State of MIDA-ACV Project

Lorenzo Lastrucci, Daniele Paletti

Politecnico di Milano

2021-05-04

Section 1

Data Exploration

Scooter Path

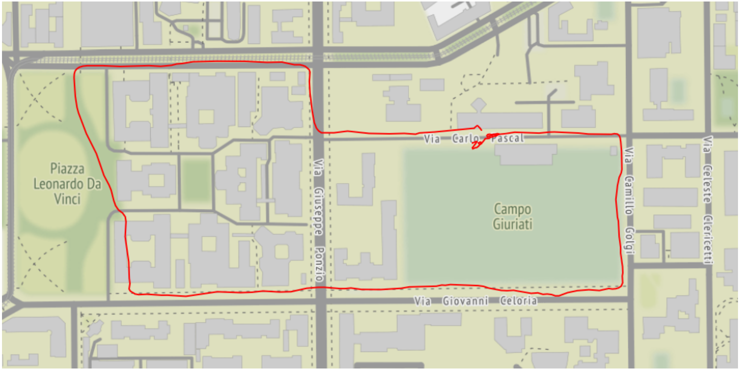


Figure: Path from Coral's Run (stem recording)

Speed: one driver, low weight

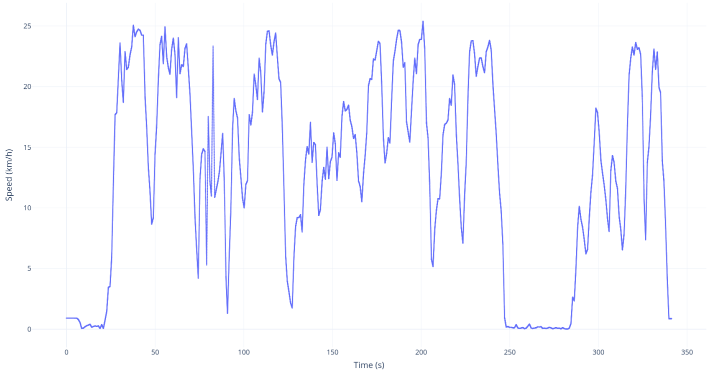


Figure: Speed from Leoni's recording

Speed: one driver, high weight

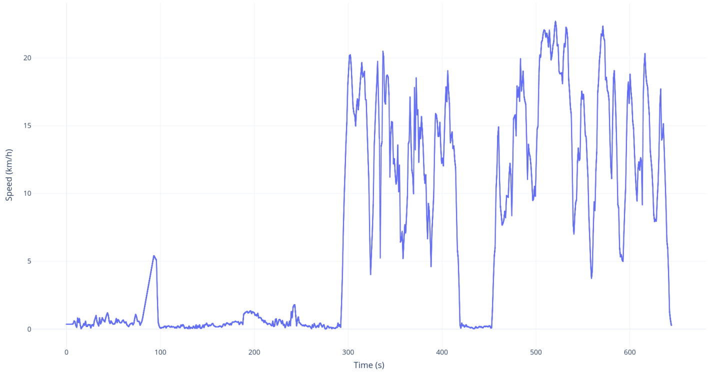


Figure: Speed from Didonato's recording

Speed: two drivers

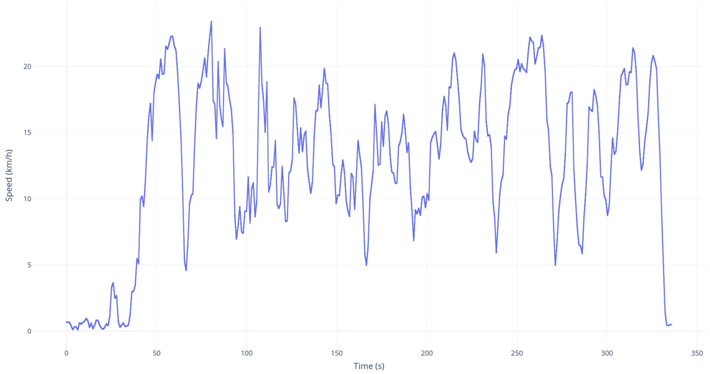


Figure: Speed from Didonato and Leoni's recording

Acceleration: one driver, low weight

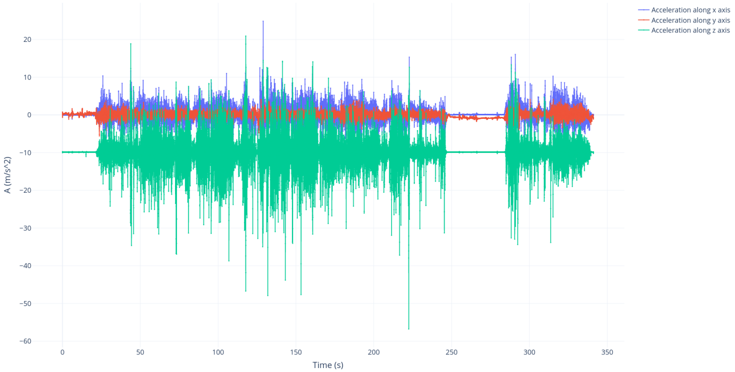


Figure: Acceleration from Leoni's recording

Acceleration: one driver, high weight

Figure: Acceleration from Didonato's recording

Acceleration: two drivers

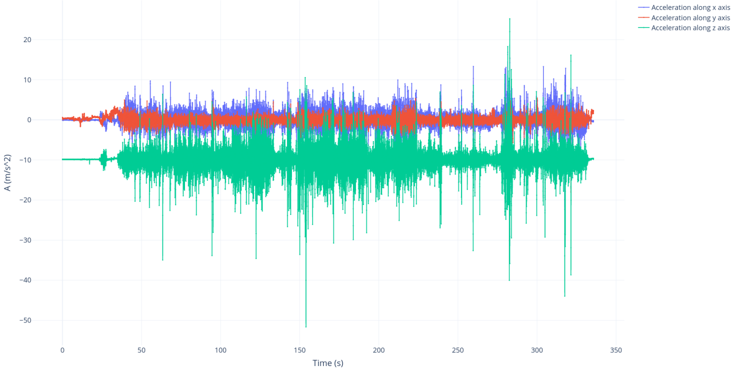


Figure: Acceleration from Didonato and Leoni's recording

Jerk: one driver, low weight

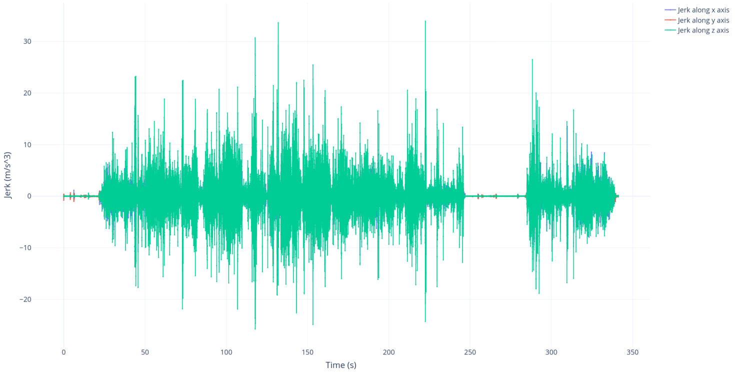


Figure: Jerk from Leoni's recording

Jerk: one driver, high weight

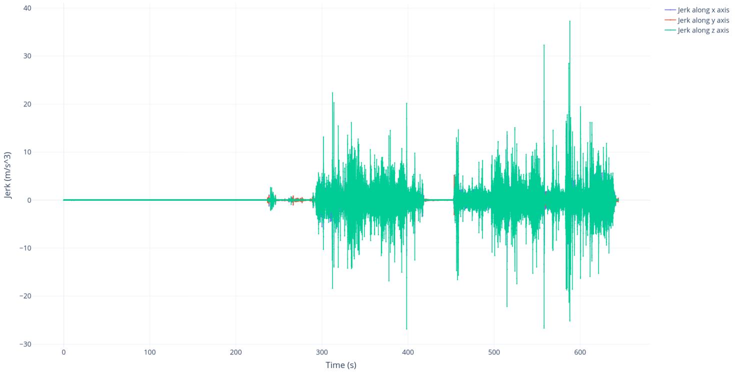


Figure: Jerk from Didonato's recording

Jerk: two drivers

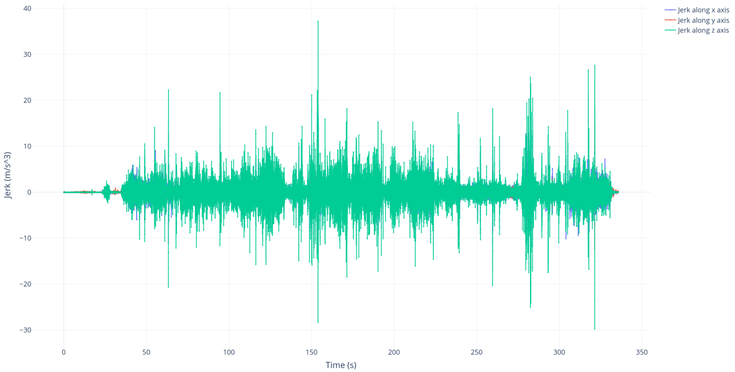


Figure: Jerk from Didonato and Leoni's recording

Orientation: one driver, low weight

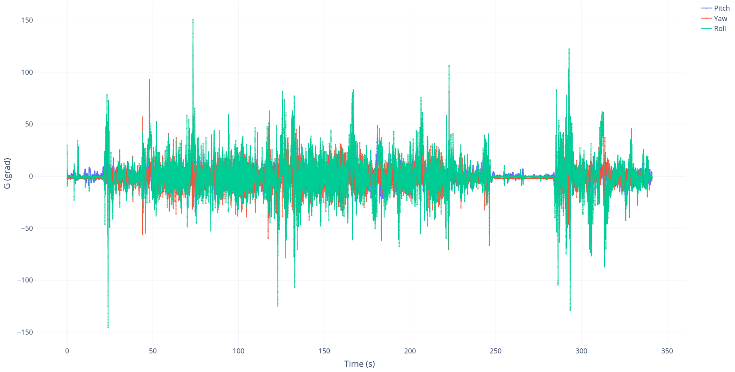


Figure: Orientation from Leoni's recording

Orientation: one driver, high weight

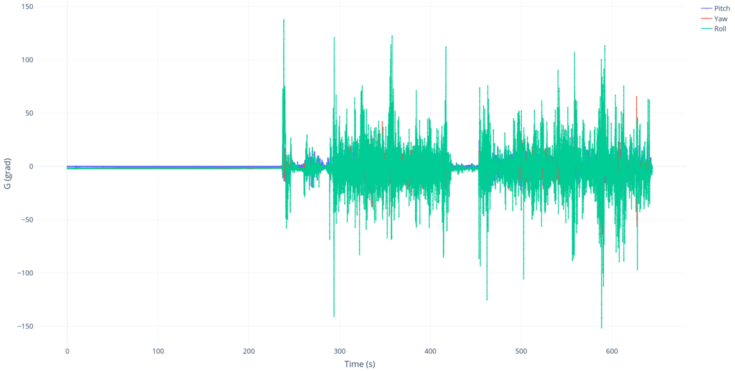


Figure: Orientation from Didonato's recording

Orientation: two drivers

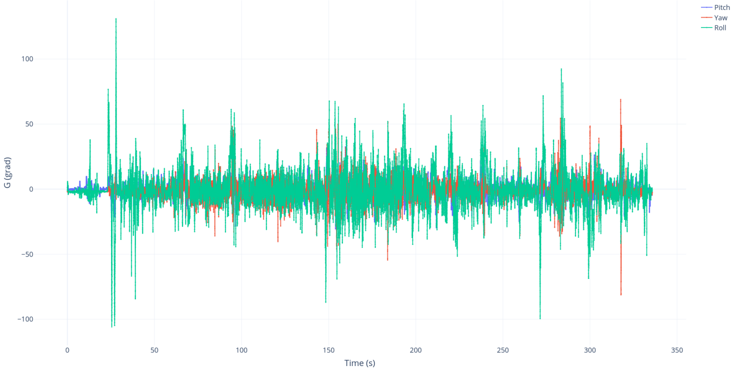


Figure: Orientation from Didonato and Leoni's recording

Section 2

Data Cleaning

Removing low speed

When speed goes lower than 5 km/h the rider is not on board, so those points can be removed.

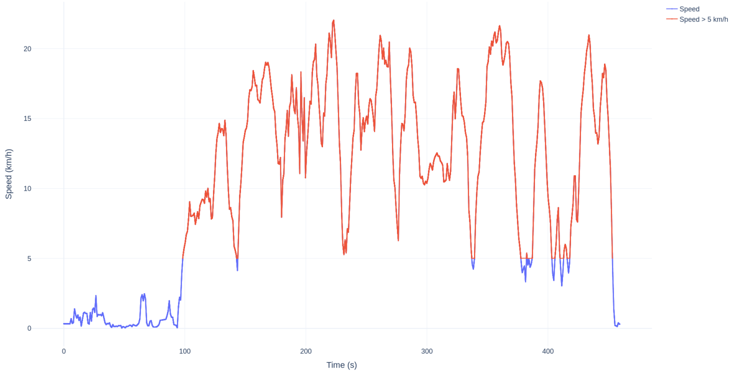


Figure: Speed from Renzo and Jessica's recording with speed > 5 km/h

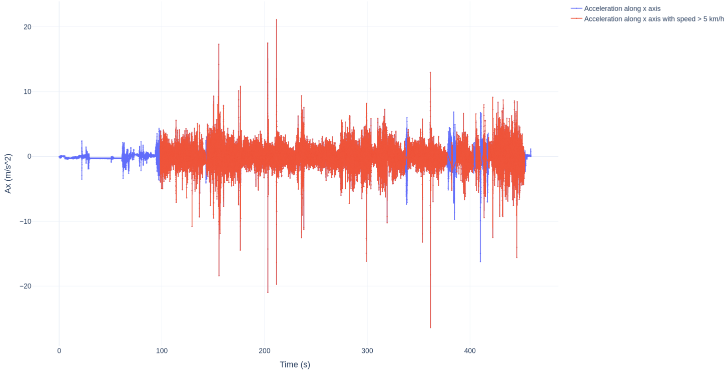


Figure: Acceleration along x axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: acceleration along y axis

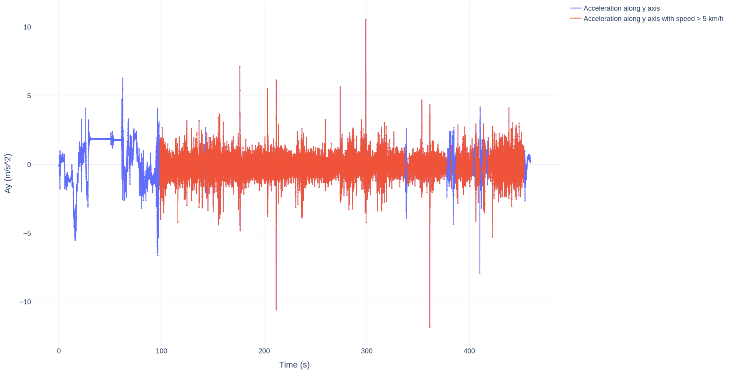


Figure: Acceleration along y axis from Renzo and Jessica's recording with corresponging speed $>$ 5 km/h

Removing low speed: acceleration along y axis

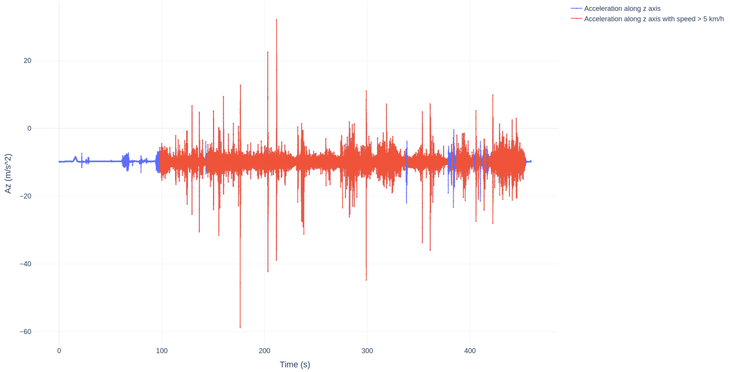


Figure: Acceleration along y axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: jerk along x axis

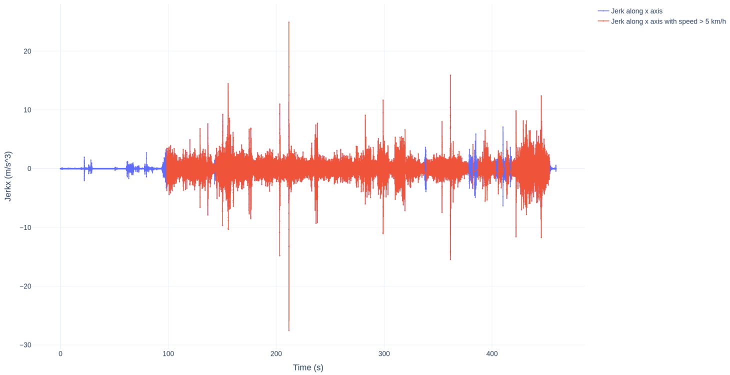


Figure: Jerk along x axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: jerk along y axis

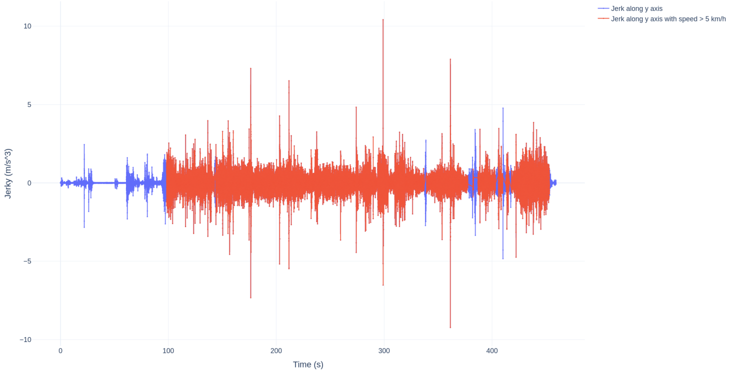


Figure: Jerk along y axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: jerk along z axis

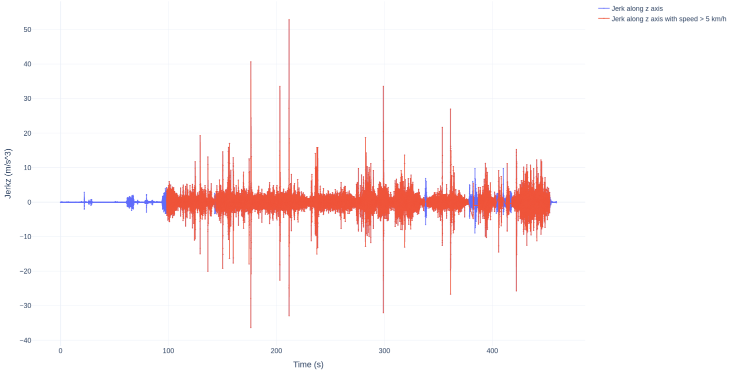


Figure: Jerk along z axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: orientation along x axis

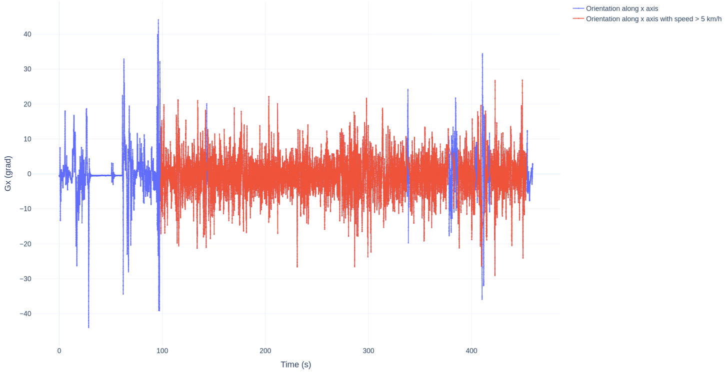


Figure: Orientation along x axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: orientation along y axis

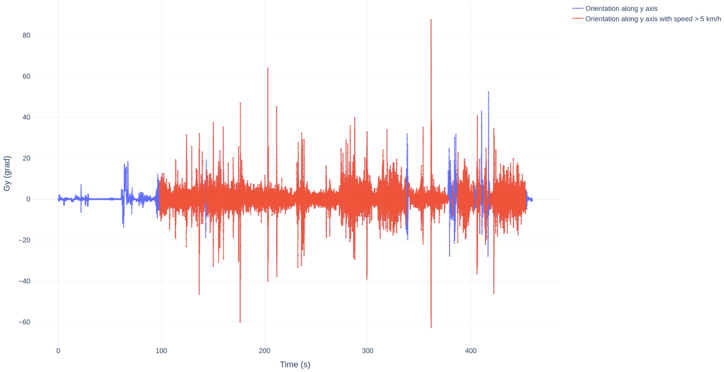


Figure: Orientation along y axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Removing low speed: orientation along z axis

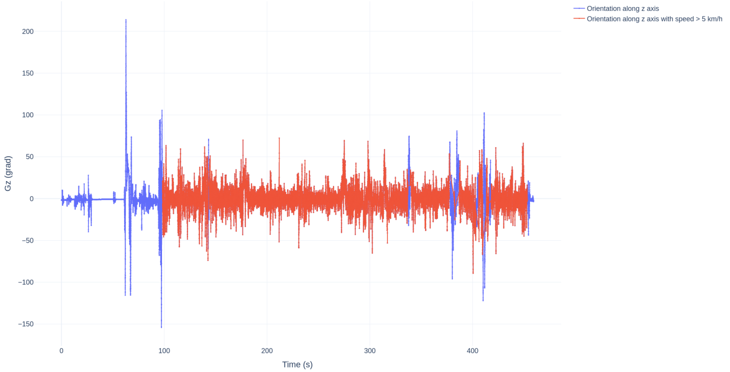


Figure: Orientation along z axis from Renzo and Jessica's recording with corresponging speed > 5 km/h

Filtering at 1 Hz

When extracting features in the time domain we are not interested in information contained above 1Hz.
Our data is already filtered at 20 Hz.

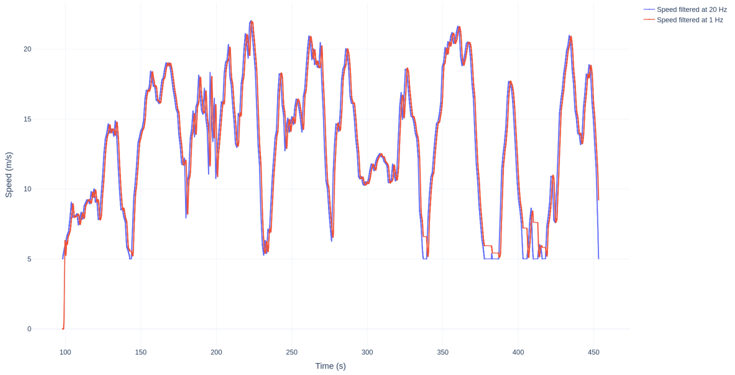


Figure: Speed filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: acceleration along x axis

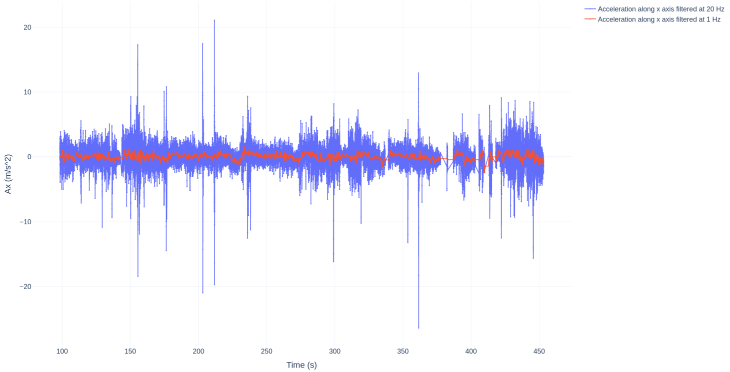


Figure: Acceleration along x axis filtered at 1 Hz from Renzo and Jessica's recording

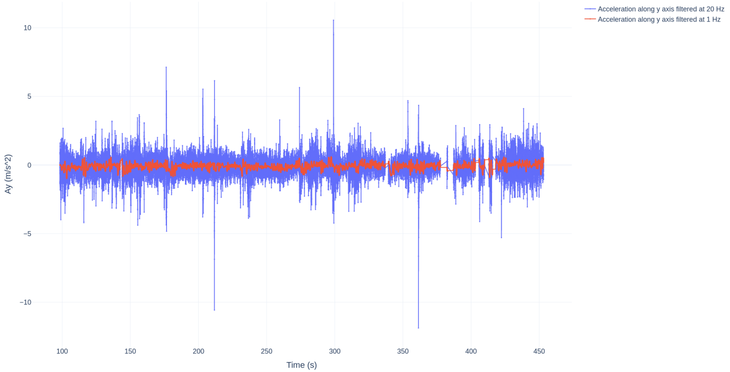


Figure: Acceleration along y axis filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: acceleration along z axis

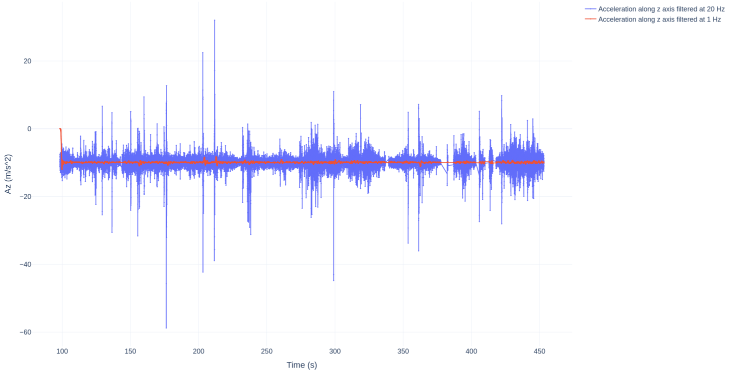


Figure: Acceleration along z axis filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: jerk along x axis

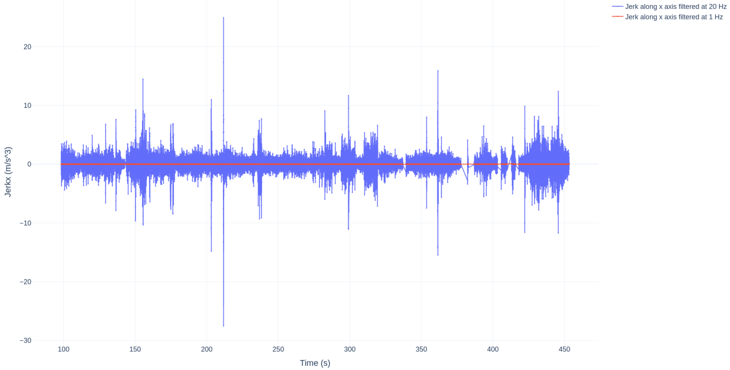


Figure: Jerk along x axis filtered at 1 Hz from Renzo and Jessica's recording

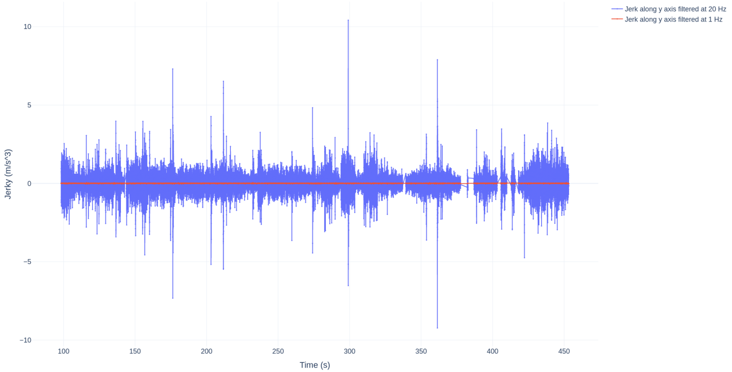


Figure: Jerk along y axis filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: jerk along z axis

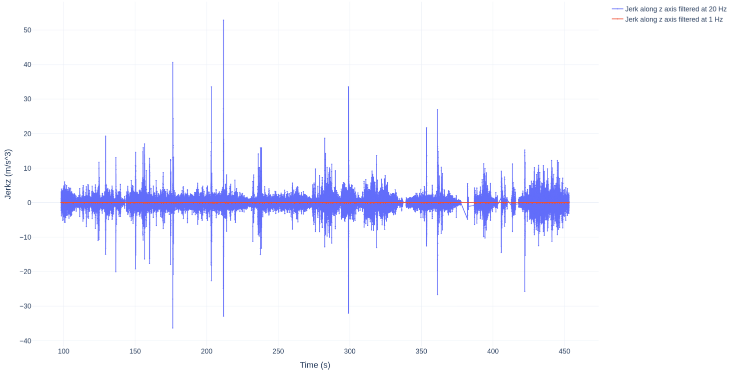


Figure: Jerk along z axis filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: orientation along x axis

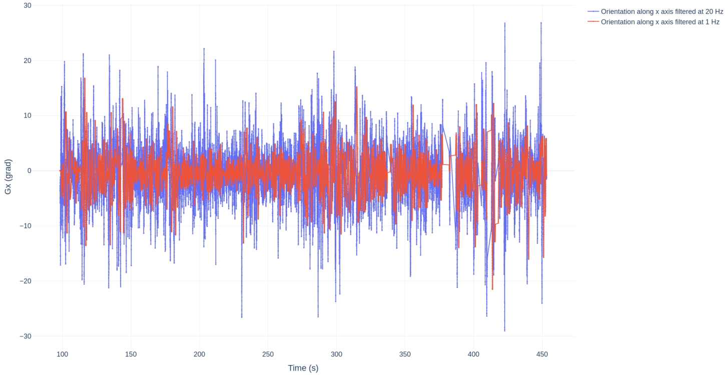


Figure: Orientation along x axis filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: orientation along y axis

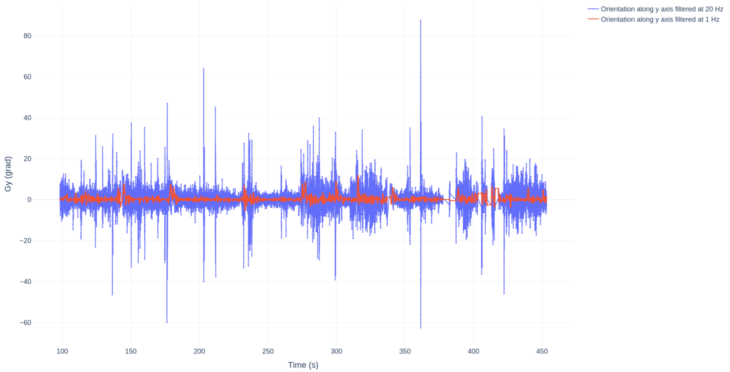


Figure: Orientation along y axis filtered at 1 Hz from Renzo and Jessica's recording

Filtering at 1 Hz: orientation along z axis

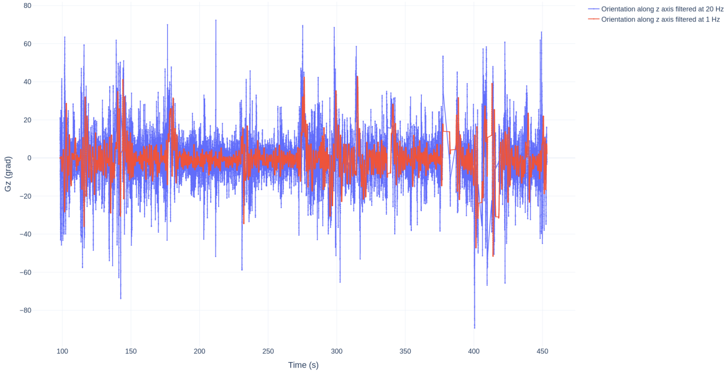


Figure: Orientation along z axis filtered at 1 Hz from Renzo and Jessica's recording

Section 3

Data Selection

Path simplification

Ramer-Douglas-Pecker algorithm can be used to downsample input data to remove redundant points on straight lines.
The only algorithm parameter is epsilon, the lower the more points are retained.
We decided non to use this approach because in the frequency domain a lot of relevant information is lost.

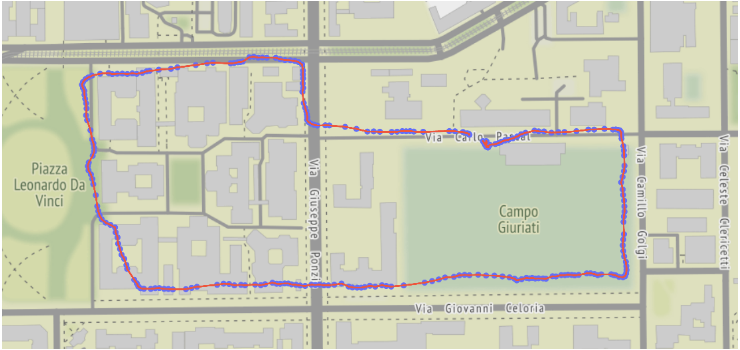


Figure: Simplified Path of Jessica and Alberto's recording (epsilon=1e-6)

Path simplification time domain: 0.7% samples

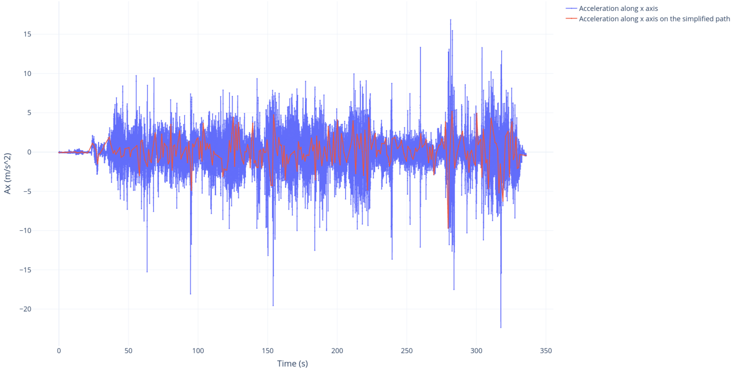


Figure: Simplified Acceleration of Jessica and Alberto's recording (epsilon=1e-6, 6000/768000 samples on the whole stem dataset)

Path simplification frequency domain: 20% samples

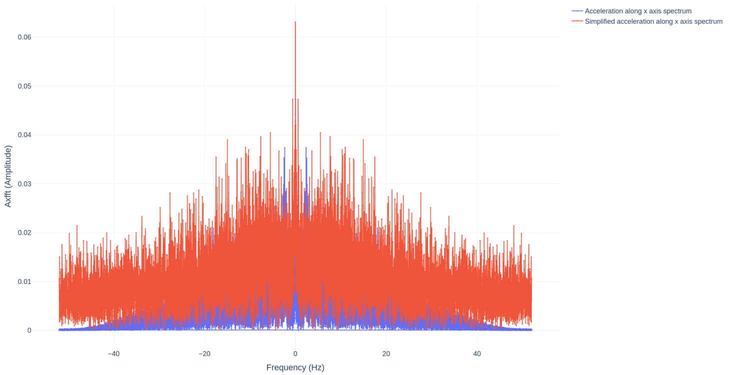


Figure: Simplified Acceleration of Jessica and Alberto's recording (epsilon=1e-14)

Path simplification frequency domain: 80% samples

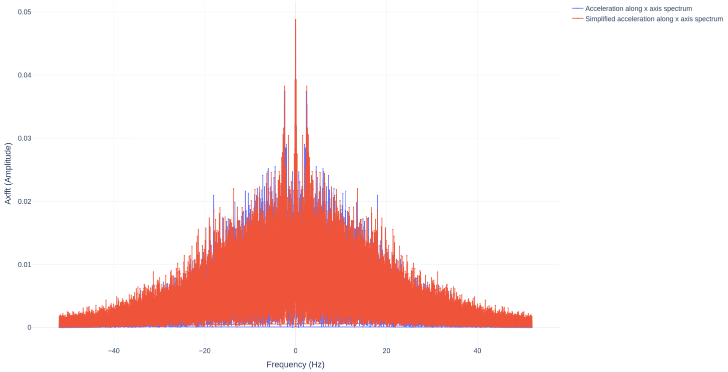


Figure: Simplified Acceleration of Jessica and Alberto's recording (epsilon=1e-16)

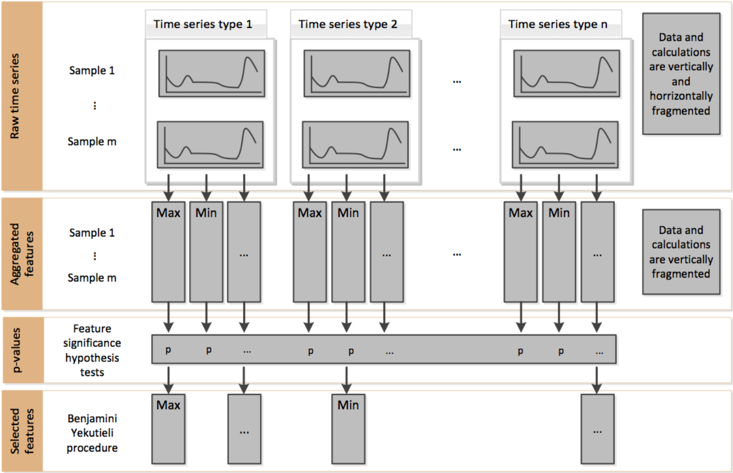
Downsampling

Regular downsampling, taking one every two samples, allows to halve the original sampling frequency while it does not effect too badly the signal in the frequency domain.

Section 4

Feature Engineering

Overview



Raw time series
Time series type 1
Time series type 2
Time series type n
Sample 1
⋮
Sample m
...
Data and calculations are vertically and horrizontally fragmented
Aggregated features
Max
Min
...
Sample 1
⋮
Sample m
Data and calculations are vertically fragmented
p-values
Feature significance hypothesis tests
p
p
...
Selected features
Benjamini Yekutieli procedure
Max
Min
...

Feature extraction

We employ a prominent python library for time series analysis: tsfresh.
Feature extraction can be carried out for each measurement through the tools provided by the library.
Extracted features:

- mean
- variance
- maximum
- minimum
- root mean square
- fast fourier transform area

Feature Selection: Significance

Each feature vector is individually and independently evaluated with respect to its significance for predicting the target under investigation.
Statistical tests employed:

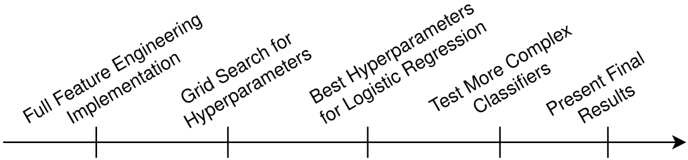
- Kolmogorov-Smirnov: binary valued feature (confidence) and real valued target (weight class)
- Kendall's tau: real feature and real valued target
- Two-sided univariate Fisher test: binary feature and binary valued target (two riders detection)
- Mann-Whitney U: real valued feature and binary valued target

Once the tests are applied we get a vector of p-values quantifying the importance of each feature.
Benjamini-Yekutieli procedure is applied to decide which feature to keep

Section 5

Roadmap

Roadmap


Full Feature Engineering Implementation
Grid Search for Hyperparameters
Best Hyperparameters for Logistic Regression
Test More Complex Classifiers
Present Final Results