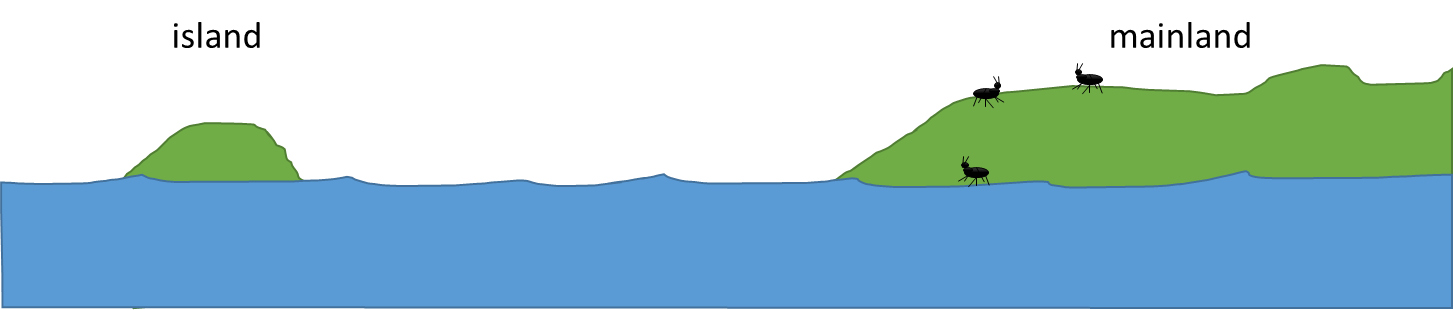
Exercise 13: Niche versus Distribution Modeling

Adam B. Smith, Danielle Svehla, & Camilo Sanín

July 31, 2017

For the past many years **species distribution modeling** has been synonymous with **ecological niche modeling**, though there have been debates over which is the better term ([Warren 2012](http://www.dx.doi/org/10.1016/j.tree.2012.03.010), [McInery & Etienne 2013](http://dx.doi.org/10.1016/j.tree.2013.01.007), [Warren 2013](http://dx.doi.org/10.1016/j.tree.2013.02.003)). However, concurrent with this debate is a parallel dialog that makes a working distinction between ecological niche models (ENMs) and species distribution models (SDMs; [Barve et al. 2011](http://www.dx.doi.org/10.1016/j.ecolmodel.2011.02.011); [Peterson et al. 2011](http://press.princeton.edu/titles/9641.html%5D). Under this distinction, S**D**Ms model the *distribution* of a species, while E**N**Ms model the *niche*. The two are related, but they are also fundamentally different for reasons that appear esoteric at first but have important consequences for modeling.

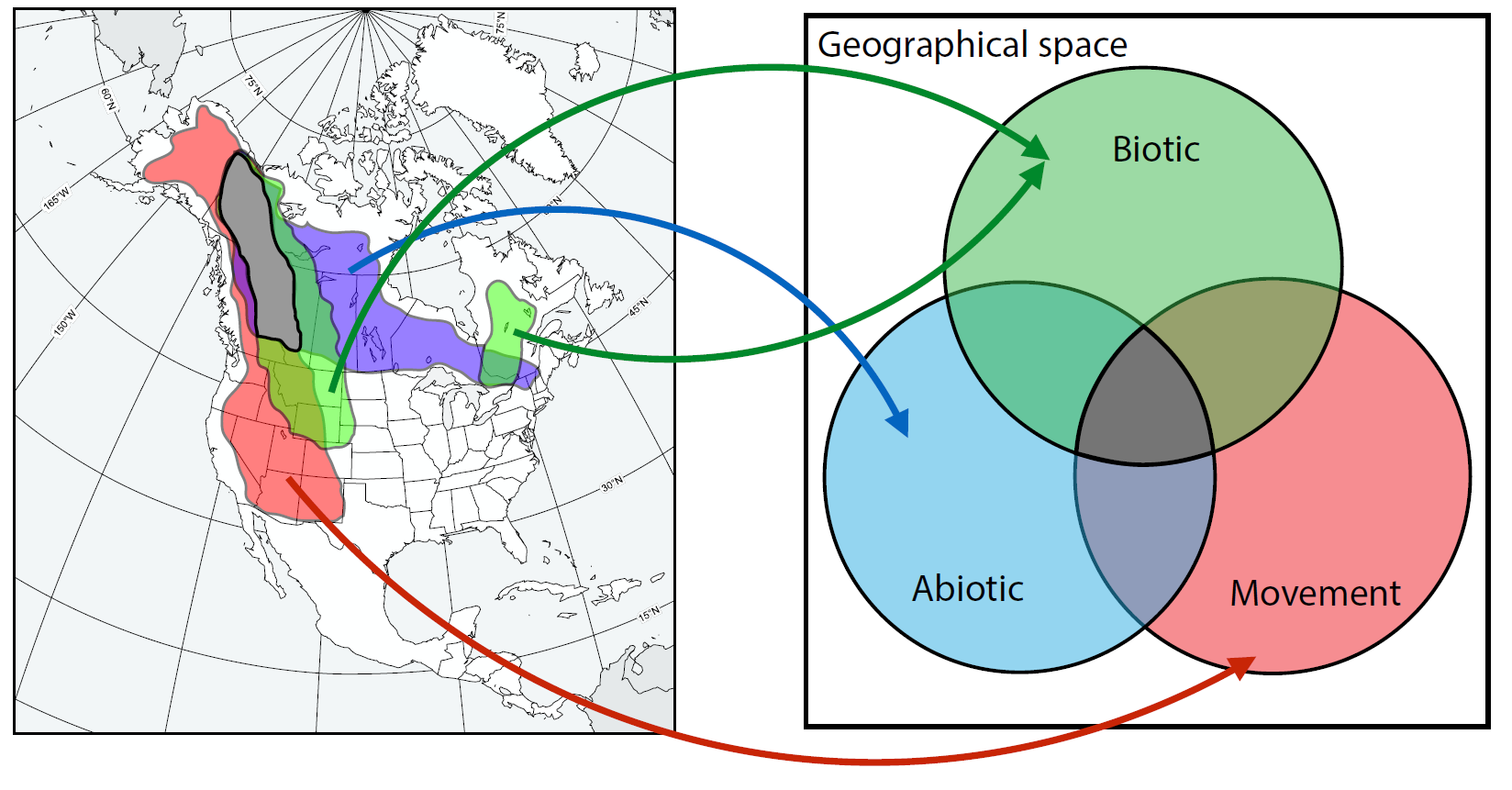
Consider a species that only existed on the mainland:



No species is an island.

Suppose we wanted to model the distribution of a species that only occurred on the mainland. Should we include the island as part of the region from which we draw background sites? It depends! If we are interested in modeling the *distribution* of the species, including the island would inform the model that this environment is available to the species but the species has been unable to inhabit it. Thus, the model would be more likely to predict the species is absent from the island. If we were interested in modeling the environmental *niche* of the species, then we should probably not include the island in the background set if the species has not been able to disperse there to "sample" its environment.

Although I've illustrated the issue with and island/mainland situation, the same kind of situation can occur when non-insular/mainland systems--i.e., sites should be considered in the background only if the species has been able to "sample" (through dispersal) their environments. Jorge Soberón ([Soberón 2007](http://www.dx.doi.org/10.1111/j.1461-0248.2007.01107.x); [Soberón & Nakamura 2009](http://www.dx.doi.org/10.1073/pnas.0901637106) has conceptualized this issue using what he calls the **biotic/abiotic/movement** or **BAM diagram**.



BAM Diagram

Here, each circle represents either 1) the **b**iotic conditions under which the species can thrive; 2) **a**biotic conditions under which it can thrive; or 3) the geographic areas to which the species has the capability of **m**oving. Note that all three conditions are necessary for the species to exist in a place. Although simple, this conceptualization has some deeply important aspects:

1. The fundamental niche is represented by **A**.
2. The realized niche is represented by the intersection of **A** and **B**.
3. The place where one can actually find the species is represented by the intersection of **B**, **A**, and **M**.
4. The species could live in sink populations that are found within **M** but outside either **A** or **B**.

Soberón & Nakamura originally proposed the BAM diagram as a conceptual device (i.e., the right side of the figure), but Camilo Sanín and Rob Anderson have demonstrated that each of its components has a real-world counterpart that can be mapped (the left side of the figure).

So if we are interested in modeling the fundamental niche, then we need to include in the background only places represented by **M**. (We might also think about how to avoid including sink populations in our model--but honestly this is generally ignored, though it probably ought not be!)

Before we proceed further, let's address **A**, the conditions under which biotic conditions allow the species to exist. How do we account for this in our ENM? You could, for example, use the presence/absence of any other species with which your focal species interacts as a predictor layer--but this would require unequivocal knowledge of other species' distributions, which are also generally unknown. Accounting for interspecific interactions is an evolving issue in distribution/niche modeling, and although a wide variety of papers have attempted to address the issue, to date there is no widely-applicable accepted method. Perhaps as a result, most niche modeling is conducted on the basis of the **Eltonian noise hypothesis**, which posits that biotic interactions determine whether a species is present at fine scales, but not coarse scales. The general reasoning is that within a large area a species an usually find some habitats that are not also occupied by a competitor/predator/parasite and so persist and so be counted as "present" when the spatial grain of analysis is coarse. Obliviously this assumption requires more research!

So far we have been modeling the distribution of our species--i.e., our focus has been on creating a SDM with a believable and well-performing spatial output. The niche of a species can also be expressed in geographic space since, after all, occupied areas (usually) have environmental conditions in which a species can thrive. Of course, we can make a distinction between the realized and fundamental niche, but in general we can assume that occupied places fall within both kinds of niches. So let's see how our species' SDM and ENM differ.

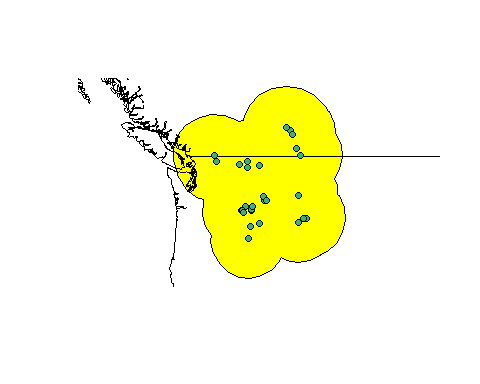
# ENM versus SDM

Recall that the practical distinction between an ENM and a SDM is that the former only includes background sites within **M** of a BAM diagram. So we need to know to where the species has (and has had) the ability to disperse. How do we do that? In general there is no shortcut since every species has its own mode of dispersal. However, you could choose background sites from:

1. An area defined by a buffer of a set distance (e.g., 250 km) from each known presence.
2. All ecoregions known to be occupied by the species.
3. An area defined by the known presences of the species plus sister species which presumably share niche characteristics of the focal species.

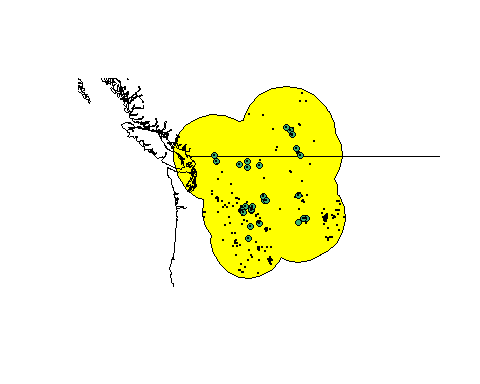
Here we'll explore just the first option. It may not be the best option for our species, but it will illustrate the point.

# project the species' records into planar coordinates... we'll use a Lambert  
# equal-area conical projection  
recordsSpatialPlanar <- spTransform(recordsSpatial, CRS('+proj=lcc +lat\_1=49 +lat\_2=77 +lat\_0=0 +lon\_0=-95 +x\_0=0 +y\_0=0 +datum=WGS84 +units=m +no\_defs +ellps=WGS84 +towgs84=0,0,0'))  
  
# create a 200-km buffer around each known presence site... express buffer size in meters  
buff <- gBuffer(recordsSpatialPlanar, width=250000)  
  
# back-project buffer into unprojected (WGS84) coordinate system  
buff <- spTransform(buff, CRS('+proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0'))  
  
# plot  
plot(buff, col='yellow')  
plot(countries, add=TRUE)  
points(records$longitude, records$latitude, pch=21, bg='mediumseagreen')



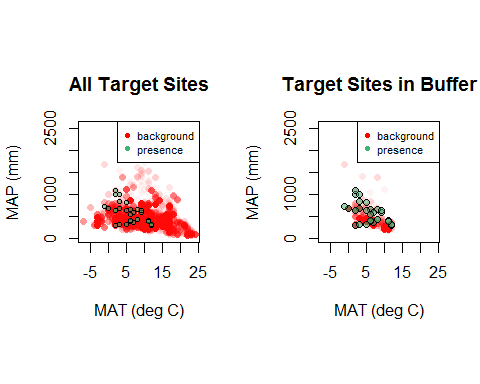
Now let's select every target background site that falls within the buffer.

# coerce target background sites into a spatial object  
targetBgSpatial <- SpatialPointsDataFrame(  
 coords=cbind(targetBg$longitude, targetBg$latitude),  
 data=targetBg,  
 proj4string=CRS('+proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0')  
)  
  
# get just target sites in buffer  
targetInBuffer <- as.data.frame(targetBgSpatial[buff, ])  
  
# plot  
plot(buff, col='yellow')  
plot(countries, add=TRUE)  
points(recordsSpatial, bg='mediumseagreen', pch=21)  
points(targetInBuffer$longitude, targetInBuffer$latitude, pch=16, cex=0.3)



What effect does drawing background sites from the area immediately surrounding presence sites have on the background that the model sees? It's easiest to examine this in environmental space:

par(mfrow=c(1, 2), pty='s')  
  
# vs all target sites  
plot(  
 targetBg$WC01, targetBg$WC12,  
 pch=16, col=alpha('red', 0.05),  
 main='All Target Sites',  
 xlab='MAT (deg C)',  
 ylab='MAP (mm)',  
 xlim=c(min(targetBg$WC01), max(targetBg$WC01)),  
 ylim=c(0, max(targetBg$WC12))  
)  
  
points(records$WC01,  
 records$WC12,  
 pch=21,  
 bg=alpha('mediumseagreen', 0.5),  
 cex=0.7  
)  
  
legend('topright',  
 legend=c('background', 'presence'),  
 col=c('red', 'mediumseagreen'),  
 pch=16,  
 cex=0.7  
)  
  
# vs restricted target sites  
plot(targetInBuffer$WC01, targetInBuffer$WC12,  
 pch=16, col=alpha('red', 0.05),  
 main='Target Sites in Buffer',  
 xlab='MAT (deg C)',  
 ylab='MAP (mm)',  
 xlim=c(min(targetBg$WC01), max(targetBg$WC01)),  
 ylim=c(0, max(targetBg$WC12))  
)  
  
points(records$WC01,  
 records$WC12,  
 pch=21,  
 bg=alpha('mediumseagreen', 0.5)  
)  
  
legend('topright',  
 legend=c('background', 'presence'),  
 col=c('red', 'mediumseagreen'),  
 pch=16,  
 cex=0.7  
)



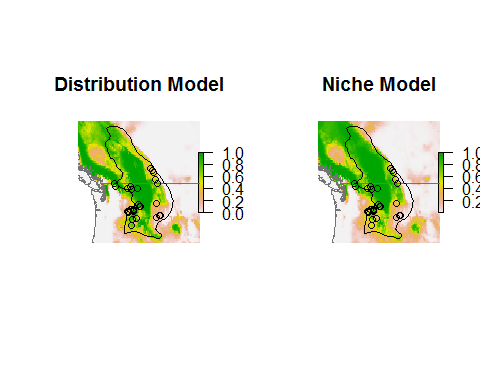
You can see that the environmental range of the background observable to the model has been truncated greatly--but that's appropriate since we're assuming the species has not been able to disperse to environments outside this. Let's train our **niche** model!

# make training data frame with predictors and vector of 1/0 for presence/background  
envData <- rbind(  
 records[ , predictors],  
 targetInBuffer[ , predictors]  
)  
  
presBg <- c(rep(1, nrow(records)), rep(0, nrow(targetInBuffer)))  
  
trainData <- cbind(presBg, envData)  
  
# create output directory for model object and rasters  
dirCreate('./Models/Model 11 Niche Model')  
  
# smooth model  
nicheModel <- trainMaxNet(  
 data=trainData,  
 regMult=c(0.5, 1:3),  
 classes='default',  
 verbose=TRUE,  
 out='model'  
)

# save model  
save(nicheModel,  
 file='./Models/Model 11 Niche Model/Model.Rdata',  
 compress=TRUE)

Now what? As before, we could write the projection raster for the niche model. Let's go ahead and do that, but keep in mind that the output is indicative of locations that fall inside the *niche* of the species, not the distribution of the species.

# predict to raster  
nicheMap <- predict(  
 climate,  
 nicheModel,  
 filename='./Models/Model 11 Niche Model/maxentPrediction1970to2000',  
 format='GTiff',  
 overwrite=TRUE,  
 type='cloglog'  
)  
  
par(mfrow=c(1, 2), pty='s')  
  
# plot newest \*distribution\* model  
plot(rangeMap, main='Distribution Model')  
plot(tunedMap, add=TRUE)  
sp::plot(countries, add=TRUE, border='gray45')  
plot(rangeMap, add=TRUE)  
points(records$longitude, records$latitude)  
  
# plot niche model  
plot(rangeMap, main='Niche Model')  
plot(nicheMap, add=TRUE)  
sp::plot(countries, add=TRUE, border='gray45')  
plot(rangeMap, add=TRUE)  
points(records$longitude, records$latitude)



## Reflection

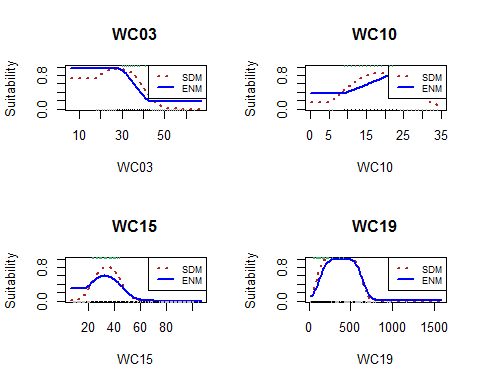
1. How do the geographic projections of the species distribution model and ecological niche model differ?
2. Why is the ENM "overpredicting" in geographic space?

Now, let's look at the niche in environmental space. We'll graph the response functions for each variable as we hold all other variables constant (at their median value).

# get min/max value of each predictor across study region  
minPred <- minValue(climate)  
maxPred <- maxValue(climate)  
names(minPred) <- names(maxPred) <- names(climate)  
  
# get median value of each predictor across species' presences  
medianPred <- apply(records[ , predictors], 2, median)  
  
# make data frame with median value of each predictor  
env <- as.data.frame(medianPred)  
env <- t(env)  
env <- env[rep(1, 100), ]  
rownames(env) <- 1:nrow(env)  
head(env)

## WC03 WC10 WC15 WC19  
## 1 36 15 36 135  
## 2 36 15 36 135  
## 3 36 15 36 135  
## 4 36 15 36 135  
## 5 36 15 36 135  
## 6 36 15 36 135

par(mfrow=c(2, 2))  
  
# for each predictor  
for (pred in predictors) {  
  
 # make copy of data frame  
 thisEnv <- env  
   
 # now vary focal predictor from min to max value  
 # all other predictors keep median value  
 thisEnv[ , pred] <- seq(minPred[pred], maxPred[pred], length.out=100)  
   
 # make prediction using this data frame  
 predictionSDM <- predict(tunedModel, thisEnv, type='cloglog')  
 predictionENM <- predict(nicheModel, thisEnv, type='cloglog')  
   
 # plot  
 plot(x=thisEnv[ , pred],  
 y=predictionSDM,  
 ylim=c(0, 1),  
 xlab=pred,  
 ylab='Suitability',  
 main=pred,  
 type='l',  
 col='firebrick',  
 lty='dotted',  
 lwd=2  
 )  
   
 lines(x=thisEnv[ , pred], y=predictionENM, col='blue', lwd=2)  
   
 legend('topright',  
 legend=c('SDM', 'ENM'),  
 lty=c('dotted', 'solid'),  
 col=c('firebrick', 'blue'),  
 lwd=2,  
 cex=0.7  
 )  
   
 # add species' presences (top rug)  
 rug(records[ , pred], side=3, col='mediumseagreen')  
   
 # add background sites (bottom rug)  
 rug(targetBg[ , pred], side=1, col='gray')  
 rug(targetInBuffer[ , pred], side=1, col='black')  
   
}



Here the gray "ticks" on the button (the rug plot) represent the location of all target background sites, while the black ticks represent target background sites within the buffered region. You can see that the range of target background sites in the buffered region is smaller than the range across all target sites. As a result, the response curves have different forms.

## Reflection

1. How does your ENM contrast with your SDM in geographic space and in environmental space? Is there something striking about the differences between the models in environmental space?
2. If you were projecting a species' response to climate change, which model type (SDM or ENM) should you use?
3. How else might you estimate **M** (in the BAM diagram)? Do you think it is possible to estimate **A**?