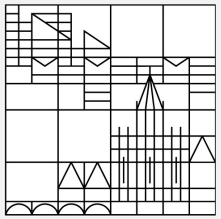
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

Universität Konstanz



A Study on ETA Prediction using Machine Learning and Recovered Routes

Arjanit Arifi Panagiotis Bouros Theodoros Chondrogiannis

Introduction

ETA Prediction

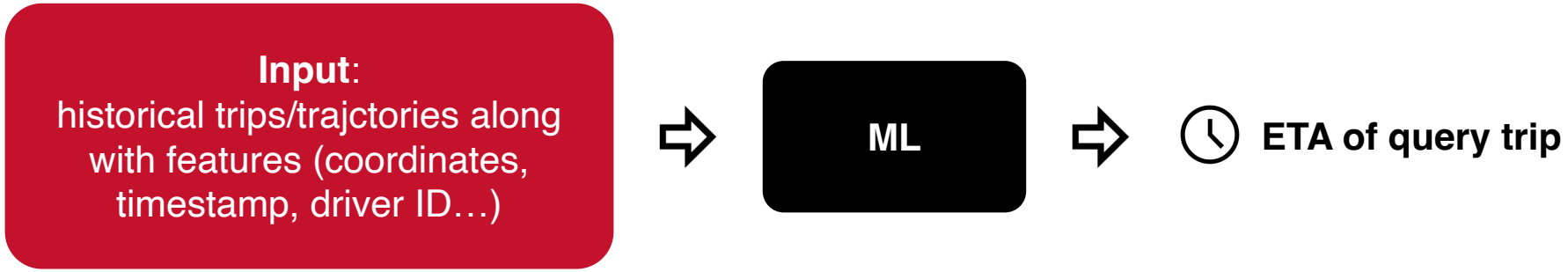
Estimated **T**ime of **A**rrival

Why is an accurate ETA prediction important?

Introduction

ETA Prediction

⇨ **Machine learning models** required for accurate ETA prediction using historical data



Problem

Limited availability of large training datasets with full trajectory/trip data

- **Train ML models with trips that only have start/destination points?**

Introduction

Route Recovery

- Use **Route Recovery** to recover the route the driver has most probably taken
  - Based on the start/destination point and travel time

- History Oblivious Route Recovery on Road Networks (Chondrogiannis et al., SIGSPATIAL'22)

Introduction

Contributions

Main Objective: Train ML models using recovered routes instead of full trajectories.

- **Implementation of two workflows**, which consist of training ETA models with (1) original trajectories and (2) recovered routes
- **ML models with different learning algorithms** (statistical, ensemble and deep learning methods)
- Use of **single route-recovery methods** proposed by in previous work
- **Performance evaluation and comparison** between ML models trained with original trajectories and recovered routes

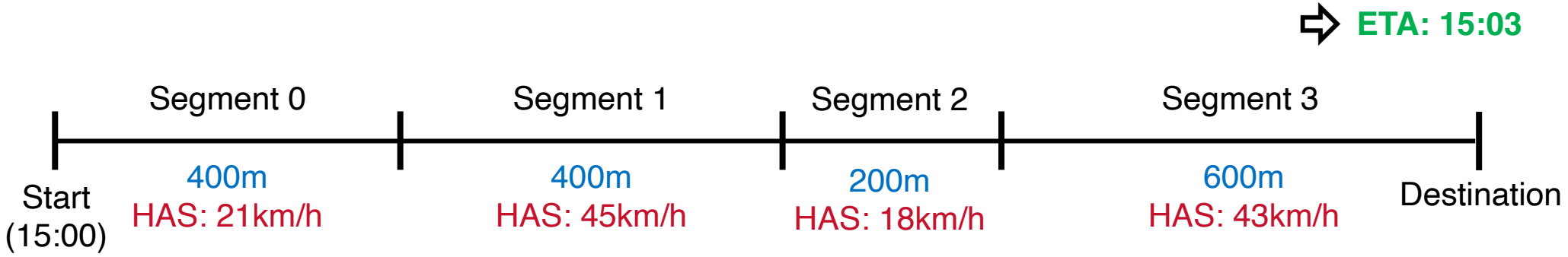
Background

ML Models for ETA

Statistical Models

Historical Average Speeds

- Historical data based real time prediction of vehicle arrival time (Maiti et al., ITSC'14)



Other factors? (weather, holiday…)

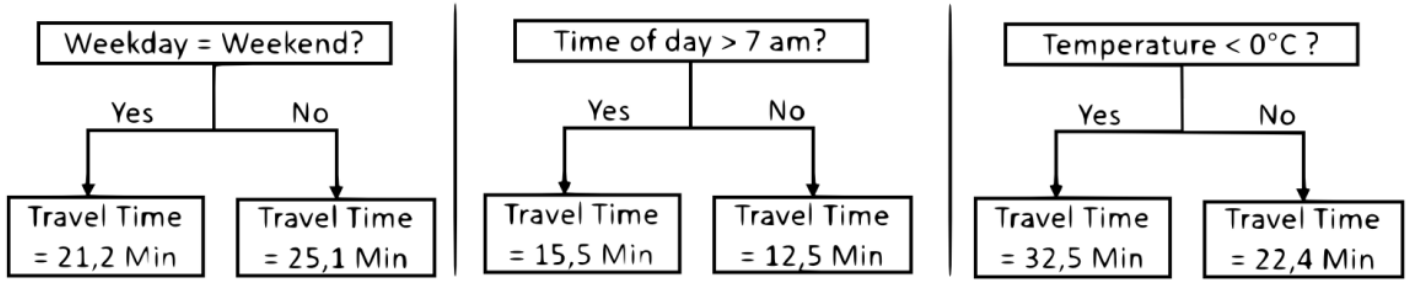
Background

ML Models for ETA

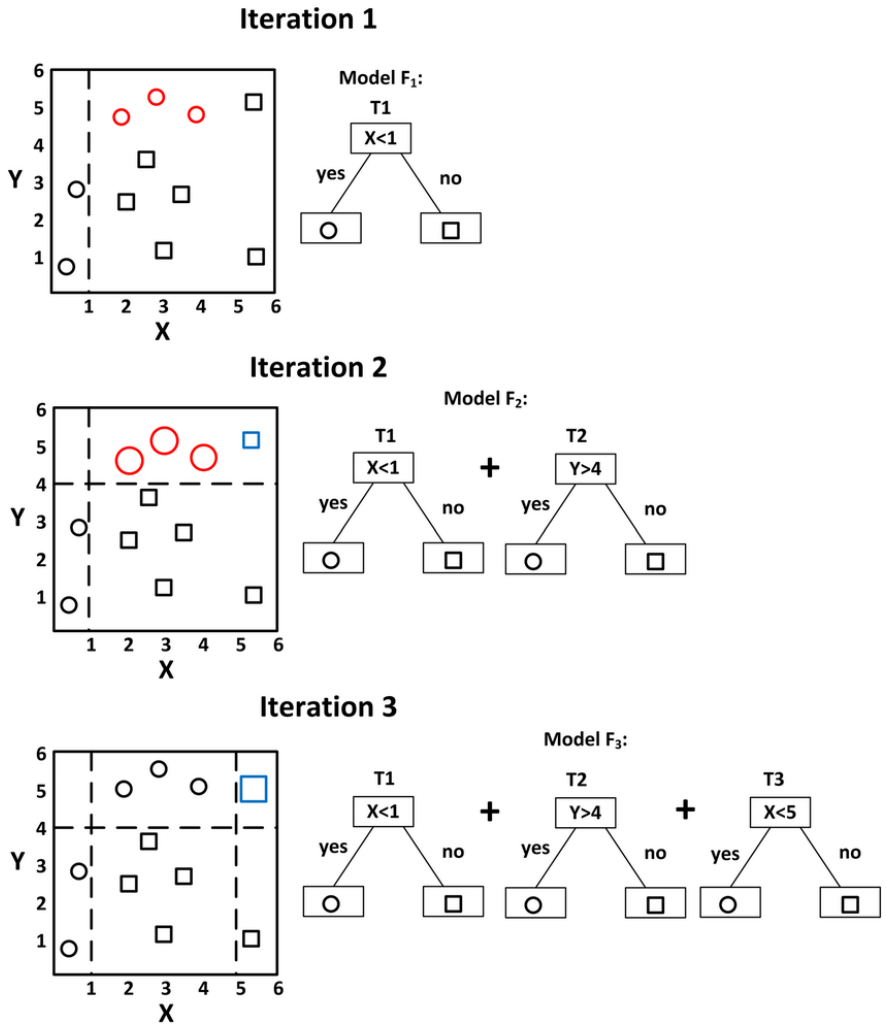
ETA models increasingly powered by machine learning

Ensemble Models

- ETA Prediction with **Random Forest** and **Gradient Boost Regression** (Gupta et al. BDCloud'18)

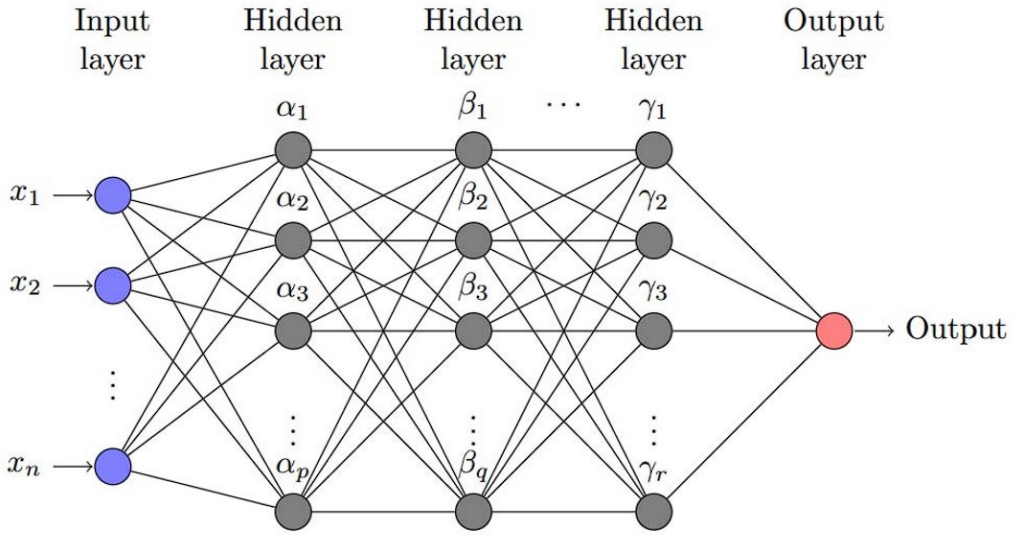


f (Weekday = Weekend, Time of day = 10 am, Temperature = -1 °C) = (21,2 + 15,5 + 32,5)/3 = 23,1 Min



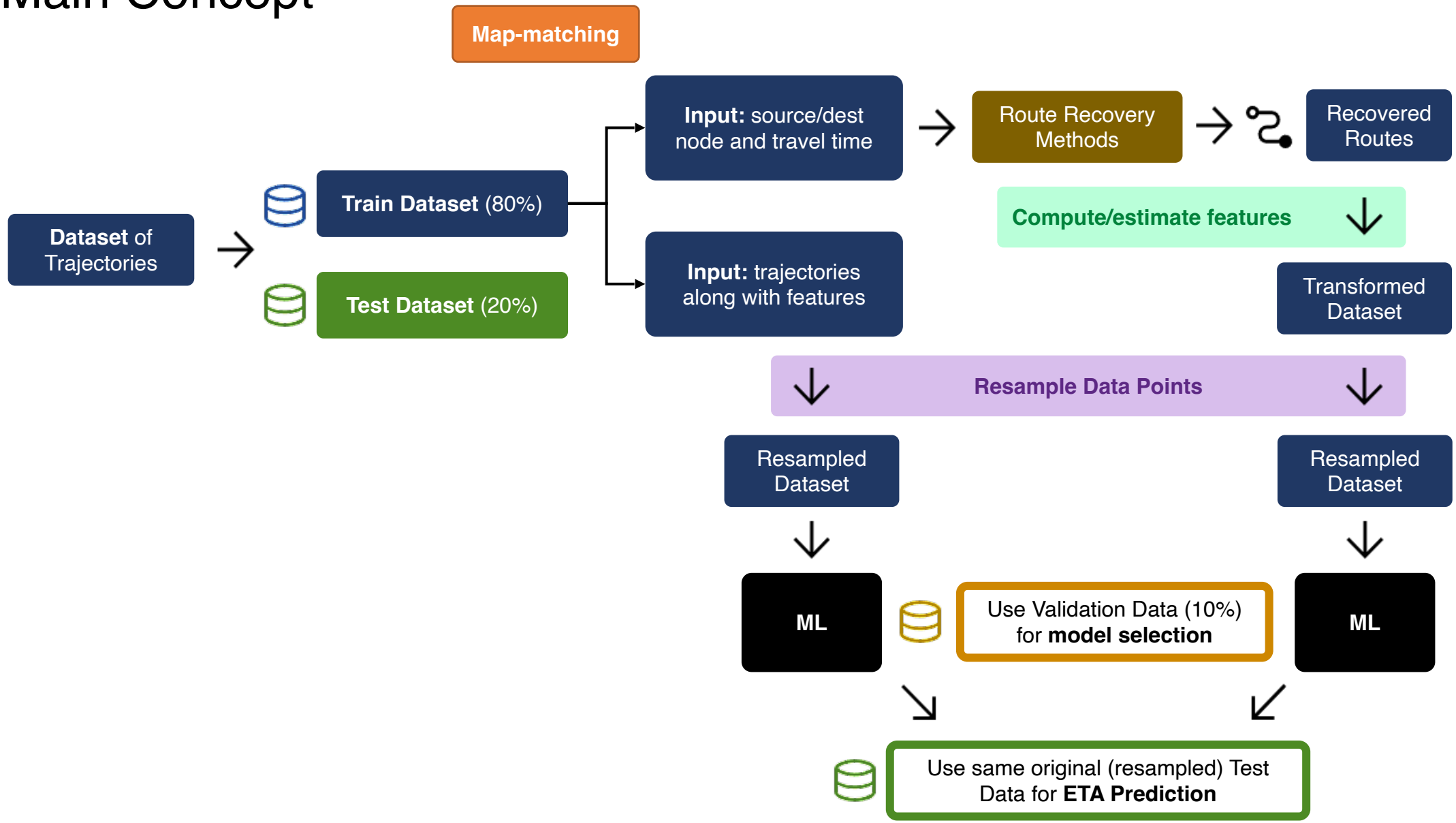
Background

ML Models for ETA



- **ETA models increasingly powered by machine learning**
  - **Deep Learning Models**
    - ✓ **DeepTTE** (Wang et al., AAAI'18)
      - **Combination of:**
        - **Geo-Convolutional Neural Network** (2D-CNN, for spatial dependencies)
        - **Recurrent Neural Network** (RNN, for temporal dependencies)
        - **Attribute Component** (for other features such as DriverID, Weather…)

Main Concept



Main Concept

1st Workflow

⇨ Data Preparation

Problem

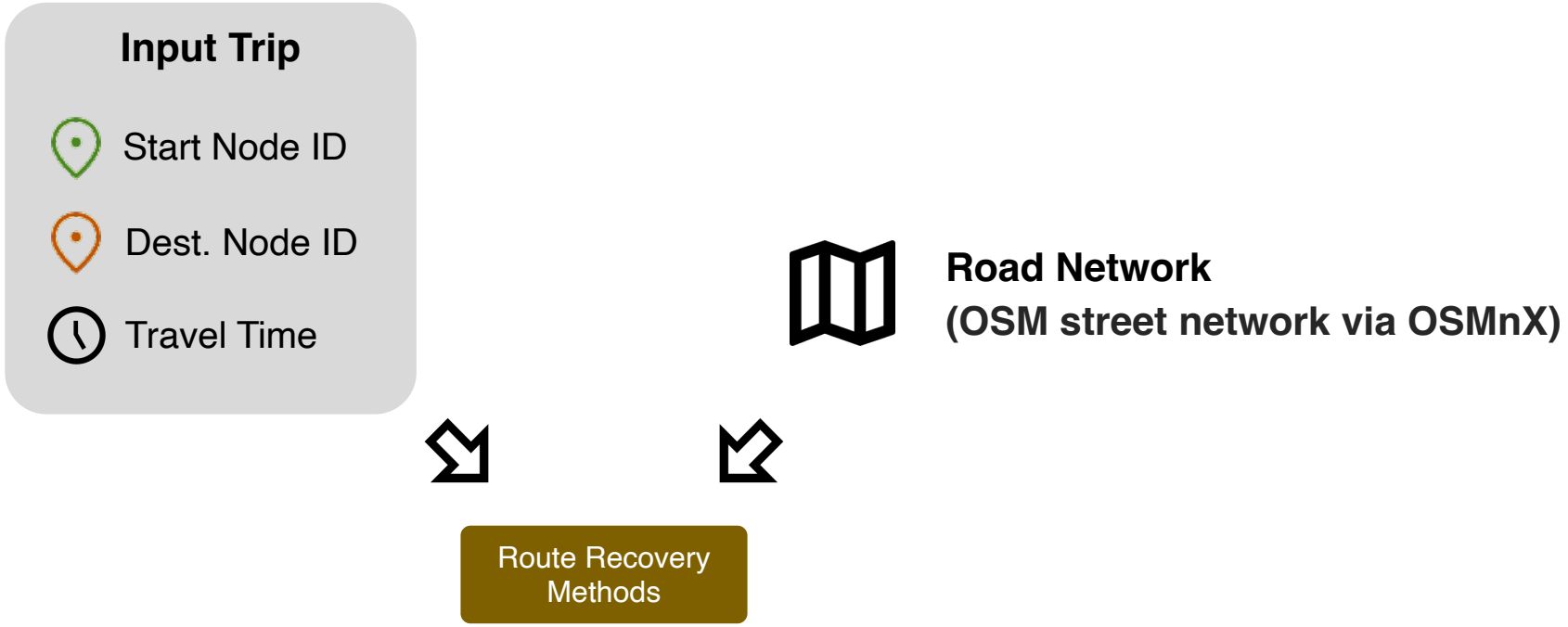
- GPS Points are often noisy and many trajectories may be short and not very useful
- Some approaches require start and destination Node IDs for route recovery

⇨ Use **map matching** to assign GPS points to the road network (and store the path's node IDs)

Main Concept

2nd Workflow

⇨ Input and Preprocessing



Main Concept

2nd Workflow

⇨ Route Recovery

⇨ Single Route Recovery

Fastest Paths (FP)

Shortest Paths (SP)

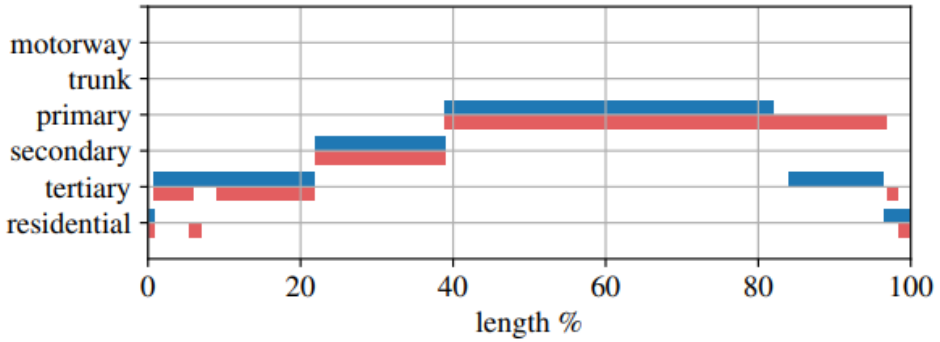
Compute **shortest/fastest** path between a *s* and *t*

Minimum Turns (Min_T)

Compute path with **minimum number of turns**

Minimum Hierarchical Peaks (Min_HP)

Compute path with **minimum** number of hierarchical peaks



Main Concept

2nd Workflow

⇨ Transformation

Problem

Recovered routes have node IDs but lack other features such as travel time

⇨ Estimate/compute relevant features by extracting them from the road network

Main Concept

2nd Workflow

⇨ Resampling

Problems

- **Sequence of nodes and coordinates after Route Recovery too dense**
- **Some methods require sample coordinates at a fixed rate**

⇨ **Resample Routes by applying a distance gap between each coordinate** (e.g. 100m), resulting in a deletion of intermediate points

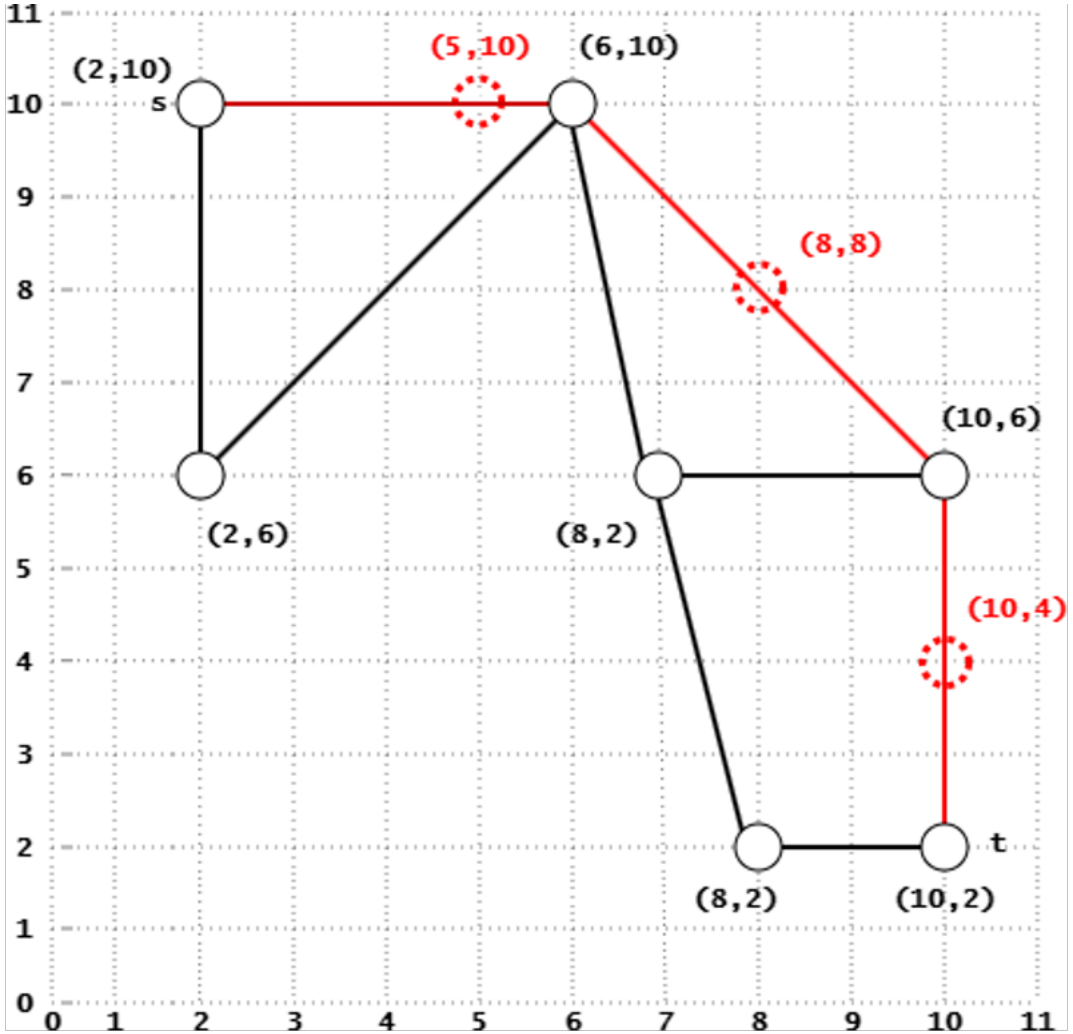
Main Concept

2nd Workflow

⇨ Resampling

- Used-defined sampling rate
- Considering edge geometry

  .. or straight line connection

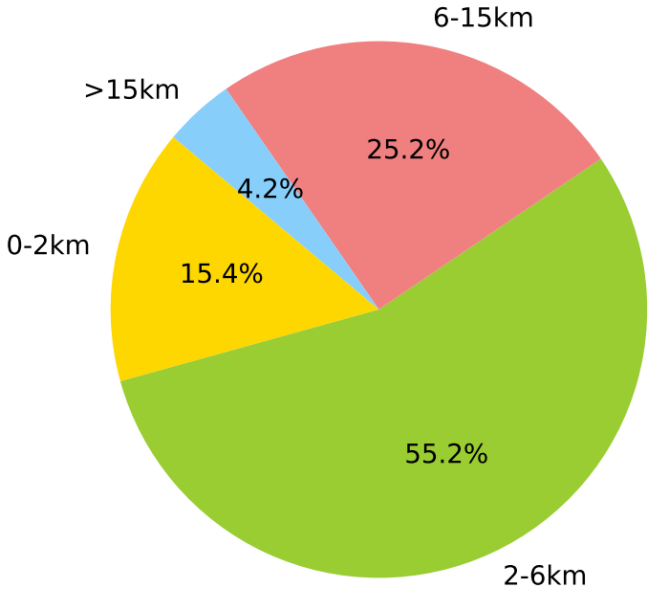


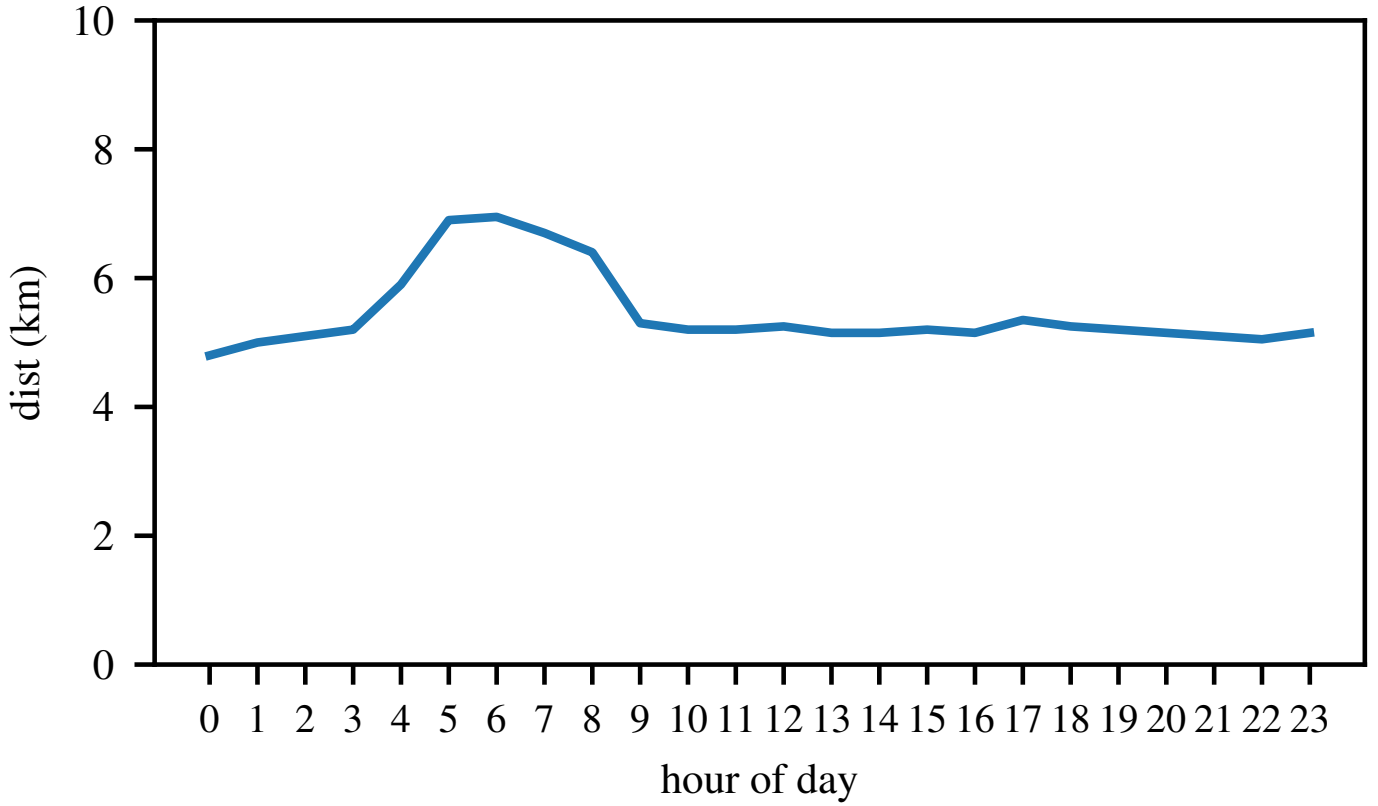
Case Study

➪ Porto Dataset

Publicly available on Kaggle.[1]

- **1,710,670 trajectories** recorded by 442 taxis



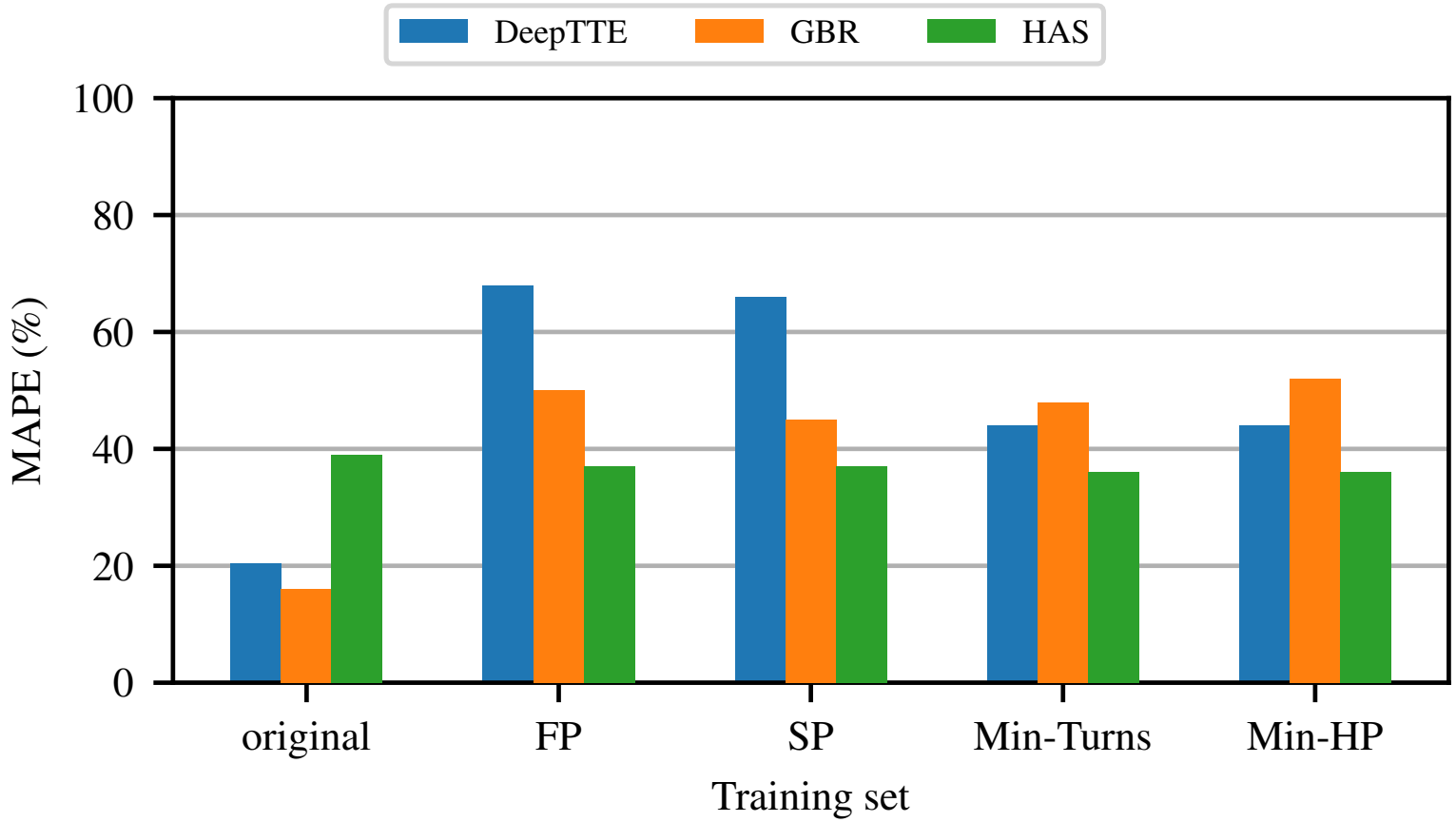


1) https://www.kaggle.com/competitions/pkdd-15-predict-taxi-service-trajectory-i/overview

2) https://github.com/RamiNaim/starline-traffic-data

Evaluation

Results

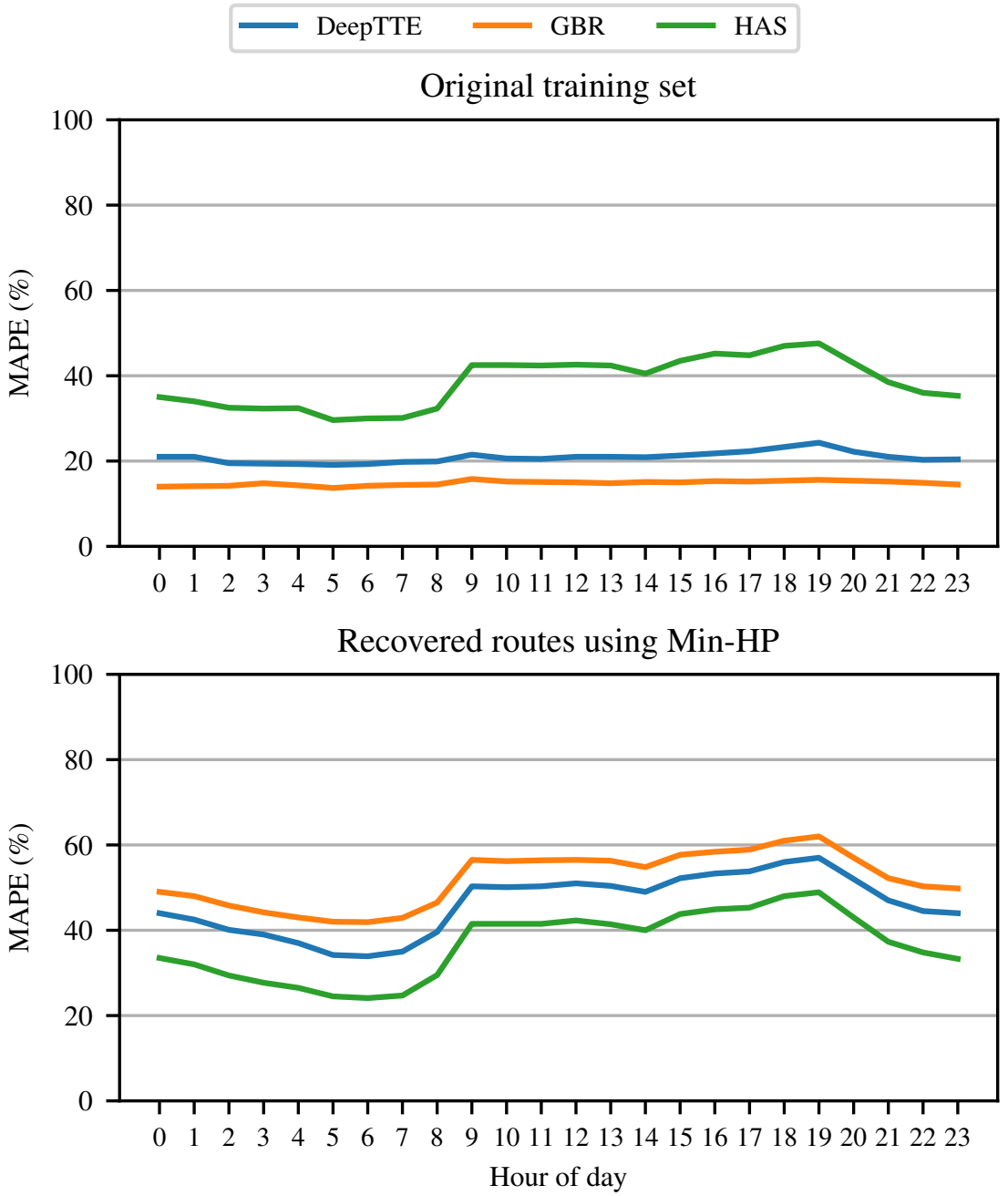


Comparison of MAPE between ML-models trained on original trajectories and recovered routes

Evaluation

Results

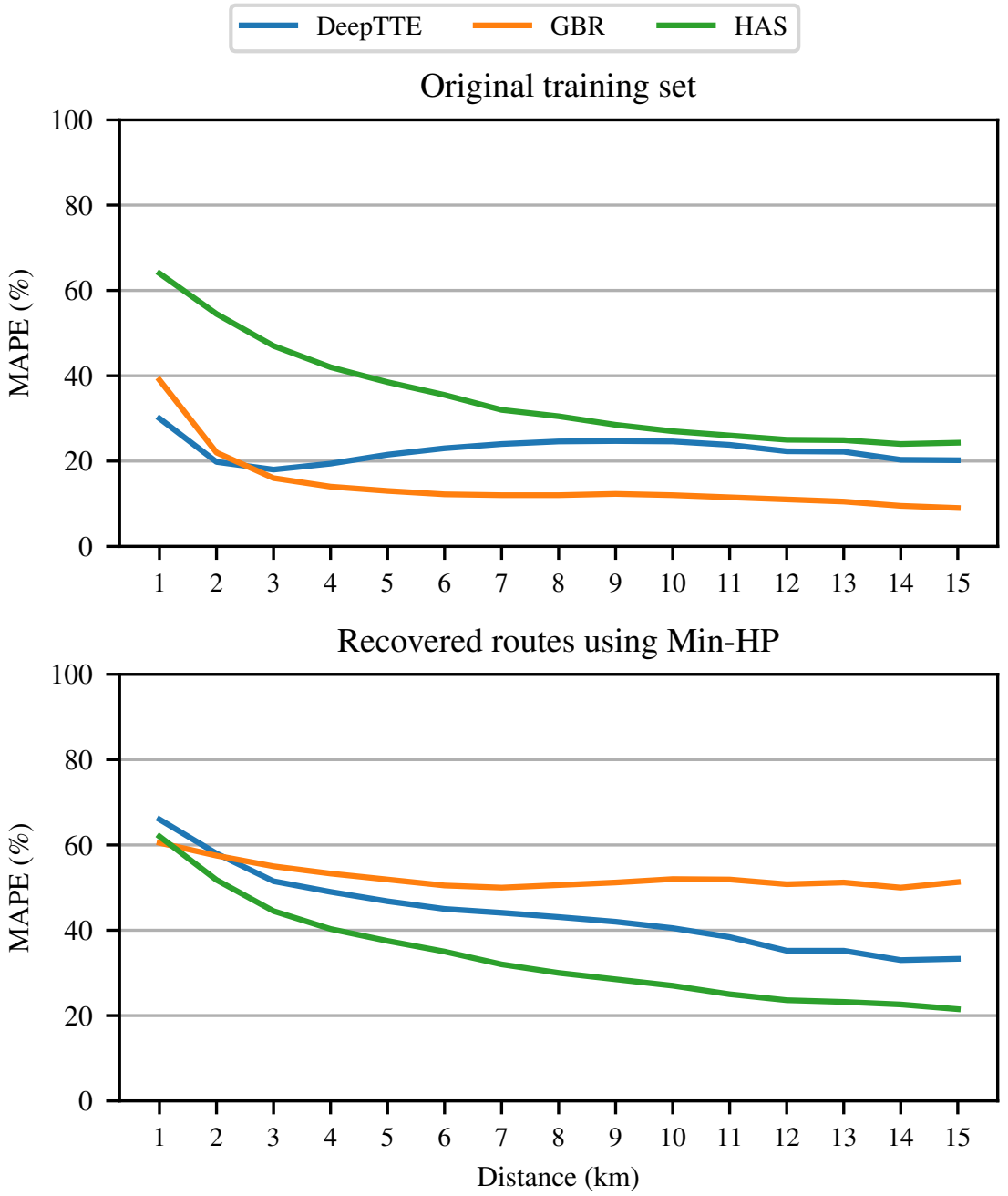
Varying time of day



Evaluation

Results

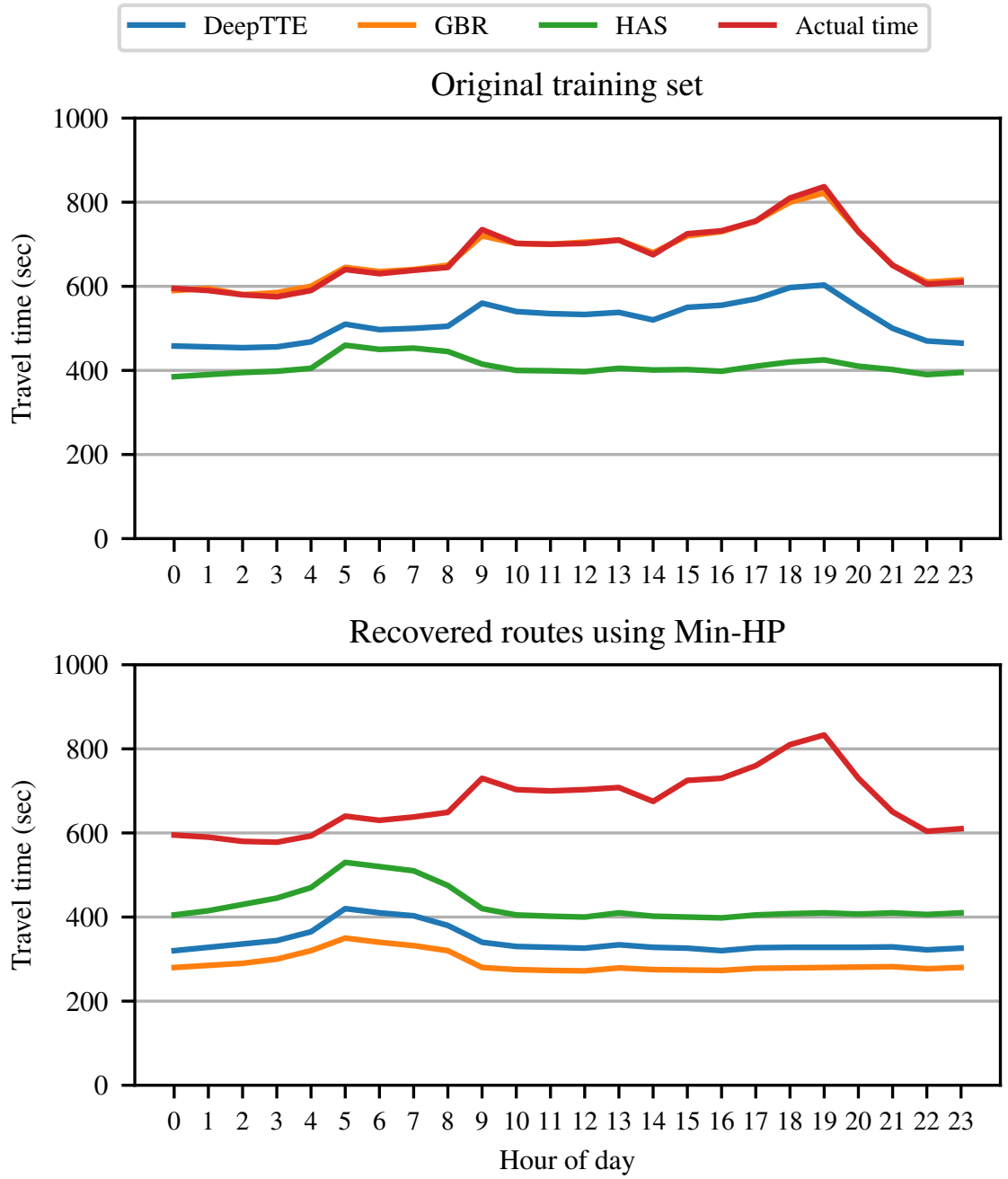
- **Varying distance**
- **What is the problem?**



Evaluation

Results

- **Varying distance**
- **What is the problem?**
- **Underestimating travel time results in high MAPE on recovered routes**



Discussion and Conclusions

Problem

Underestimation of travel time leads to a high error of ML models trained with recovered routes

⇨ Reexamine **transformation process** of recovered routes

- Solely using segment length and speed limit for calculating travel time of recovered routes
- In practice: suboptimal traffic conditions (congestion, traffic lights, etc.)
- Accuracy of single route recovery methods is not enough

- **Problem: Fitting ML models to optimal traffic conditions!**

Future Work

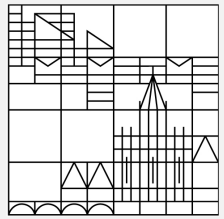
1. Training ETA models with Recovered Routes from **Region Recovery / Group-Based Region Recovery**
2. Use **sparse trajectories** as ML / RR input
3. Apply same approach to **traffic prediction and other problems**

Transformed Routes → Resampling → Resampled Routes → ML-ready

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

Universität Konstanz



Thank you!