

# Mini-Project (ML for Time Series) - MVA 2023/2024

Firstname Lastname [youremail1@mail.com](mailto:youremail1@mail.com)

Firstname Lastname [youremail2@mail.com](mailto:youremail2@mail.com)

January 5, 2024

**What is expected for these mini-projects?** The goal of the exercise is to read (and understand) a research article, implement it (or find an implementation), test it on real data and comment on the results obtained. Depending on the articles, the task will not always be the same: some articles are more theoretical or complex, others are in the direct line of the course, etc... It is therefore important to balance the exercise according to the article. For example, if you have reused an existing implementation, it is obvious that you will have to develop in a more detailed way the analysis of the results, the influence of the parameters etc... Do not hesitate to contact us by email if you wish to be guided.

**The report** The report must be at most FIVE pages and use this template (excluding references). If needed, additional images and tables can be put in Appendix, but must be discussed in the main document. The report must contain a precise description of the work done, a description of the method, and the results of your tests. Please do not include source code! The report must clearly show the elements that you have done yourself and those that you have reused only, as well as the distribution of tasks within the team (see detailed plan below.)

**The source code** In addition to this report, you will have to send us a Python notebook allowing to launch the code and to test it on data. For the data, you can find it on standard sites like Kaggle, or the site <https://timeseriesclassification.com/> which contains a lot of signals!

**The oral presentations** They will last 10 minutes followed by 5 minutes of questions. The plan of the defense is the same as the one of the report: presentation of the work done, description of the method and analysis of the results.

**Deadlines** Two sessions will be available :

- **Session 1**
  - Deadline for report: December 18th (23:59)
  - Oral presentations: December 20th and 22th (precise times TBA)
- **Session 2**
  - Deadline for report: January 9th (23:59)
  - Oral presentations: January, 11th and 12th (precise times TBA)

# 1 Introduction and contributions

The Introduction section (indicative length : less than 1 page) should detail the scientific context of the article you chose, as well as the task that you want to solve (especially if you apply it on novel data). **The last paragraph of the introduction must contain the following information:**

- Repartition of work between the two students

Esteban : Inpainting, outlier detection and ringing removal.

Alexi : Detrending, rereferencing, step removal.

- Use of available source code or not, percentage of the source code that has been reused, etc.

The source code available is in matlab. This is a shitty language and releasing a paper with only a matlab implementation is equivalent to not releasing source code. Furthermore, it was poorly written and unreadable without any damn documentation. We only looked at the code to get the details missing in the paper because the paper lacks details. (freaking learn to write dammit). AND FOR GOD'S SAKE  $d(t) \neq d$ . Writing things like  $d(t)/std(d(t))$  makes you look like a five years old.

- Use of existing experiments or new experiments (e.g. test of the influence of parameter that was not conducted in the original article, application of the method on a novel task/data set etc.)

TODO

- Improvement on the original method (e.g. new pre/post processing steps, grid search for optimal parameters etc.)

TODO

## 2 Method

The Method section (indicative length : 1 to 2 pages) should describe the mathematical aspects of the method in a summarized manner. Only the main steps that are useful for understanding should be highlighted. If relevant, some details on implementation can be provided (but only marginally).

A subsection for each method. Short paragraph to introduce it and then pseudo code + one small figure as exemple. We each do the methods we implemented.

### 2.1 Robust detrending

---

```
1 detrending(signal):  
2     weights = [1, 1, ..., 1] # weights for outlier detection (0 means outlier)  
3     iterate  $n_{\text{iter}}$  times:
```

```

4     projected_signal = fit_to_basis_using_weights(signal, weights)
5     error_on_projection = abs(signal - projected_signal)
6     for t in len(signal):
7         if error_on_projection[t]/std(error_on_projection) > threshold:
8             weights[t] = 0
9         else:
10            weights[t] = 1
11     return signal - fitted_signal # detrended signal

```

---

## 2.2 Inpainting

## 2.3 Outlier detection

## 2.4 Robust rereferencing

```

1 rereferencing(signal):
2     weights = 1 - outlier_detection(signal) # 1 : valid data / 0 : outlier
3     weighted_mean = mean_using_weights(signal, weights)
4     return signal - weighted_mean

```

---

## 2.5 Step removal

```

1 # Recursively look for steps in the signal
2 step_detection(signal, depth)
3      $\forall(t, T) \ M[t][T] = \text{mean}(x[t:T+1])$ 
4      $\forall(t, T) \ V[t][T] = \sum_{i=t}^T (x_i - M_t^T)^2$ 
5
6     t0 = argmin(V[1][t] + V[t][T], t) # most likely position to be a step
7     if not(is_step(signal, t0)): # Check if t0 is indeed the position of a step
8         return [] # no steps
9
10    # Check for steps on both sides of t0
11    steps_left = step_detection(signal[:t0], depth - 1)
12    steps_right = t0 + step_detection(signal[t0:], depth - 1) # + t0 because
    ↪ signal si shifted
13
14    return concatenate(steps_left, [t0], steps_right)
15
16 step_removal(signal, depth):
17     steps = step_detection(signal, depth)
18
19     for step in steps:
20         signal[steps:] -= difference_mean_before_and_after_step(signal, step)
21
22     return signal # Without steps

```

---

## 2.6 Ringing removal

## 3 Data

The Data section (indicative length : 1 page) should provide a deep analysis of the data used for experiment. In particular, we are interested here in your capacity to provide relevant and thoughtful feedbacks on the data and to demonstrate that you master some "data diagnosis" tools that have been dealt with in the lectures/tutorials.

TODO. At least use data from the original paper

## 4 Results

The Result section (indicative length : 1 to 2 pages) should display numerical simulations on real data. If you re-used some existing implementations, it is expected that this section develops new experiments that were not present in the original article. Results should be discussed not only based on quantitative scores but also on qualitative aspects. In particular (especially if your article focuses on black box methods), please provide some feedbacks whether the method was adapted to the data or not and whether the hypothesis behind the approach you used were validated or not.

TODO.