Zurich University of Applied Sciences
Institute of Computational Life Sciences

Optimisation

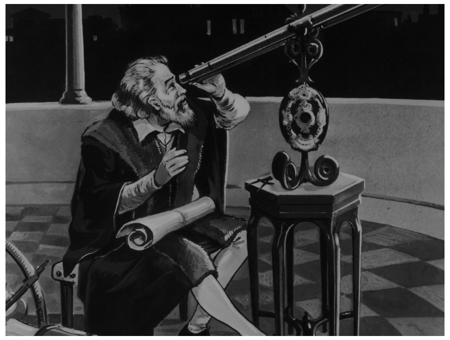
Module Optimisation and High Performance Computing

Introduction to final project

A bit of context

The magnetic activity of the sun

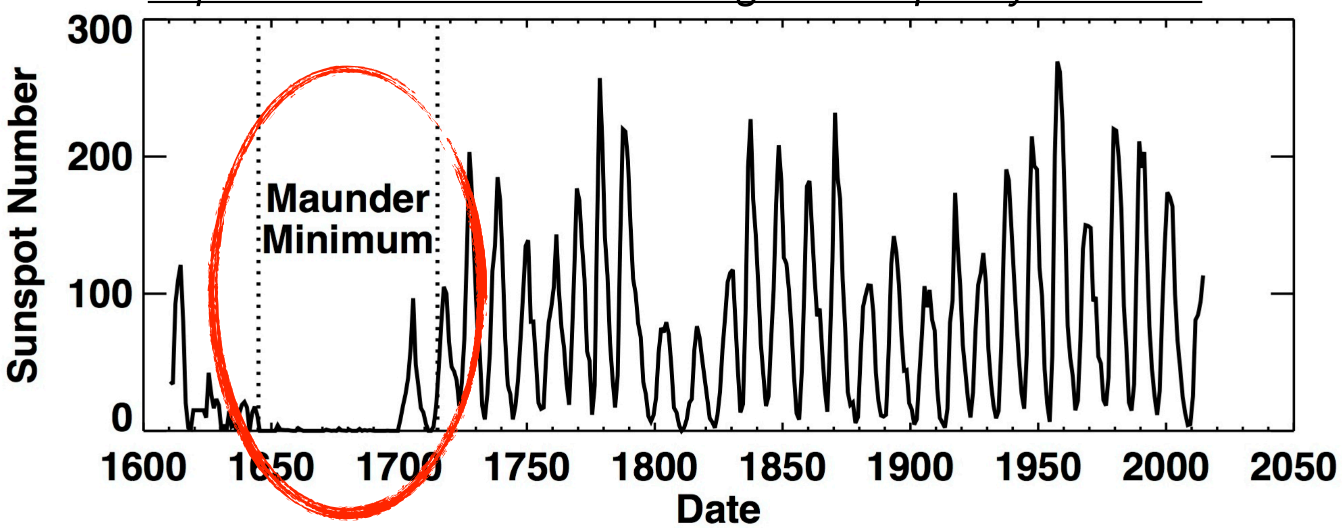
SUNSPOTS provide the longest **DIRECT record of solar magnetic activity**.

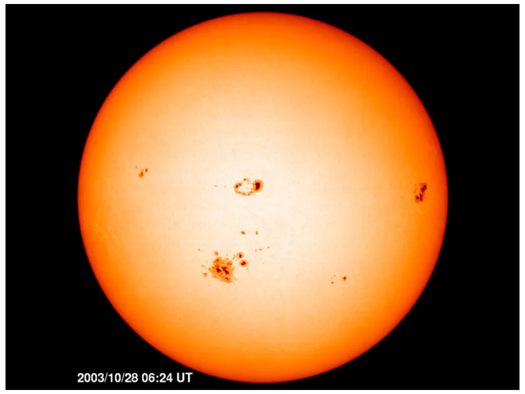


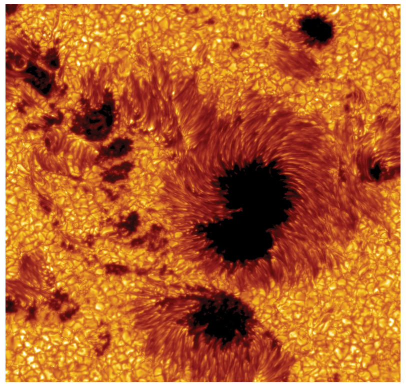
In 1610, shortly after viewing the sun with his new telescope, Galileo Galilei made the first European observations of **sunspots**.

Continuous daily observations were started at the Zurich Observatory in 1849 and earlier observations have been used to extend the official record back to 1610.

https://solarscience.msfc.nasa.gov/SunspotCycle.shtml







- Well-known **11-year cycle**, **long-period modulations** and **Grand Minima.**
- Many physics mechanisms underlying such phenomena are still **fundamentally obscure**.
- Understanding such mechanisms may boost our predictive capabilities of the solar activity, which could be of great practical importance considering the **impact that the sun can have on the Earth climate**.

The magnetic activity of the sun

Active regions are areas on the Sun with very strong magnetic fields (several thousands times stronger than the Earth's magnetic field).

Sunspots ("black" spots on the surface of the Sun) mark the places where the magnetic field is strong enough to hold back the plasma flows that transport heat and radiation from the interior.

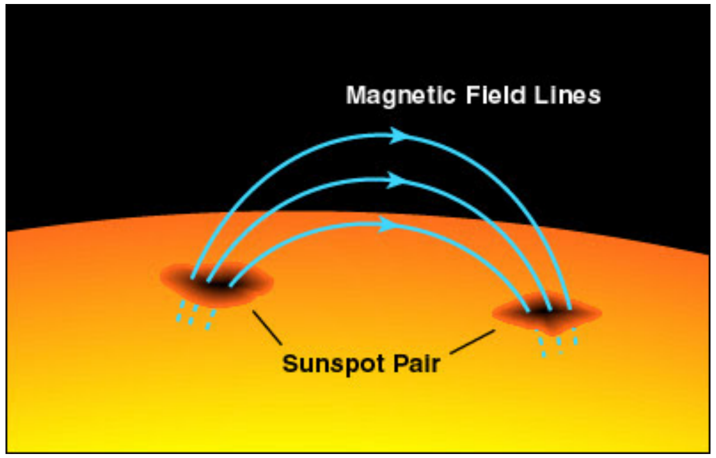


Table 2: Dates and values for sunspot cycle minima. The value is always the value of the 13-month mean of the International Sunspot Number. The dates differ according to the indicator used.

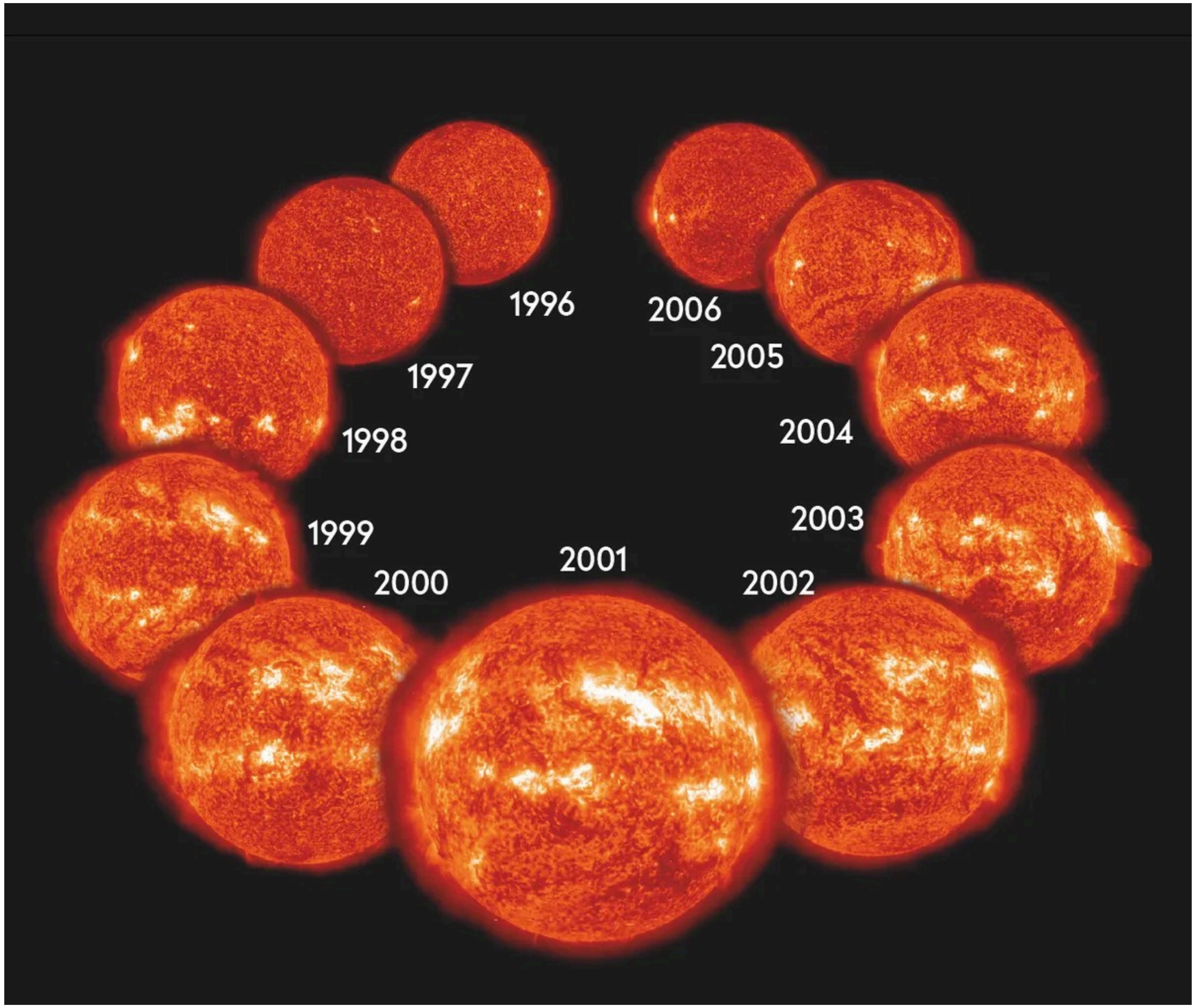
Cycle	13-month Mean Minimum Date	13-month Mean Minimum Value	Waldmeier/ McKinnon Date	Spotless Days Maximum Date	New > Old Date	Average Date
1	1755/02	8.4	1755.2			1755/02
2	1766/06	11.2	1766.5			1766/06
3	1775/06	7.2	1775.5			1775/06
4	1784/09	9.5	1784.7			1784/09
5	1798/04	3.2	1798.3			1798/04
6	1810/08	0.0	1810.6			1810/08
7	1823/05	0.1	1823.3	1823/02		1823/04
8	1833/11	7.3	1833.9	1833/11		1833/11
9	1843/07	10.6	1843.5	1843/07		1843/07
10	1855/12	3.2	1856.0	1855/12		1855/12
11	1867/03	5.2	1867.2	1867/05		1867/04
12	1878/12	2.2	1878.9	1878/10	1879/01	1878/12
13	1890/03	5.0	1889.6	1890/02	1889/09	1890/01
14	1902/01	2.7	1901.7	1902/01	1901/11	1901/12
15	1913/07	1.5	1913.6	1913/08	1913/04	1913/06
16	1923/08	5.6	1923.6	1923/10	1923/09	1923/09
17	1933/09	3.5	1933.8	1933/09	1933/11	1933/10
18	1944/02	7.7	1944.2	1944/02	1944/03	1944/02
19	1954/04	3.4	1954.3	1954/04	1954/04	1954/04
20	1964/10	9.6	1964.9	1964/11	1964/08	1964/10
21	1976/03	12.2	1976.5	1975/09	1976/08	1976/03
22	1986/09	12.3		1986/03	1986/10	1986/07
23	1996/05	8.0		1996/07	1996/12	1996/08
24	2008/12	1.7		2008/12	2008/09	2008/11

We are now in Solar Cycle 25 with peak sunspot activity expected in 2025

Solar Cycle 24 had the 4th-smallest intensity since regular record keeping began with Solar Cycle 1 in 1755

D. H. Hathaway, The Solar Cycle, Living Rev. Solar Phys., 12, 2015, https://doi.org/10.1007/lrsp-2015-4

The magnetic activity of the sun



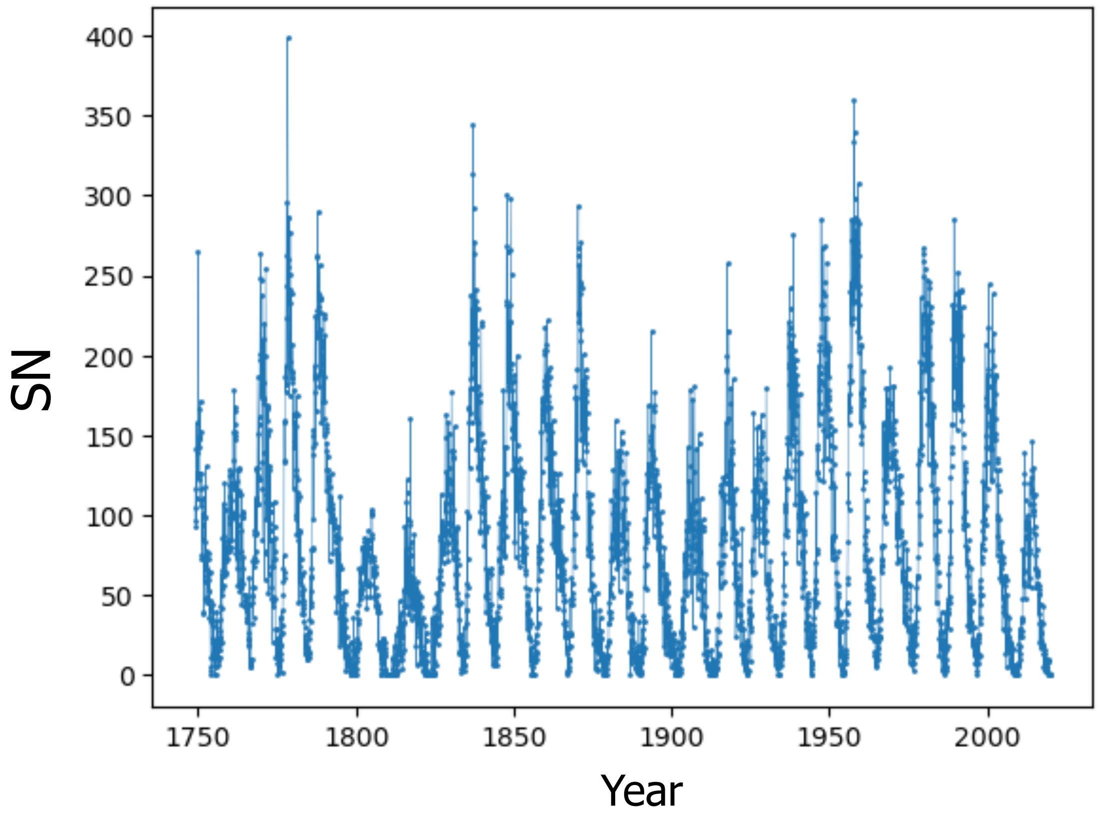
Cycle 23

Now, practically, what should you do?

- You will study the **Solar Cycle**,
- optimise a model describing the Cycle using **sunspots data**,
- derive Solar Cycle parameters and study their correlation,
- compare your results with those reported in the astrophysics literature

Sunspots data

- You will use the official Sunspot Number (SN) record:



D. H. Hathaway, The Solar Cycle, Living Rev. Solar Phys., 12, 2015, https://doi.org/10.1007/lrsp-2015-4

Sunspots data

- You will use the official Sunspot Number (SN) record:

- Each group will be assigned (randomly) a dataset corresponding to **TEN** cycles (about **40'000 points**, with **daily** resolution).

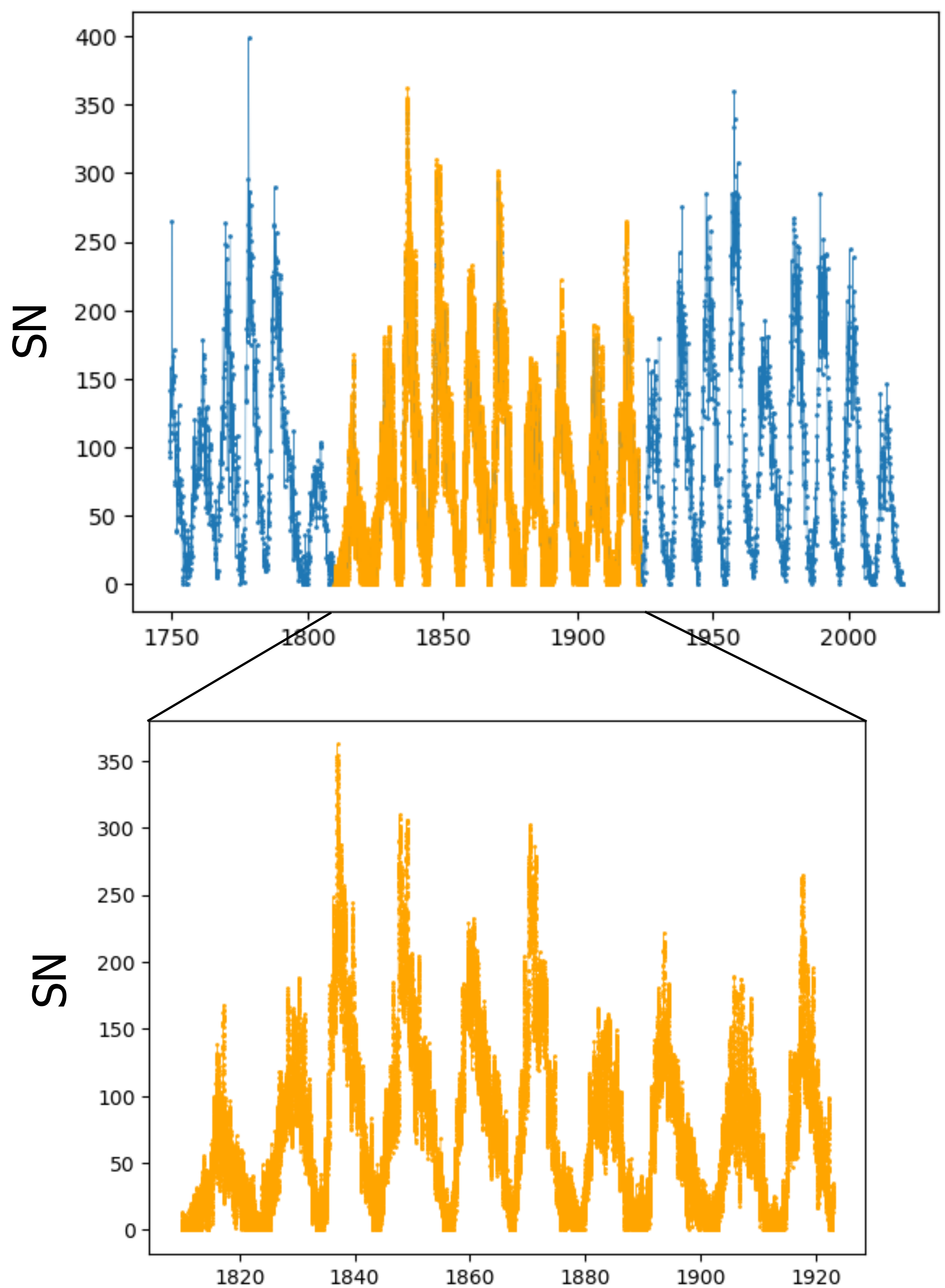
- You will implement a **Simulated Annealing (SA)** algorithm in Python.

  IMPORTANT: the project must be carried out using **your own SA function**, **NOT** a Python library.

- You will carry out **TWO tasks**:

1. **Optimise SA hyper parameters**. Run SA for different values of the algorithm hyper parameters, namely, $T0$ and σ, and identify their best values.

2. **Calibrate (= optimise) model parameters** using your SA algorithm and the assigned SN dataset.



D. H. Hathaway, The Solar Cycle, Living Rev. Solar Phys., 12, 2015, https://doi.org/10.1007/lrsp-2015-4

Model and Loss functions

Model function: The shape of **each** solar cycle is described by a unique functional form:

$$x_k(t) = \left(\frac{t - T0_k}{Ts_k}\right)^2 e^{-\left(\frac{t-T0_k}{Td_k}\right)^2}$$

where the index k refers to the cycle number, the parameters Ts_k and Td_k define its amplitude and rising time, and $T0_k$ is its initial time (according to Hathaway, 2015).

D. M. Volobuev, The Shape of the Sunspots Cycle: A One-Parameter Fit, Solar Physics, 258, 2009

So, for 10 cycles: $$x(t) = \sum_{k=1}^{10} x_k(t) = \sum_{k=1}^{10} \left(\frac{t - T0_k}{Ts_k}\right)^2 e^{-\left(\frac{t-T0_k}{Td_k}\right)^2}$$

$\longrightarrow$ **30 free parameters to be optimised!** $T0_1, Ts_1, Td_1, T0_2, Ts_2, Td_2, \ldots T0_{10}, Ts_{10}, Td_{10}$

Loss function (to be minimised): **MSE**

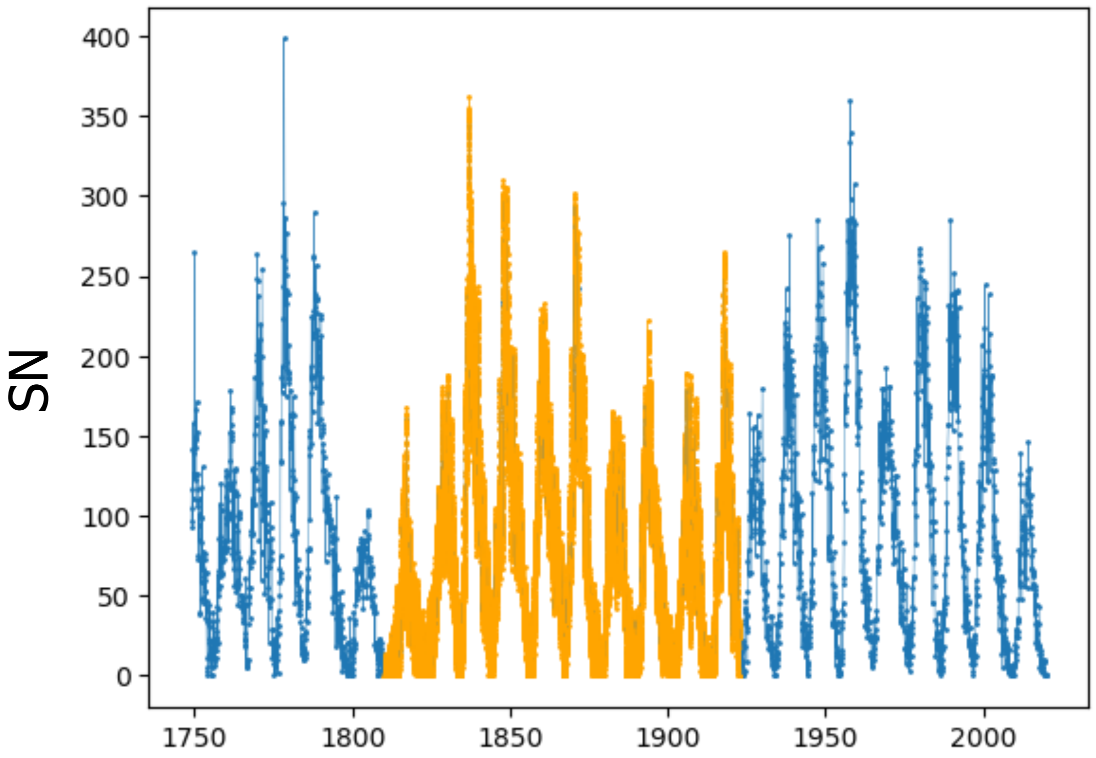
$$\frac{1}{N}\sum_{i=1}^{N} \epsilon_i^2 = \frac{1}{N}\sum_{i=1}^{N} \left(y_i - x(t_i)\right)^2$$

where N is the number of points in your dataset ($\approx 40'000$), y_i is a data point at time t_i and $x(t_i)$ the model prediction.

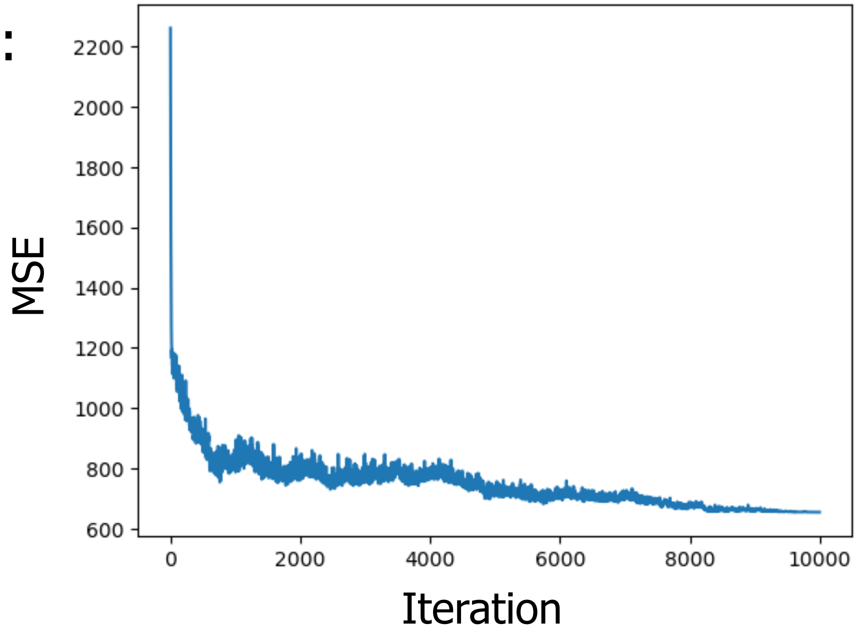
Hyper parameters optimisation

1. **Optimise SA hyper parameters**.

- Run SA for different values of the algorithm hyper parameters, namely, $T0$ and σ, and identify their **best** values.
- This task must be **parallelised on the cluster**.
- This is a **very high-dimensional problem**, and it is **crucial** to choose good starting values for the problem variables. As initial conditions for the starting times of the cycles, $T0_k$, use the times reported in Hathaway 2015. As initial conditions for the other parameters, Ts_k and Td_k, you can use 0.3 and 5, respectively.
- For each run (= each set of hyper parameters) plot the evolution of the loss function, and look at its final value. Compare the different results. We will learn how to choose the "best" hyper parameters.



Example:



Model calibration (optimisation)

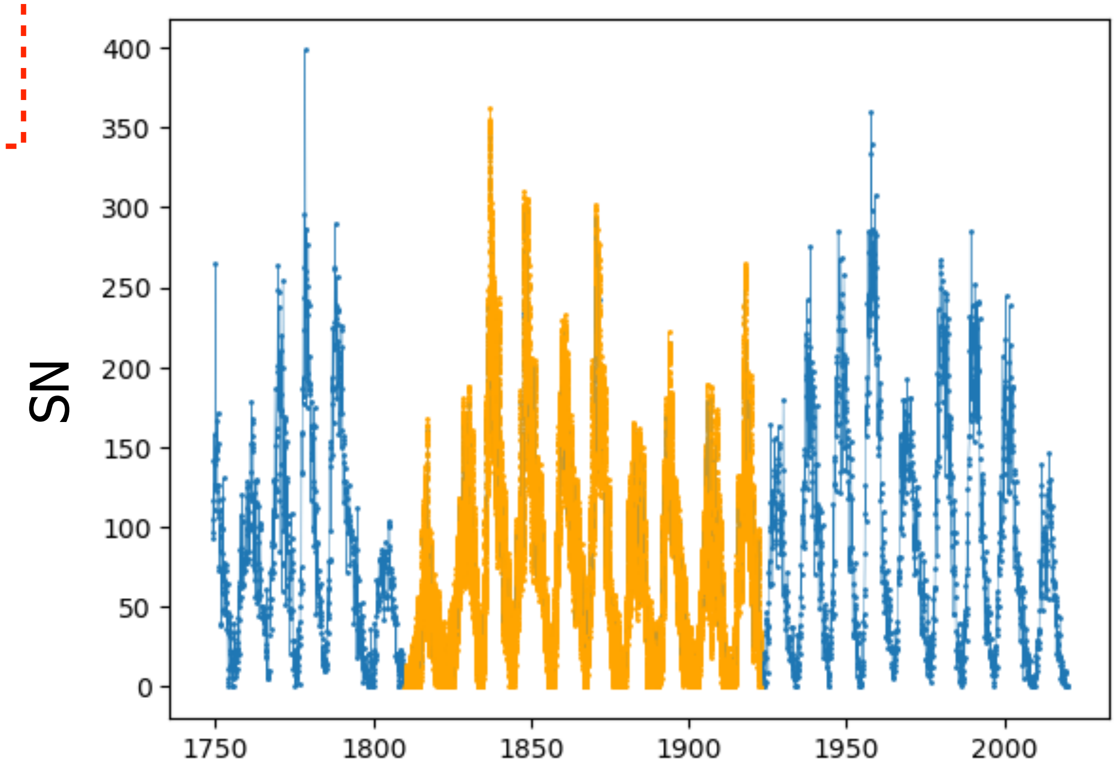
2. **Calibrate (= optimise) model parameters** using your SA algorithm and the assigned SN dataset.

- Using the best hyper parameters $T0$ and σ from task 1, run multiple independent SA optimisations (e.g., 10) using (slightly) different initial conditions.

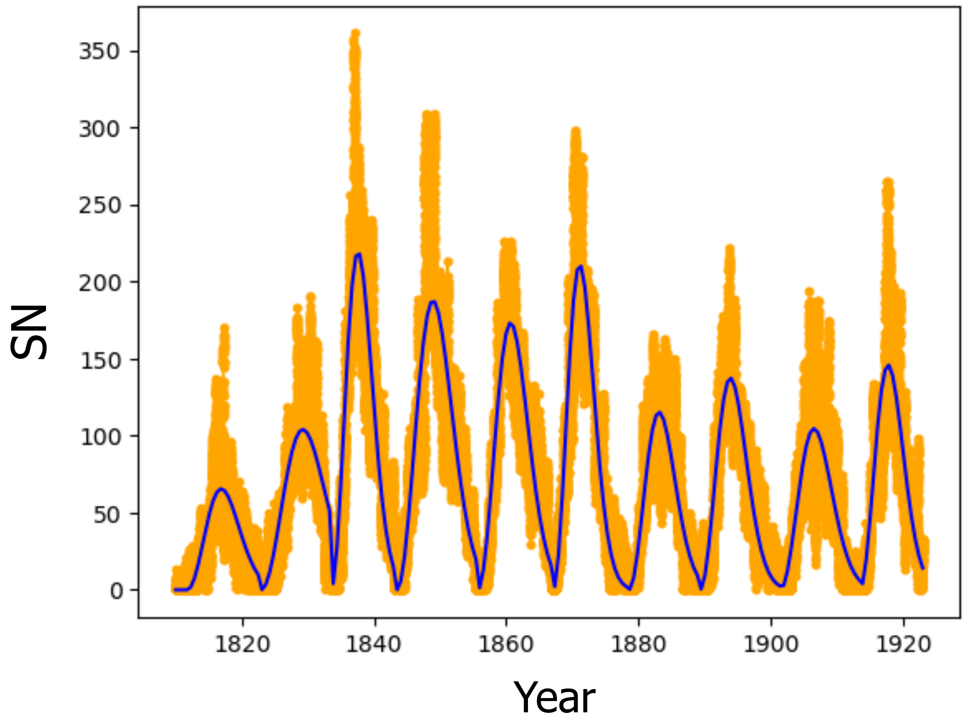
- As initial condition, you can use the result of the best run of task 1 and add a little bit of noise on top of it.

- This task must be **parallelised on the cluster**.

- At the end, collect the results of the independent runs into one numpy `array` and combine them calculating for each variable its Center of Mass. **This will be your best estimation for the 30 model parameters!**

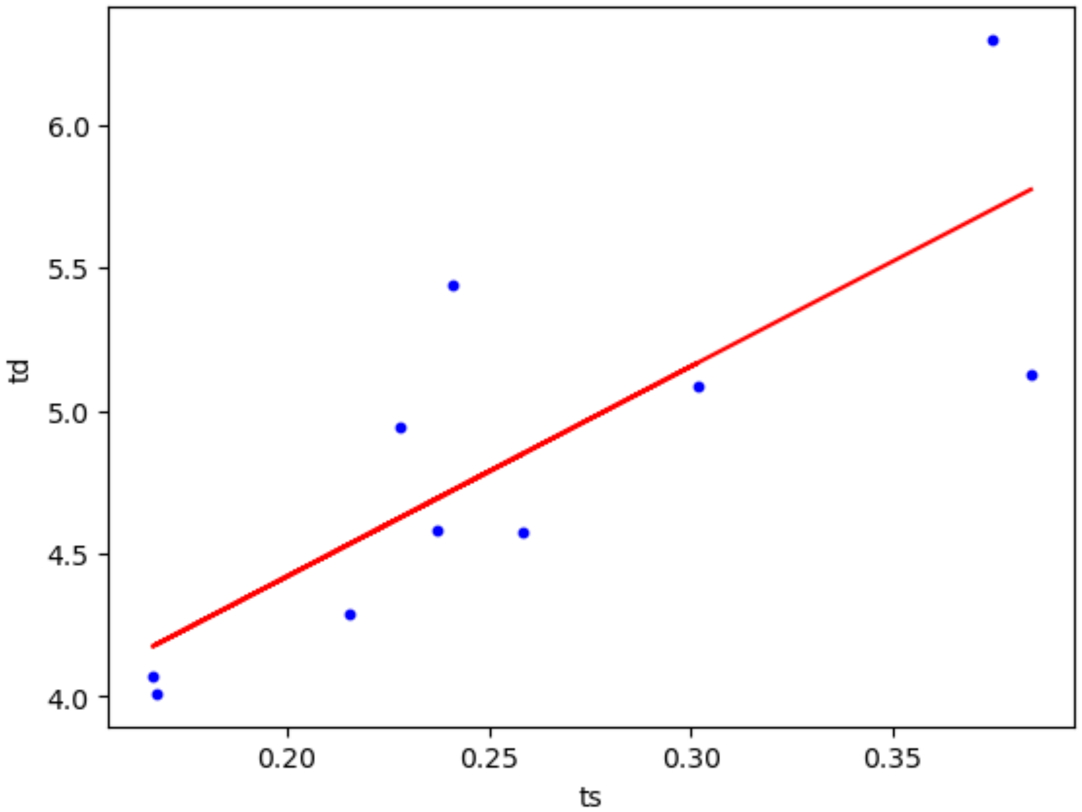


You should obtain a fit like this:



Final Fit and Linear Regression

- Study the **linear correlation** between Ts_k and Td_k



Equation (4) in [Penza et al. 2024]

$$\boxed{Td_k = s_1 \cdot Ts_k + s_2}$$

$$\longrightarrow \quad s_1 = 0.02 \pm 0.01 \text{ year}$$

- Carry out your own liner fit:

```
slope, intercept = np.polyfit(ts, td, 1)  # 1 means linear fit
td_fit = slope * ts + intercept
```

- What value do you obtain for the slope s_1? Is it in agreement with the literature?

N.B.: the value 0.02 reported in the astrophysics literature is in 'years'. Your dataset has a daily resolution, therefore your slope is likely to be in units of 'days'. Remember to convert your result in the correct units!

Practical details

- Deliverables:
  1. Presentation slides (Moodle)
  2. Presentation video (TBD)
  3. Code (Moodle)

- Provided material:
  1. Datasets
  2. Project description (this presentation)
  3. Outline for your presentation

- Recommendations:
  1. Use GIT in your team, e.g., on ZHAW GitHub: https://github.zhaw.ch
  2. Take notes during your project work regarding what you want to present at the end.
  3. Create a Conda environment and share the environment file with your team members, e.g., via your Git repository.

Final presentation

- It should be **15-20** minutes long.
- Every team members must present a part of the project. You are free to decide how to split the tasks, but **everyone must speak** during the presentation.
- In the following slide we provide an outline for your presentation. **All points must be addressed**.

Outline for the presentation

1. **Introduction**
   - Briefly introduce the team members
   - Provide an overview of the optimisation problem (scientific background, available data, goals)
2. **Describe the optimisation problem**
   - Describe the optimisation problem: objective function, model, hyper parameters, problem variables (to be optimised), optimisation algorithm
   - Explain any constraints or limitations
   - Outline your initial solution (non-parallel)
3. **Parallel programming techniques**
   - Discuss parallelisation opportunities in the problem
   - Explain the parallelisation techniques that you have used (e.g., data partitioning, multiprocessing, MPI, …)
   - Explain why you chose such approach(es)
   - Highlight any challenges faced during parallelisation

Outline for the presentation

4. **Results and Performance Metrics**
   - Compare performances of the initial (non-parallel) solution and the parallelised version
   - Give quantitative results (wall time)

5. **Conclusions**
   - Present your conclusions regarding the scientific problem and briefly discuss them. E.g., quality of the fit, linear correlation fo the parameters, agreement with literature.

6. **Lessons learned and personal experience**
   - Share insights gained during the project
   - Discuss teamwork dynamics, divisions of tasks
   - Mention any unexpected challenges

7. **Wrap-up**
   - Final considerations

Evaluation

- Quality of the fit
- Code quality (understandable for others)
- Overall runtime of the Simulated Annealing model calibration part (**task 2**)

  Use **ONE compute node**, **32 cores**, **partition Earth-3**
- Quality and completeness of the presentation and slides

Good luck!

… and don’t worry, we are here to help