

Data Dictionary for /data Directory

This document provides detailed descriptions of the files included in the /data directory of the GitHub repository <https://github.com/likun-stat/XVAE>, to support reproducibility of the analyses presented in the manuscript. All files are in R's .rda binary format and are intended to be loaded using R.

example_X.rda

Type: Simulated (artificial) data

Generated by: ./Simulation Study/Autodiff_w_torch (optim).R, corresponding to Model III in the manuscript.

Contents:

- `x`: 2000×100 numeric matrix. Rows represent spatial locations; columns represent simulated time replicates.
- `stations`: 2000×2 data frame. Geographic coordinates (longitude, latitude in degrees) for each location in `x`.

copulas.rda

Type: Simulated (artificial) data for emulation comparison

Generated by: ./Simulation Study/Autodiff_w_torch_Flex_time_extGAN_comp.R, corresponding to Model III in the manuscript but with fewer locations to accommodate the limited power of extGAN.

Contents:

- `U_sim_grid`: 480×500 data frame of simulated data (uniform scale).
- `U_xvae_grid`: 480×500 data frame of XVAE-emulated data (uniform scale).
- `U_gan_grid`: 480×500 data frame of extGAN-emulated data (uniform scale).
- `stations_grid`: 2000×2 data frame. Geographic coordinates (longitude, latitude) for each location.

mon_max_allsites.rda

Type: Real observed data

Contents:

- `mon_max_allsites`: 16703×372 data frame. Rows = locations, columns = monthly max temperatures ($^{\circ}\text{C}$) over 31 years.

new_loc.rda

Type: Real observed data (location metadata)

Contents:

- `new_loc`: 16703×2 data frame. Longitude and latitude (degrees) for locations in `mon_max_allsites`.

fitted_gev_par.rda

Type: Real observed data (fitted model parameters)

Contents:

- `fitted_gev_par`: 16703×4 data frame. Each row corresponds to a location, with columns:
 1. Location: GEV location parameter ($^{\circ}\text{C}$)
 2. Scale: GEV scale parameter ($^{\circ}\text{C}$)
 3. Shape: GEV shape parameter (unitless)
 4. P-value: Chi-squared goodness-of-fit test p-value (unitless, between 0 and 1)

Notes:

- All .rda files are in R's native binary format and can be loaded using the `load()` function.
- Artificial data sets were generated independently and used for validation or emulation method comparison.
- Real datasets were used in the manuscript's application and model validation sections.