**Monday in class**

**Video screening and voting/judging**

**Intro new unit**

Now we begin the second half of this class, in which we are studying the *Research Uses of collections*

This is a review article from Trends in Ecology and Evolution, fondly referred to as TREE. It has a **glossary**. I went through and highlighted the first time each of the terms in the glossary appears in the paper. You need to look at the glossary, because the terms have particular scientific meanings in this paper that may be different from what you are used to, but are critical to understanding the paper.

This is more dense and maybe even complicated than what we’ve been doing before, so I will do a thorough reading guide for you. Take NOTES.

This comes from MVZ, so the goal here is to get people using museum specimens—in good ways where good means statistically valid ways

Also, to influence collectors of contemporary specimens and data to collect in ways that make the data maximally useful for looking at range change over time

1. Museum specimen data are like the lord of the rings

**2. GLOSSARY** Important definitions that I might test you on: covariate, false absence, non-detection, occupancy, occupancy modeling, season

3. IN a review, a box gives you the basic biology, theory or other background knowledge and assumptions that an expert would know and that are often critical for interpreting the papers in the field

Q1. What information must be available to make the historical occurrence data useful in understanding species ranges (location and time)

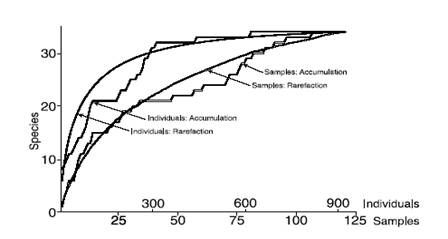
4. This section sets out to **first define the four different types of historical occurrence data and some of the problems specific to each type**—these are summarized in Table 2, you should mark all over this table and know it really well. Specifically, know the issues with each type of data

Q2—NHM specimens are what type of historical data

NHM data issues,

* they show presence only
* If you have a spot with no records, surrounded by records, you don’t know if the species is absent from that location, or simply wasn’t recorded🡪that is a non-detection
* So a museum database has no information about non-detections
* That means you can only use the locations of where species were found and look forward in time for extinctions in those sites. If you have a site with no museum records and then you suddenly see the species in that location, you don’t know if it was historically absent or just not collected from that location.
* People have generated random pseudo-absence data sets, where the total study area is divided into sites, or “communes”, and any un-recorded site is a possible site for an absence. There are more sophisticated ways than random, where you use prior knowledge about the site, including environmental/habitat characteristics and presence-absence of similar species

Presence and non-detection: where a site is visited once and only once, but multiple sites are visited

* An absence in this dataset can be false, because sampling intensity is related to species recovery
* Here is a [species accumulation curve:](http://www.biosym.uzh.ch/modules/models/Biodiversity/MeasuresOfBioDiversity.html)**A species accumulation curve**(collector’s curve)  records the total number of species revealed, during the process of data collection, as additional individuals are added to the pool of all previously observed or collected individuals or samples (Colwell and Coddington 1994). This curve is not smooth.  
  In contrast**, a rarefaction curve**is produced by repeatedly re-sampling the pool of N individuals, at random, plotting the average number of species represented by 1,2,…,N individuals or samples (figure 3).

**Qx:** Another type of data is **presence and non-detection**🡪how do you get estimable absence data? (historical surveys must have been repeatd within short time periods , repeated surveys basically

**Why is it difficult to use historical abundance data?** (use of outdated or nonstandard methodologies in historical abundance data)

5 **Make a list of the problems with using historical data**

* Differentiating between true absence and non-detection
* False presences or misclassifications of species (worse with surveys)
* Bias from studies conducted in different ways by different people over time (worse with longer times between surveys)
* Geographical precision
  + comes from vague accounts, dishonest reporting, human error,
  + affects estimates of species richness and community composition
  + can lead to false conclusions on extinction and colonization events!
    - i.e. Inflates turnover estimates

**False presences—what are they, in what types of data are they a bigger problem?** (generally when a species ID is wrong, worse problem in surveys than in museum specimens because they can’t be detected after the fact)

9. Equation reads:

**Probability of false absence** is where you calculate the probability of detecting the species over all of your repeated surveys at the site, take 1 – that probability (the 1 – turns your probability of finding it into a probability of not finding it when you should have), and then do the same thing for every site in your study and add those (1 – probabilities) all up. And that gives you the probability of a false absence, or the probability you didn’t find it, but you should have.

Look at Figure 1 in Box 3. They marked down every presence and every non-detection at each sight and trapped multiple nights in a row at each site.

**What do you use a Pfa test for**?

* Test for changes in the limits of distributions
* What is happening at the limit/edge of a distribution—vagrant individuals can be detected outside of their fundamental niche (fundamental niche means the full range of env conditions and resources an organism can possibly occupy and use), this could be be a sink population, that is a population that survives by immigrants bolstering its numbers, without immigration, it would decline, things are in flux at the boundaries and these populations can have high turnover
* So its better to look at the whole distribution of the species, not just its limits when modeling

What is a **species optimum**—see box 3 and last paragraph in section 9

* The maximum probability of occupancy in a distribution along a gradient is the species optimum
* Example is in box 3, where the optimum elevation is 320m higher in modern day compared to historic times

What does **box 3** tell you:

* that the range has expanded to encompass lower sites AND
* that the optimum site is a higher elevation—this may seem contradictory, but it is telling you that you need to look at both the optimum site and at the edges of the distribution to get a full picture of the changes in species distribution

10 C**onclusions**

* biases can be overcome with occupancy modeling and other techniques like species optimum from max probability approaches
* but, you need repeated visits within survey periods and data from enough sites
* Emphasize the importance of observational and occurrence data (although earlier they said that is a problem if it doesn’t have a specimen because of false IDs!), encourage NHMs to make non-vouchered observation data available publicly along with specimen occurrences
* Encourage repeated site visits for increasing from presence-only to presence and estimable absence data
  + This is not something we normally do, we usually visit as many habitats and locations as possible to capture as much diversity as possible. We quit early on the species accumulation curve at single sites and focus on multiple sites instead
* Historical data must be used because it expands both the geographic and temporal scales of inference on change