Exercise 3: Obtaining and Formatting Environmental Variables

Adam B. Smith, Danielle Christianson, & Camilo Sanín

July 18, 2017

In this exercise you will learn **where to obtain climate data**.

There are *many* kind of environmental data that can be used for modeling, many different places from which to obtain it, and many different kinds of data formats. Also, the files tend to be very large. As a result, we won't be able to cover every aspect of obtaining environmental data that may be relevant to your research--showing you how to obtain and process one particular data set won't show you how to do the same for another data set. Hence, this part of the workshop is somewhat "pre-packaged" in the sense that we've already downloaded some of the data and pre-processed it for you.

**Take home**: Obtaining and preparing predictor variables is often a highly manual process that can't be easily generalized. There are many places to obtain this data and many formats in which it can be obtained. In the end, though, you'll want to get it all into the same format with the same resolution and coordinate system.

# Predictors for distribution/niche models

Many factors can influence a species' range and shape its niche:

* Climate
* Soils and geology
* Disturbances like fire and floods
* Other species and diseases/pathogens
* Anthropogenic land cover that can be remotely sensed like agriculture, development, forestry, and infrastructure
* Anthropogenic land use activities that are difficult to sense remotely like recreation, hunting/harvest, and selective logging

We've listed these in rough order of the availability of this kind of data (climate data is easy to find, others less so).

If you're dealing with aquatic systems, some other factors may also be relevant:

* Dissolved minerals and gases (e.g., oxygen), pH, flow rate, depth, sunlight penetration, habitat type (lotic, lentic, estuary, marine)
* In lotic systems flood frequency and intensity, plus dams and impoundments
* In lentic systems spring/fall turnover and depth
* And in marine systems salinity, storms, tides, pressure, currents, and upwelling

Finally, there are many factors "intrinsic" to a species that interact with the factors listed above to determine range and niche properties:

* Dispersal ability
* Natural selection
* Demographic processes
* Competitors, mutualists, predators, prey, parasites and diseases

In general, distribution/niche modeling ignores these latter aspects and therefore assumes that species are in equilibrium with their environment. Newer techniques enable incorporating these processes into the modeling framework, but as of yet there is no overarching set methodology for doing this, and so models tend to be species-specific.

For the sake of expediency this workshop we will focus only on climate-based predictors. However, it is *very* worth pursuing other kinds of predictors if you feel they are relevant to modeling your species.

# Climate variables

## The BIOCLIM system

We often speak of "temperature" and "precipitation" as if there were only one way to measure each, when in fact there are probably hundreds of ways to measure each one of these aspects in a manner relevant to distribution/niche modeling. There are several semi-standardized systems for calculating versions of these variables. One of the most popular is the "BIOCLIM" system of variables, which was (confusingly) named after a model algorithm of the same name.

These variables are:

BIO1 = Annual Mean Temperature  
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))  
BIO3 = Isothermality (BIO2/BIO7) (\* 100)  
BIO4 = Temperature Seasonality (standard deviation \*100)  
BIO5 = Max Temperature of Warmest Month  
BIO6 = Min Temperature of Coldest Month  
BIO7 = Temperature Annual Range (BIO5-BIO6)  
BIO8 = Mean Temperature of Wettest Quarter  
BIO9 = Mean Temperature of Driest Quarter  
BIO10 = Mean Temperature of Warmest Quarter  
BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation  
BIO13 = Precipitation of Wettest Month  
BIO14 = Precipitation of Driest Month  
BIO15 = Precipitation Seasonality (Coefficient of Variation)  
BIO16 = Precipitation of Wettest Quarter  
BIO17 = Precipitation of Driest Quarter  
BIO18 = Precipitation of Warmest Quarter  
BIO19 = Precipitation of Coldest Quarter

This list was taken from the [WORLDCLIM](http://www.worldclim.org/bioclim) website and is easy to find if you just search for "BIOCLIM" and "WORLDCLIM"--it's nearly always the first link shown in the results. The first 11 variables pertain to temperature and the next 8 to precipitation. In the BIOCLIM system "quarters" can span any three contiguous months of the year (e.g., December-January-February).

The BIOCLIM variables attempt to encapsulate the major climatic forces that might affect a species' range. For example, many temperate plants are only able to withstand cold temperatures up to a certain point, so BIO06 (minimum temperature of the coldest month) is an appropriate variable to use for modeling their distributions. Likewise, during the winter many temperate animals depend on food stored in the summer (either as fat or external reserves), so summer primary productivity is likely relevant to their ability to accumulate food stores and body fat. In this case summer rainfall (BIO18 precipitation of the warmest quarter) is probably relevant, though BIO10 may be as well.

## Obtaining climate data

As mentioned, there are *many* sources of climate data, though a few are used commonly in niche/distribution modeling. Please see <http://www.earthskysea.org/climate-weather/> for an incomplete listing. Below we'll describe two commonly used data sets plus a newer "up-and-coming" data set.

### WORLDCLIM

**Link**: <http://www.worldclim.org/>  
**What**: Gridded climate reconstructions from 1-km to ~20-km resolution  
**Where**: All of the world's land area (sans Antarctica) plus any year up to 2100 (projected)  
**When**: Climate averaged across 1970-2000 (historic "measured") plus paleo (reconstructed) plus future periods (projected) **Description**: WORLDCLIM was generated by interpolating data from tens of thousands of weather stations. The interpolation algorithm considers geographic position and elevation. The results are climate rasters (=gridded data) available with a cell size from 1-km to ~18-km, depending on which version you download. WORLDCLIM has been used in thousands of studies.  
**Citation**: [Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978.](http://dx.doi.org/10.1002/joc.1276)

In this workshop we will be using the WORLDCLIM data set.

### PRISM (Parameter Regression on Independent Slopes Model)

**Link**: <http://www.prism.oregonstate.edu/>  
**What**: Weekly climate at 4-km (free) and monthly/daily at 30-arcsec (~800 m; pay for US) resolution  
**Where**: Coterminous United States and British Columbia  
**When**: 1895-current (historic) for monthly/weekly data and 1981-current for daily data  
**Description**: Like WORLDCLIM, PRISM was generated by interpolating weather station data. Also like WORLDCLIM, PRISM accounts for elevation and position in its interpolation algorithm, but also attempts to capture many other kinds of factors that can affect weather/climate (e.g., distance to coast, temperature inversions, windward/leeward side of mountains, etc.). The 1-km resolution version of PRISM is the industry "standard" climate data set for work in the coterminous United States and costs about $2500 + $25 per year of data requested. The 4-km version is free.  
**Citation**: [Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., and Pasteris, P. 2002. A knowledge-based approach to the statistical mapping of climate. Climate Research 22:99-113.](http://dx.doi.org/10.3354/cr022099)

### Climate*XX*

**Link**: <http://www.ualberta.ca/~ahamann/data.html>, but see also <http://cfcg.forestry.ubc.ca/projects/climate-data/climatebcwna/>.  
**Where**: North America, South America, Europe  
**When**: 1901-2009 (historic) plus any year up to 2100 (projected) **Description**: "ClimateXX" stands for ClimateNA (North America), ClimateWNA (Western North America), ClimateSA (South America), and ClimateEU (Europe). The ClimateXX system uses PRISM 4-km data as a starting point then interpolates using high-order polynomials. The dates covered are from 1901 to 2009, plus future years up to 2100. Some standardized periods are available for download (e.g., 1981-2010) at 1-km resolution, but savvy users can generate their own custom climate coverages using the free ClimateNA software (for North America).  
**Citation**: [Wang, T., Hamann, A., Spittlehouse, D.L., and Murdock, T.Q. 2012. ClimateWNA - High resolution spatial climate data for Western North America. Journal of Applied Meteorology and Climatology 51:16-29.](http://dx.doi.org/10.1175/JAMC-D-11-043.1)

### CHELSA

**Link**: [http://chelsa-climate.org/](http://chelsa-climate.org)  
**Where**: World  
**When**: 1979-2013  
**Description**: Monthly temperature and precipitation plus BIOCLIM variables at 30-arcsec (~1 km)-resolution.  
**Citation**: Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E, Linder, H.P. & Kessler, M. (Accepted) Climatologies at high resolution for the earth’s land surface areas. Scientific Data.

We will download and process the climate data in subsequent exercises.