This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

The raw data set used in this paper consists of two parts; the first part contains observed daily precipitation in millimeters and daily temperature averages in degrees Celsius from 125 monitoring stations in the Danube river basin (Europe) and from 2229 monitoring stations in the Mississippi river basin (North America), which is publicly available from the Global Historical Climatology Network (GHCN) for the period 1965–2015. The second part consists of the predicted daily tempearture output from the sixth Coupled Model Intercomparison Project (CMIP6) between 2016–2100, which is publicly available on the Copernicus data base.

Availability

☒ Data **are** publicly available.
☐ Data **cannot be made** publicly available.

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

☒ Data are available online at: Global Historical Climatology Network (GHCN) and Copernicus data base. The R data files as a product of our analysis are available at the Google drive.

☐ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Description

We provide the data in R Data format, which are hosted on the Google drive as mentioned above. One should download the data first to proceed with the code accompanying this paper.

File format(s)

☐ CSV or other plain text.
☒ Software-specific binary format (.Rda, Python pickle, etc.): .RData
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (please specify):

Data dictionary

☒ Provided by authors in the following file(s):

- marginal_fit.RData
  - date.ts: vector of dates for the observed precipitation and temperatures in the period 1965–2015.

  - est.shape.gpd: estiamted shape parameter in the marginal GPD model for those two regions, danube and mississippi.
  - precip.ts.df: list of the precipitation data in the two regions.
  - station.df: information about the monitoring stations including altitude, latitude, longitude.
  - tep: a list of vectors of tempearture coviarite derived from observed tempeatures in each region.
  - U.data: pesudo-uniform scores of the precipitation data based on the marginal fit.
- marginal_model_for_danube(or mississippi).RData
  - results.bin: object of the binary GAM model in each region.
  - results.gam: object of the Gamma GAM model in each region.
  - results.gpd: object of the GPD GAM model in each region.
- result_pot_10_moving(2).Rdata
  - estimates.bootstrap.list: list of the estimates derived from the bootstrap scheme in the two regions.
  - estimates.list: list of the estimates based on the original data
  - result_pot_10_moving.Rdata contains the estimates where θ=$\hat{\xi}$.
  - result_pot_10_moving.Rdata contains the estimates where θ=1$.
- temperature_covariate.Rdata
  - data.mean.obs.hist.day: list of temperature daily averages for each region based on the observed temperatures.
  - data.mean.tep.245.day: list of temperature daily averages for each region based on the future (2016–2100) predictions from the climate model outputs under the Shared Socioeconomic Pathways (SSPs) 2-4.5.
  - data.mean.tep.hist.day: list of temperature daily averages for each region based on the historical runs of the climate model outputs
  - data.mean.tep.585.day: list of temperature daily averages for each region based on the future (2016–2100) predictions from the climate model outputs under the Shared Socioeconomic Pathways (SSPs) 5-8.5.

☐ Data file(s) is(are) self-describing (e.g., netCDF files)
☐ Available at the following URL:

Additional Information (optional)

Part 2: Code

Abstract

Description

Code format(s)

☒ Script files
  ☒ R
  ☐ Python
  ☐ Matlab
  ☐ Other:
☒ Package
  ☒ R
  ☐ Python
  ☐ MATLAB toolbox
  ☐ Other:
☐ Reproducible report
  ☐ R Markdown
  ☐ Jupyter notebook
  ☐ Other:
☐ Shell script
☐ Other (please specify):

Supporting software requirements

Version of primary software used R version 4.2.1

Libraries and dependencies used by the code

- R Packages
  - parallel version 4.2.1
  - evd version 2.3-6
  - fields version 13.3
  - mgcv version 1.8-40
  - evgam version 1.0.0
  - mvPotST (developed based on mvPot version 0.1.5)
  - lubridate version 1.8.0

Supporting system/hardware requirements (optional)

Parallelization used

☐ No parallel code used
☒ Multi-core parallelization on a single machine/node
  - Number of cores used: 28

☐ Multi-machine/multi-node parallelization
  - Number of nodes and cores used:

License

☒ MIT License (default)
☐ BSD
☐ GPL v3.0
☐ Creative Commons
☐ Other: (please specify)

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper
☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))
☒ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below:

Workflow

Location

The workflow is available:

☐ As part of the paper's supplementary material.
☒ In this Git repository: https://github.com/PangChung/SpatialScalePrecipExtremes
☐ Other (please specify):

Format(s)

- ☐ Single master code file
- ☐ Wrapper (shell) script(s)
- ☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach
- ☐ Text file (e.g., a readme-style file) that documents workflow
- ☐ Makefile
- ☒ Other (more detail in *Instructions* below)

Instructions

The main analysis are performed using the code in the file main.R. To fit the dependence model presented in the paper, we provide the R script Dependence_Model_Fit.R. The file extra_functions.R contains functions used in the file, main.R, and file, Dependence_Model_Fit.R.

The package mvPotST includes functions and methods to perform the dependence fit and run simulations. The reader should install the package, mvPotST,first in order to proceed the analysis presented in our R code.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

- ☐ $<$ 1 minute
- ☐ 1-10 minutes
- ☐ 10-60 minutes
- ☐ 1-8 hours
- ☐ $>$ 8 hours
- ☒ Not feasible to run on a desktop machine, as described here: The analysis includes performing nonparametric bootstrap to produce confidence intervals, where we used computer clusters. Moreover, To fit the marginal model as well as the dependence model, we strongly advice readers to use a workstation to run the code.

Additional information (optional)

Notes (optional)