**Areas For Improvement**

**June 17, 2021**

*Reminder*: Our main research question is, *how did statistical techniques change over time in the ecological sciences?* This should be central to how our decisions are made when searching and then filtering for papers to read.

**A) Search Criteria**

Web of Science (WoS) Core Collection Help: [link](https://images.webofknowledge.com/images/help/WOS/hs_advanced_fieldtags.html)

Help Center WoS: [link](https://mjl-clarivate-com.proxy.uchicago.edu/help-center)

**TS = Topic** (this means that the article must contain any word or form of the word in the article)

**SU = Subject Area**

**CU= Country/Region** (each author has an address in WoS, so based on where authors lived and studied, authors will be filtered by country/region)

The search criteria has been updated. Another prominent journal, the *Journal of Ecology*, has been added. Additionally, WC=”ECOLOGY” on Web of Science was removed in the search criteria and replaced with the following in the advanced search:

**((TS=("eco\*" OR “evo\*” NOT "econo\*" NOT "ecofem\*" NOT “ecotour\*” NOT “paleo\*” NOT "bio\*" NOT “evok\*” NOT “evoc\*”) AND SU=Environmental Sciences & Ecology AND CU=USA))**

When you search that criteria and select 1900-1999, “English”, and “Article”, first you are left with 18,247 articles, but once you select only for *Ecology* and the *Journal of Ecology* under the “Source Titles” drop down menu, you are left with a total of 1,169 articles. This pool of papers is manageable and contained to the USA (authors’ address; also, all authors represent only US institutions). This also better reflects a process suggested during a past meeting:

It was asked, why not just include all articles offered by the two prominent journals? When you go to the Web of Science Basic Search page and you select “Publication Name” in the drop down menu next to the Search button, type in “Journal of Ecology” in the search bar and then click on search. You get 7,211 hits. For “Ecology”, you get 17,080 hits. However, when you remove the years 2000-2021 for each, you then get 3,997 and 10,067, respectively. When you select only for articles you then filter down to 3,520 and 8,408, respectively. Selecting for English reduces each slightly to 3,520 and 8,406, respectively.

Finally, when you select for USA as the Country/Region, you are left with 561 and 4,677 respectively. Notably, if you select for a Research Area (SU) of Environmental Sciences and Ecology, then there is no change in the number of hits. This then leaves you with a total of 5,238.

In turn, our current hypothesis as to why that difference exists is that maybe once you filter for the topics (TS) you lose about 4k hits.

This can be tested in code. How? Start by seeing if all 1,169 articles overlap.

**Pre-Analysis:**

* + Each time you determine your search criteria on Web of Science, SAVE that search criteria and then compare which articles stay or are removed in code.
  + A code\_catalogue.docx has been created as a dictionary of all our possible search engine results. This will help with organizing the data and naming the data.
    - Helpful functions to use in R:
      * duplicated( )
      * intersect()
      * groupby( ) # compute group counts by year
      * hist( ) # for plotting

**Pending:** Add unpublished data (i.e. conference abstracts, master/PhD theses).

**B) Filtering**

What do you want the library to look like? Repeating authors? Most popular authors across the century? Across each decade? Do we want to prioritize authors who were also major editors of ecology associations? Explore each combination in the get\_masterlist.Rmd code.

**C) Spreadsheet**

Add “Article Title” column. The ideal would be to have the exact same article title (capitalization and syntax) as exported by Web of Science. Later down the line, Anastasia can constrain all the syntax and capitalization to be the same no matter how it gets input, so datasets can be merged.