How I Detect and Remove Artifacts in EEG Signals

Steven Cao

11/27/2022

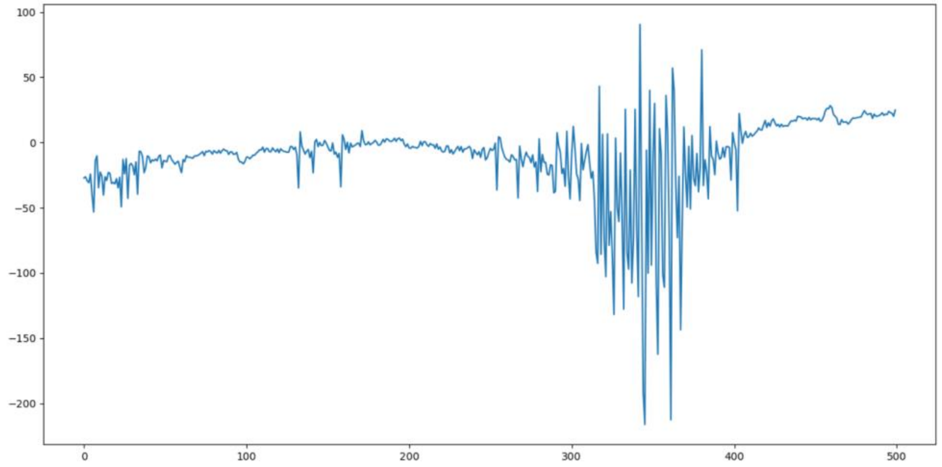
What is an artifact in EEG?

Definition

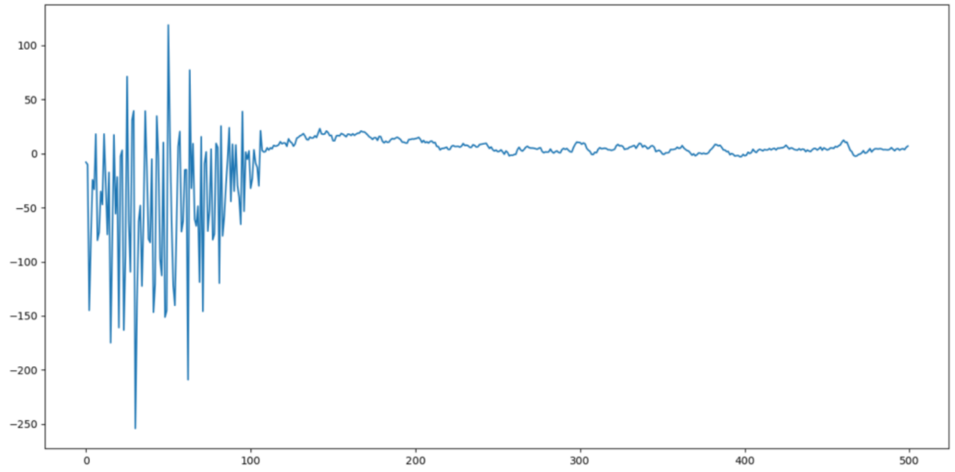
- Artifacts are <u>**signals recorded**</u> by EEG that are <u>**not generated**</u> by the brain

Examples of what can cause artifacts

- Body movement
  - For example, <u>**arm movement**</u>
- Electrodes <u>**detaching**</u> from the scalp
- Electrodes <u>**not fully connected**</u> to the person
- <u>**Sweating**</u>, which can create electrical disturbances by short-circuiting electrode pairs
  - Causes abnormal connections between circuit nodes, since impurities in sweat can cause erratic current flow

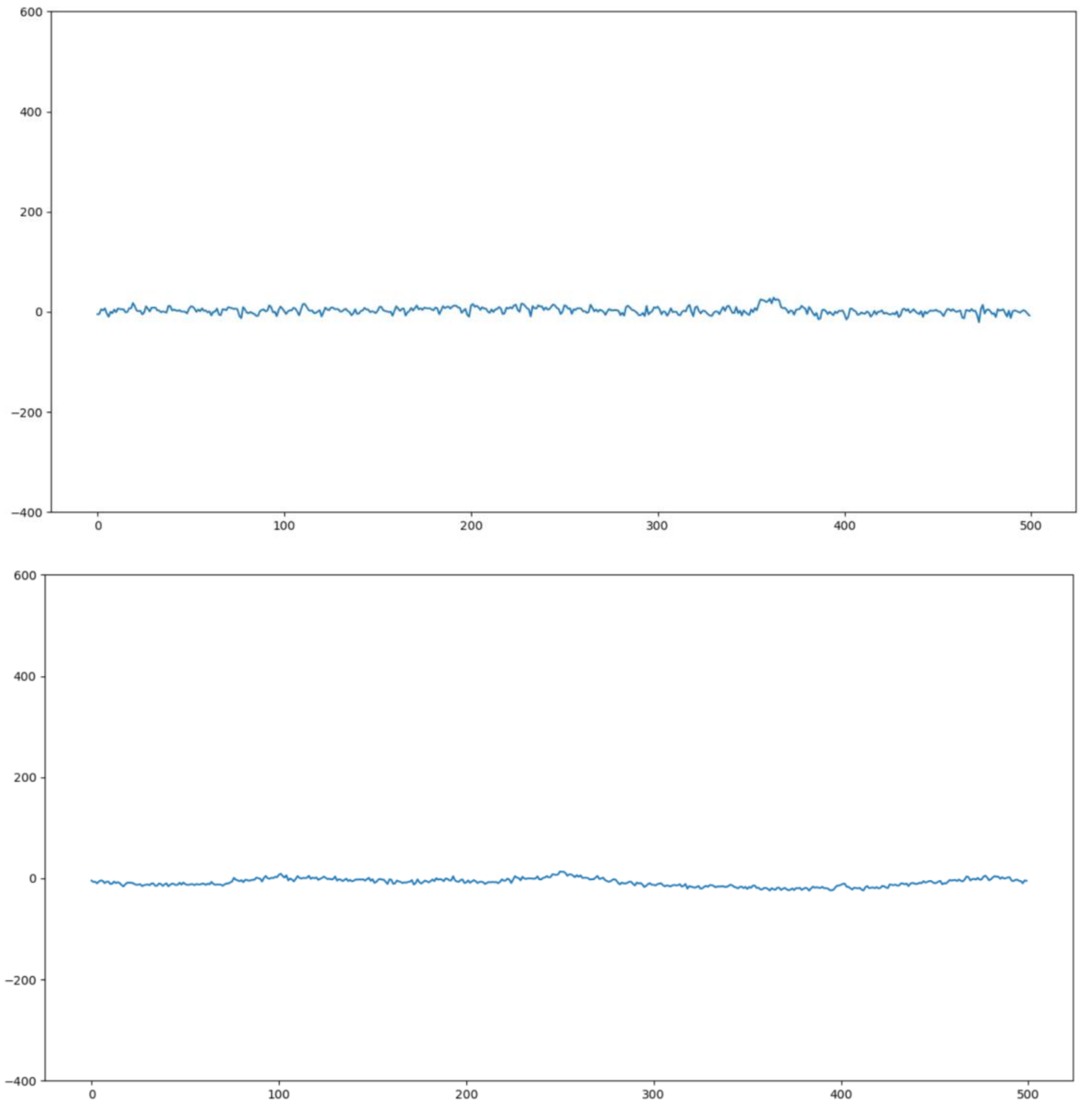


Subject#8: Epoch #11

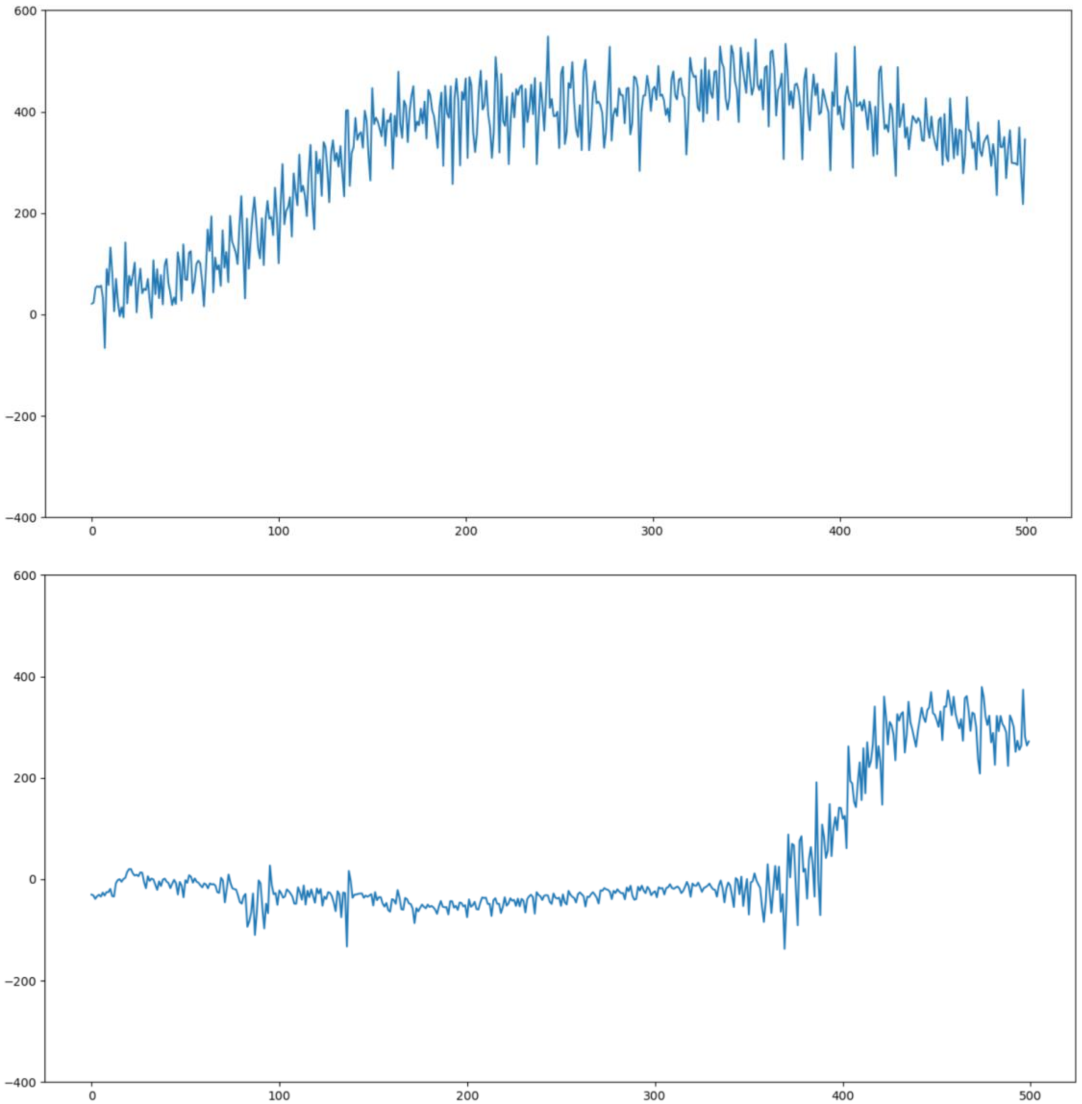


Subject#8: Epoch #57

Non-Artifact vs. Artifact Signals (scaled)



Good signals

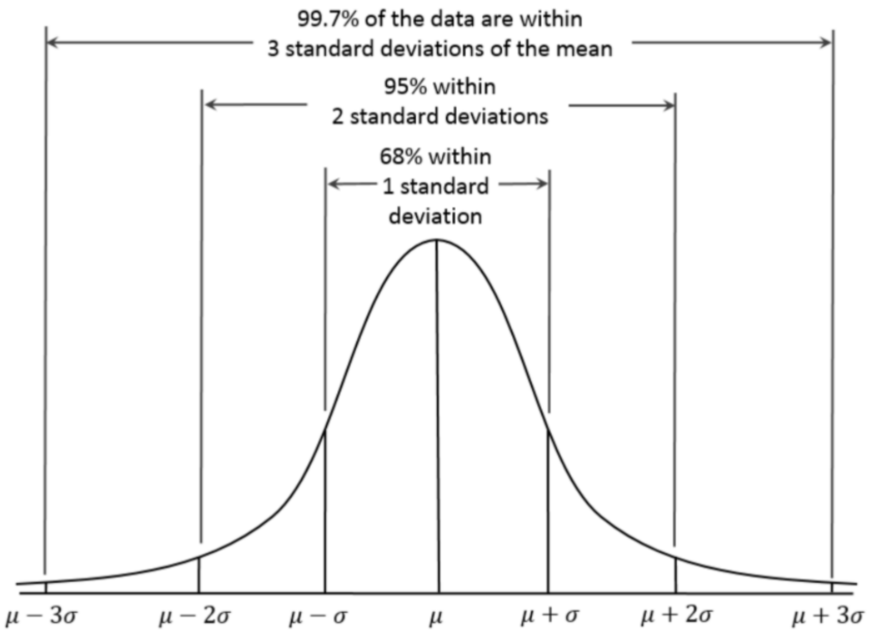


Bad signals

Inspiration: Central Limit Theorem (CLT)

Definition

- If we repeatedly take large random samples from a population, the sample averages <u>tend to form a bell curve (normal distribution)</u>, even if the original data is <u>messy</u>.
  - Center of that bell curve = true population mean (μ)
  - Spread gets smaller with larger sample size (σ / sqrt(n))



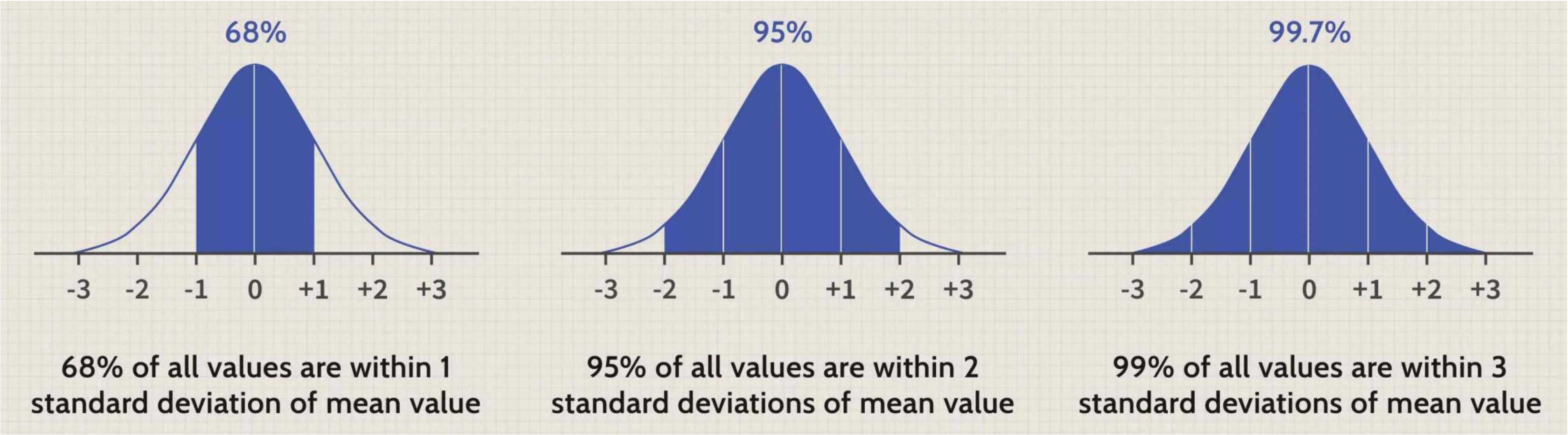
Why this matters?

- It lets <u>us trust average-based</u> methods like:
  - confidence intervals
  - t-tests for comparing two groups

From CLT to an artifact threshold

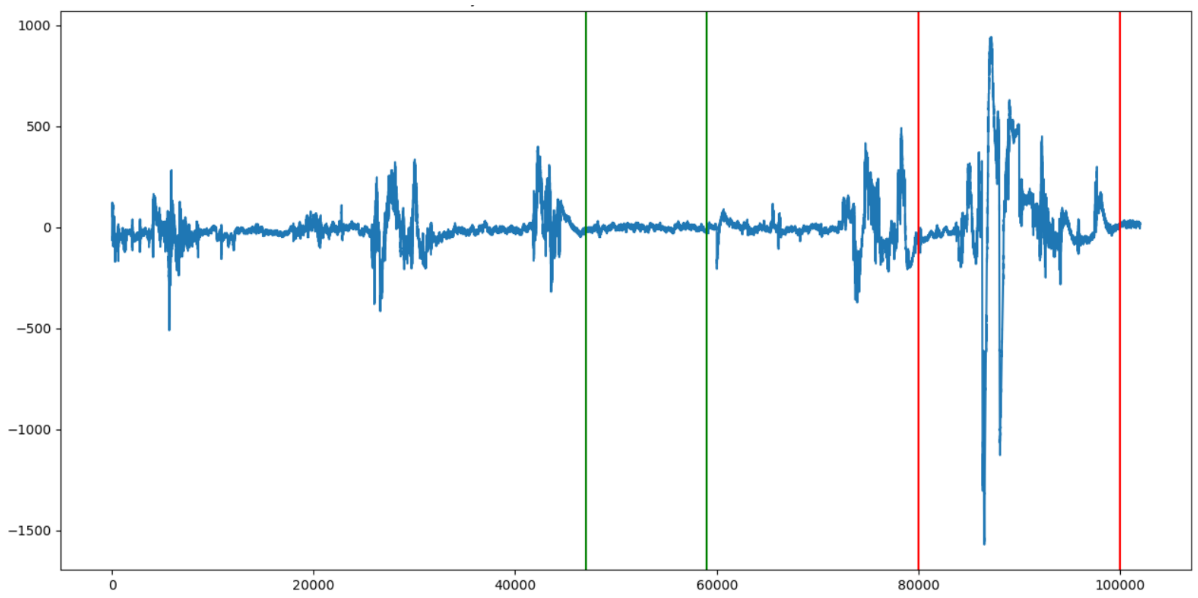
Idea

- For each channel, compute the standard deviation of each epoch.
- The distribution of these epoch-wise standard deviations is approximately normal (bell-shaped).
- Using this, values beyond mean ± 3*std are very unlikely (about 0.3% total probability).
  - So, I treat epochs with standard deviation beyond this range as artifact candidates.

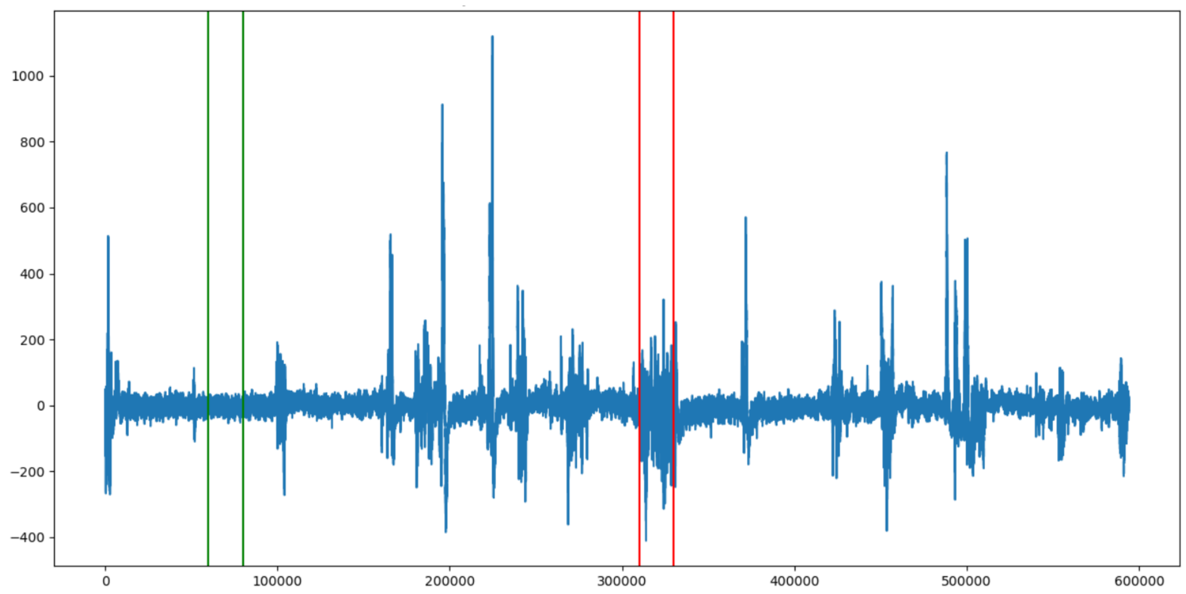


Step 1: Manually select good epoch indexes

For each subject, I plot the signal and manually choose indexes that correspond to a clean data segment.



Dataset#1 : Subject#2



Dataset#1 : Subject#3

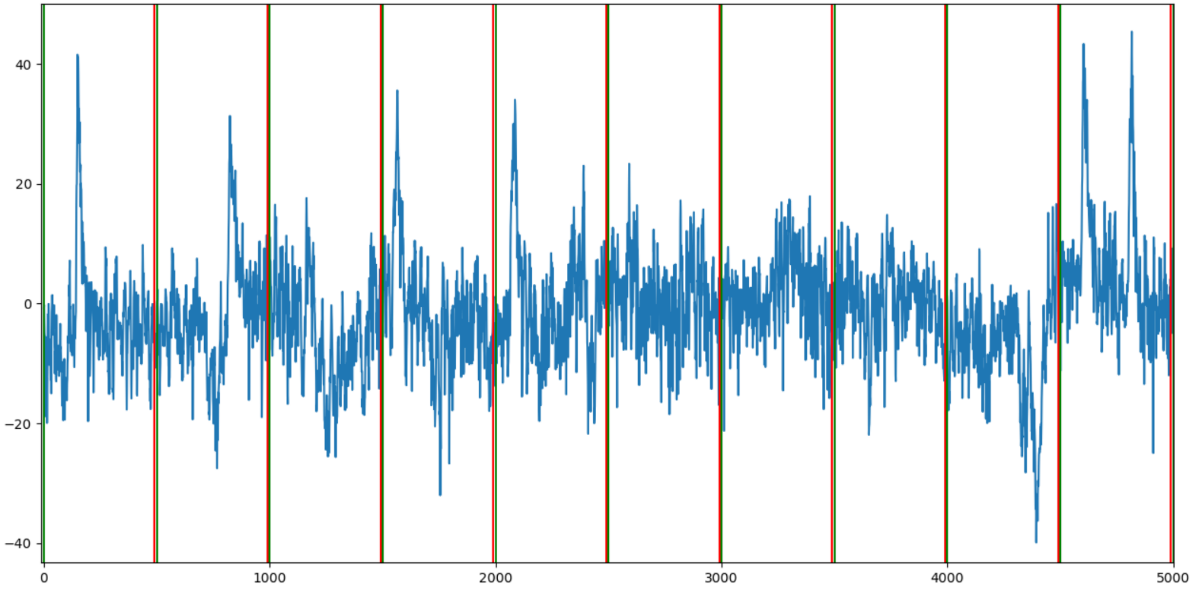
Green lines mark the start and end indexes of the <u>**selected good segment**</u> for that subject.

Step 2: Split into epochs

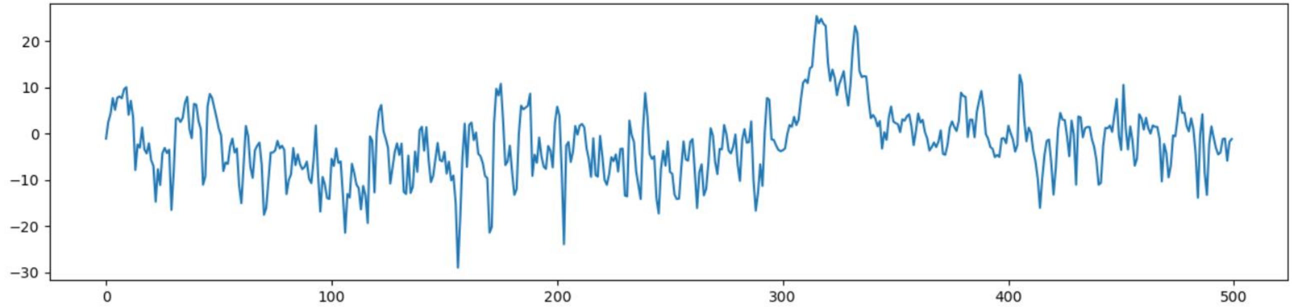
<u>Split</u> each subject's raw EEG signal into epochs.

In this run, each epoch contains 500 samples of raw EEG data.

- At <u>256 Hz</u>, that is about 1.95 s per epoch (500 / 256).
- Epoch length is a <u>**tunable parameter,**</u> so this value can be changed for other runs.



Subject#7



Subject#7: <u>Epoch #118</u>

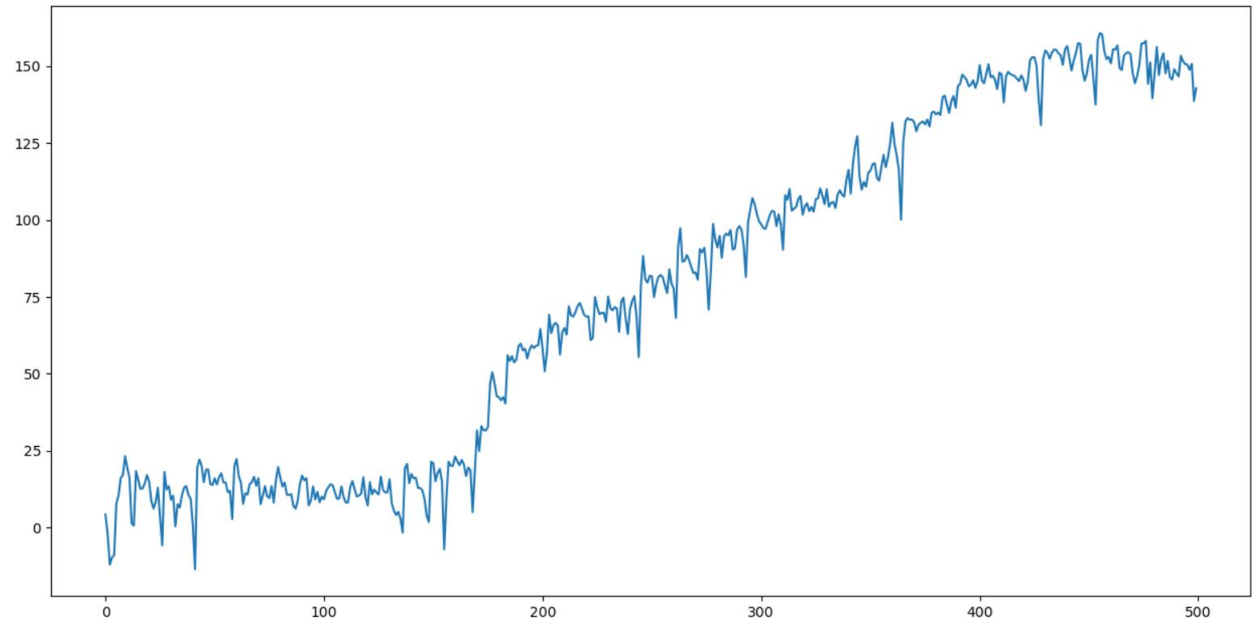
Step 3: Remove trend

For each epoch, first **remove the linear trend**.

- Fit a least-squares regression line to the epoch:
  - y = mx + b
- **Detrend** by taking residuals for each sample:
  - residual = actual value - predicted value (from the fitted line)

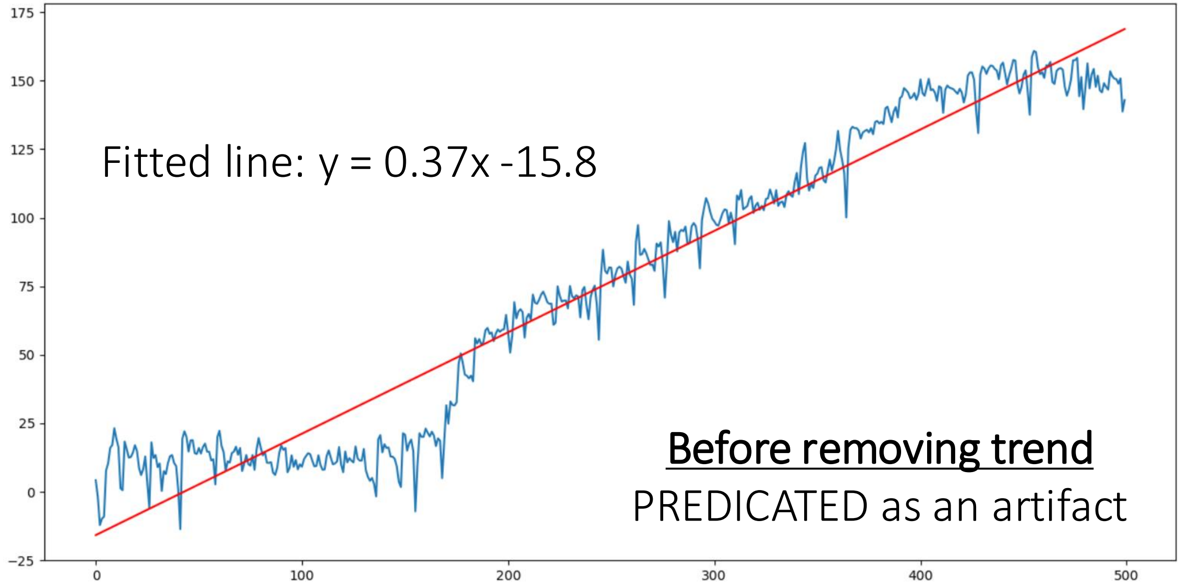
Why:

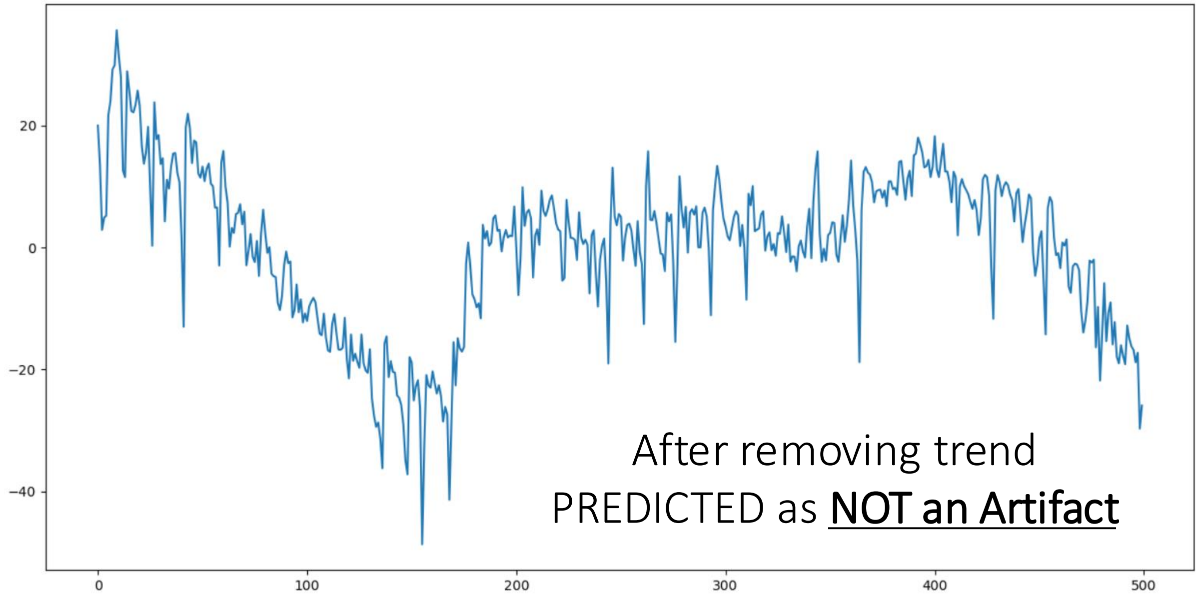
- This **removes slow drifts**, so artifact detection focuses on **abnormal fluctuations** rather than **baseline slope**.



Subject#7: Epoch #1010

Step 3: Remove trend – visual example





Subject#7: Epoch #1010

Step 4: Compute artifact threshold

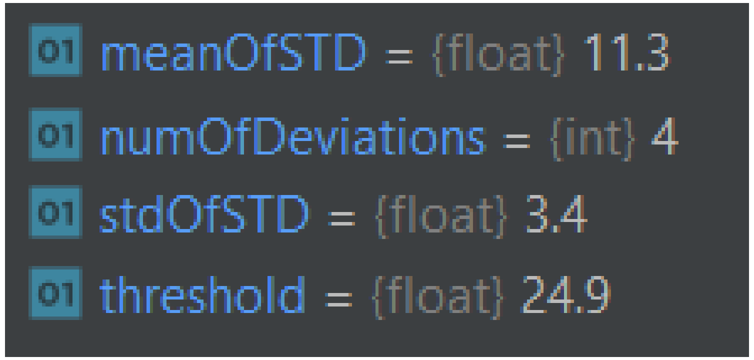
1) For one subject, compute the standard deviation of each epoch in the manually selected good segment.

2) From those epoch std values, compute:

- meanOfStds
- stdOfStds

3) Set the subject-specific threshold:

- threshold = meanOfStds + 3 * stdOfStds



Example in code (here, k = 4 for number of deviations)

4) Mark epochs with std above this threshold as artifact candidates (outliers).

5) Repeat steps 1-4 for each subject.

Step 5: Threshold check

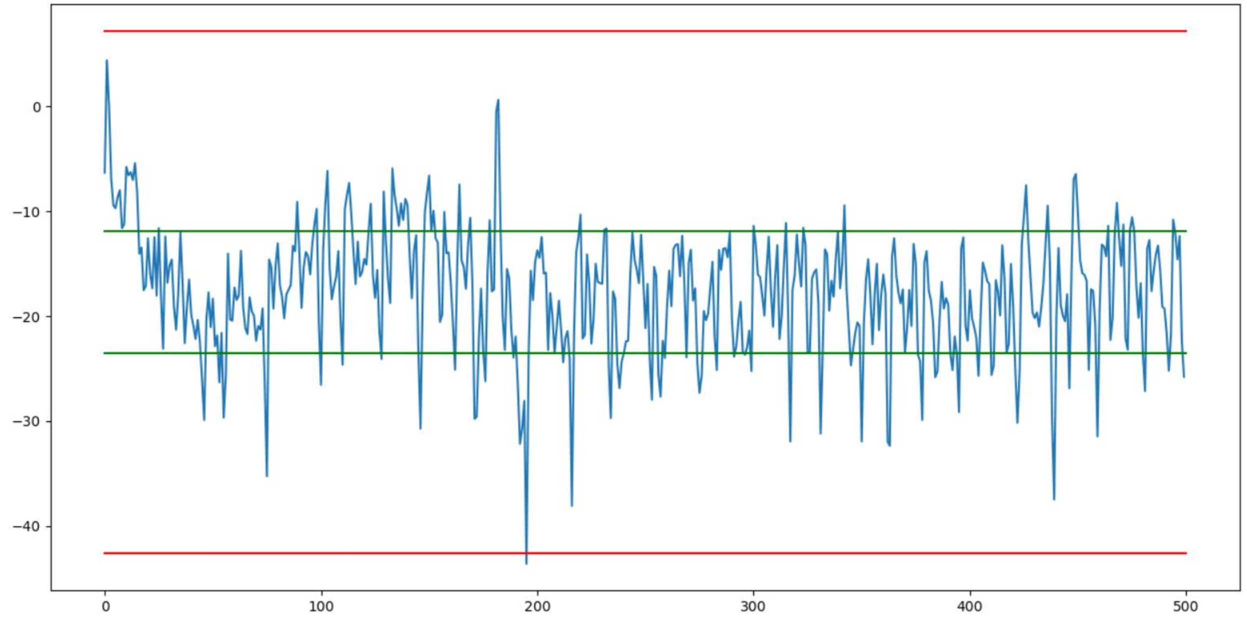
1) Compute the standard deviation of the epoch.

2) Compare it to the threshold:

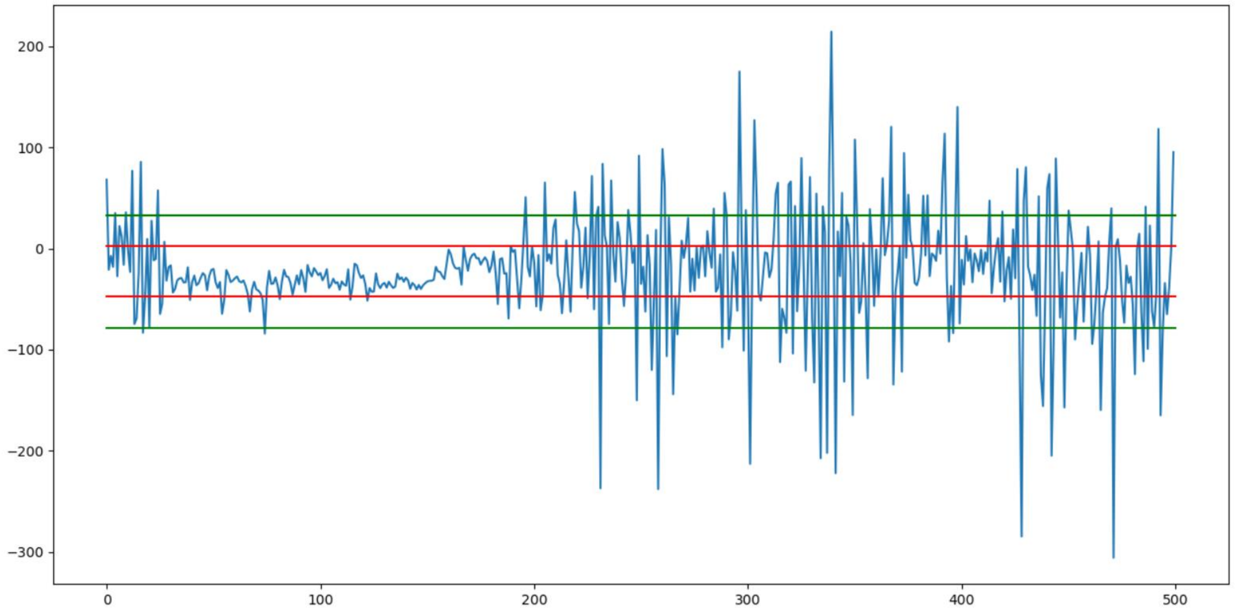
- If epoch_std > threshold -> artifact candidate
- Otherwise -> keep as clean epoch

Green lines: epoch standard deviations
Red lines: threshold

If green is **above** red -> artifact candidate
If green is **below** red -> likely clean



Subject#7: Epoch #1



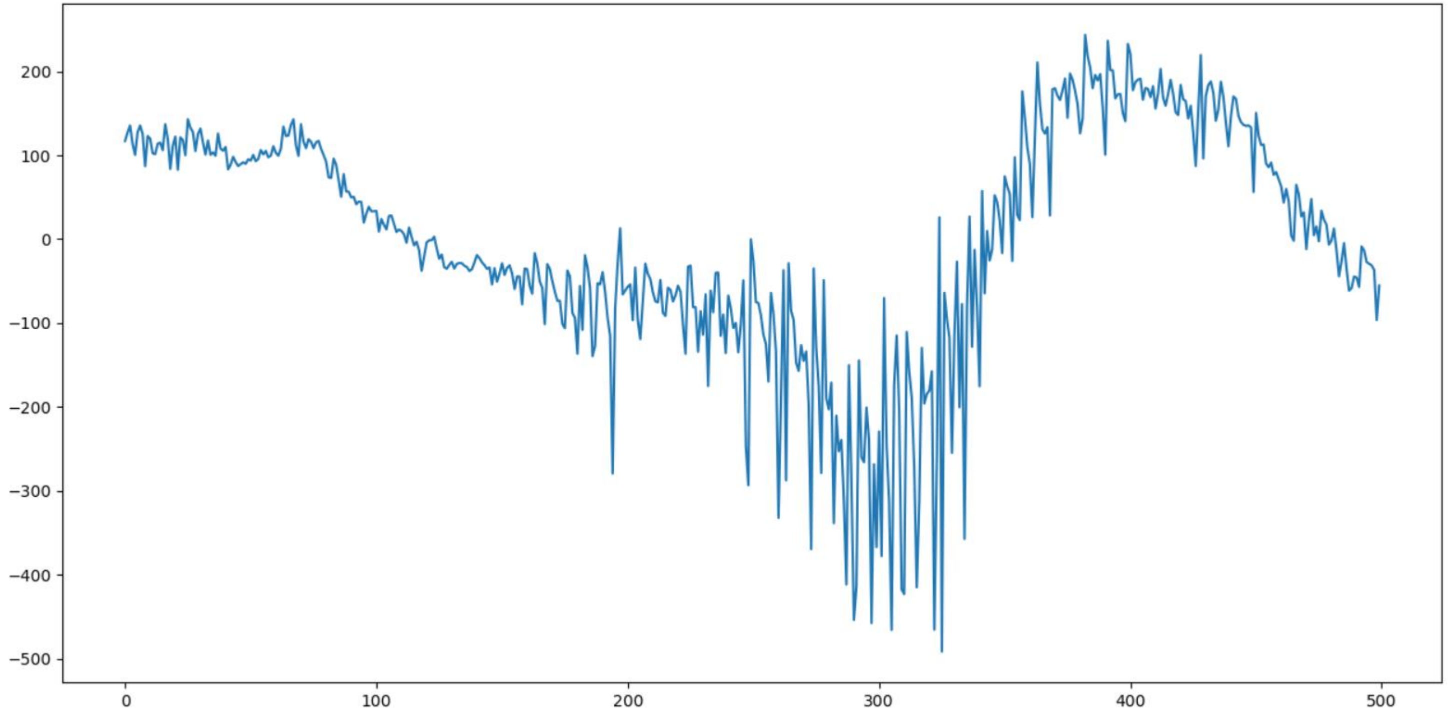
Subject#7: Epoch #58

Step 6: Amplitude check

1) Set maximumValue from plotted EEG signals (manual cutoff for **implausible** amplitudes).

2) For each epoch, compute max(abs(epoch)).

3) If max(abs(epoch)) **> maximumValue** -> artifact candidate.



Subject #16 : Epoch #524

Summary: Artifact Removal Algorithm

1) Select **clean reference** segment (manual)
- For each subject, plot EEG and mark a clean segment.

2) **Split sig**nal into epochs
- Segment each subject’s EEG into epochs (500 samples per epoch in this run).

3) **Detrend** each epoch
- Fit y = mx + b, then use residuals (actual - predicted).

4) Compute **subject-specific threshold**
- From clean-segment epochs:
  - threshold = meanOfStds + 3 * stdOfStds

5) **Std-based** artifact check
- If epoch_std > threshold -> artifact candidate.

6) **Amplitude-based** artifact check
- If max(abs(epoch)) > maximumValue -> artifact candidate.