Result Reproduction

1. Installation and Setup

Set up the correct work directory and install all the required packages using the following code.

```r
setwd("~/ExtremePrecip/")
install.packages(c("lubridate","evd","mgcv","evgam"),
            dependencies=TRUE,repos="https://cloud.r-project.org")
install.packages(c("gridExtra","ggpubr","ggplot2"),
            dependencies=TRUE,repos="https://cloud.r-project.org")
install.pacakges("code/archived/mvPotST/mvPotST_0.0.0.9000.tar.gz",
            repos=NULL,type="source")
```

2. Data and Objective

The dataset consists of observed daily precipitation data in millimeters (mm) from 125 monitoring stations in the Danube river basin (Europe) and from 2229 monitoring stations in the Mississippi river basin (North America) over the period from 1965 to 2020. We divided the Mississippi river basin into 7 subregions according to Watershed Boundary Dataset, see the figure below. The temperature covariate (Celsius degree) used to fit the model in this project was derived from the ERA5-Land reanalysis data for the corresponding region we selected, which is a global land-surface dataset with a high spatial resolution of 9km. The projected temperature covariate is derived from climate models outputs of the sixth Coupled Model Intercomparison Project (CMIP6), namely AWI, MIROC, and MPI.

We have saved the data used in our marginal modeling and dependence modeling in the format of R data file (.RData), and we host those data on Github[https://github.com/PangChung/ExtremePrecip] under the folder "data'', which are public accessible. Though the raw data of the EAR5-Land Reanalysis data from which we derived the temperature covariate in Celsius degree can be downloaded from the website https://www.copernicus.eu/en/access-data, this raw data will be provided upon email request as the size of it is hundreds of gigabytes and it will take great efforts for someone to download directly from the Copernicus website. The email address is peng[dot]zhong[at]unsw[dot]edu[dot]au. The descriptions of each R data file are following:

- data/precip.RData: The precipitation data for the eight subregions, which contain R objects:
  - precip: Lists of lists, raw data of the precipitation in millimeters (mm) in 8 subregions. This is the response variable used in the marginal fit.
  - region.name: a vector, names of the 8 subregions
  - region.id: a vecotr, regional ID number that corresponding to the regional number in the shape files.
  - station: a data frame, contains geograhical information about the monitoring stations, where the precipitation data were recorded. "X" column is latitude degrees, "Y" column is the longitude in degrees, "start" column is the start measuring year, "end" is the last measureing year, "elev" is the elevation of the monitoring stations in meters, and "group.id" corresponds to the region.id for identifiying each of the 8 subregions.

  - START.date and END.date: dates, between which the data were used in our analysis.
- data/temperature.RData: The derived temperature covariate over the period 1965–2020.
  - date.df: data frame, contains the date between 1965–2020, and its corresponding season of the year.
  - temperature.covariate: list of vectors, the derived temperature covariate (30 day moving averages) used in marginal modeling and dependence modeling.
  - temperature: list of vectors,daily spatial temperature avarages in Celsius degree, used only during data preprocessing.
  - loc_df: geographical information about the locations of the ERA5-Land temperature data, used only during data preprocessing.

- data/temperature_pred.RData: The derived temperature covariate from the climate models over the period 2015–2100 under different shared socioeconomic pathways (SSP 2-4.5 or SSP 5-8.5).
  - date.245, date.585: vector of dates, corresponding to the temperature covariate from the climate models.s
  - idx.models: 5 climate models, we used the first, the 3rd, and the 4th for future projections, which are denoted by AWI, MIROC, and MPI.
  - temperature.245.avg, temperature.585.avg: lists of list, each list contains the derived temperature covariate in Celsius degree from one climate model for future projections before realign with the temperature covariate derived from the ERA5-Land data under SSP 2-4.5 (or SSP 5-8.5). SSP 2-4.5 is denoted by 245, and the same goes for SSP 5-8.5.

- data/marginal_fit_quantiles.RData: Transformed margins based on the marginal fit.
  - theretical.quantiles: list of matrixs, each matrix corresponding to the fitted marginal data based on the marginal model fit for each subregion, this is only used to generate the QQplots in the manuscript.
- data/dep.fit.boot.results3.RData: Fitted results from the bootstrap scheme for the dependence model.
  - boot.result.df: data frame, contains the 95% confidence interval (e.g, the column low.shape, high.shape) based on the bootstrap for each season, each region and each risk functional.
  - boot.result.list: lists of lists for each risk functional (first dim), each season (second dim) and each subregion (third dimension). In each list, it contains standard error of the three parameter estimates, the estimates using the whole data, and the 300 nonparametric boostrap estimates, named by jack.

- data/era5_geoinfo.RData: Shape files for the eight subregions.
  - shape1: shape file of the US river-basins
  - shape3: shape file of Danube river-basin
- data/transformed_coordinates.RData: Transformed coordinates for the eight subregions, which transformed the latitude/longitude coordinate to the Eculidean coordinate.
  - loc.trans.list: list of matrix: transfomred coordinates for each of the 8 subregions in meters.
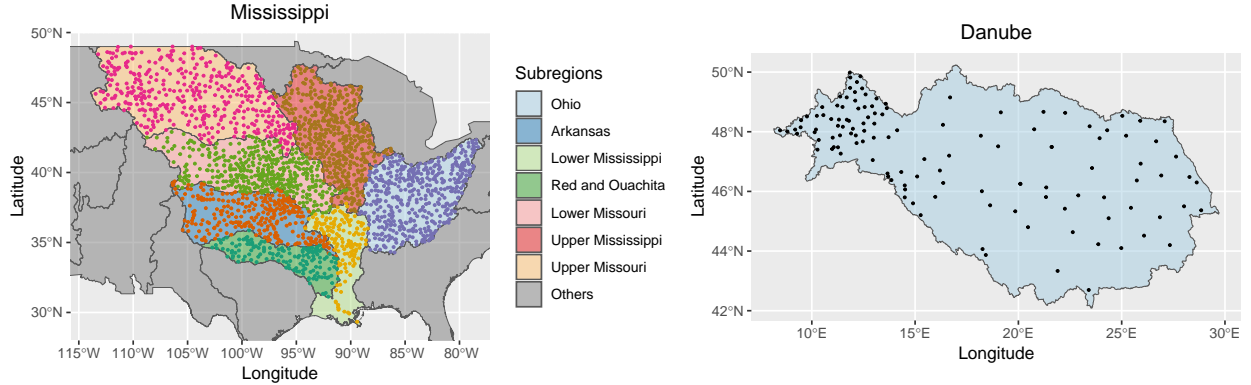
3. Fit the marginal model and the dependence model

Run the following code in your local terminal to fit the marginal model and the dependence model for the region Danube by setting `idx.region=1`. You can fit the model for other region by setting

- `idx.region=2` ⇒ Ohio region
- `idx.region=3` ⇒ Arkansas
- `idx.region=4` ⇒ Lower Mississippi
- `idx.region=5` ⇒ Red and Ouachita
- `idx.region=6` ⇒ Lower Missouri
- `idx.region=7` ⇒ Upper Mississippi

- `idx.region=8` ⇒ Upper Missouri

`bootstrap.ind=0` means we use all the data to fit the model instead of doing nonparametric bootstrap. If we set `bootstrap.ind` to any positive integer, it will use the block resampled data to implement the nonparametric bootstra and `bootstrap.ind` will serve as the index of the 300 nonparametric bootstrap fittings.

```
Rscript code/bootstrap.R "idx.region=1;bootstrap.ind=0;computer=\"local\""
```

After running the above code in shell terminal, the marginal fit results will be stored in the R data file, `/data/marginal_fit_<bootstrap.ind>_<idx.region>.RData`, and the 4 dependence fit for the four seasons will be stored in R data file, `/data/fit_<bootstrap.ind>_<idx.region>.RData`.

4. Implement nonparametric bootstrap

As this nonparametric bootstrap requires significant computing resources, we recommend the user to use HPC computing facilities to do it. One can simply load the bootstrap results from the R data file `data/dep.fit.boot.results3.RData`. If HPC is available with OpenPBS management system, one can submit the computing job via the following bash script.

```bash
#!/bin/bash

#PBS -l ncpus=16
#PBS -l mem=124gb
#PBS -l walltime=12:00:00
#PBS -j oe
#PBS -m a
#PBS -J 0-2399
#PBS -M your_email_address

cd ~/R/ExtremePrecip/ # change it to your work directory
module load r/4.3.1

region=$(( PBS_ARRAY_INDEX % 8 + 1 ))
boot_ind=$(( PBS_ARRAY_INDEX / 8 + 1))

Rscript code/bootstrap.R "idx.region=${region};bootstrap.ind=${boot_ind};computer=\"hpc\""
```

5. Plot the results