Deadline: 27/02/2025

1 Description of the problem

In this assignment, we will explore one of the most significant challenges facing our planet: long-term climate change. Through exploratory data analysis, we will examine climatological trends over the past century, seeking to understand the key patterns and variations that have shaped our environment. This investigation builds upon the context provided in the final project introduction and serves as our first step in analyzing these critical environmental patterns.

1.1 Data

As a first step, you will focus on the temperature data coming from GISTEMP [2]. The series provided represents changes in the global surface temperature over time. It is derived from a global network of land and ship weather stations, which track daily surface temperatures. Measurements are then adjusted for urban heat effects and other biases, aggregated monthly and averaged across stations. Lastly, since the quantity of interest is *the variation* of the global surface temperature, the final measurements are adjusted by subtracting the average global surface temperature over the period 1951-1980, which serves as a reference value.

The data set provides a reliable long-term record of temperature anomalies, offering valuable information on climate trends and variability.

The data is provided in a csv file named `gistemp.txt`. Each row refers to a calendar year (starting from 1880, up to 2024) and contains the following variables.

- 1st column: calendar year;
- 2nd to 13th: monthly temperature difference with respect to reference period;
- 14th to 15th: annual average of temperature difference taken as Jan-to-Dec (`J-D`) or Dec-to-Nov (`D-N`);
- 16th to 19th: seasonal average for winter (`DJF`), spring (`MAM`), summer (`JJA`) and autumn (`SON`).

You will need to extract the relevant variable (the monthly data) and convert it to a properly formatted `ts` object.

2 Task 1

Perform a time series decomposition on the GISTEMP data and provide the appropriate graphs to assess it. Use both `decompose` and `stl` functions. The decomposition will extract the following:

- **Trend**: Identifying long-term warming or cooling patterns.
- **Seasonality**: Recognizing periodic fluctuations.
- **Noise** (Residuals): Examining irregular variations.

Research Questions

Answer the following questions, supporting your answers with valid analysis when necessary (at your discretion).

Trend Analysis:

- What long-term trends are visible in the temperature data?
- Are there periods of acceleration or deceleration in warming?

Seasonality:

- Can you interpret the seasonal pattern?
- Is the seasonality constant or time–varying? Explain how different `R` functions tackle the problem; then try all possible corresponding decomposition and pick the one you reckon is the most suitable for the data.

Noise and Uncertainty:

- Briefly comment on the variation that is unexplained by the decomposition. Are there anomalies in the residuals?

3 Task 2

You will now focus on the trend extracted from data in section 2, specifically on the window ranging from the beginning up to December 1930. Imagine the data arrive sequentially. Fit an exponential smoothing algorithm on this subset of data.

- Comment on the value of the α parameter used in the `R` function `HoltWinters`. How is it determined? Try another small set of values, showing how the resulting fitted values change.
- In particular, plot the one-step-ahead predictions for this time interval and compare them to the observed values.
- Specify the analytic expression of the *forecast function* of the exponential smoothing algorithm for different time steps.
- Now assume that you are a statistician from December 1930 and you want to predict (through exponential smoothing) the temperature difference for the following 95 years (i.e. up to 2024). Plot the forecasts along with the observed data and compute the prediction error at 1 day, 1 year, 5 years, 10 years, 50 years and 90 years. Do you notice a pattern? Support your answer, both if your answer is positive or negative.

Guidelines

Remember that the assignments (as the final project) are also a useful exercise of **presentation**. Below you find suggestions on how your analysis should be presented (they were posted, and are still available, on BBoard).

Submission

- Report in pdf (if from rmarkdown: Knit to PDF. do not export HTML and then print)
- Code in .R or .rmd
- Name of file = group name

Length The PDF file must be no longer than **3** pages.

First page includes

- Group name
- Names of group components

Format Remember: you are supposed to send your code so the report should not include any!

- NO R console output: use tables
- NO R messages
- NO R code anywhere ever
- NO code chunks
- NO mention of the functions you use, and no explanation of your code
- Can the report be read 100% the same way if the code was not written in R? If yes, then good; if not, then make it independent of the code. The analyses and your interpretations are important, not the specifics of your code. Good code will lead to more elegant analyses, plots and overall presentation
- NO screenshots

Contents

- All models are written in formulas
- Notation is consistent

- Estimates for all unknowns are reported in tables/plots or discussed in the text and interpreted
- Model comparisons are meaningful

Plots, figures and tables

- All plots have short description/title/caption and are numbered
- All figures numbered sequentially
- All figures are mentioned in text, in the order in which they appear
- All plots have meaningful axis titles (if not redundant e.g. in the title)
- All plots are well positioned in the page (centered)
- All text in the plots is readable without zooming in
- No text is too big in the plot
- Plots are not ”warped”
- No plot is pixelated or blurry or with jpeg artifacts
- All plots are useful for the purpose of answering the research question
- All plots are explained and interpreted (not just described passively)
- Report quantities with names meaningful to the application and not with generic ones (e.g., State1, State2, ecc)

General

- Spacing is used efficiently: no excessive white spaces
- Borders are normal, line spacing is standard, no other weirdness to fit everything within the page limit
- English: spelling mistakes? Too verbose? Concise enough? We’re not the British Council but you don’t want to be sloppy.
- Report does not look hastily made or sloppy
- Report looks professional
- Text is concise and to the point

Code: we will randomly pick some groups for a code check. Or, we may check the code when figures or values look funny (as it happens).

- Submitted code can be compiled/run without error generating all figures and tables in the report, with the same numbers
- Code is easily readable and it is possible for anybody to understand what is going on
- Variables are named to improve readability (i.e. avoid calling things "a1" "x9534", "asdfa", but rather use names such as "user_speed", "daily_price", "log_returns".
- The code would work with minor modification on different data

Riferimenti bibliografici

[1] Menne, M.J., I. Durre, B. Korzeniewski, S. McNeill, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E.Gleason, and T.G. Houston, 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.31. NOAA National Climatic Data Center. Dataset accessed 2025-02-03 at http://doi.org/10.7289/V5D21VHZ .

[2] GISTEMP Team, 2025: GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 2025-02-03 at https://data.giss.nasa.gov/gistemp/.

[3] Lenssen, N., G.A. Schmidt, M. Hendrickson, P. Jacobs, M. Menne, and R. Ruedy, 2024: A GISTEMPv4 observational uncertainty ensemble. J. Geophys. Res. Atmos., 129, no. 17, e2023JD040179, doi:10.1029/2023JD040179.