Assignment 1

Reconstructing the past: estimating solar radio flux F10.7 from sunspot records

Performance - Tuesday, September 30, 2025

Due to submit a performance report – Monday, October 6, 2025, 23:59 p.m.

The aim of this assignment is to investigate the connection between two fundamental indicators of solar activity - the sunspot number and the solar radio flux F10.7 (2800 MHz), and to learn how past radio flux can be reconstructed from historical sunspot observations. Using 3d order linear regression and the Least-Squares Method (LSM), you will smooth and analyze the data, build a regression model, and apply it to extend radio flux records backward in time. Such reconstructions are essential for developing long-term prediction models of solar activity, a task of high importance for space agencies in forecasting long-term space weather impacts.

A brief overview of sunspot numbers and solar radio flux at 10.7 cm

Sunspot numbers and solar radio flux at 10.7 cm are key indicators of solar activity. Sunspot numbers are measured daily by observatories worldwide and calculated using the formula $R = k(n + 10g)$, where n is the number of observed sunspots, g is the number of sunspot groups, and k is the telescope's calibration coefficient.

Solar radio flux at 10.7 cm represents the radio emission at a wavelength of 10.7 cm (2800 MHz) from all sources on the solar disk, including sunspots, solar flares, solar proton events, and white light faculae fields.

Here is the recommended procedure:

1. The code (Python/Matlab) should be well-commented and include the following:
   - Title of the assignment, e.g.,
     ```
     %Reconstructing the past: estimating solar radio flux F10.7 from
     sunspot records
     ```
   - Team members' names, Skoltech affiliation, and the date, e.g.,
     ```
     % Tatiana Podladchikova, Shantanu Jain, Zaina Abu-Shaar, Skoltech, 2025
     ```
   - Comments for main procedures, e.g.,
     ```
     % 13-month running mean
     ```
     …followed by the code.

2. Download the monthly mean sunspot numbers and solar radio flux F10.7 cm measurements from Canvas. Each team will have different files to work with.

   ```
   Radio_flux_monthly_mean.txt
   Sunspot_number_monthly_mean.txt
   ```

   Format of data:

   ```
   Radio_flux_monthly_mean.txt
   1 column - year
   2 column – month
   3 column – monthly solar radio flux at 10.7 cm

   Sunspot_number_monthly_mean.txt
   1 column - year
   2 column – month
   3 column – monthly sunspot number
   ```

3. Plot the monthly mean sunspot numbers and solar radio flux F10.7 cm for visual representation at one plot. Each plot should include a title, labels for the x-axis and y-axis,

and a legend for the lines. Comment on similarities and differences in the sunspot numbers and F10.7 dynamics.

4. Apply a 13-month running mean to smooth the monthly mean data for sunspot numbers and solar radio flux F10.7. Smoothing the data before applying linear regression helps reduce noise and variability, allowing for a clearer relationship to emerge between the variables. Please note that the use of built-in moving mean functions is not allowed for learning purposes. You must write your own code. Then, plot the results.

   13-month running mean $\overline{\boldsymbol{R}}$

$$\bar{R} = \frac{1}{24}R_{i-6} + \frac{1}{12}(R_{i-5} + R_{i-4} + \cdots + R_{i-1} + R_i + R_{i+1} + \cdots + R_{i+5}) + \frac{1}{24}R_{i+6}$$

   Comment:
   For the first six months of available data, the values are averaged to obtain the smoothed estimates. The same applies to the last six months of data.

5. Visual analysis is important for understanding data relationships. Identify the time interval where both sunspot numbers and F10.7 measurements overlap and create **a scatter plot** of the smoothed sunspot and F10.7 data. Each plot should include a title, labels for the x-axis and y-axis, and a legend for the lines. Conclude whether a clear relationship between the solar activity indicators exists.

6. **Construction of multi-dimensional linear regression.**

   - The time interval you identified in the previous step will serve as training data for building your regression model.

   Multi-dimensional linear regression model:

$$F_i = \beta_0 + \beta_1 R_i + \beta_2 R_i^2 + \beta_3 R_i^3 + \varepsilon_i \qquad (1)$$

   Where,

   - F is vector of dependent variables (regressand), representing the smoothed solar radio flux at 10.7 cm.

$$F = \begin{vmatrix} f_1 \\ f_2 \\ \dots \\ f_N \end{vmatrix}$$

     and $f_1,\ f_{2,\dots,}f_N$ is flux at different times.

   - R is matrix of independent variables (regressors)

$$R = \begin{vmatrix} 1 & r_1 & r_1^2 & r_1^3 \\ 1 & r_2 & r_2^2 & r_2^3 \\ \dots & \dots & \dots & \dots \\ 1 & r_N & r_N^2 & r_N^3 \end{vmatrix}$$

     and $r_1, r_2, \dots, r_N$ are the smoothed sunspot numbers at different times.

- β is vector of coefficients to be estimated

$$\beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{vmatrix}$$

- ε_i is error term

7. Define the vector of regressands (dependent variables), the matrix of regressors (independent variables), and the vector of coefficients in the code using the radio flux and sunspot data.

8. Determine the vector of coefficients using the Least Squares Method (LSM). The vector of coefficients is calculated according to the following equation

$$\beta = (R^T R)^{-1} R^T F$$

9. Reconstruct the smoothed solar radio flux at 10.7 cm using the smoothed sunspot numbers and the estimated regression coefficients based on Equation (1)

$$F_i = \beta_0 + \beta_1 R_i + \beta_2 R_i^2 + \beta_3 R_i^3$$

10. Determine the root-mean square error (RMSE) of the reconstructed solar radio flux at 10.7 cm.

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (f_i - \hat{f}_i)^2}$$

    Please use the equation provided above to calculate the RMSE. For learning purposes, note that the use of built-in functions is not allowed.

11. Apply your model to earlier periods where only sunspot numbers are available. Add error bounds estimated as $\pm 3\sigma$, when extending reconstructions backward in time. Plot the reconstructed flux together with error ranges (e.g. shaded areas, error bars, etc).

12. **Make conclusions about the Assignment.**
    Conclusions should be presented in the form of a learning log. Learning log should capture the key takeaways from the assignment. The learning log is not simply a diary of tasks completed but a reflective journal showcasing your personal learning, experiments, and critical insights gained throughout the process.

13. **Prepare a performance report and submit to Canvas:**
    Each assignment will be graded on a 0-100% scale and should be submitted as a PDF report. Any accompanying code in Python or Matlab must also be converted to PDF.

    The breakdown is as follows:

- ✓ 50%: Successful completion of all steps, thoroughly documented in the report.
- ✓ 20%: Clear presentation—figures must include captions, labeled axes, legends for curves, and concise conclusions or discussions.
- ✓ 30%: Learning log.
- ✓ **Quizzes.** Regular written quizzes will be held throughout the course, covering key concepts and specific assignments discussed in class. Laptops and phones will not be allowed during quizzes, so please review the relevant materials and complete assigned tasks in advance. Your performance in these quizzes will contribute to your assignment grade.