Signal Quality and Inversion Detection Summary

Repository: github.com/mkucukos/physio-qc-toolkit

Overview

The physio-qc-toolkit provides automated quality control (QC) pipelines for physiological signals such as ECG and airflow data extracted from EDF recordings. The toolkit evaluates multiple metrics per epoch (e.g., 30 s) to flag segments with poor quality, missing data, or polarity inversion.

The toolkit provides:

- End-to-end EDF reading and channel parsing
- Epoch-based QC computation (30 s, configurable)
- Visualization of quality flags and signal integrity overlays
- JSON-based summaries for downstream analysis or batch monitoring

1. ECG Inversion Detection

Description: The inversion detection algorithm automatically identifies lead polarity errors by evaluating waveform morphology and autocorrelation.

- **Green:** Normal (non-inverted) segments
- **Red:** Inverted signal periods
- **Outcome:** The algorithm successfully identified a simulated inversion where the first half of the signal was flipped and appended to the end, with minimal false positives.
-

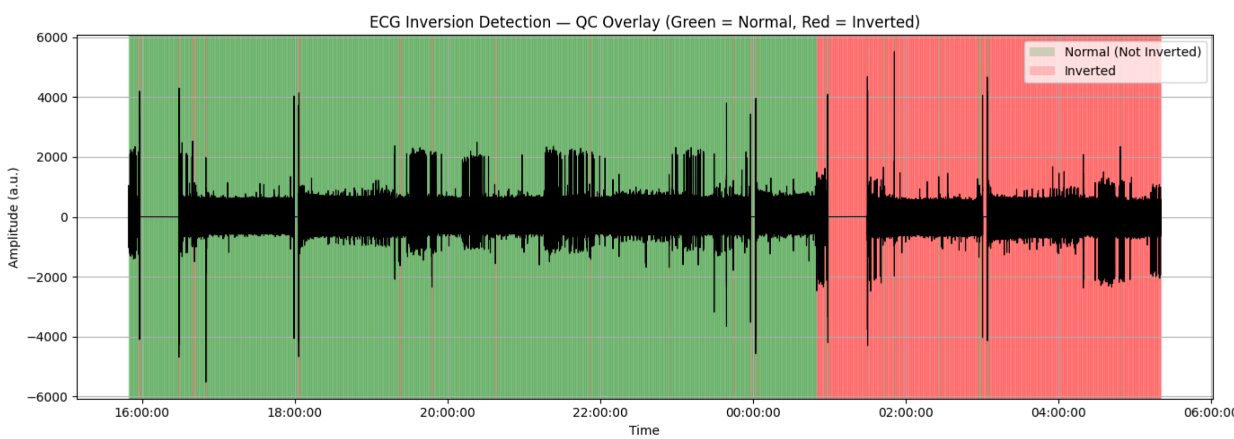

Figure 1. ECG Inversion Detection QC overlay. The system automatically identifies polarity inversion in ECG channels.

2. ECG Baseline Wander QC

Description: Evaluates low-frequency baseline drift in the ECG signal, typically caused by electrode movement, poor contact, or respiration artifacts. Segments with more than 15% power below 0.3 Hz are flagged as poor quality.

- **Green:** Stable baseline
- **Red:** Excessive low-frequency drift
- **Outcome:** Most of the recording maintained stable baselines except for short periods.

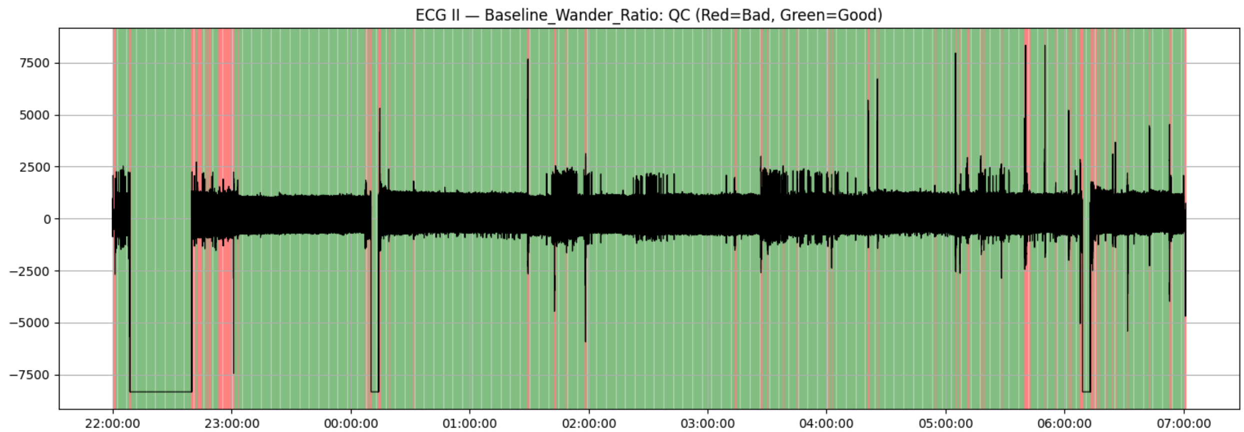

Figure 2. ECG baseline wander QC overlay showing stable (green) and motion-affected (red) segments.

3. ECG Clipping Ratio QC

Description: Quantifies the proportion of signal values that reach the ADC's digital limits, indicating potential hardware saturation or data loss.

- **Green:** Within acceptable amplitude range (clipping ratio < 0.5)
- **Red:** Clipped or saturated epochs (clipping ratio ≥ 0.5)
- **Outcome:** Clipping detected only during short calibration intervals; overall ECG amplitude range remained valid.

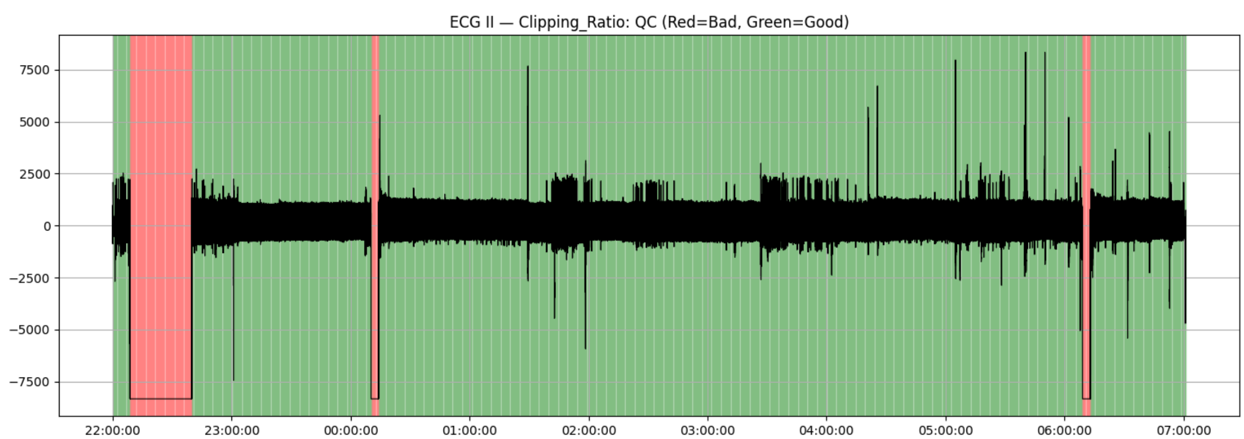

Figure 3. ECG clipping ratio QC plot showing saturated segments at amplitude extremes.

4. ECG Signal-to-Noise Ratio (SNR) QC

Description: Measures the ratio of cardiac signal power to background noise using NeuroKit2-based R-peak detection and cleaned waveform subtraction. During feature extraction, ±0.1 s windows centered around each detected R-peak are isolated.

- **Green:** SNR > 6 dB (good signal quality)
- **Red:** SNR ≤ 6 dB (noisy or motion-corrupted)
- **Outcome:** Most epochs exceeded the 6 dB threshold, confirming high-quality ECG recordings.

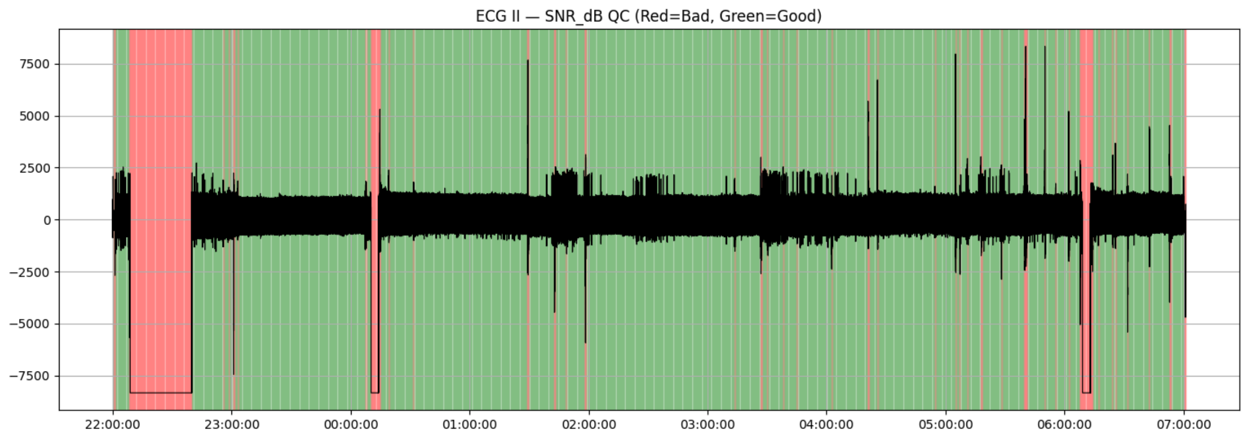

Figure 4. ECG SNR-based QC visualization highlighting noisy (red) epochs.

5. ECG Flatline Ratio QC

Description: The flatline ratio quantifies how much of the ECG signal remains nearly constant over time. It identifies loss of signal due to electrode detachment, hardware failure, or frozen input buffers.

- **Green:** Dynamic signal (flatline ratio ≤ 0.5)
- **Red:** Flatlined or unchanging signal (flatline ratio > 0.5)
- **Outcome:** The ECG II channel showed minimal flatlining, indicating stable electrode connectivity throughout most of the recording.

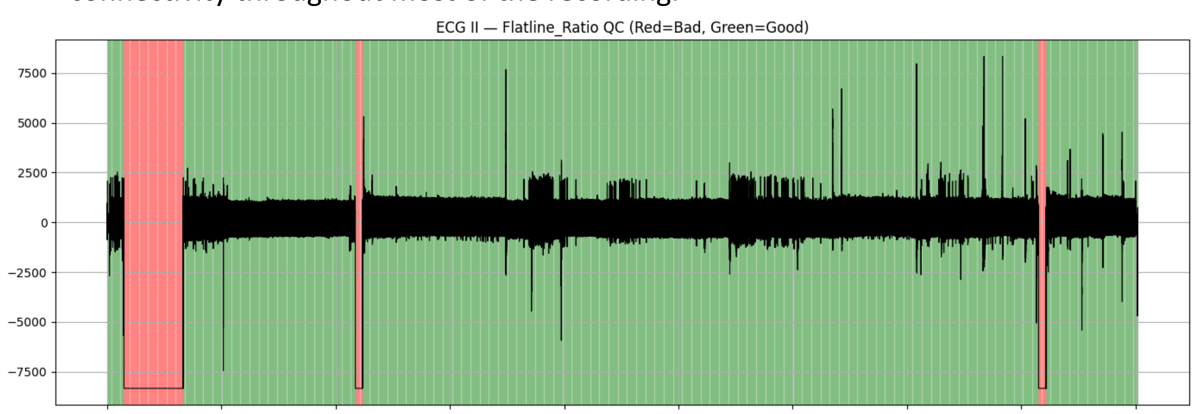

Figure 5. ECG Flatline Ratio QC plot showing static (red) vs. dynamic (green) signal epochs.

Example Output Summary — ECG QC

The following JSON summary represents the automated ECG quality control (QC) output generated by the toolkit after processing an EDF file. Each epoch (30 s) is evaluated across multiple QC metrics, including clipping, flatline, baseline wander, and SNR. **The inversion results are not included here, as the inversion signal was artificially generated for validation purposes.**

{'total_epochs': 1083, 'good_epochs': 820, 'bad_epochs': 263, 'good_ratio': 0.757, 'bad_ratio': 0.243}

This output indicates that **75.7%** of epochs passed all QC checks, while **24.3%** were flagged as poor-quality due to one or more issues (e.g., noise, clipping, or baseline drift). The structured format allows easy integration into batch pipelines or downstream analytics dashboards.

6. Thermistor Flow — Autocorrelation QC

Description: Evaluates the periodicity and regularity of respiratory cycles using autocorrelation within a 10 s lag window. Low autocorrelation values indicate irregular or noisy breathing patterns.

- **Green:** Regular and rhythmic respiration (autocorrelation > 0.5)
- **Red:** Irregular or unstable cycles (autocorrelation ≤ 0.5)
- **Outcome:** The autocorrelation quality was generally poor across large portions of the recording, as indicated by extended red regions. This suggests reduced rhythmicity and higher noise levels in the thermistor signal, likely due to motion artifacts or transient airflow disruptions.

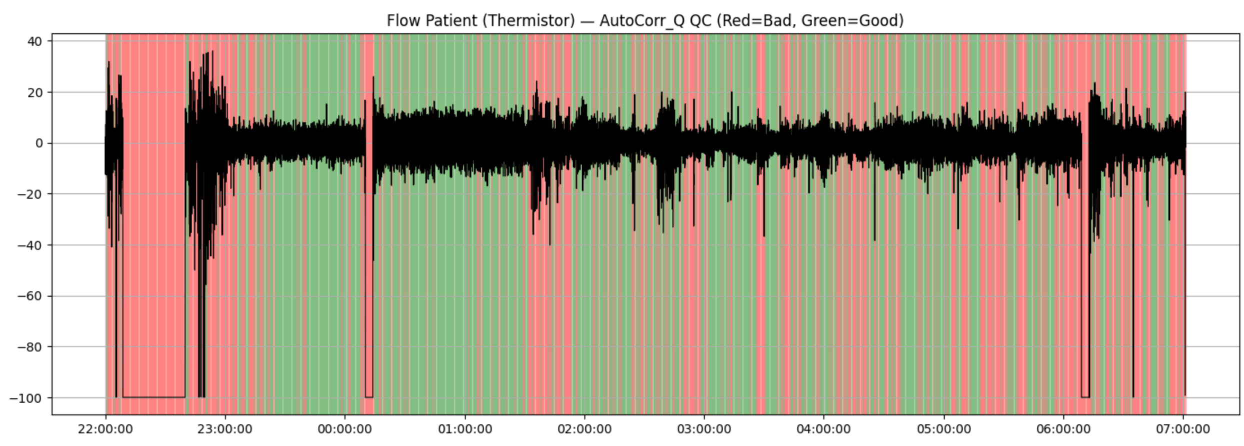

Figure 6. Autocorrelation QC showing consistent respiration (green) versus irregular epochs (red).

7. Thermistor Flow — Clipping QC

Description: Assesses amplitude saturation in the thermistor signal that can occur when airflow peaks exceed the recording range. This QC ensures reliable amplitude capture during breathing cycles.

- **Green:** Within acceptable amplitude range (clipping ratio < 0.5)
- **Red:** Clipped or saturated epochs (clipping ratio ≥ 0.5)
- **Outcome:** Clipping was detected only during short transitions at the start and close to the end of the recording; overall thermistor signal amplitude was stable.

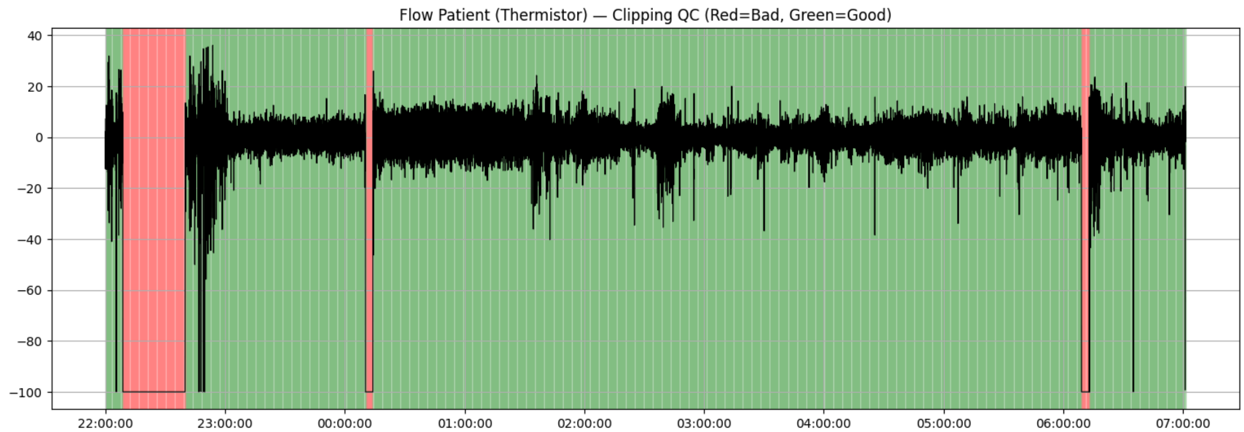

Figure 7. Flow Clipping QC plot showing normal amplitude (green) segments and short saturated (red) segments

8. Thermistor Flow — Flatline QC

Description: Measures periods where the thermistor signal remains constant, indicating sensor disconnection or freezing.

- **Green:** Dynamic airflow signal (flatline ratio < 0.5)
- **Red:** Flatlined or unchanging signal (flatline ratio ≥ 0.5)
- **Outcome:** Flatline detection confirmed short periods of signal inactivity at transitions, consistent with brief pauses or sensor resets.

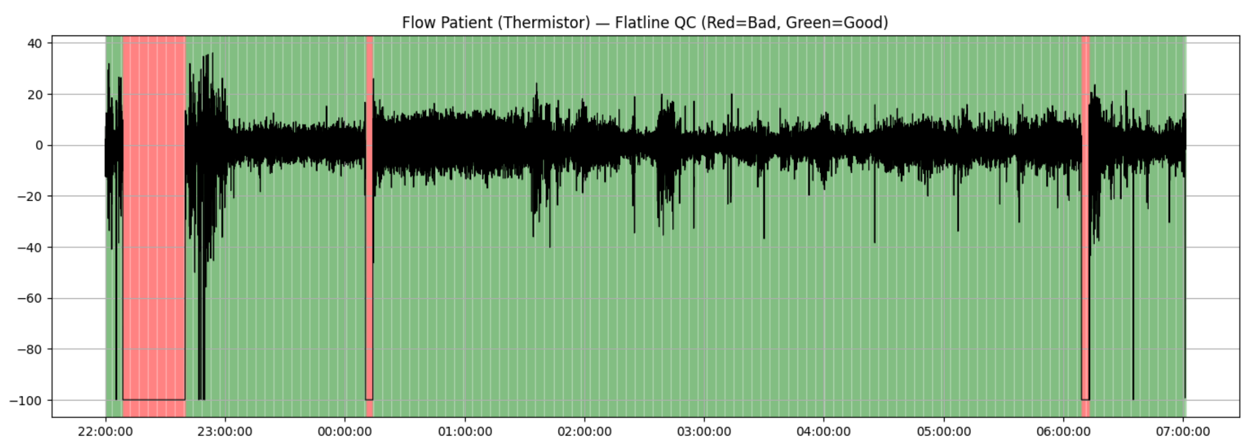

Figure 8. Thermistor Flow Flatline QC visualization showing dynamic (green) vs. static (red) regions.

Example Output Summary — Thermistor Flow QC

The following JSON summary represents the **automated Thermistor Flow quality control (QC)** output generated by the toolkit after processing an EDF file. Each 30-second epoch was evaluated across multiple QC metrics, including **autocorrelation regularity**, **amplitude clipping**, and **flatline detection**.

{"total_epochs": 1083, "good_epochs": 572, "bad_epochs": 511, "good_ratio": 0.528, "bad_ratio": 0.472}

This report indicates that **52.8% of thermistor flow epochs** passed all QC checks, while **47.2%** were classified as poor-quality due to irregular breathing patterns, motion-related signal disruptions, or intermittent sensor inactivity. The structured JSON format enables easy integration into automated batch pipelines and visual summary dashboards for respiratory signal quality assessment.

9. Pressure Flow — Autocorrelation QC

Description: Evaluates the periodicity and regularity of respiratory cycles using autocorrelation within a 10 s lag window. Low autocorrelation values indicate irregular or noisy breathing patterns.

- **Green:** Regular and rhythmic respiration (autocorrelation > 0.5)
- **Red:** Irregular or unstable cycles (autocorrelation ≤ 0.5)
- **Outcome:** The autocorrelation quality was generally poor across large portions of the recording.

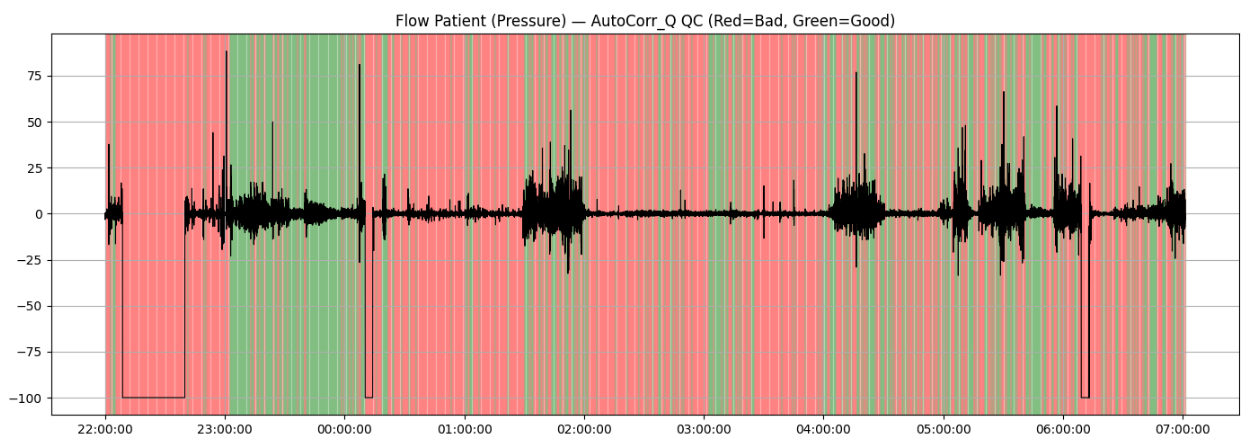

Figure 9. Autocorrelation QC showing consistent respiration (green) versus irregular epochs (red).

10. Pressure Flow — Clipping QC

Description: Assesses amplitude saturation in the pressure flow signal that can occur when respiratory pressure peaks exceed the sensor's dynamic range. This QC metric ensures reliable measurement of inhalation and exhalation amplitudes throughout the recording.

- **Green:** Within acceptable amplitude range (clipping ratio < 0.5)

- **Red:** Clipped or saturated epochs (clipping ratio ≥ 0.5)
- **Outcome:** Clipping was detected only during short transitions at the start and close to the end of the recording; overall pressure signal amplitude was stable.

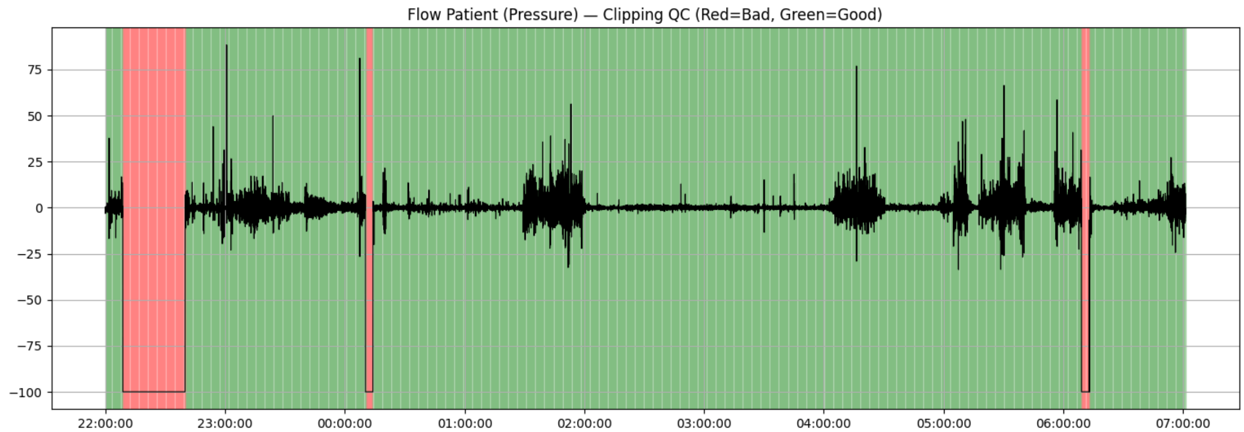

Figure 10. Pressure Flow Clipping QC plot showing normal amplitude (green) segments and short saturated (red) segments

11. Pressure Flow — Flatline QC

Description: Identifies periods where the **pressure flow** signal remains nearly constant, typically indicating sensor freezing, disconnection, or lack of respiratory effort.

- **Green:** Dynamic airflow signal (flatline ratio < 0.5)
- **Red:** Flatlined or unchanging signal (flatline ratio ≥ 0.5)
- **Outcome:** Flatline detection revealed short intervals of inactivity, mostly corresponding to calibration or brief sensor pauses, confirming overall stable pressure signal continuity.

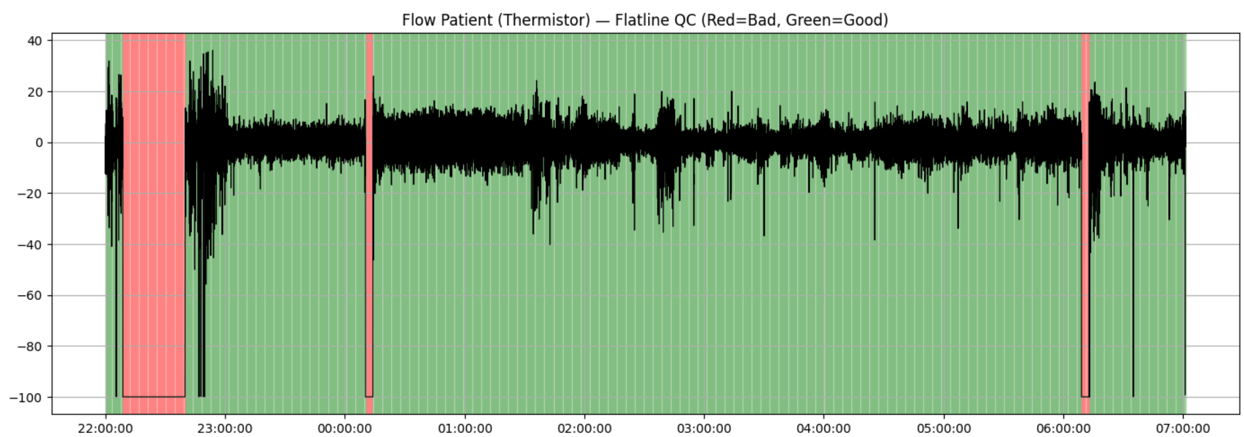

Figure 11. Pressure Flow Flatline QC visualization showing dynamic (green) versus static (red) epochs.

Example Output Summary — Pressure Flow QC

The following JSON summary represents the **automated Pressure Flow quality control (QC)** output generated by the toolkit after processing an EDF file. Each 30-second epoch was evaluated across multiple QC metrics, including **autocorrelation regularity**, **amplitude clipping**, and **flatline detection**.

{"total_epochs": 1083, "good_epochs": 374, "bad_epochs": 709, "good_ratio": 0.345,"bad_ratio": 0.655}

This report indicates that 34.5% of pressure flow epochs met all QC criteria, while 65.5% were classified as poor-quality. The higher proportion of low-quality epochs reflects significant irregularity and instability in the pressure flow signal, likely due to motion artifacts, transient signal saturation, or sensor pauses.

12. EEG Artifact Detection QC

Description:
The EEG artifact detection algorithm automatically identifies epochs contaminated by low-frequency drift, movement, or high-frequency electromyographic (EMG) noise. Each 30-second epoch is transformed into the frequency domain using a fast Fourier transform (FFT), and spectral amplitudes are normalized by the mean power between 1 and 40 Hz to reduce inter-channel variability. The normalized spectra are then log-transformed (base 10) prior to artifact evaluation.

Artifact likelihood is computed using an empirically derived **triple-ratio metric** that captures both drift and EMG contamination:

$$AR = (SP_{0.1\text{-}2Hz} / SP_{2\text{-}5Hz}) \times 1/SP_{40\text{-}50Hz}$$

where *SP* denotes mean spectral power within each frequency band.
Epochs with **AR > 6** are classified as artifactual.

This approach was derived empirically through manual inspection of labeled EEG epochs from large-scale sleep datasets. By comparing the spectral profiles of contaminated and clean segments, consistent patterns emerged—artifactual epochs displayed elevated power at both low (0.1–2 Hz) and high (40–50 Hz) frequencies relative to mid-range activity (2–5 Hz). A systematic grid search across band combinations confirmed this ratio as the strongest discriminator between clean and noisy epochs.

The resulting formula operationalizes these observations into a reproducible, interpretable, and computationally efficient QC method that accurately identifies both drift- and EMG-related artifacts at the full 30-second epoch level. Importantly, this ratio was derived and validated using sleep EEG recordings, where the predominant activity lies in the delta and sigma ranges and where motion and EMG artifacts present distinct spectral signatures. Consequently, epochs corresponding to awake periods—characterized by elevated alpha or broadband muscle activity—may occasionally be flagged as artifactual by this sleep-calibrated threshold.

Visualization:
The EEG channel **C4** is displayed as a log-power spectrogram using the Jet colormap, where cleaner spectral activity spans blue to green to red hues.
Detected artifacts are overlaid as **semi-transparent magenta regions**, providing an intuitive temporal and spectral view of contamination without obscuring underlying signal structure.

Color legend:

- **Jet colormap:** Clean EEG spectral power (relative log scale)
- **Magenta overlay:** Artifact-flagged epochs (AR > 6)

Outcome:
The triple-ratio method reliably detected both slow-drift and EMG-related noise while preserving physiologic EEG activity. Across most channels, only short artifact intervals were observed, confirming high-quality data capture and stable electrode contact.

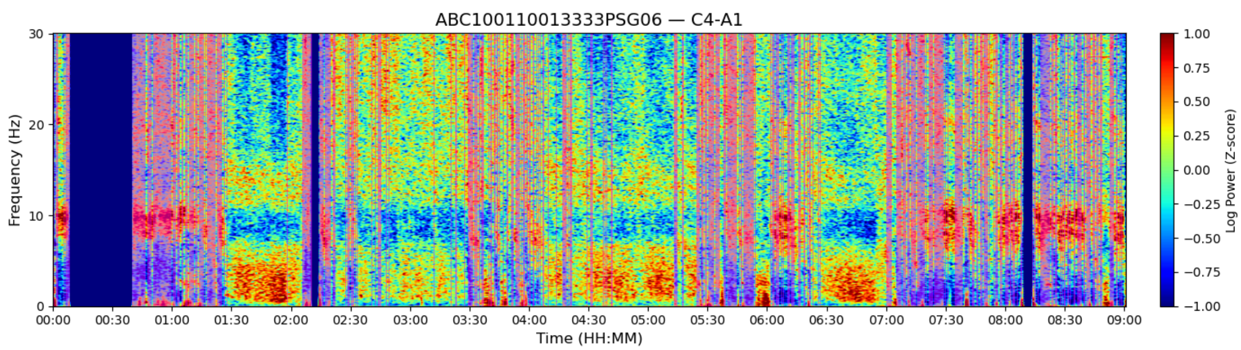

Figure 12. EEG Artifact Detection QC visualization of **channel C4** showing clean spectral dynamics (Jet colormap) with semi-transparent **magenta** overlays marking artifact-flagged epochs.

13. EEG Flatline Detection QC

Description:
The EEG Flatline Detection QC identifies epochs where the signal exhibits abnormally low amplitude or complete loss of physiological variability, typically resulting from electrode disconnection, high impedance, or amplifier saturation. Each 30-second epoch is evaluated using three complementary metrics:
(1) **Signal variance**,
(2) **Peak-to-peak amplitude**, and
(3) **Proportion of repeated consecutive samples**.

Epochs with exceptionally low variance and amplitude, below 20% of the 5th percentile of all values, are considered potential dropouts. Additionally, epochs in which more than 98% of samples remain identical are flagged as complete flatlines. These criteria jointly detect both subtle amplitude suppression and total channel failure.

Flatline detection is particularly important in long-duration sleep recordings, where brief disconnections or cable motion can produce signal loss without immediately visible noise. By quantifying the degree of signal constancy, this QC ensures that silent or saturated epochs are excluded from downstream spectral and statistical analyses.

Visualization:
Channel C4is shown as a log-power spectrogram (Jet colormap), where normal EEG activity spans the full spectral range.
Flatlined periods are overlaid as **semi-transparent magenta regions**, indicating intervals of near-zero variance or repeated samples while retaining the visibility of surrounding spectral context.

Color legend:

- **Jet colormap:** Clean EEG spectral power (relative log scale)
- **Magenta overlay:** Flatline or signal-loss epochs

Outcome:
Flatline detection reliably isolated brief intervals of signal loss, primarily corresponding to transient electrode disconnections or impedance spikes. Across most channels, flatlines were sparse and short, confirming stable contact throughout the recording.

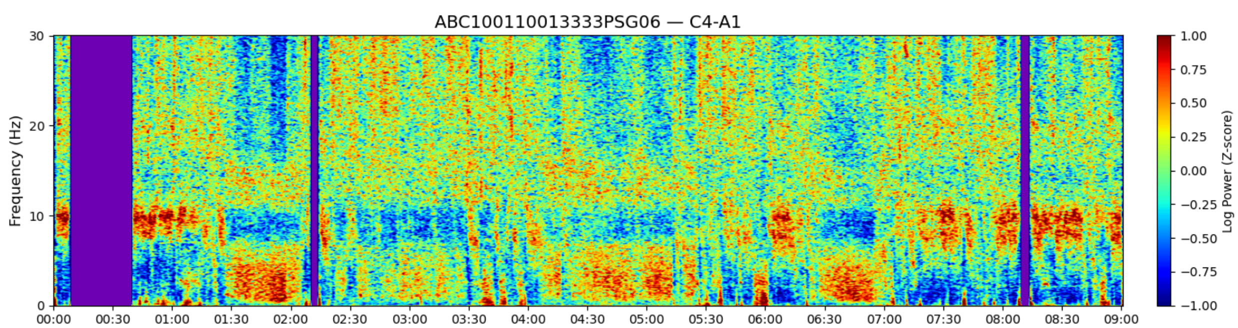

Figure 13. EEG Flatline QC visualization for channel C4 showing clean spectral activity (Jet colormap) and semi-transparent magenta overlays marking epochs with signal dropout or near-zero variance.

Example Output Summary — EEG QC

Description:
The following JSON summary represents the automated EEG quality control (QC) output generated by the toolkit after processing an EDF file looking only at channel **C4**. Each 30-second

epoch was evaluated across multiple QC metrics, including spectral artifact detection using the triple-ratio method and flatline detection based on signal variance and repeated-sample ratios.


```
{
  "total_epochs": 1082,
  "clean_epochs": 592,
  "noisy_epochs": 416,
  "flatline_epochs": 74,
  "clean_ratio": 0.547,
  "artifact_ratio": 0.453,
  "noisy_ratio": 0.384,
  "flatline_ratio": 0.068
}
```


This report indicates that for channel C4, **54.7%** of EEG epochs met quality criteria, while **38.4%** were classified as noisy and **6.8%** were identified as flatlined or dropped-signal segments. The relatively small proportion of flagged epochs suggests stable electrode contact and minimal contamination during recording. In cases where the percentage of flagged epochs is higher, these typically correspond to periods of movement, elevated muscle tone, or transient electrode disconnection.