

JASA ACS Reproducibility Initiative - Author Contributions Checklist Form

Title: Spatially-dependent multiple testing under model misspecification, with application to detection of anthropogenic influence on extreme climate events

Authors: Mark Risser, Christopher Paciorek, and Daithi Stone

Data

Abstract (Mandatory)

The data used for the analyses in Section 4 of the paper consist of simulations of precipitation and temperature over land from the Community Atmospheric Model, version 5.1, run in its ~1 degree configuration. The simulations are provided as part of the C20C+ Detection and Attribution project, which is a subproject of the World Climate Research Programme's (WCRP) Climate Variability Programme's (CLIVAR) Climate of the 20th Century Plus Project (C20C+). Full details on the simulations are available at [url blinded].

Availability (Mandatory)

The data are publicly available for download via the online data portal at [url blinded].

Description (Mandatory if data available)

The data are available at [url blinded]. No registration is required. All output from this project published online is available according to the conditions of the Creative Commons License (<https://creativecommons.org/licenses/by-nc-sa/2.0/>).

Climate simulations are aggregated monthly (from 01/1959 to 06/2015, for 678 total months) for each of two climate scenarios, with one file for each of 400 model runs, separately for surface temperature (tas) and precipitation (pr). The data from each run is on a longitude/latitude grid of size 288 x 192.

The analyses in the paper use these monthly variables aggregated over a large set of land regions (denoted WRAF05). A separate file of region weights identifies which grid cells belong in each land region.

Code is provided (see below) to download the raw data from the portal and aggregate the grid cell data to the land regions. However, this process is extremely time-consuming; therefore, we also provide a pre-processed .RData file in the Bitbucket repository [url blinded].

Optional Information (complete as necessary)

Not applicable.

Code

Abstract (Mandatory)

All of the data processing and analysis for this paper were done in R. The corresponding code is provided to download the climate model data from the C20C+ data portal (see above); aggregate the grid-cell data to land regions; conduct various pre-processing steps; simulate artificial data for use in the simulation study; fit a set of Bayesian statistical models to the simulated data via Markov chain Monte Carlo (MCMC) methods; carry out a set of multiple testing procedures conditional on MCMC posteriors; and generate descriptive plots used in the paper.

Separate code files are provided to conduct both a statistical simulation study and analyses on real data sets for the case studies.

Description (Mandatory)

All of the R scripts used in the paper are available in a public repository on Bitbucket [url blinded]. The MIT license applies to all code, and no permissions are required to access the code.

Optional Information (complete as necessary)

R version 3.3.3 (2017-03-06, “Another Canoe”) was used for the analyses in this paper. The necessary R libraries for the code used for data processing and analysis are:

- ncdf4, version 1.16 (<https://CRAN.R-project.org/package=ncdf4>)
- nimble, version 0.6-8 (<https://CRAN.R-project.org/package=nimble>)
- geoR, version 1.7-5.2 (<https://CRAN.R-project.org/package=geoR>)
- MASS, version 7.3-48 (<http://www.stats.ox.ac.uk/pub/MASS4>)
- ggplot2, version 2.1.1 (<http://ggplot2.org>)
- ggmap, version 2.6.1 (<http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>)
- ggalt, version 0.4.0 (<https://CRAN.R-project.org/package=ggalt>)
- geosphere, version 1.5-7 (<https://CRAN.R-project.org/package=geosphere>)
- colorspace, version 1.3-2 (<http://CRAN.R-project.org/package=colorspace>)
- RColorBrewer, version 1.1-2 (<https://CRAN.R-project.org/package=RColorBrewer>)
- maptools, version 0.9-2 (<https://CRAN.R-project.org/package=maptools>)
- rgeos, version 0.3-26 (<https://CRAN.R-project.org/package=rgeos>)
- plyr, version 1.8.4 (<http://www.jstatsoft.org/v40/i01/>)
- gridExtra, version 2.3 (<https://CRAN.R-project.org/package=gridExtra>)

- gdata, version 2.18.0 (<https://CRAN.R-project.org/package=gdata>)
- checkpoint, version 0.4.3 (<https://CRAN.R-project.org/package=checkpoint>)

Instructions for Use

Reproducibility (Mandatory)

All data preparation, simulation, and analyses are reproduced, as well as all Figures in the paper.

All workflow information is contained in the MASTER_reproducibility.R script. The general steps are:

1. Download and aggregate the CAM5.1 runs to the WRAF05 regions.
2. Conduct data processing/preparation for the analyses.
3. Generate simulated data.
4. Compile Bayesian models in C using nimble and run the simulation study.
5. Case studies: model fitting, summary.
6. Generate all plots in the paper draft.

Replication (Optional)

Not applicable.

Notes

Not applicable.