Safety Assessment on Roslagsbanan Rail System, 2021-2022

Authors

- **Cecilia Karina Volpe Baridon**
  · 🕽 [cvolpebaridon](#)

- **Elie Roudiere**
  · 🕽 [elieroudiere](#)
  Railway Group, KTH Royal Institute of Technology - Stockholm, Sweden

- **Berkan Usta**
  · 🕽 [berkanusta](#)

Dataset Decription

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 latittude 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 pay closer attention to certain areas, thus doing

more frequent measurements. In particular the section between kilometre point 24 and 34 comprises over 20 distinct runs (often going northbound and then southbound on the same day, on this single track section)
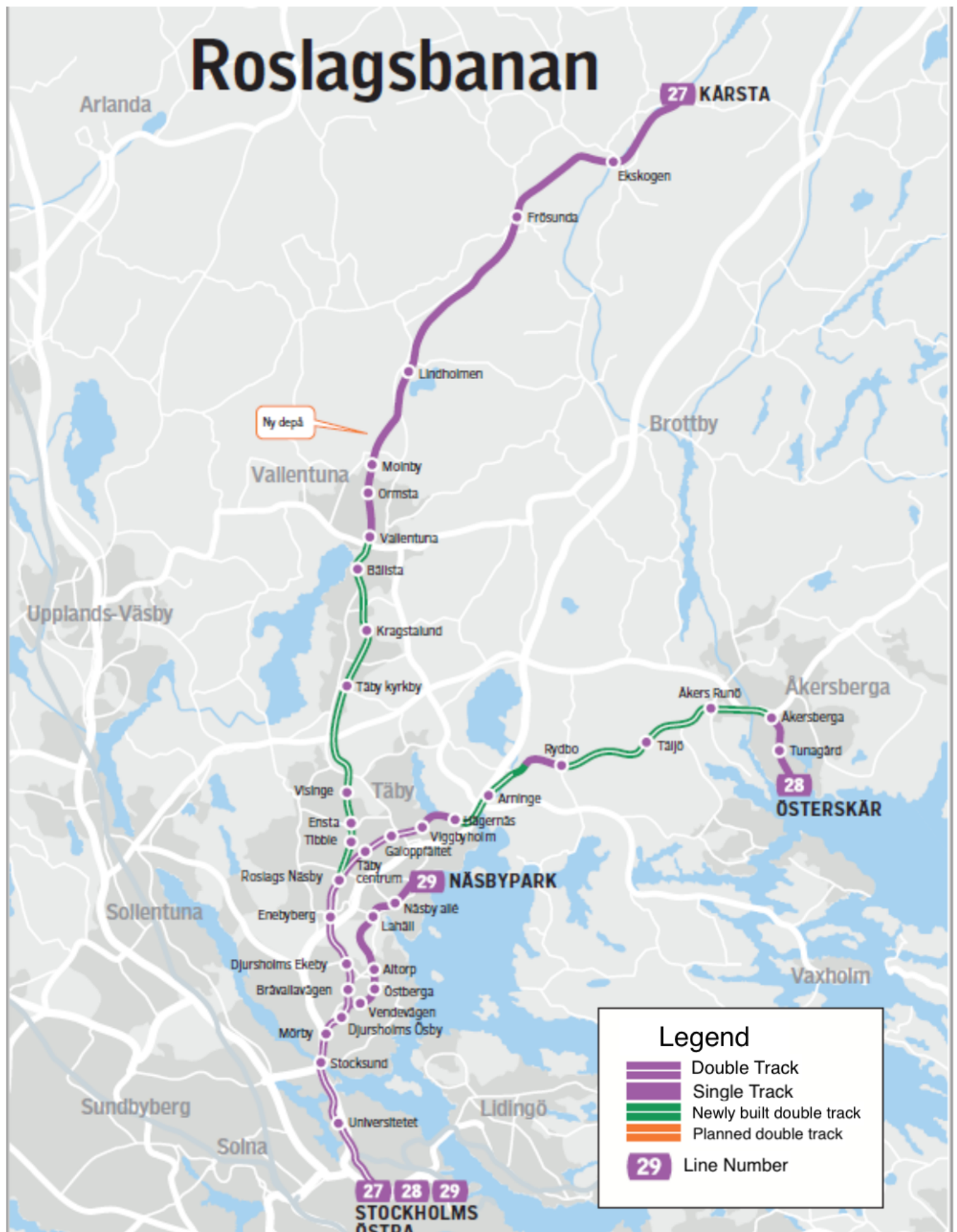

Figure 1: Line map overview of the Roslagsbanan. This study will focus on line 27 to Kårsta

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?]

Statistic	Value
Earliest measurement	21st October 2021
Latest measurement	10th May 2022
Line Length	41.700 km (25.91 miles)
Data points per file	20,000 to over 100,000
File size	

Table 1: Summary Table. {#tbl:sum-table-1}

Predictive modelling

Due to the somewhat limited amount of independent variables already present inside the measurement set, there may be challenges. Correlation plots can help us determine links between variables and what model to create.
The first part of the predictive model will try and find locations where the track degradation seems to form a clear trend (typically a linear increase in a given direction) over time. Here, time is the independent variable. From an initial analysis of the dataset, it is likely this predictive model will yield some results (picture of a steady trend?) The second part of the predictive model we would like to create would take into account external factors, in particular temperature/weather changes and soil properties if obtainable. This could give us locations prone to developing trends in the short to medium term, which could be investigated by a maintenance team. However we have envisioned some possible issues, namely that obtaining and translating soil data into something usable by the model may be a challenge, and no link is guaranteed to appear (what if all soils are similar?). A possible approach could be to grade soils based on properties, for example a very stable soil could get a score of 1 and a worse soil from the point of view of building a railway line could get a score of 4 (or higher/lower), with appropriate intermediate values. We could look into D30/D10/D60 values.