Predicting response times of the Paris Fire Brigade vehicles

Antoine Moulin

Master student
Department of Applied Mathematics
ENS Paris-Saclay

January 13th

Contents

I. Introduction

II. Data Analysis

III. Feature engineering

IV. Model

V. Conclusion

Introduction

Quick briefing

- Response time: turnout + travel

- Besides the IDs, 3 groups of features:
  - Vehicle features
  - Intervention features
  - Spatial-temporal features

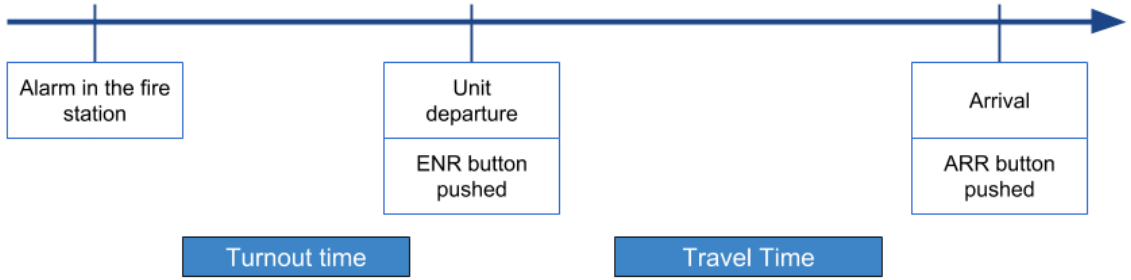


Figure 1. Definition of the response time.

Data Analysis

Response time distribution

- After cleaning, we notice both the turnout and travel times seem to follow a Rayleigh distribution
- Average turnout time: 2:18 min
- Average travel time: 5:48 min

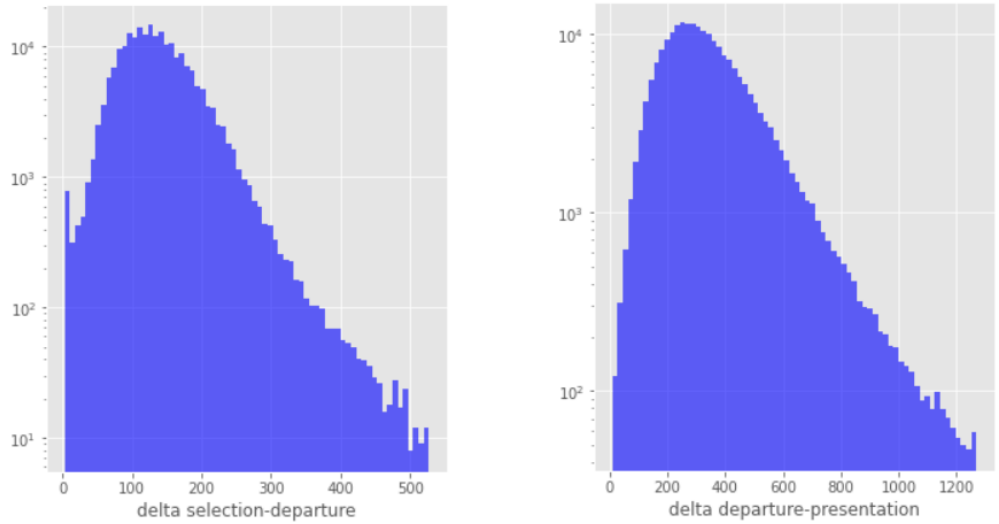


Figure 2. Distribution of the response time.

Vehicle features

- Emergency vehicle type (category)
- Status preceding selection (boolean)
- Departed from its rescue center (boolean)

Vehicle features

- 70 types of vehicle
- Can we distinguish slow and fast vehicles?

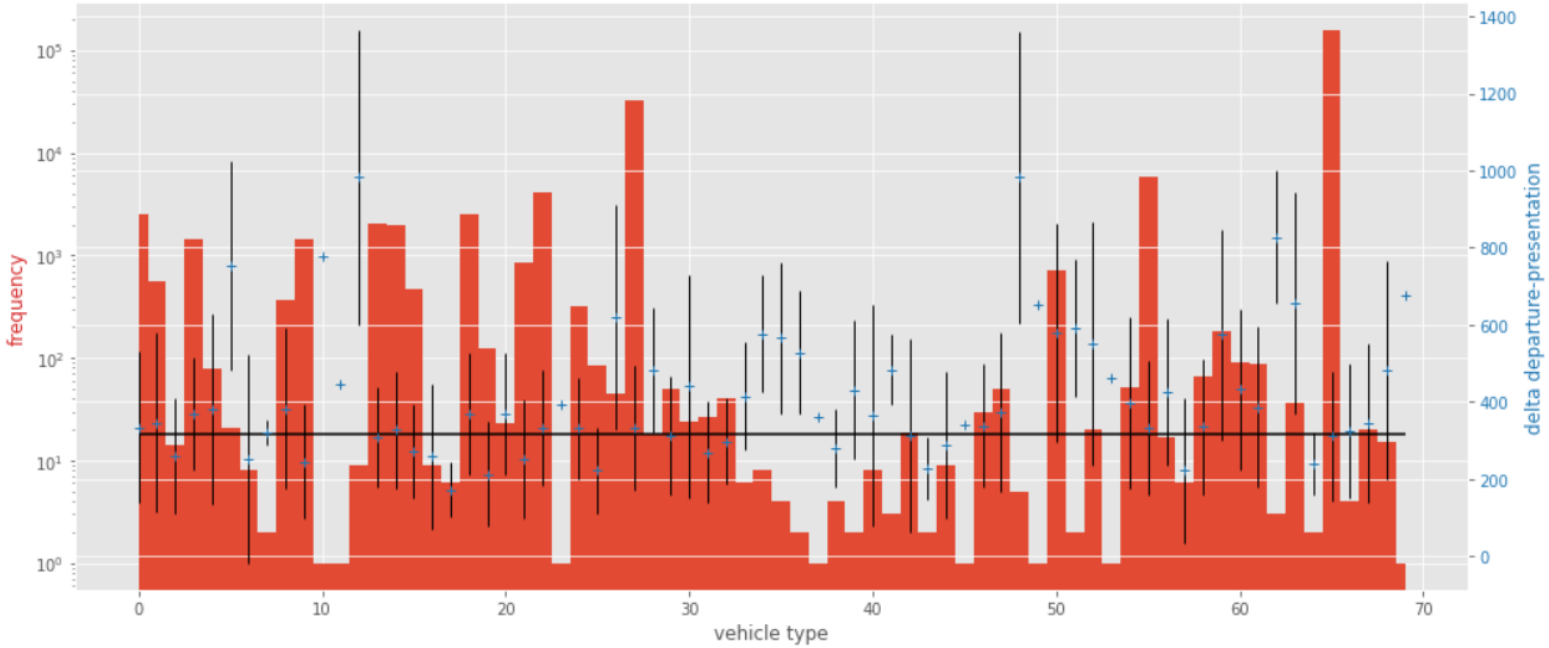


Figure 3. Travel time depending on the vehicle type.

Vehicle features

- Reduce to 50 types when using the first word
- Still issues with some types...
- When reducing to 30, 20 or 10 types, the problem remains

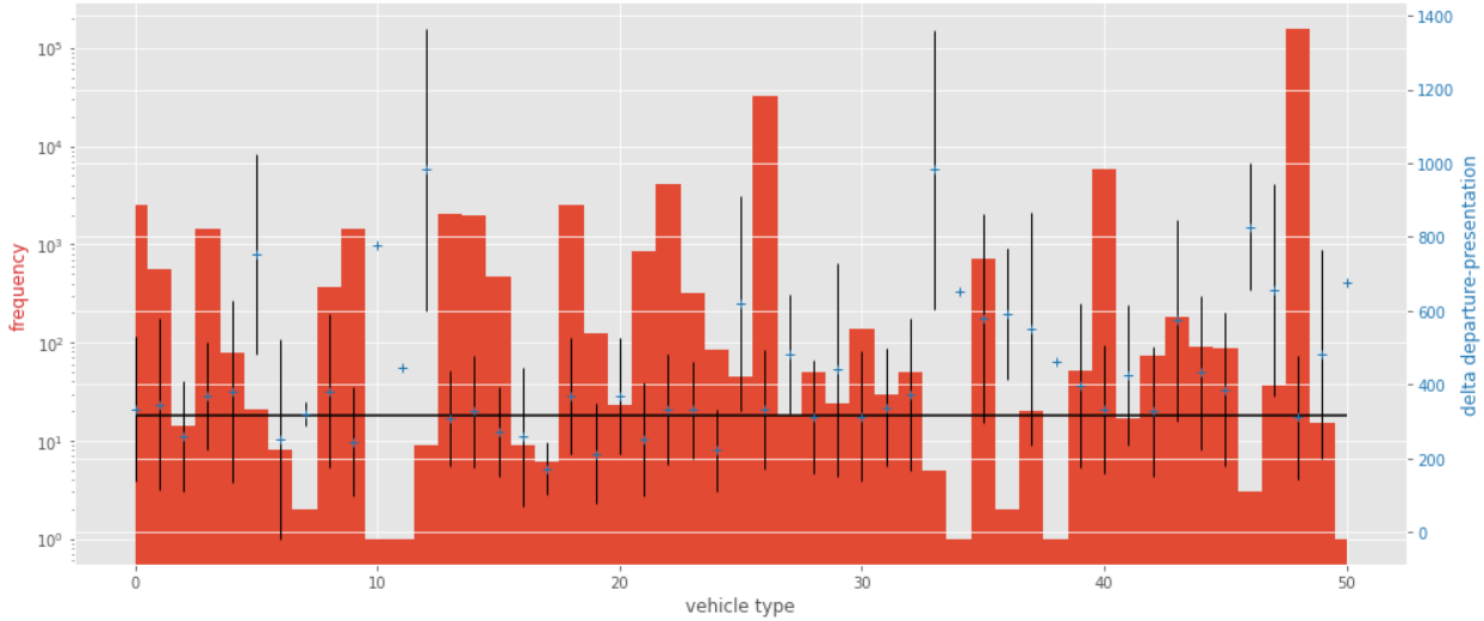
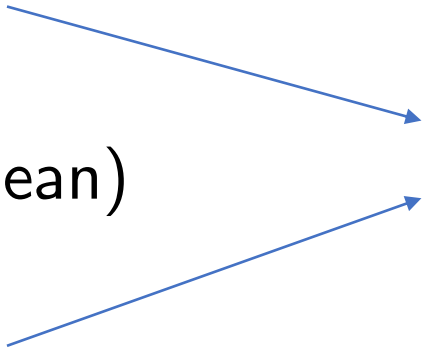


Figure 4. Travel time depending on the vehicle type after grouping.

Intervention features

- Alert reason (category)
- Alert reason category (category)
- Intervention on public roads (boolean)
- Location of the event (category)
- Floor (int)



Too much values to be useful...

Intervention features

- We notice a difference between the 9 categories of the alert reason feature

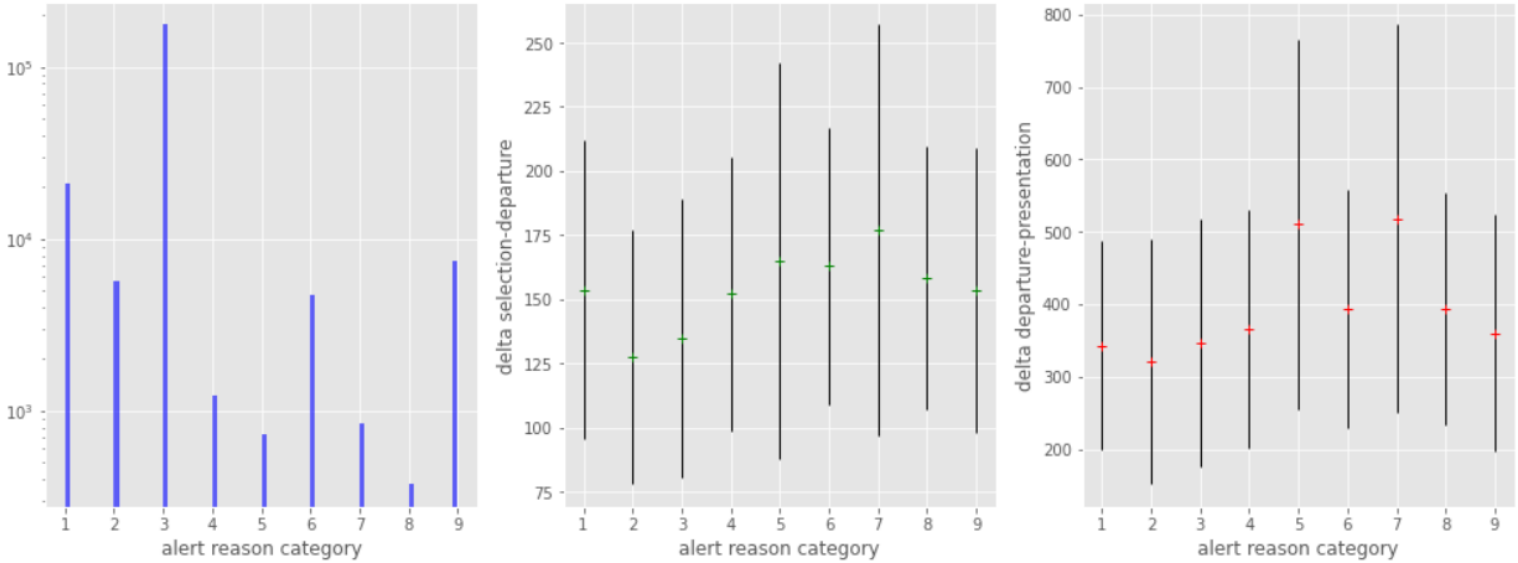


Figure 5. Times depending on the alert reason category.

Spatial-temporal features

- Longitudes/latitudes (float)
- GPS tracks (list of positions or dates)
- OSRM response (estimation of distance and duration)

Spatial-temporal features

- The interventions span most of the region Ile-de-France
- Centers are well distributed
- The vehicle selected is not always in the closest center

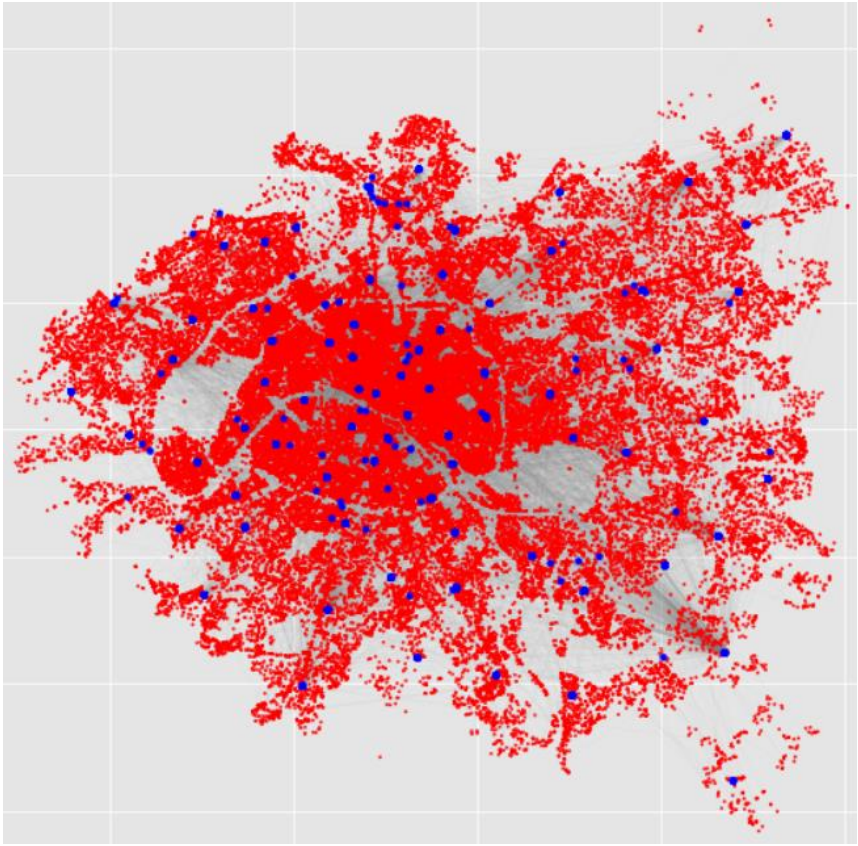


Figure 6. Repartition of the interventions and the rescue centers.

Spatial-temporal features

- The OSRM estimations are very useful but not always accurate
- Additional data can be used (see next section)
- GPS tracks can be used (see next section)

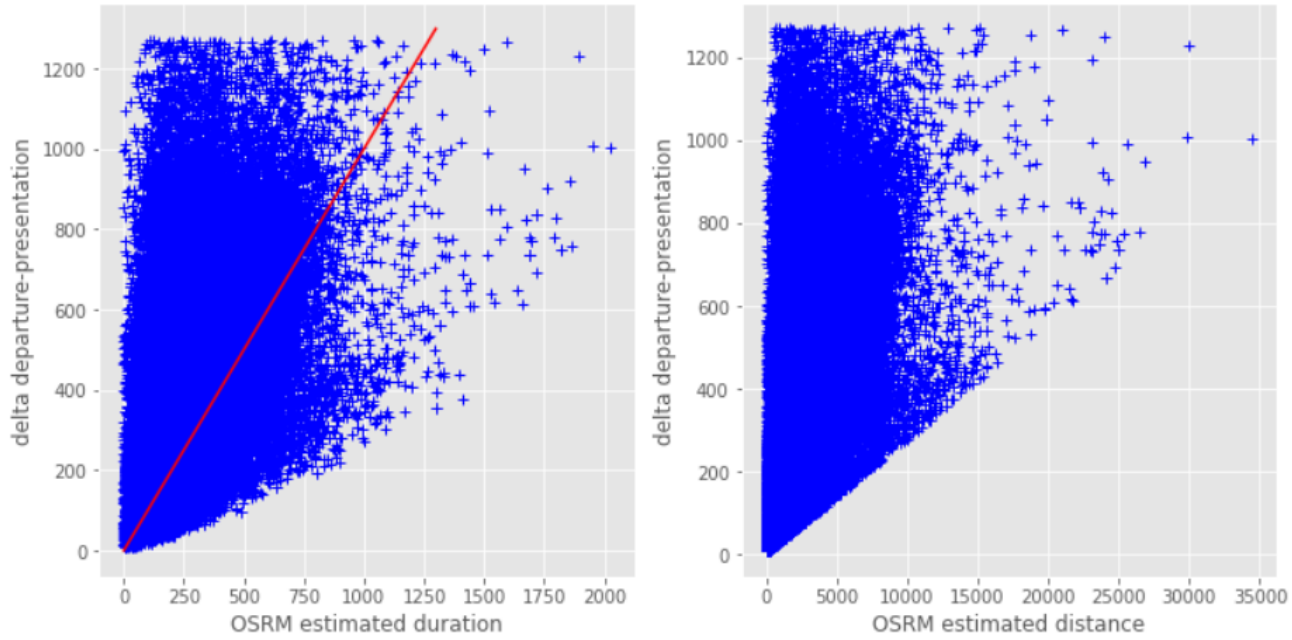


Figure 7. Travel time as a function of the OSRM estimates.

Feature engineering

Seasonality

- Retrieve the hour, the day and the month of the intervention
- Working hours
- Weekend or not
- Hollidays or not

Figure 8. Travel time depending on the hour, day or month, respectively.

Meteorological data

- Weather influences both times
- Scrapped the data for 2018

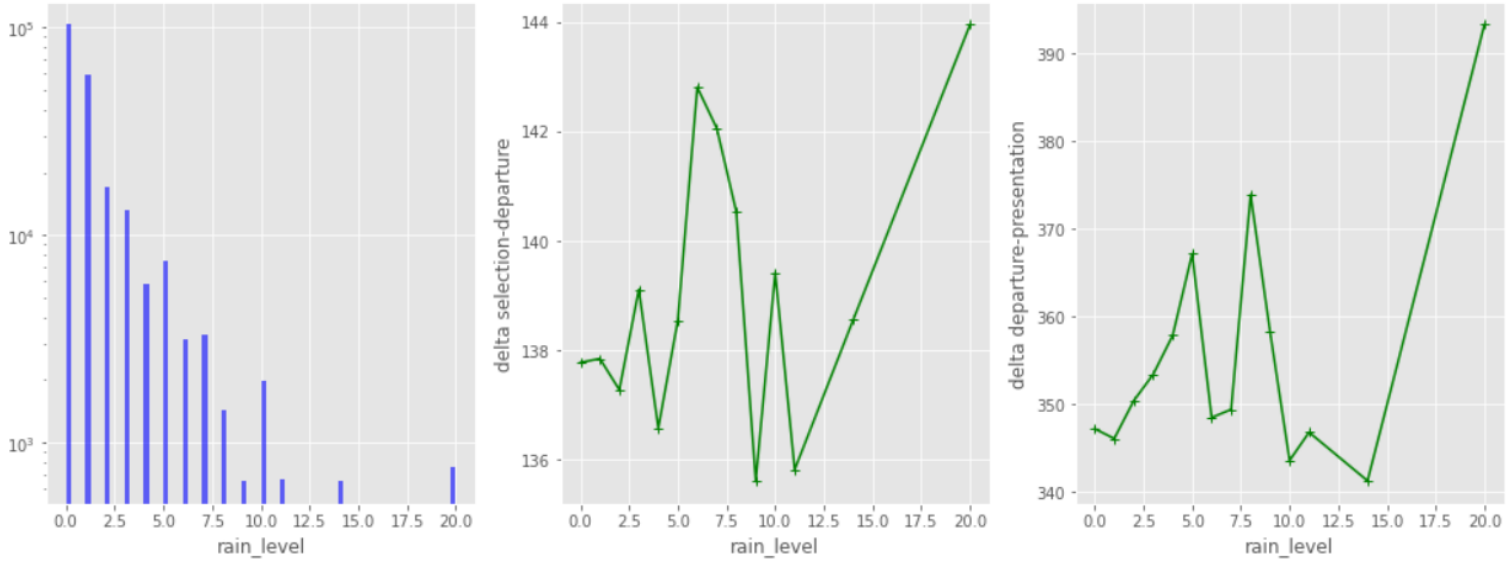


Figure 9. Travel time depending on the rain level.

Using longitudes and latitudes

- Compute Haversine and Manhattan distances
- Compute the direction
- Compute the average speed
- Cluster to find neighborhoods

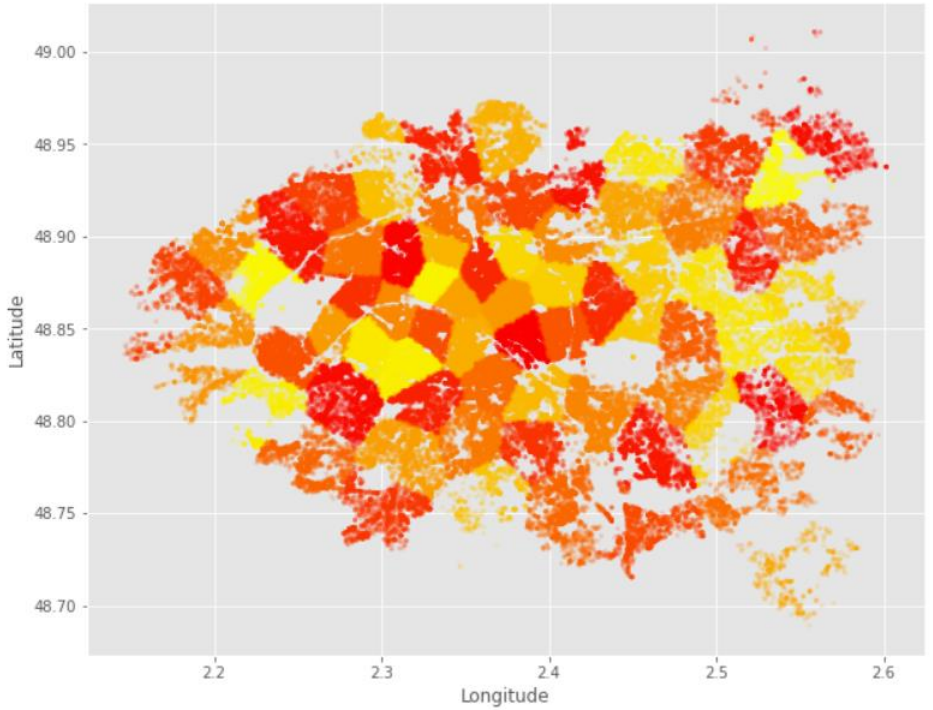


Figure 10. Kmeans on the longitudes/latitudes.

Refine OSRM estimation

- Use the additional data available for 30% of the dataset to update the OSRM estimation
- Use the GPS tracks to compute an estimation of the travel time for 30% of the dataset

Refine OSRM estimation

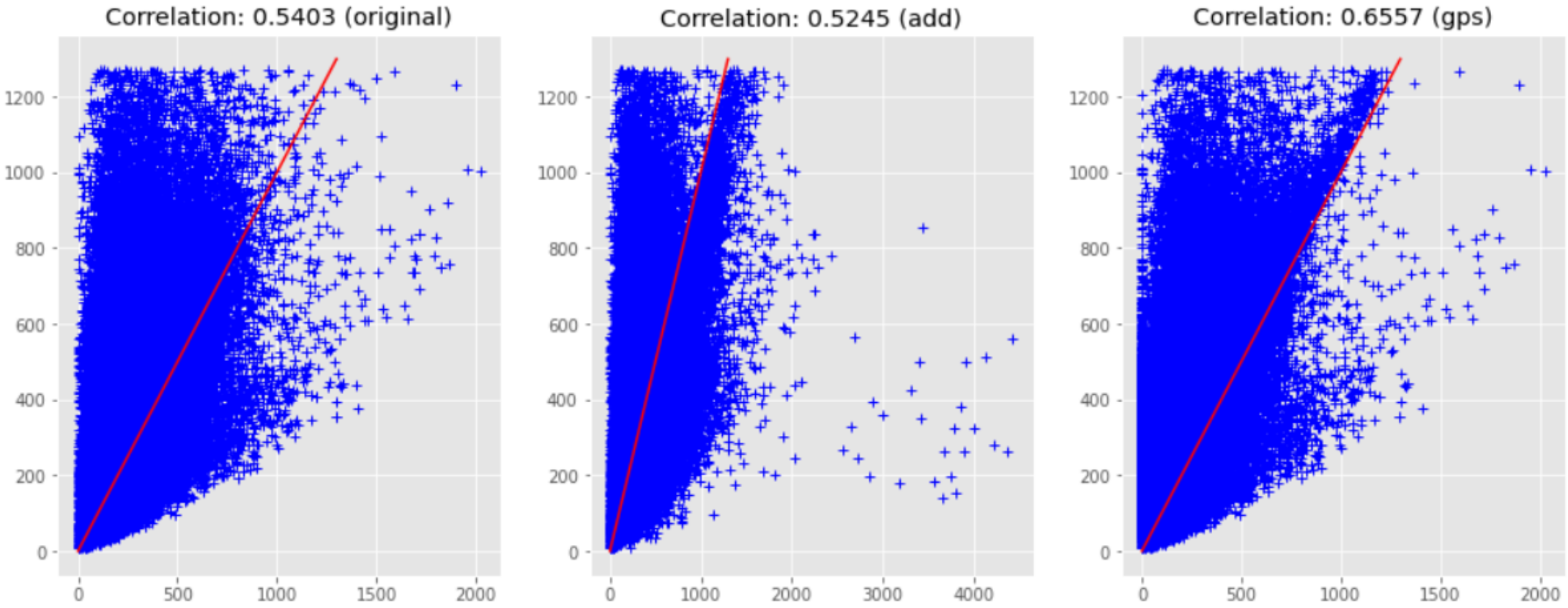


Figure 11. OSRM time estimations. Original, with the additional data or with the GPS tracks, respectively.

Refine OSRM estimation

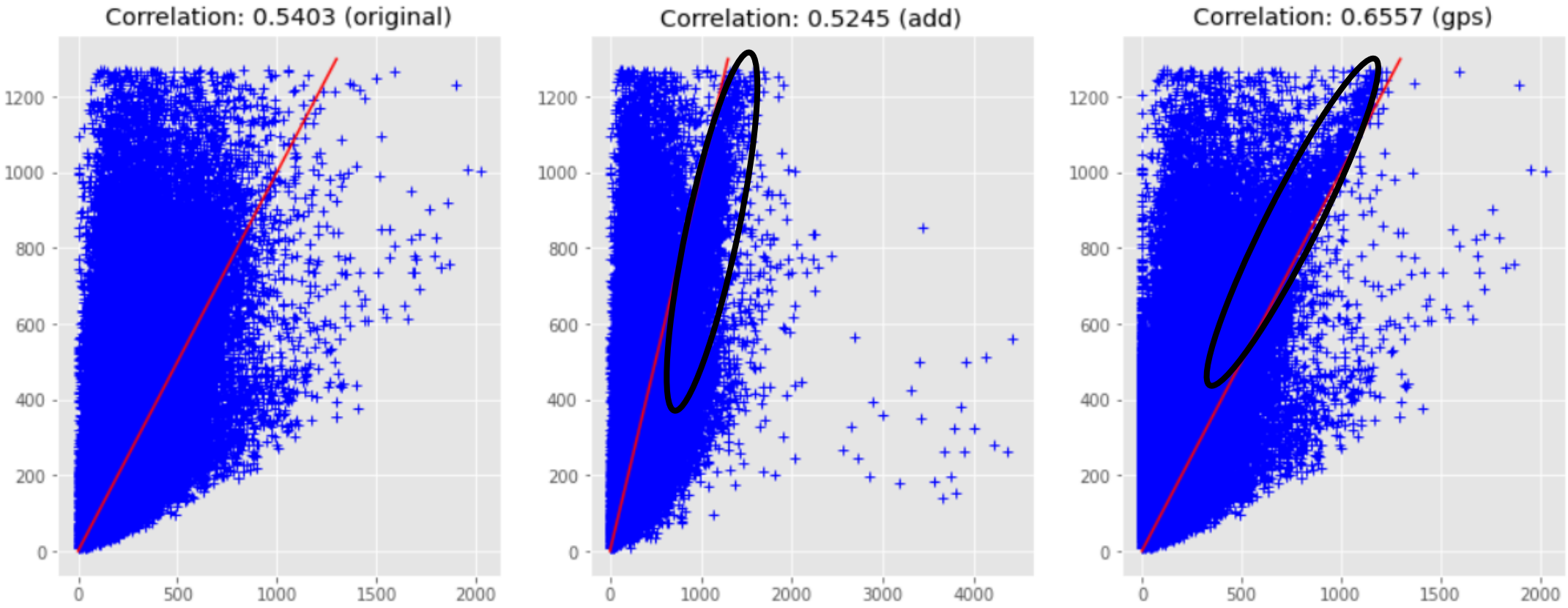


Figure 11. OSRM time estimations. Original, with the additional data or with the GPS tracks, respectively.

Refine OSRM estimation

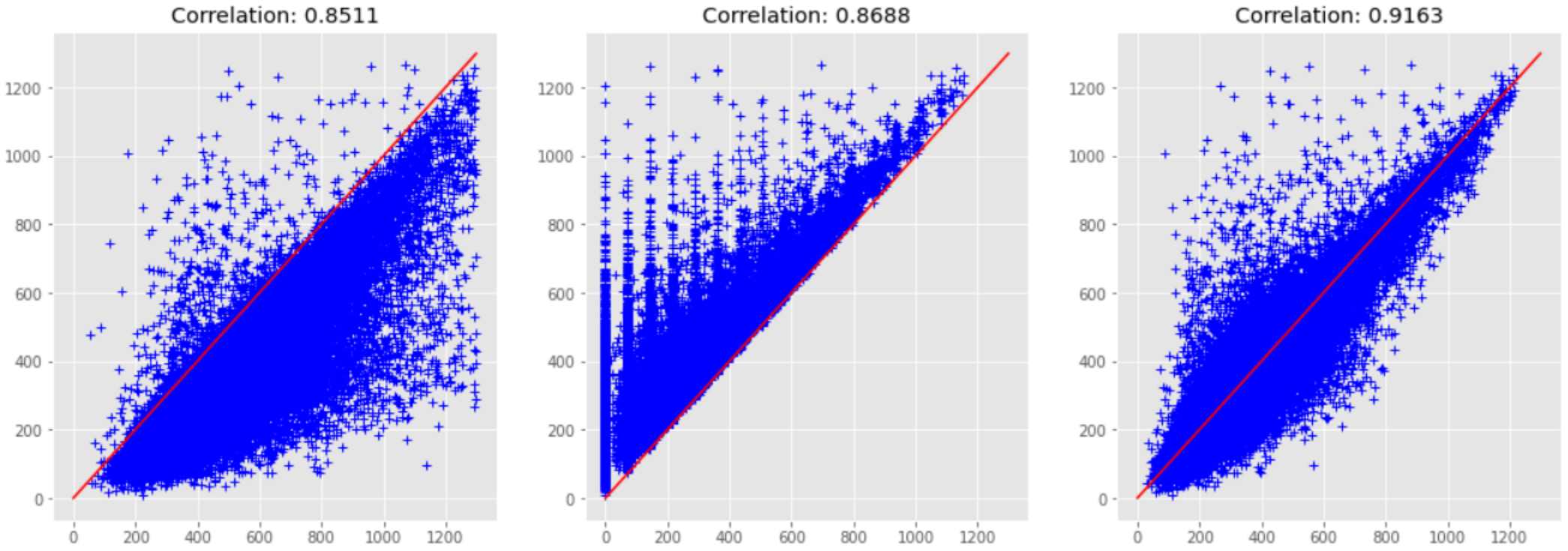


Figure 12. OSRM time estimations. With the additional data, with the GPS tracks, or both, respectively.

Model

Models considered

- Linear regression with L1 or L2 regularization (i.e. Ridge or LASSO), polynomial regression
- Issue: there are a lot of categorical features
- Solution: use decision trees and boosting (XGBoost)
- Neural networks are likely to overfit

Models considered

- Fine-tuned the maximum depth, regularization parameters etc.
- Plot the importance of each feature

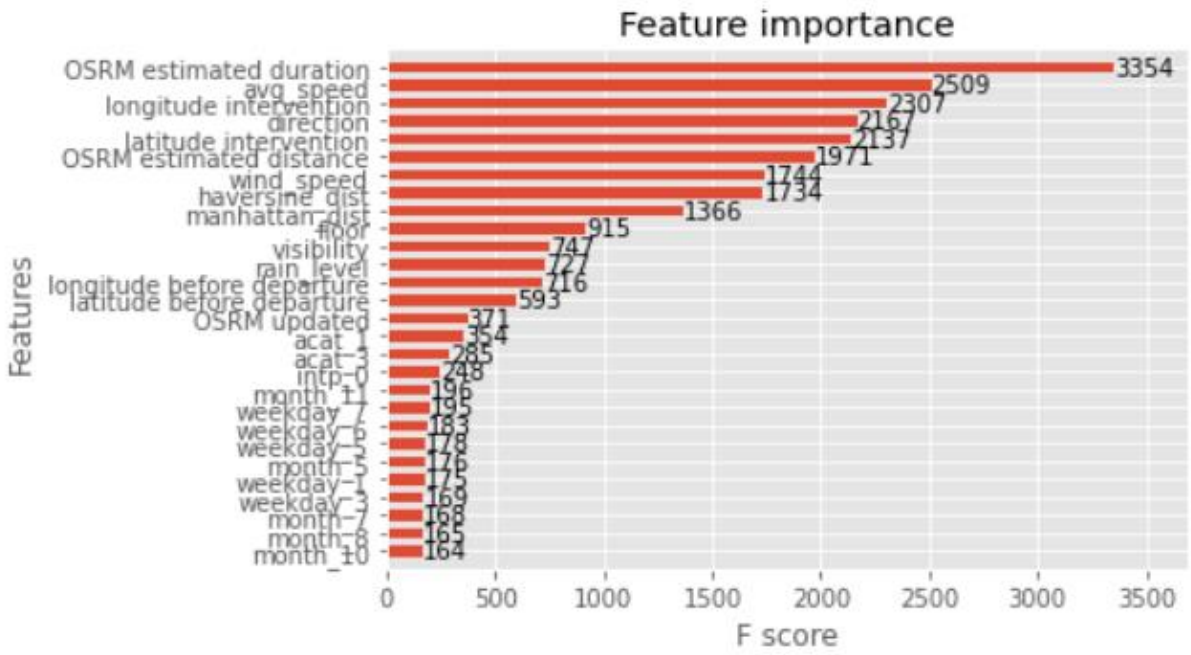


Figure 13. Feature importance of the XGBoost.

Conclusion

Thank you!