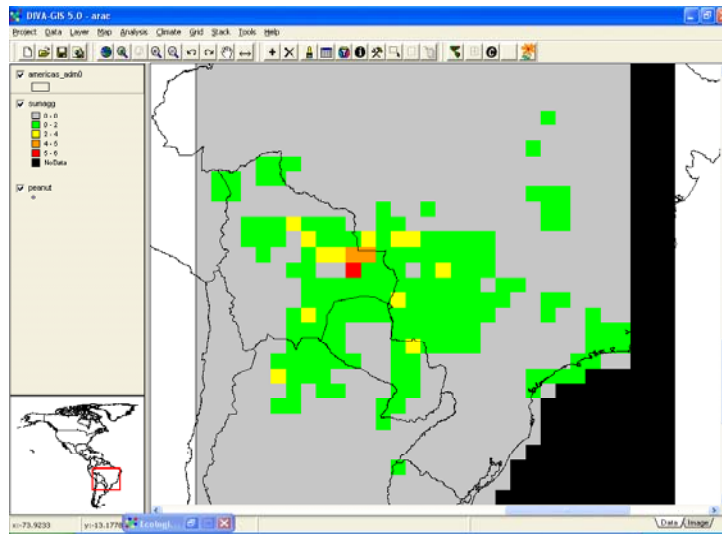
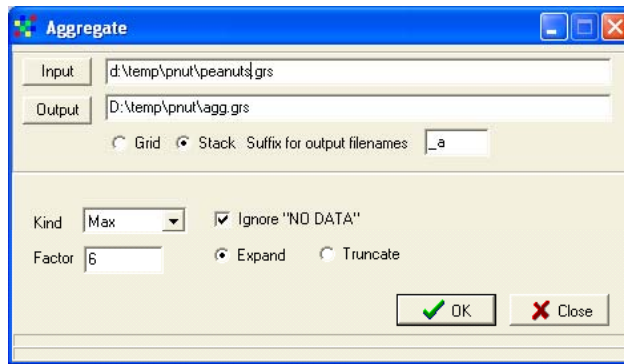


## 7) Modeled versus Observed Diversity

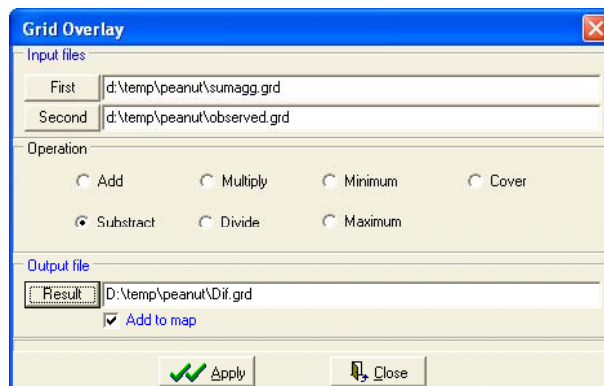
- a. Add all the gridfiles (*Stack/Calculate /Sum*). Use the “*Sum as Present/Absent*” and “*> 0*” option. Only very few areas are predicted to have more than one species. Are all areas included that were on the prediction for the genus as a whole? With other model settings, the results can change quite a bit. For example, repeat the models and summing only using bioclimatic variables 1, 5, 6, 12, 15.



- b. Aggregate (*Grid / Aggregate*) the stack to a 1 degree resolution (6x) using the “*MAX*” option (why?) and sum the result again.



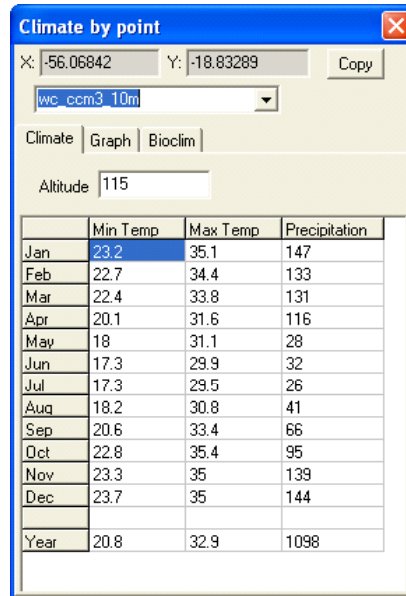
- c. Make a map with observed species richness maps at the same resolution as the previous map. Use *Analysis/Point to Grid/Richness* (use *Define Grid*, and Use *parameters from another grid* to assure having exactly the same grids).
- d. Compare modeled and observed richness before and after aggregation. First use *Analysis/Regression*. Then use *Grid / Overlay*.



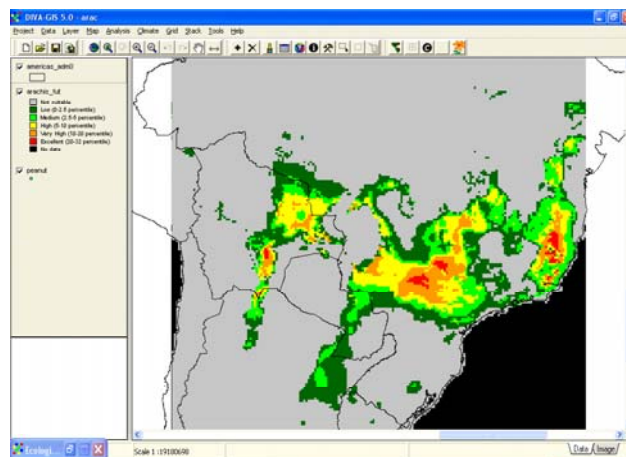
## 8) Climate Change



- a. Now consider the potential effect of climate change on the distributions of wild *Arachis*. First check if you have the predicted climate data. (the wc\_ccm3\_10m data) on the *Climate/Point* window.



- b. Now let us consider the potential effect of climate change on the distributions of wild *Arachis*. First check if you have the predicted climate data for around 2050 (2x CO<sub>2</sub>). Run Bioclim for the whole *Arachis* genus, as in section 4, but now use the projected future climate data. Note that the input climate data (input tab) should be for current conditions (climatic adaptation is inferred from these data), whereas the future conditions should be selected in the predict tab.



If you compare the result with the map obtained in section 4, or simply with the observed localities, you will note that the areas that are predicted suitable in the future are almost completely shifted away from where the *Arachis* species are now.

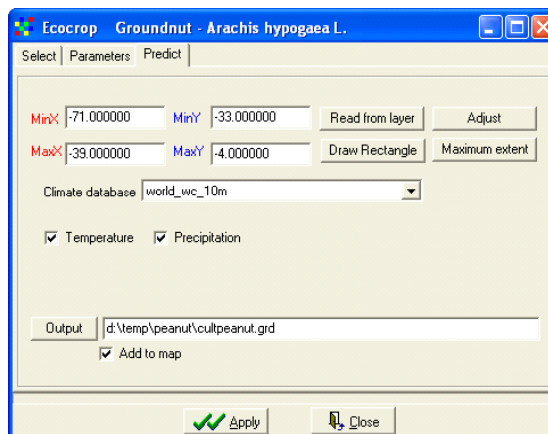
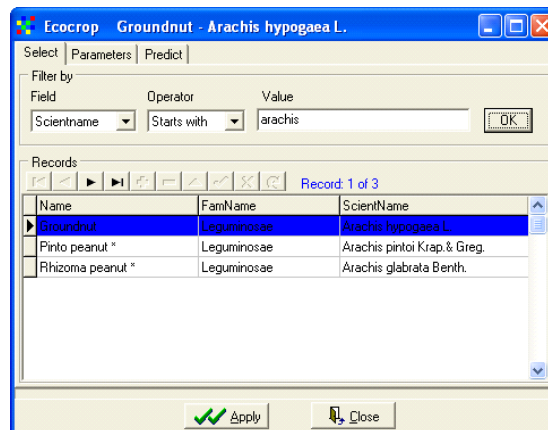
One must be very careful not to jump to conclusions from this result. There are a number of reasons for this. Particularly important is that we do not know the real

climatic adaptation of the genus. All we know is where it currently occurs, and that is a result of past climate change, current climate, dispersal limitation, and interactions with other species including humans. Whereas it is quite reasonable to assume a steady state primarily determined by the climate for the current distribution, it is highly speculative to extend that to the climatic conditions predicted for the future.

Nevertheless, this type of predictive modeling may be of some use, for example, to identify the areas or species that are most likely to be affected. It is probably better, when making predictions for the future potential distribution, to select fewer climate variables, and perhaps not all “tails” of these variables.

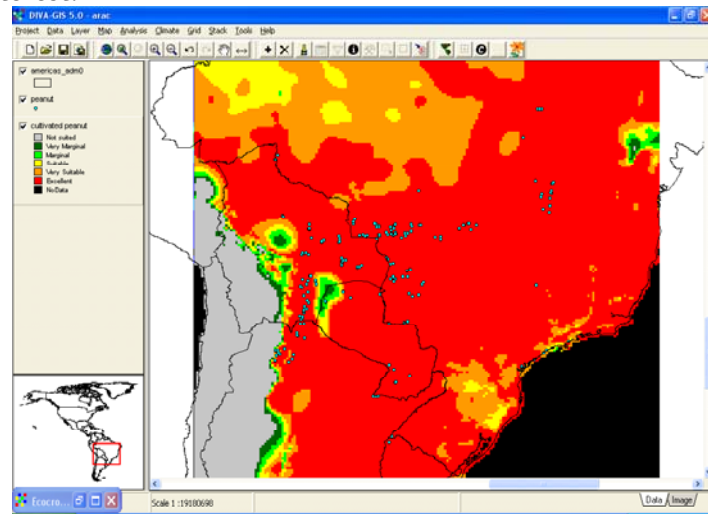
Ideally, to predict future distributions one would use models that are not simply based on current distribution but that are based on physiological knowledge about the taxon. Such an approach, albeit rudimentary, can be taken using the Ecocrop module.

- c. Go to *Climate/Ecocrop* and on the first tab search for scientific names that start with *Arachis*. None of the species we were dealing with is in the list. So let's use the cultivated peanut (groundnut) *Arachis hypogaea*. Select that record, go to the predict tab, select the area used for the Bioclim models.

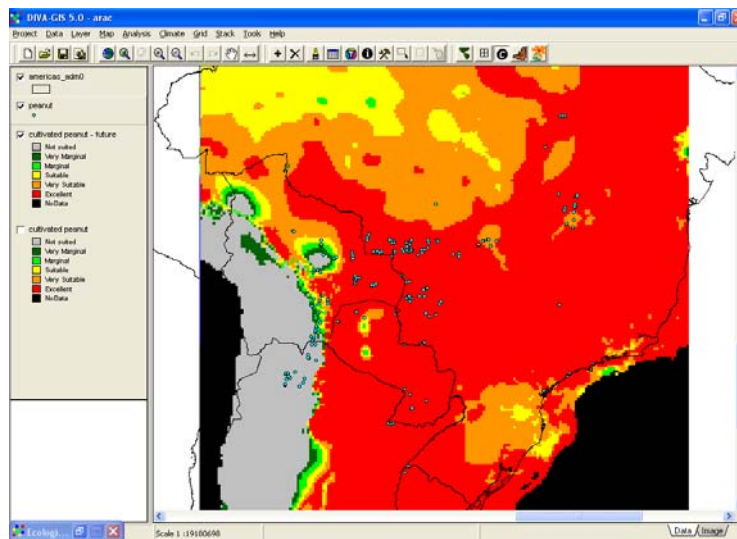


Note that the area predicted to be highly suitable for cultivated peanut includes all or the area of our wild *Arachis* records, except for three records of *A. batizocoi*. If you look at this species in the Frequency tab of the Ecological Niche Modeling module you will see that these records are outliers in the rainfall distribution, it is much drier where they are than in the other locations where the species has been observed. Again,

in the first place it would be good to check the validity of the geographic coordinates of the data, and if these appear to be correct, then these populations might be of special interest.



Now repeat but the above steps in the Ecocrop module but using the future climate data to model the climatic suitability for cultivated peanut in this area. The modeling is in both cases based on the same physiological parameters that are shown in the second tab of the Ecocrop window. Note how little overall change there is between the two results (but there are some major changes in the western end of the suitable area).



## 9) Finding the ancestors

The cultivated peanut is an allotetraploid ( $2n=4x=40$ ), while its closest wild relatives are essentially all diploids ( $2n=2x=20$ ). The species is postulated to have arisen from a fortuitous natural crossing of two diploid species, which produced a tetraploid interspecific hybrid with noticeably larger plants and fruits. Attempts to identify the peanut's progenitor species and experimentally recreate the original interspecific hybridization event using combinations of known species as parents have thus far been unsuccessful.

Cultivated peanut is notoriously susceptible to a long list of diseases and pests that can significantly reduce its yield (or lead to undesired use of pesticides). For many peanut diseases and pests, no good sources of resistance have been found within the cultivated species. Many of the related wild species screened for those same diseases show high levels of resistance, although interspecific fertility barriers and differences in genomic composition prevent the easy transference of this resistance to the crop. The progenitor species, if identified, could serve as a direct gene donor or as "bridge" species through which desirable traits in other *Arachis* species could be transferred to the cultigen using conventional breeding methods. Recent collecting missions in Bolivia have discovered new species closely related to the cultivated species, giving hope to the possibility that the progenitor species may still exist. Unfortunately, various development activities in remote, fragile environments in eastern Bolivia threaten to extirpate populations of wild peanuts, and may even destroy the progenitor species before they can be collected, described and conserved.

Efforts to clarify the origin of the cultivated peanut are focused around the genomic makeup of the crop and its progenitor species. There is a genetic separation between those diploid species in the section *Arachis* that have an "A" genome and those that have a "B" genomes (Smartt et al., 1978). The cultigen and the wild *Arachis monticola* are tetraploid and have an AB genome. Experimental crosses between *A. batizocoi* (BB) and various AA species from the section *Arachis* created synthetic AABB amphidiploids but were not successful in producing plants with the same genomic configuration of the cultigen. Other experimental crosses between the cultigen and likely diploid progenitor species, with synthetic AABB amphidiploids, and with synthetic auto-tetraploids, all showed varying degrees of incomplete chromosome pairing. These results lead scientists to conclude that either the diploid ancestors or the cultigen have diverged genetically over time, or that one or both of the true progenitor species were missing from the experiments conducted. Given the incomplete exploration of these species' natural range and the fact that new section *Arachis* species continue to be discovered as plant collectors penetrate the more inaccessible areas where they occur, it is not unreasonable to suspect that the actual progenitors of *A. hypogaea* may still exist but have yet to be collected. In light of the above, the most logical places to look for undiscovered wild peanut progenitors would be those areas where populations of both A and B genomes species are believed to occur sympatrically.

10) Predict the areas where the cultivated peanut may have arisen, based on the likely co-occurrence of the two genome types.