Investigations of road traffic and delays on the railroad

Kristofer, Maria, André, Joël

University of Oslo

October 20, 2020

Project Goal

- Model and predict train delays and road traffic
- Investigate the relation between them
- In and around Oslo

Motivation

- Many people rely on trains
- Changed travel behavior?
- Negative effect on the environment from car use
- Get from A to B efficiently

Potential business partners

- Transport and Environment Department of Oslo kommune
- Jernbaneverket and NSB
- Taxi companies
- The public

Datasets

- Weather Data
- Train Data
- Road traffic data

Datasets: Weather data

- Scraped from the yr.no website's historical data using BeautifulSoup and regular expressions.
- We used the Blindern weather station which had the most detailed data for the Oslo region.
- Data gathered into MongoDB, then transformed to CSV.

Datasets: Train data

- SQLite database file received from Jernbaneverket.
- One entry each time a train stops at a station, from 2012 to 2016.
- When the train should have arrived, when it did arrive, and where.
- Too large to handle directly, so we aggregated delay data per day and hour.

Datasets: Road traffic data

- Received from Vegvesenet.
- Hourly counts of number of cars passing in each lane of the road.
- Twenty one different locations.
- Also counts by car length categories.

Map of counting points

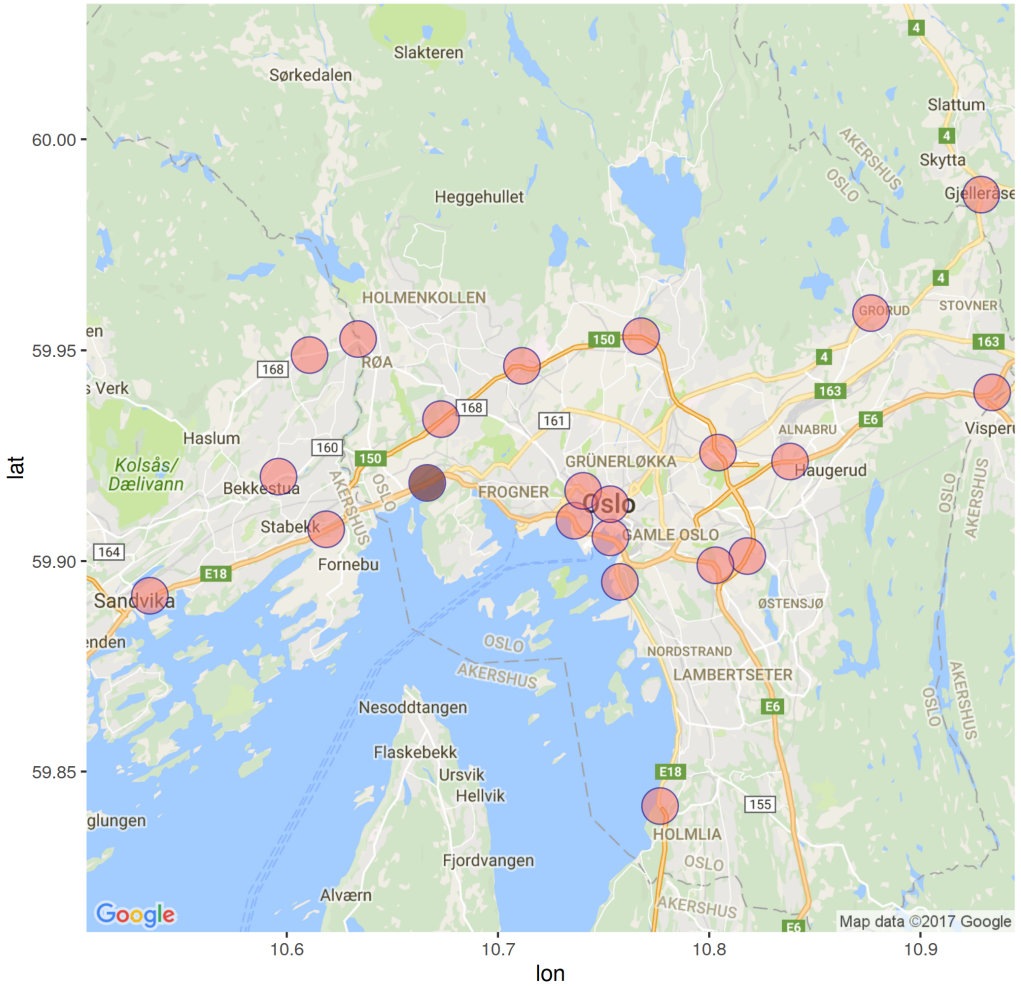
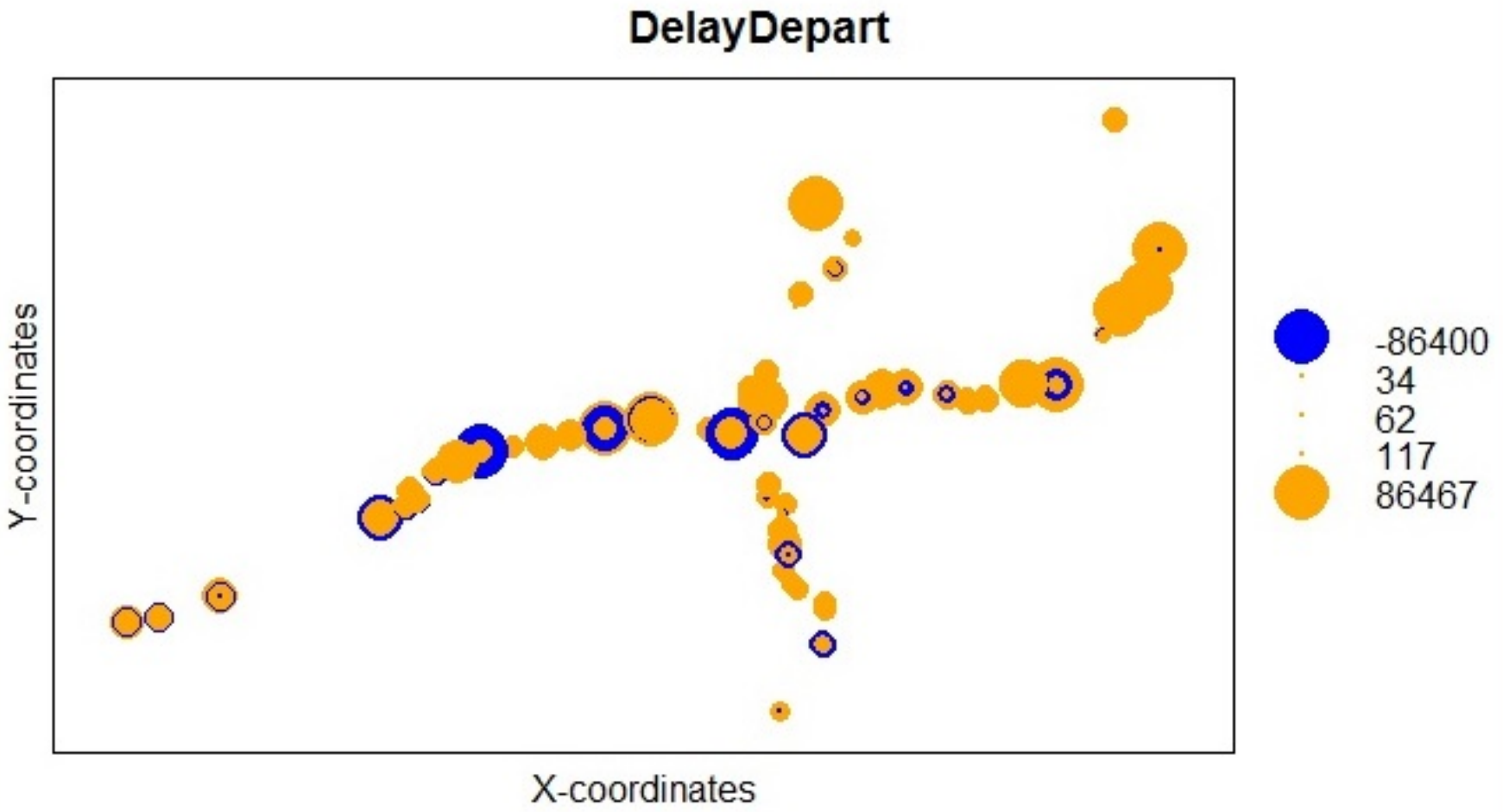


Figure: Road traffic measurement points.

Bubble plot of train delay.



Train delay data

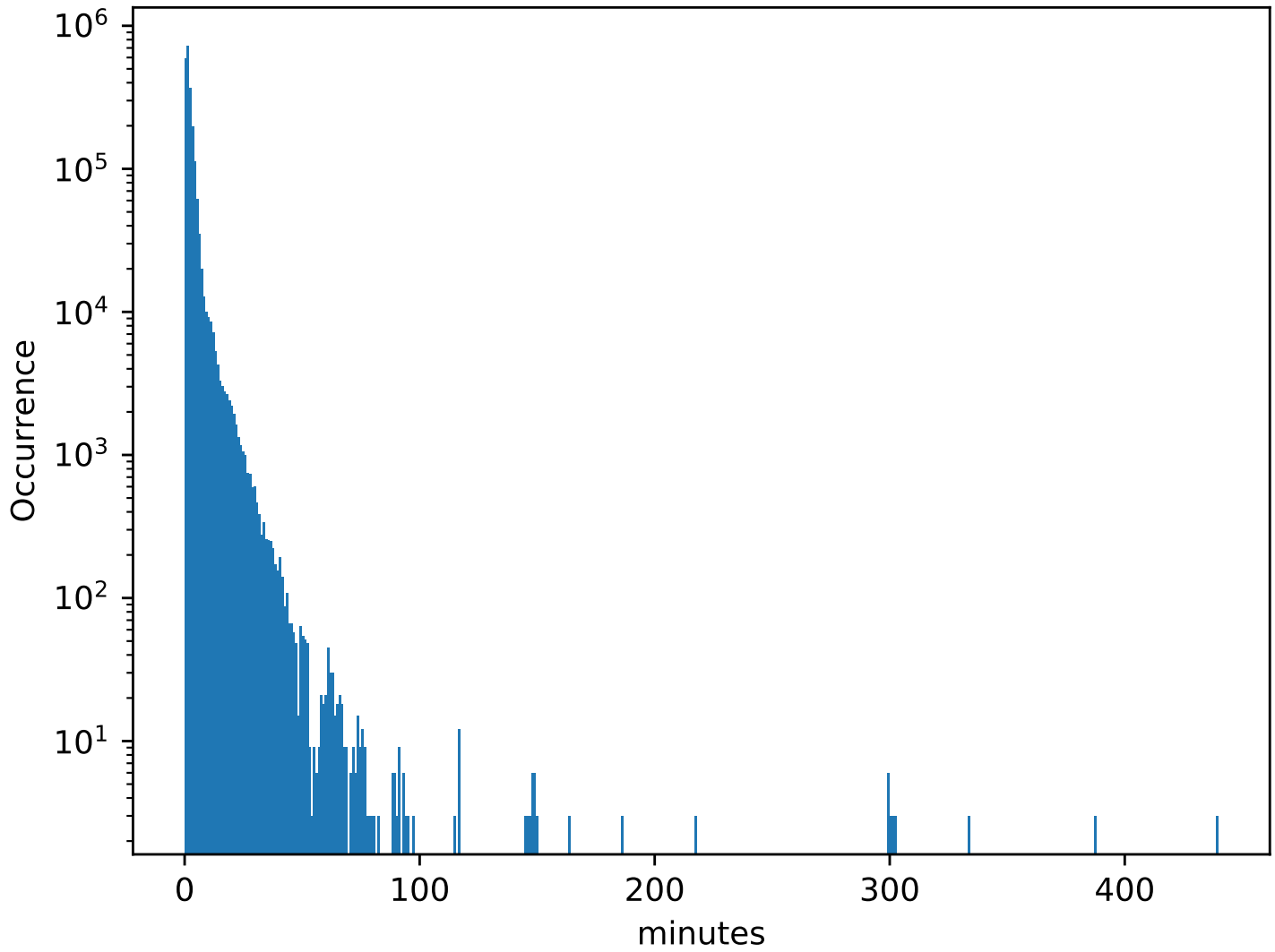


Figure: Distribution of train delay on the whole network.

Predicting road traffic from train delays

- Recap: Try to predict road traffic at certain detectors ("counting positions").
- I found that median daily train delays weren't useful, but time of year was.
- But what about using only train delay data from stations near the road in question?
- I started doing this, which entailed quite a bit of aggregation work.
- First attempt: mean of train delays among closest four stations, for each counting point
- Second attempt: number of delays in intervals [-2...0) hrs, [0...1) min, [1...5), [5...15), [15...∞), as well as number of cancellations. Lower AIC.

Predicting road traffic from train delays

- Counted the number of delays in total for each interval for all train stations within 8 km
- Final models found by thinking and using subset selection:
  - Baseline model: same as in midterm (only weekday, month, holiday, easter effects)
  - Model with weather: Baseline, plus amount of snow in the day in question, weather type, middle temperature
  - Full model: Model with weather, plus number of delays within 8 km
- Fitted model to 2012-2015 and predicted for 2016.
- Weather model much lower RMSE than baseline (5860 vs 6210)
- Full model had worse RMSE than both at 6450
- But likelihood ratio test shows significance between full and weather-only models with better likelihood for the full model

Traffic prediction, plot 1

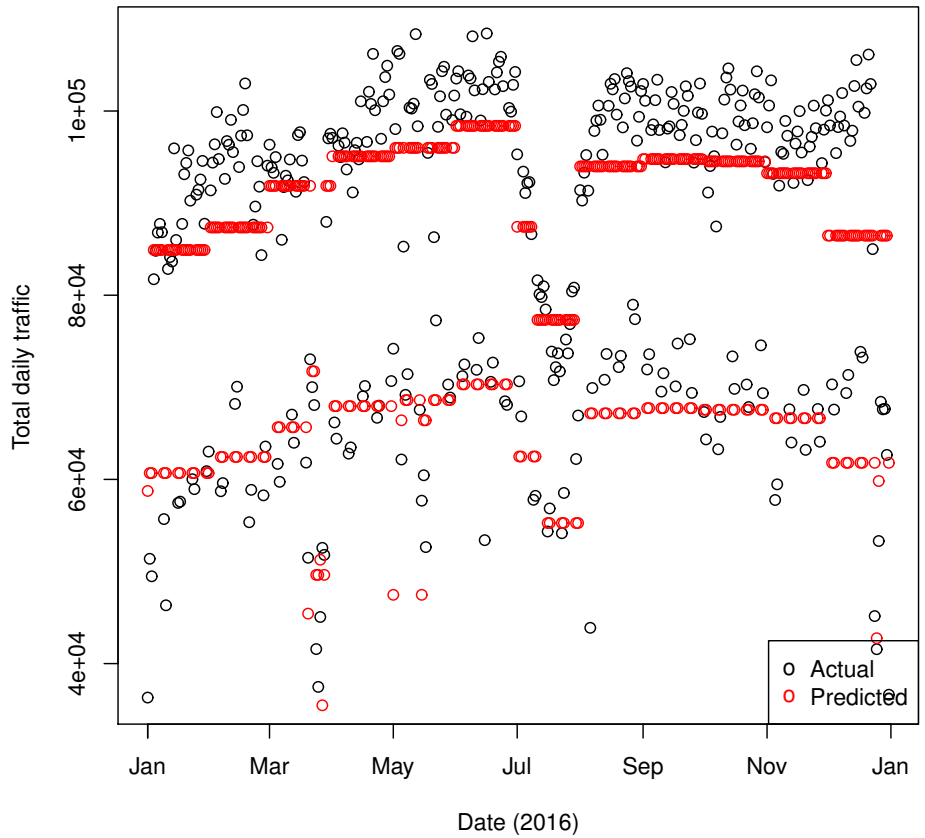


Figure: **Predicted and actual traffic for Maritim 250B, base model**

Traffic prediction, plot 2

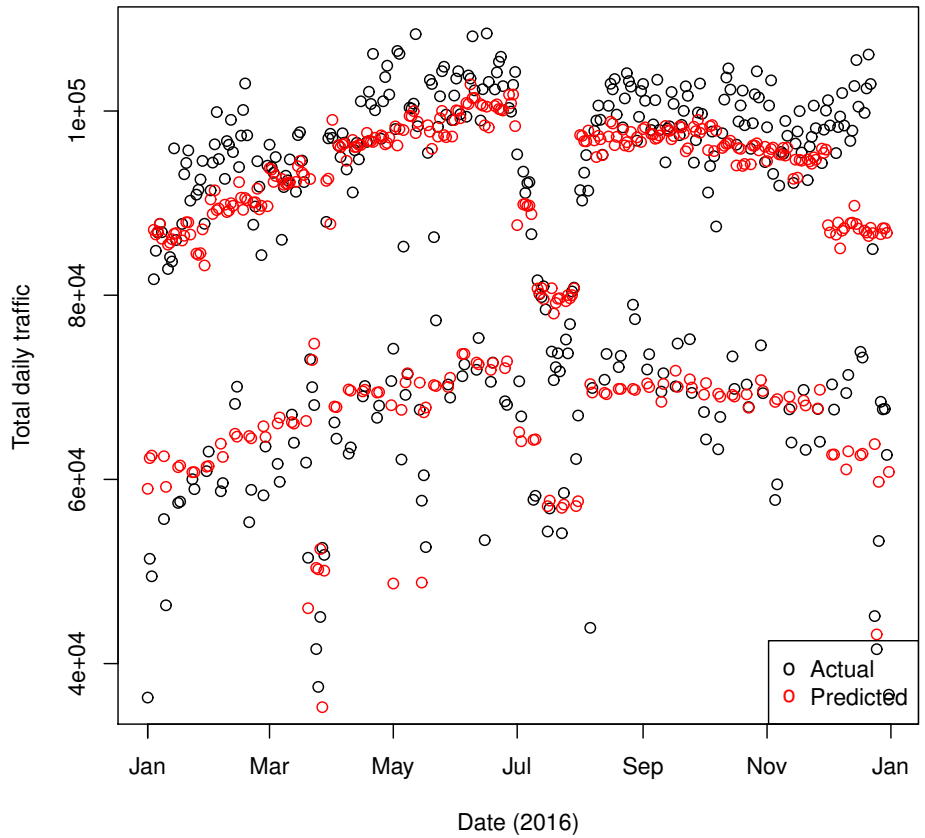


Figure: **Predicted and actual traffic for Maritim 250B, weather model**

Traffic prediction, plot 3

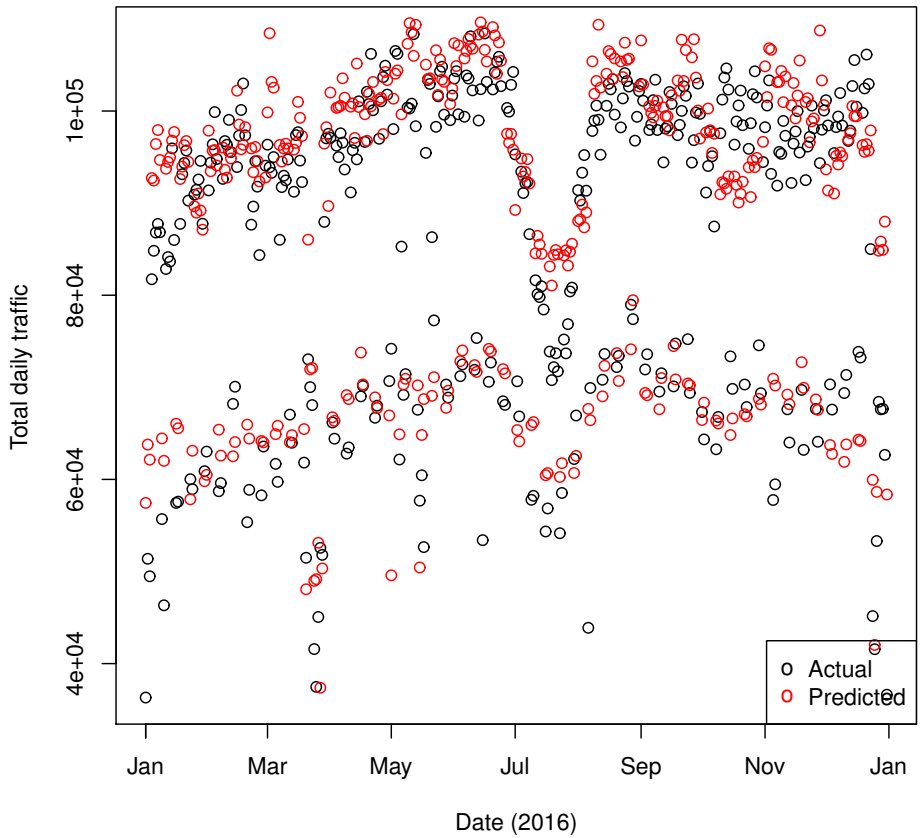


Figure: **Predicted and actual traffic for Maritim 250B, full model**

Predicting road traffic from train delays

- Small effects
  - One additional cancellation: 0.033% increase in traffic.
  - In contrast: Weekend: 27% reduction compared to weekdays
- Possible reasons
  - Roads have finite capacity (traffic jams).
  - Time lag?
- Difficult to use the model for prediction when the effects are small.
- Easier with large effects like weather and holidays!

Predicting delays on the railway

- Predicting delays within zone 2.
- Predicting average number of minutes delayed.
- Hourly prediction.
- Splitting the network into 6 groups.

Train line map

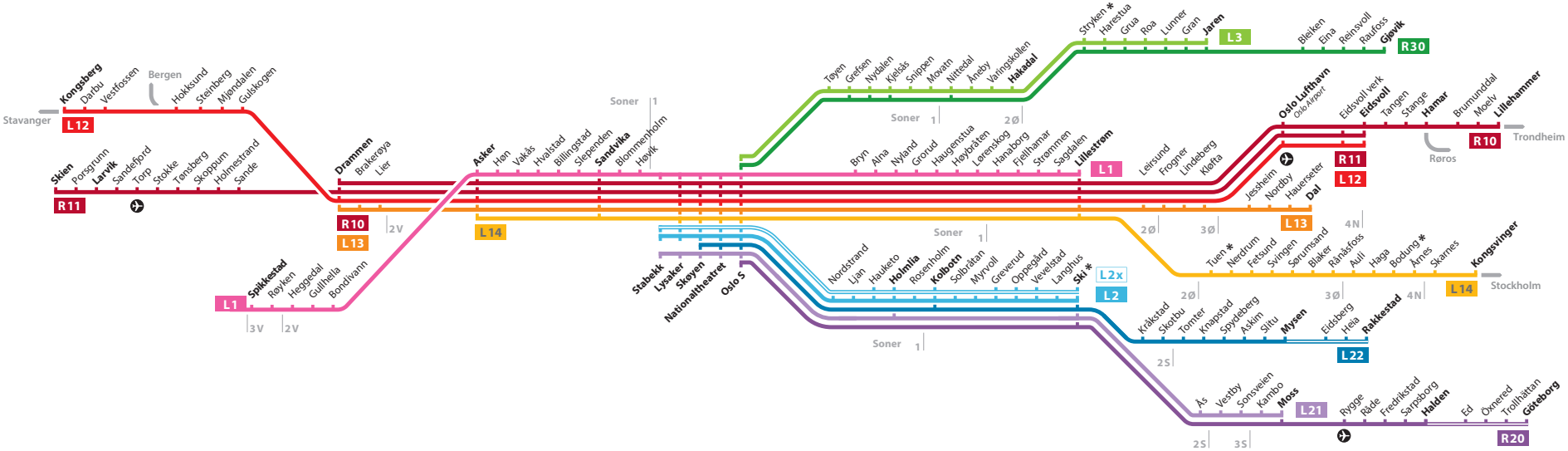


Figure: Visualization of train lines in the Oslo region. Lines are categorized into three main branches.

Finding delay till next train

- 2/3 of the trains in the data have line numbers.
- Increased number to 91%, by matching visiting stations with lines.
- Direction is found by which order trains visits the stations.
- Finds delay till next train arrives at a station, if a train is delayed.

Train delay prediction model

- Tried two models: Gradient boosted decision trees and neural network.
- Training on data from 2012 to 2015 and testing on 2016.
- Input data:
  - One hot encoded weather data.
  - Historical data periodicities: 1 hour, 1 day and 1 week.
  - Historical data: max and mean delay, number of trains.

Decision trees vs. neural networks

Model	1 w	1 e	2 w	2 e	3 w	3 e
MLP	606	455	410	404	512	337
GBT	605	473	417	410	518	337

Table: L_2 norm of the error(in minutes) of testing data for a multilayered perceptron model and a gradient boosted tree model.

Prediction of train delays

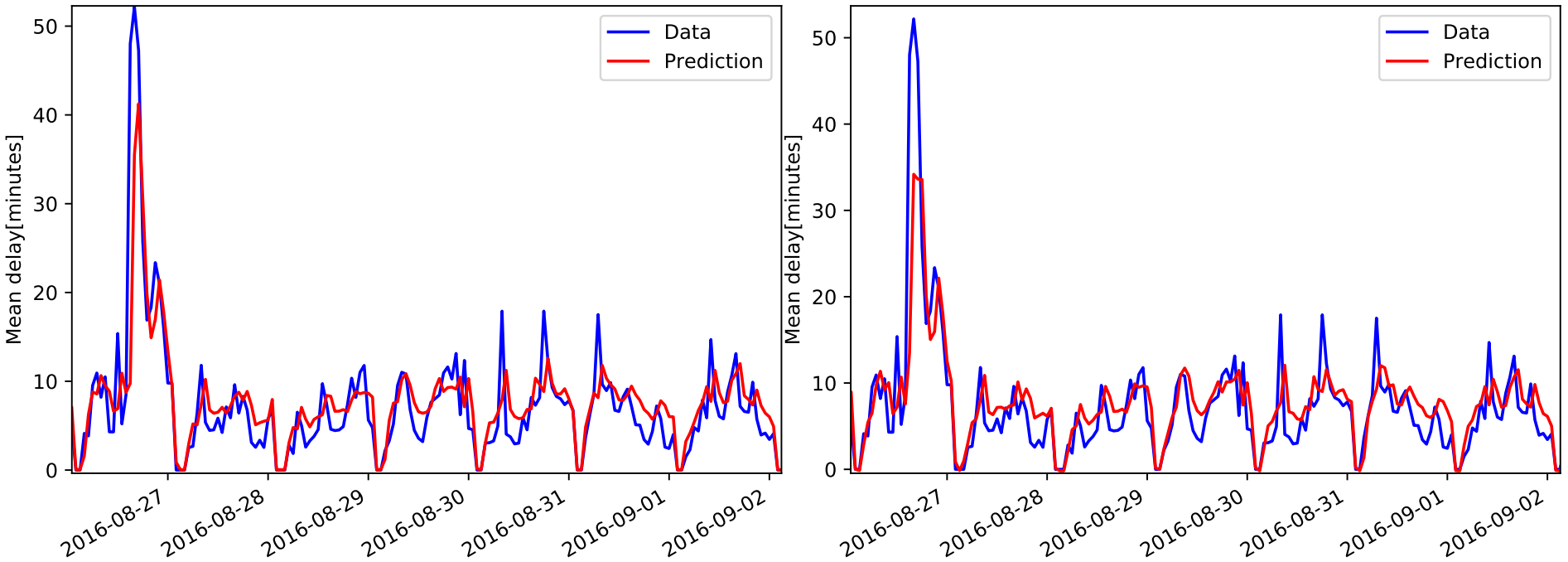


Figure: Predicted delay vs. measured data for line 1 heading west. Neural network to the left and gradient boosted trees on the right.

Did we meet our goals?

- Models to predict train delays
- Models to predict road traffic
- Investigated the impact of train delays on road traffic
- Compare the two models to suggest mode of transportation

Thank you!