

Homework 3: Time Series and Extreme Values

BEE 4850/5850, Spring 2026

Due Date

Friday, 3/14/25, 9:00pm

Tip

To do this assignment in Julia, you can find a Jupyter notebook with an appropriate environment in [the homework's Github repository](#). Otherwise, you will be responsible for setting up an appropriate package environment in the language of your choosing. Make sure to include your name and NetID on your solution.

Overview

Instructions

The goal of this homework assignment is to practice developing and working with probability models for data.

- Problem 1 asks you to diagnose autocorrelation and develop an AR model for the weather-based variability at the Sewell's Point tide gauge.
- Problem 2 asks you to fit a sea-level rise model under both assumptions of Gaussian i.i.d. residuals and autocorrelated residuals.
- Problem 3 asks you to conduct an analysis of extreme temperatures in Ithaca under both stationary and non-stationary assumptions.
- Problem 4 asks you to model how often Ithaca temperatures exceed thresholds.

Load Environment

The following code loads the environment and makes sure all needed packages are installed. This should be at the start of most Julia scripts.

```
import Pkg
Pkg.activate(@_DIR_)
Pkg.instantiate()
```

The following packages are included in the environment (to help you find other similar packages in other languages). The code below loads these packages for use in the subsequent notebook (the desired functionality for each package is commented next to the package).

```
using Random # random number generation and seed-setting
using DataFrames # tabular data structure
using DataFramesMeta # some other dataframes interaction macros
using CSV # reads/writes .csv files
using Distributions # interface to work with probability distributions
using Plots # plotting library
using StatsBase # statistical quantities like mean, median, etc
using StatsPlots # some additional statistical plotting tools
using Optim # optimization tools
using Dates # date-time interface
```

Problems

Scoring

- Problem 1 is worth 5 points.
- Problem 2 is worth 8 points.
- Problem 3 is worth 7 points.
- Problem 4 is worth 5 points.

Problem 1 (4 points)

Tide gauge data is complicated to analyze because it is influenced by different harmonic processes (such as the linear cycle). In this problem, we will develop a model for this data using [NOAA data from the Sewell's Point tide gauge](#) outside of Norfolk, VA from `data/norfolk-hourly-surge-2015.csv`. This is hourly data (in m) from 2015 and includes both the observed data (`Verified (m)`) and the tide level predicted by NOAA's sinusoidal model for periodic variability, such as tides and other seasonal cycles (`Predicted (m)`).

Problem 1.1

Load the data file. Take the difference between the observations and the sinusoidal predictions to obtain the tide level which could be attributed to weather-related variability (since for one year sea-level rise and other factors are unlikely to matter). Plot this data. What autoregressive model order would you pick?

Problem 1.2

Fit an AR(1) model for the weather-related variability in the Norfolk tide gauge. Simulate 10,000 replications of this model, add the simulations back to the sinusoidal harmonics, and plot a histogram of the maximum and median tide levels from your simulations. How well does this model do in capturing these statistics from the observed dataset?

Problem 2

Consider the following sea-level rise model from [Grinsted et al \(2010\)](#), which models sea-level rise based on a linear relationship between global mean temperature and an “equilibrium” sea level:

$$\frac{dS}{dt} = \frac{S_{\text{eq}} - S}{\tau}$$
$$S_{\text{eq}} = aT + b,$$

where

- $S(t)$ is the global mean sea level (in mm) at time t ;
- τ is the response time of sea level (in yrs);
- S_{eq} is the equilibrium sea-level (in mm) at temperature T (in $^{\circ}\text{C}$);
- a is the sensitivity of S_{eq} to T (in mm/ $^{\circ}\text{C}$);
- b is the intercept of S_{eq} , or the S_{eq} when $T = 0^{\circ}\text{C}$ (in mm).

As discussed in class, we can use a simple numerical integration method (forward Euler integration) to convert this differential equation into a difference equation, modeling an updated sea level state $S(t + \Delta t)$ as a function of $S(t)$ and the model parameters. We will also need an additional uncertain parameter $S_0 = S(0)$ to initialize the integrated model.

Problem 2.1

Load the data from the `data/` folder and, following Grinsted et al (2010), normalize both datasets to a 1880–1900 mean (subtract that mean from the data). * Global mean temperature data from the HadCRUT 5.0.2.0 dataset (<https://hadobs.metoffice.gov.uk/hadcrut5/data/HadCRUT.5.0.2.0/download.html>) can be found in `data/HadCRUT.5.0.2.0.analysis.summary_series.global.annual.csv`. This data is averaged over the Northern and Southern Hemispheres and over the whole year. * Global mean sea level anomalies (relative to the 1990 mean global sea level) are in `data/CSIRO_Recons_gmsl_yr_2015.csv`, courtesy of CSIRO (https://www.cmar.csiro.au/sealevel/sl_data_cmar.html). The standard deviation of the estimate is also added for each year.

Plot the resulting datasets.

Problem 2.2

Write a function to simulate global mean sea levels under a set of model parameters after discretizing the equations above with a timestep of $\delta t = 1$ yr. You will need to subset the temperature data to the years where you also have sea-level data.

Fit the model under the assumption of Gaussian i.i.d. residuals (include both an uncertain model error term and the standard deviation of the observations in the probability model specification) by maximizing the likelihood. Report the parameter estimates. What can you conclude about the relationship between global mean temperature increases and global mean sea level rise rates? How appropriate was the Gaussian i.i.d. probability model for the residuals? Use any needed quantitative or qualitative assessments of goodness of fit to justify your answer. If this was not an appropriate probability model, what would you change?

Problem 2.3

Now redo Problem 2.2 but with an AR(1) model for the residuals. You can assume that the residual AR(1) model has mean zero (so $\alpha = 0$). Did your model parameter estimates change? If so, what changed about your conclusion(s) about the relationship between global mean temperature increases and global mean sea level rise?

Problem 3

Let's look at how (modeled) annual maximum temperatures have (or have not) increased in Ithaca from 1850–2014. Model output from NOAA's GFDL-ESM4 climate model (one of the models used in the latest Climate Model Intercomparison Project, [CMIP6](#)) is available in `data/gfdl-esm4-tempmax-ithaca.csv`. While this model output has not been bias-corrected,

we won't worry about that for the purposes of this assignment. The temperature data is in degrees Celsius, and the dates are provided as calendar dates.

Problem 3.1

Load the (daily maximum) temperature data from `data/gfdl-esm4-tempmax-ithaca.csv`. Find the modeled annual maximum temperatures and plot them.

Problem 3.2

Fit a stationary GEV model to the annual maximum series and report your parameter estimates. Plot the fitted distribution along with the data. What type of GEV distribution did you end up with? What is the 100-year return level?

Problem 3.3

Now fit a nonstationary GEV with respect to time (only the location parameter should be non-stationary), $y \sim \text{GEV}(\mu_0 + \mu_1 t, \sigma, \xi)$ (for scaling purposes, begin with $t = 0$ corresponding to 1850). Plot how the 100 year return levels change over time. Do you get the same return level today as in Problem 3.2?

Problem 4

Now let's analyze how often the Ithaca temperature data exceeds certain thresholds.

Problem 4.1

Suppose that we were interested in looking at temperature exceedances over 28°C. Decluster these occurrences and plot the number of exceedances by year. Have they increased over time?

Problem 4.2

Fit a stationary Poisson-GPD model for the exceedances. What does the GPD distribution look like?

Problem 4.3

From your model, estimate the 100-year return level. How does it differ from the 100-year return period from the stationary model in Problem 3? Why do you think it agrees or disagrees?

References