

# Safety Assessment on Roslagsbanan Rail System, 2021-2022

This manuscript ([permalink](#)) was automatically generated from [uiceds/cee-492-term-project-fall-2022-time@9e91bad](#) on October 22, 2022.

## Authors

---

- **Cecilia Karina Volpe Baridon**  
· [cvolpebaridon](#)
- **Elie Roudiere**  
· [elieroudiere](#)  
Railway Group, KTH Royal Institute of Technology - Stockholm, Sweden
- **Berkant Usta**  
· [berkanusta](#)

## Dataset Description

The Roslagsbanan is a railway system operating in Stockholm. As part of an improvement in the transportation system, Stockholm Municipality invested in a new fleet of trains switching from X10p trains to the new X15p. In order to analyze the safety of the new trains, brake tests were conducted. During these tests, the Latronix Track Measurement (LTM) system was mounted on the train. The LTM uses laser beams, and an onboard computer to perform measurements over a railroad track while the train drives on the track at normal speed.

The data that will be used in this study was collected using the LTM system affixed to a Roslagsbanan vehicle. While the train was running over the sections, the level of each rail, alignment, curvature, cant, gauge, and twist were measured every 256mm. The data was compiled in csv files and was collected between October 2021 and May 2022 with an approximate interval of one month. In other words, a given section of track typically has around six measurement files. The total amount of data is 60 CSV files of 30,000 to 80,000 rows each, for a total of 1.5GB of data.

*These data and measurements are property of Latronix AB of Sweden, and are used with their permission for research and educational purposes only.*

In the CSV files the structure of columns is as follows:

- **Marker and Marker\_Offset:** The kilometre post of the data point, where Marker is in km and Marker\_offset is in m. In other words, if Marker=20 and Marker\_offset=250.5, the kilometre post is 20,250.5
- **Coordinate\_northing, Coordinate\_easting, Latitude, Longitude:** As their name suggests, they are the coordinates of the data point in northings and eastings, as well as latitude and longitude.
- **Speed:** records the speed in km/h at which the measurement was taken. This is especially relevant as different operating speeds give different tolerances in terms of the measured properties.

- Level\_left and Level\_right D0, D1, D2, D3: The deviation of the longitudinal level (height) of the left and right rail in mm, compared to its reference position. D0, D1, D2, D3 correspond to different measurement wavelengths of the same data.
- Alignment\_left and Alignment\_right D0, D1, D2, D3: The deviation in longitudinal alignment ("straightness") of the left and right rail in mm, compared to its reference position. D0, D1, D2, D3 correspond to different measurement wavelengths of the same data.
- Track\_gauge: The deviation in the spacing between the rails in mm, from a reference of 891mm.
- Curvature: The curvature of the track at that location, in  $m^{-1}$ .
- Cant: The cant (inclination) of the track at the location, in mm from the horizontal.
- Twist\_3m, Twist\_6m: The rate of change in cant over a 3m or 6m distance.
- Level\_versine\_left\_5m, Level\_versine\_right\_5m: Derived from Level\_left and Level\_right, this level deviation in mm corresponds to longitudinal versine measurement of the top of rail over a 5m distance. This historical metric is more commonly used in railway engineering and is thus provided.
- Alignment\_versine\_left\_10m, Alignment\_versine\_right\_10m: Similarly to the versine level, this measure of alignment in mm uses versines over a 10m distance.

Finally, Track\_name, Track\_class, Station\_flag and Link\_name, Pos\_unfiltered Pos\_filtered NTP\_sync, Cant\_D1, Std\_level, Std\_alignment, Std\_cant, Quality\_1 are mainly for internal use and not part of the analysis.

## Project Proposal


The main goal of this project is to identify patterns in the data, especially the track degradation of certain sections, and thus to analyze the change in railway safety during this time. The first approach will be to define degradation trends and build a model to process the data that finds them automatically. Then, the study will attempt to find external factors which caused the observed trends. The analysis will include graphics of the parameters over time as well as comparing them over the sections to provide a better understanding of the data visually.

This research will help assess system performance, identify possible maintenance schedules, and provide useful data for further research in this area. The study also falls within the scope of developing predictive maintenance in civil engineering and railways. Creating and improving analysis and predictive maintenance models such as this one can contribute to increase safety of rail networks and a higher quality of service for passengers. They will benefit from fewer unplanned service interruptions due to track maintenance and even fewer unsafe tracks causing line closures.

## Exploratory Analysis

Dataset description: The main dataset is composed of 60 .csv files, dated between October 2021 and May 2022. These files correspond to measurements performed on a railway track forming part of the Roslagsbanan network in Stockholm, using a track measurement device affixed to a train running on the track during test runs. Each file corresponds to a certain measurement run, over a given section of track, and different sections of track have varying amounts of measurement runs. The reason for this

is that the engineers monitoring the track wished to monitor certain areas more closely, thus doing more frequent measurements.

Line map overview of the Roslagsbanan. In the files, the metrics listed in the introduction are recorded, but it should be noted that the files require cleaning as they comprise numerous N/A cells, as well as somewhat questionable datapoints (especially when the train was coming to a stop, some outliers seem to have been generated [plot showing what happens with 0 speed?]). The date and time of the measurement is not hardcoded inside the csv but only mentioned in the file name, as such the code for analysing will have to take into account the name of the file when making comparisons. Measurement runs are not perfectly synced with each other as shown on [figure], they will have to somehow be matched or correlated to clearly interpret the results.

[Stats summary about size, distance covered, frequency of measurements, others?]