Exercise 4: Clipping Environmental Rasters to the Study Region

Adam B. Smith, Danielle Svehla, & Camilo Sanín

July 19, 2017

In this exercise you learn **how to clip environmental rasters to the study region**. Selecting and obtaining environmental rasters is a workshop unto itself--however, in almost all cases modelers use some sort of climate variables. Less commonly they also employ land cover rasters and other types of data that they feel is relevant to the species of interest. You can browse a list of environmental data sources at [Earth::Sky::Sea](http://www.earthskysea.org/climate-weather/).

We will be using the [WORLDCLIM](http://www.worldclim.org/) climate data set. WORLDCLIM is a set of climate rasters created by Robert Hijmans (UC Davis) and colleagues. The dataset was created by interpolating records from thousands of weather stations across the world. The "base" WORLDCLIM dataset comprises rasters representing the mean precipitation, average minimum temperature, and average maximum temperature of each month of each month, averaged across many years. In June of 2017 Hijmans released version 2.0 of WORLDCLIM, but only for the present (1970-2000). The website says the future rasters are coming soon, so in the meantime we'll use Version 1.4's future rasters.

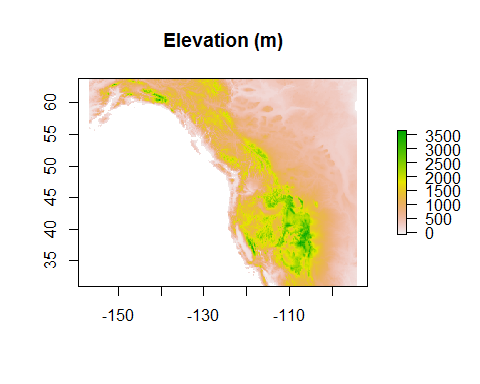
WORLDCLIM is available in several resolutions, with cell sizes from 10 arcmin, or about 17 km at the equator to 30 arcsec, or about 1 km at the equator. We will be using the 10 arcmin rasters because they are much faster to process in R. You should use rasters with resolutions that best match the spatial uncertainty in your records while at the same time best capturing important environmental gradients. For example, montane environments change sharply across short distances, so fine-resolution rasters are appropriate. Ergo, we should not be using 10 arcmin rasters for the Columbian ground squirrel! But for the sake of expediency we'll sacrifice accuracy.

Note that R can have problems processing very large rasters. If you have large rasters, clipping them manually in a program like ArcMap or using Python may be faster.

## Clip elevation

We'll start by clipping a raster that represents elevation. This raster used to be on the WORLCLIM web site, but we can't find it there now (a danger of using others' data)! So we've included it in the folder WORLDCLIM\Elevation\World.

# create output directory  
elevation <- raster('./WORLDCLIM/Elevation/World/elevation.tif')  
elevation <- crop(elevation, studyExtent)  
elevation <- setMinMax(elevation)  
projection(elevation) <-  
 '+proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0'  
  
dir.create('./WORLDCLIM/Elevation/Study Region', recursive=TRUE, showWarnings=FALSE)  
  
writeRaster(elevation, './WORLDCLIM/Elevation/Study Region/elevation',  
 format='GTiff', datatype='INT2S', overwrite=TRUE)  
  
plot(elevation, main='Elevation (m)')



What's going on? The raster() command loads the raster. The crop() function clips the raster to the extent of the study region, discarding the portion outside. setMinMax() re-sets the raster's min/max values in its metadata table. We then define the raster's projection to an unprojected (WGS84) coordinate system. Finally, we save the raster. The remainder of the scripts in this section do the exact same thing, except that they cycle through sets of rasters. They will do this very quickly because we're using coarse-resolution (10-arcmin) rasters.

## Download WORLDCLIM climate data

1. In your working folder create a folder named WORLDCLIM, and inside that create a folder named 1970-2000 and inside that one named World (ergo, ./WORLDCLIM/1970-2000/World).
2. Go to the WORLDCLIM page for [Version 2.0](http://worldclim.org/version2).
3. Scroll downward to the second table and download the file for the "bio\_10m" version of "Bioclimatic variables"." Save this in the folder ./WORLDCLIM/1970-2000/World.
4. Unzip the files into the same folder.

## Exploring WORLDCLIM data

Let's look at the current climate layers. First, let's look at mean annual temperature (MAT) for the current time period.

mat <- raster('./WORLDCLIM/1970-2000/World/wc2.0\_bio\_10m\_01.tif')  
mat

## class : RasterLayer   
## dimensions : 1080, 2160, 2332800 (nrow, ncol, ncell)  
## resolution : 0.1666667, 0.1666667 (x, y)  
## extent : -180, 180, -90, 90 (xmin, xmax, ymin, ymax)  
## coord. ref. : +proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0   
## data source : C:\ecology\Drive\Workshops\SDM from Start to Finish 2017-08 (ESA, Portland)\Production Documents\WORLDCLIM\1970-2000\World\wc2.0\_bio\_10m\_01.tif   
## names : wc2.0\_bio\_10m\_01   
## values : -53.70207, 33.26064 (min, max)

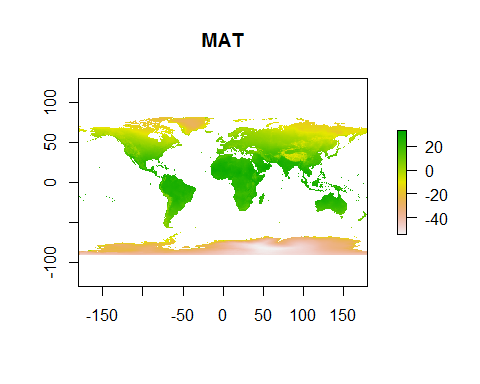
This is the contents of the mat raster--you can see that it is 900 cells high by 2160 cells wide. It also has a resolution of 0.1667 or 1/6th of a degree (= 10 arcmin). The extent covers the entire world, from -180 longitude to 180 longitude (the International Date Line) and from -60 degrees south (southern the tip of South America) to 90 north (the North Pole). The raster uses the WGS84 coordinate system (a model of the Earth for designating locations--more on this later). The name of the raster in R is WC01 (not BIO01). Finally, R shows you the presumed min and max values for the raster, but in actuality these are just the min and max values that the data format in which this raster is stored can represent.

You can see that the minimum value is about -53 and the maximum 33. If these values are not shown you can define them using

mat <- setMinMax(mat)  
mat

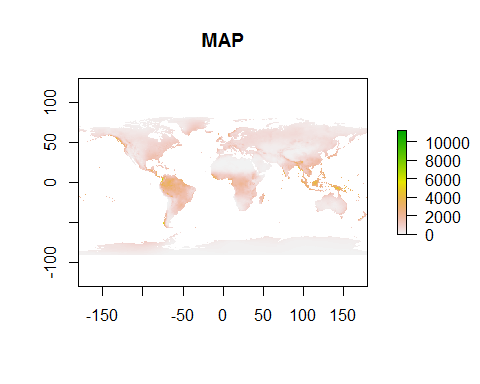
Let's plot the raster.

plot(mat, main='MAT')



Now, let's look at mean annual precipitation (MAP).

map <- raster('./WORLDCLIM/1970-2000/World/wc2.0\_bio\_10m\_12.tif')  
plot(map, main='MAP')



The units for most precipitation variables in WORLDCLIM is millimeters, so the map of MAP represents precipitation ranging from no precipitation to >10 m per year!

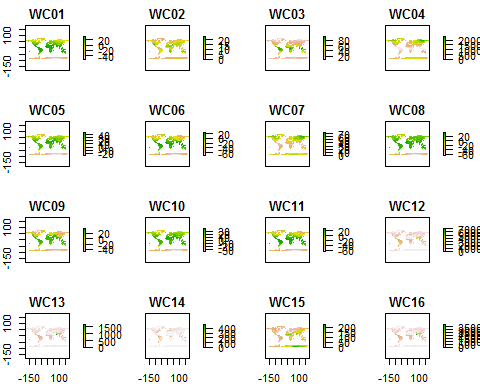
Now let's look at all of the BIOCLIM layers together (or most of them). We'll create a "raster stack" object that holds multiple rasters together that can be treated as if it were a single variable.

climate <- stack(listFiles('./WORLDCLIM/1970-2000/World'))  
names(climate) <- paste0('WC', prefix(1:19, 2))  
climate

## class : RasterStack   
## dimensions : 1080, 2160, 2332800, 19 (nrow, ncol, ncell, nlayers)  
## resolution : 0.1666667, 0.1666667 (x, y)  
## extent : -180, 180, -90, 90 (xmin, xmax, ymin, ymax)  
## coord. ref. : +proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0   
## names : WC01, WC02, WC03, WC04, WC05, WC06, WC07, WC08, WC09, WC10, WC11, WC12, WC13, WC14, WC15, ...   
## min values : -53.70207, 0.00000, 3.50136, 0.00000, -27.21700, -71.80151, 0.00000, -64.45292, -48.10492, -35.97421, -64.45292, 0.00000, 0.00000, 0.00000, 0.00000, ...   
## max values : 33.26064, 25.69088, 96.95946, 2390.73398, 47.44875, 26.50000, 74.21200, 37.70562, 38.43350, 38.85667, 31.56267, 11191.00000, 2381.00000, 484.00000, 229.00169, ...

You can see that the stack object contains 19 layers, one for each BIOCLIM variable.

plot(climate)



The world maps are nice, but our focal species does not occur everywhere! Before we model, we'll want to clip the rasters to the appropriate spatial extent, choose which amongst them we want to use, and do some further preparation.

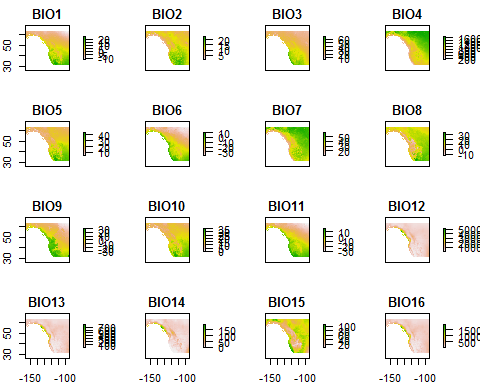
## Clip current BIOCLIM rasters

Now we will clip the rasters to the study region to make them more manageable.

dir.create('./WORLDCLIM/1970-2000/Study Region', recursive=TRUE, showWarnings=FALSE)  
   
# for each BIOCLIM raster  
for (i in 1:19) {  
  
 print(paste('Clipping current WORLDCLIM raster', i))  
 flush.console()  
  
 # get the world raster  
 rast <- raster(paste0('./WORLDCLIM/1970-2000/World/wc2.0\_bio\_10m\_',  
 ifelse(i < 10, '0', ''), i, '.tif'))  
   
 # crop the raster and update metadata  
 rast <- crop(rast, studyExtent)  
 rast <- setMinMax(rast)  
 projection(rast) <-  
 '+proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0'  
   
 # save raster  
 writeRaster(  
 rast,  
 paste0('./WORLDCLIM/1970-2000/Study Region/WC',  
 ifelse(i < 10, '0', ''), i), format='GTiff', datatype='INT2S',  
 overwrite=TRUE  
 )  
   
}

# visually check  
current <- stack(  
 list.files(  
 './WORLDCLIM/1970-2000/Study Region',  
 full.names=TRUE,  
 pattern='.tif'  
 )  
)

plot(current, main=paste0('BIO', 1:19))



What are these? The rasters represent the 19 classic "BIOCLIM" variables which attempt to capture aspects of climate that are physiologically stressful or affect resource availability. You can see there definitions by doing a web search for "WORLDCLIM BIOCLIM" (or [here](http://www.worldclim.org/bioclim)). We'll discuss the merit of these variables later. For now note that most temperature-related variables are expressed in degrees C and most precipitation-related variables are expressed in mm.

## Clip future WORLDCLIM BIOCLIM rasters

For our future climate layers we'll be using a global ensemble prediction that we calculated from 8 of the AOGCMs available on the [WORLDCLIM website](http://www.worldclim.org/cmip5_10m) as of August 3, 2017 (specifically, version 1.4 Release 4 of WORLDCLIM). We chose these 8 models because they had the full complement of data available (e.g., the AOGCM ACCESS1-0 doesn't have predictions for RCP2.6) and by eliminating all but one model in each "family" (e.g., there are three HadGEM models). When deciding between models in the same family, we chose the one that represents the full Earth-atmosphere-ocean process (some ignore components like some chemical reactions in the atmosphere).

To create the ensemble we downloaded each individual AOGCM's monthly predictions for min/max temperature and precipitation, for each variable averaged across AOGCMs for each month, then from the three sets of monthly values for min/max temperature and precipitation calculated the 19 BIOCLIM variables we examined above. The BIOCLIM variables were calculated using the bioclim() function in the dismo package.

Note that these projections are really averages for periods *centered* on 2050 and 2070 even though it's convenient to call them just "2050" and "2070". That's important because predictions for any one year in the future are sure to be very different from reality, while predictions averaged across a time period are likely more accurate.

You can find the future layers in the folder ./WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE/World. For this workshop we'll only use the projections for 2070.

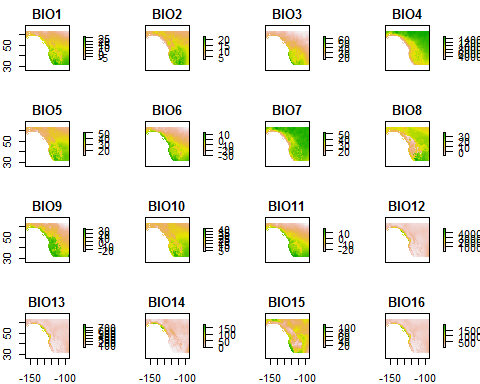
Now let's crop the rasters representing the ensemble forecast for the 2070s. You can't actually download this data--rather, you have to download each of the [8 GCM datasets](http://worldclim.org/cmip5_10m) then average them. This can take a lot of time, so we've already done this for you and have included them in the folder ./WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE. They represent climate expected under the RCP8.5 emissions scenario (the worst-case scenario).

Note that the temperature-related variables for the *future* scenario are in units of 10ths of a degree C, so we'll re-scale them to match the units of the current rasters (degrees C)

dir.create(  
 './WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE/Study Region/',  
 recursive=TRUE, showWarnings=FALSE  
)  
  
# for each BIOCLIM variable  
for (i in 1:19) {  
  
 print(paste('Clipping future WORLCLIM raster', i))  
 flush.console()  
  
 # get the world raster  
 rast <- raster(  
 paste0(  
 './WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE/World/en85bi70',  
 ifelse(i < 10, '0', ''), i, '.tif')  
 )  
   
 # crop raster  
 rast <- crop(rast, studyExtent)  
   
 # rescale units for BIOCLIM variables 1-2 and 5-11  
 if (i %in% c(1:2, 5:11)) rast <- rast / 10  
   
 # update metadata  
 rast <- setMinMax(rast)  
 projection(rast) <-  
 '+proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0'  
   
 # save raster  
 writeRaster(  
 rast,  
 paste0('./WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE/Study Region/en85bi70', ifelse(i < 10, '0', ''), i),  
 format='GTiff', datatype='INT2S', overwrite=TRUE  
 )  
   
}

# visually check  
future <- stack(  
 list.files('./WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE/Study Region',  
 full.names=TRUE, pattern='.tif')  
)

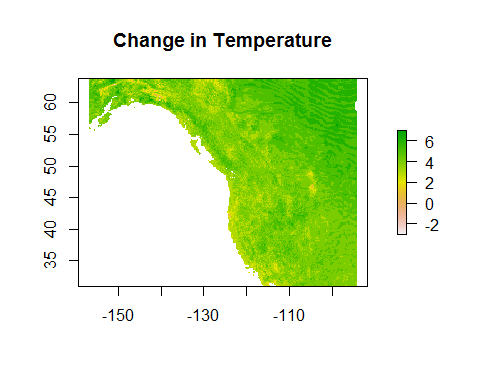
plot(future, main=paste0('BIO', 1:19))



## Delta T

How much warmer is the study region going to get under RCP8.5 in 2070? Let's calculate the difference (delta) raster.

currentTemp <- raster(  
 './WORLDCLIM/1970-2000/Study Region/WC01.tif')  
futureTemp <- raster(  
 './WORLDCLIM/2061-2080 RCP8pt5 ENSEMBLE/Study Region/en85bi7001.tif')  
  
deltaTemp <- futureTemp - currentTemp  
plot(deltaTemp, main='Change in Temperature')



You can do the same thing for mean annual precipitation (WC12) if you'd like.

# Handling rasters

## A note on raster formats

You can see the raster formats the raster package can handle by using ?writeFormats. By default the raster package in R uses the GRD format, but we've had mixed luck opening these up in other programs.

We prefer the GeoTIFF format because it is has a small size on the disk and because it can be opened by many GIS programs. Small file size reduces processing time.

Often, raster data comes in the ASCII format, which is also compatible across GIS programs. However, ASCII rasters can be very large and as a result take a lot of time for R to process.

## A note on raster processing

R is good at many things, but it's not good at handling large rasters. It can literally take weeks to manipulate a set of rasters for niche/distribution modeling. In many cases we skip my usual R-based workflow and use ArcMap or other software which can usually do things much faster. The main aspects of rasters that affect processing time are:

* Size (number of cells wide by long)
* Resolution (smaller cells = more cells)
* How many rasters are to be processed
* Number of cells that have data (are not NA)

We are using the 10-arcmin (~18-km) resolution WORLDCLIM rasters *specifically* because R can handle them quickly. Normally we suggest using rasters with smaller cell sizes to capture finer-scale local variation in climate, provided the precision of your species record coordinates are comparable to the given resolution.

# Reflection

1. For any given species in which you're interested, what climate variables (if any) would be appropriate to use for modeling? Does any empirical work inform your selection?
2. Under what conditions would you choose to work with high-resolution (small cell size) versus low-resolution (large cell size) rasters?
3. What non-climatic variables might be important to modeling any particular species in which you're interested? Are these kinds of data available?