Rangelands analog analysis: a brief summary

Hao Li, Ph.D. in Economics (SES, WSU)

(last update: Jan 5, 2021)

Concept

1. Research objective: find out current ecological analogs of a future rangeland.
    a. For example, we have a set of rangelands {A, B, C};
    b. How does A look like in 2050?
        i. is it similar to B or C in 2020 (today)?
        ii. if yes, B and/or C are analogs of A;
        iii. if not, there are novel climates.

Empirical framework

1. Methodology: principal component analysis (PCA).
    a. a "distance" based method;
    b. uses orthogonal transformation to project the original data into a new space where:
        i. the values of the variance-covariance matrix (variances) are maximized;
        ii. the off-diagonal values (covariances) are minimized.
        iii. this is to remove trivial variations.
    c. to minimize the squared average squared distance (Mahalanobis distance).
        i. without step b., it's called the Euclidean distance
    d. read more: CMU PCA Toturial

2. Variables:
    a. Climate variables:
        i. Max temperature (Sprint, Summer, Fall, Winter);
        ii. Min temperatures (Sprint, Summer, Fall, Winter);
        iii. Precipitation (Sprint, Summer, Fall, Winter).

    b. Ecological Variable: net primary productivity (npp, Annually).

3. Studied periods:
    a. Historical periods: 1980 - 2014 (in order to construct the A matrix)
    b. Future periods:
        i. 2025 - 2050 (the B1 matrix);
        ii. 2051 - 2075 (the B2 matrix);
        iii. 2076 - 2099 (the B3 matrix).
    c. Reference matrix: 1980 - 2014 (the C matrix)
4. Scenarios: RCP45 and RCP85
5. Data sources: 5 simulation models for each scenario, 10 models in total --

```
"RCP45_CGCM3", "RCP45_CM5", "RCP45_ES365", "RCP45_M",
"RCP45_RESM1", "RCP85_CGCM3", "RCP85_CM5", "RCP85_ES365",
"RCP85_MR", "RCP85_RESM1"
```

Data

1. Historical data:
    a. PRISM models:
        i. variables: max temperature, min temperature, precipitation
        ii. path (Kamiak absolute paths, same below): `/data/rajagopalan/MC2/Harddrive_MC2/ConUS/Climate/PRISM`
    b. MC2 models:
        i. variable: net primary productivity
        ii. path : `/data/rajagopalan/MC2/Harddrive_MC2/ConUS/MC2_Results/Hist/hist_nfs`
2. Future data:
    a. MACA models:
        i. variables: max temperature, min temperature, precipitation
        ii. path:
            i. RCP45: `/data/rajagopalan/MC2/Harddrive_MC2/ConUS/Climate/MACA_rcp45`

ii. RCP85: `/data/rajagopalan/MC2/Harddrive_MC2/ConUS/Climate/MACA_rcp85`

   b. MC2 models:
      i. variable: net primary productivity
      ii. path:
         i. RCP45: `/data/rajagopalan/MC2/Harddrive_MC2/ConUS/MC2_Results/Rcp45`
         ii. RCP85: `/data/rajagopalan/MC2/Harddrive_MC2/ConUS/MC2_Results/Rcp85`
3. Rangeland districts shapefile:
   a. Defines the geometric shape of all the rangeland districts in the U.S.
   b. path: `/data/rajagopalan/MC2/USRD`
4. Extracted regional data:
   a. Regional data that are filtered, masked, and extracted from the raw data with the shapefile.
   b. path: `/data/rajagopalan/MC2/sel_data`

Programming preparations

1. Tool: Python
2. Platform: Kamiak or personal computer.
3. Required packages:
   a. basics: xarray, numpy, os
   b. Data match, mask, and extraction: geopandas, GDal, regionmask, netCDF4
   c. data process and calculation: numpy, scipy
   d. plots: basemap, matplotlib

Procedures

1. Region data mask.
   a. objective: extract the exact regional data based on the rangeland districts shapefile.
   b. code: Hao Li's GitHub: regional data extraction
   c. inputs: raw netCDF4 files and the rangeland district shapefile

d. outputs: a series of numpy data files that contains regional climate and npp data per each district.
e. output path: `/data/rajagopalan/MC2/sel_data`
f. time consumption: takes 60 - 70 hours by Kamiak in total (I run a separate code for each group of data, so the total time is an estimate)

2. PCA procedures and analog finding.
   a. objective: perform PCA procedures to find spatial analogs;
   b. code: Hao Li's GitHub: analog finding, all districts, all models, all scenarios, by Kamiak
   c. input: the regional data files generated from the above step.
   d. outputs: distance tables, quantile tables, and the best analogs.
   e. analog results: `/data/rajagopalan/MC2/analog_results`
   f. time consumption: approximately 4 - 5 hours (4:43.362032 hours in the July 21 sample run)
3. Map plotting:
   a. objective: plots maps that containing targeted rangeland districts and their analogs with graduation color schemes to distinguish the similarity.
   b. code: Hao Li's GitHub: analog map plotting, all districts, all models, all scenarios, by Kamiak
   c. input: the regional data files generated from the above steps.
   d. outputs: analog maps (jpeg files)
   e. analog maps: `/data/rajagopalan/MC2/analog_map_plots`
   f. time consumption: approximately 89 hours by Kamiak (3 days, 16:01:34.409949 in the Aug 18 instance)

Findings

1. Most rangeland districts are identified to have current analogs;
2. All the distance tables are generated;
3. All the analogs are plotted (~14,000);
4. Most analogs and most of the best analogs migrate to the south along with time.