Data Analytics based Dynamic Passenger Information System

Rakesh Behera
4th year B. Tech. student
Dept. of Civil Engineering
IIT Madras

Devarsh Kumar
4th year dual degree student
Dept. of Civil Engineering
IIT Madras

Lelitha Vanajakshi
Associate Professor
Dept. of Civil Engineering
IIT Madras

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion

Public Transportation Issues

otal time aken by ssenger	=	walking time from origin to bus stop	+	waiting time at the bus stops (original / transfer)	+	actual travel time on bus	+	walking time from bus stop to destinatio

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion

Proposed Solution

- Bus arrival information system
  - Based on travel time prediction using real-time and historical data
- Trip planner
  - Recommends riders the quickest route along with bus information

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion

Literature Review

Author(s)	Input data	Prediction technique	Approach
Bo *et al.*	One month weekdays bus trip data	Linear regression	Data driven
Patnaik *et al.*	6 months weekday data	Regression	Data driven
Bhandari	7 months AVL data	Time series	Data driven
Liu *et al.*	82 days volume and flow data	ANN	Data driven
Krishnan *et al.*	15 minute aggregate flow data for 3 months	k – Nearest Neighbor	Model based
Vanajakshi *et al.*	Preceding 2 buses data	Kalman filtering	Model based
Dailey *et al.*	Different days trips at the same time of the day	Kalman filtering	Model based

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion

Raw Data

- 85 days GPS data for 2 buses chosen for demo.
- Buses belonged to two routes sharing common end points.
- Sampling rate: 6 logs/minute.
- Each day had approx. 8000 GPS logs.
- Each record had a timestamp, latitude and longitude of the location of the bus.

Sample Raw Data

Timestamp (dd-mm-yyyy HH:MM:SS)	Longitude (degrees)	Latitude (degrees)
17-09-2013 09:30:22	80.12769318	12.92300034
17-09-2013 09:30:32	80.12789917	12.92298985
17-09-2013 09:30:43	80.12844849	12.92292023
17-09-2013 09:30:53	80.12899781	12.92286968
17-09-2013 09:31:03	80.12966156	12.92282963

Routes Chosen

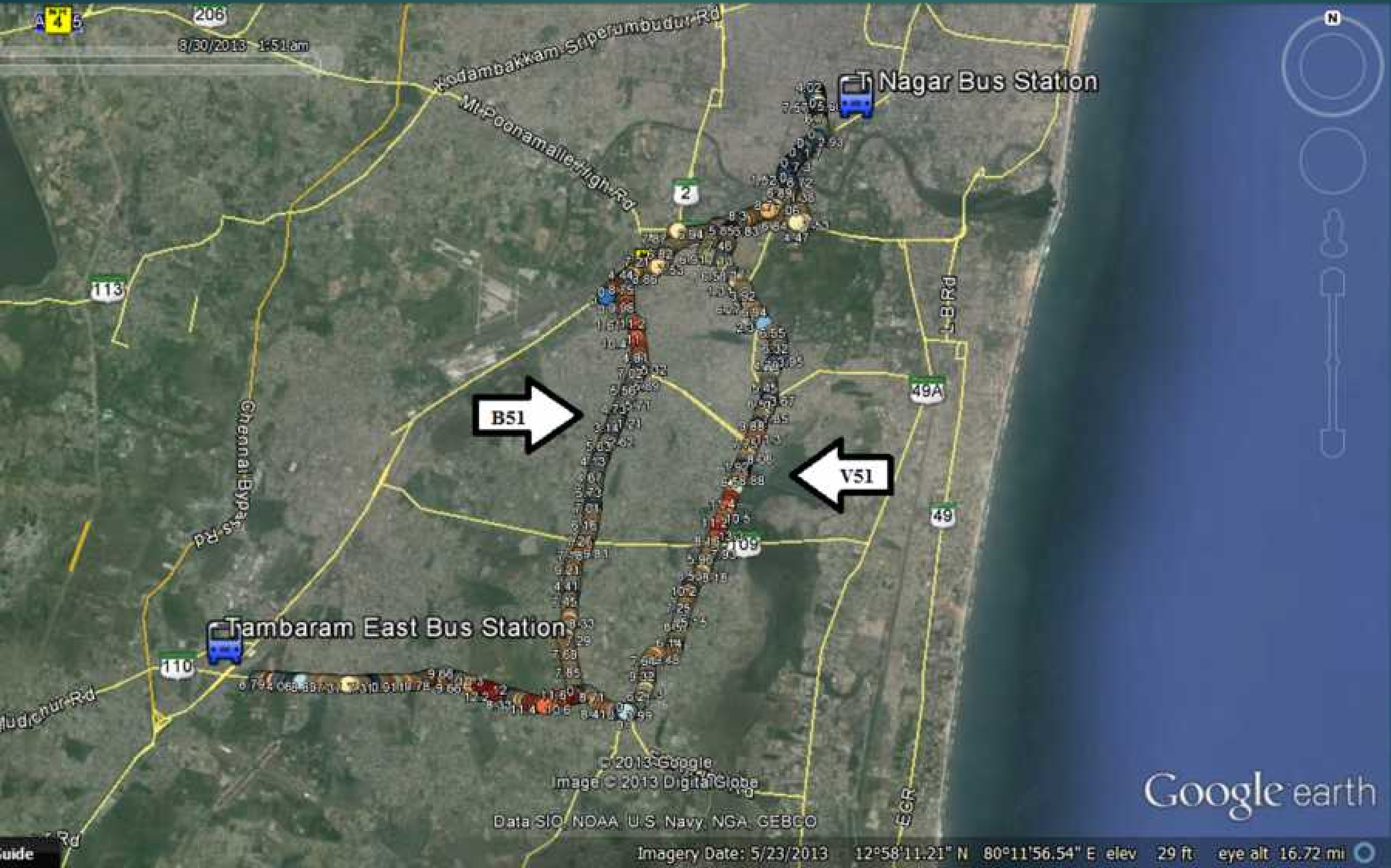


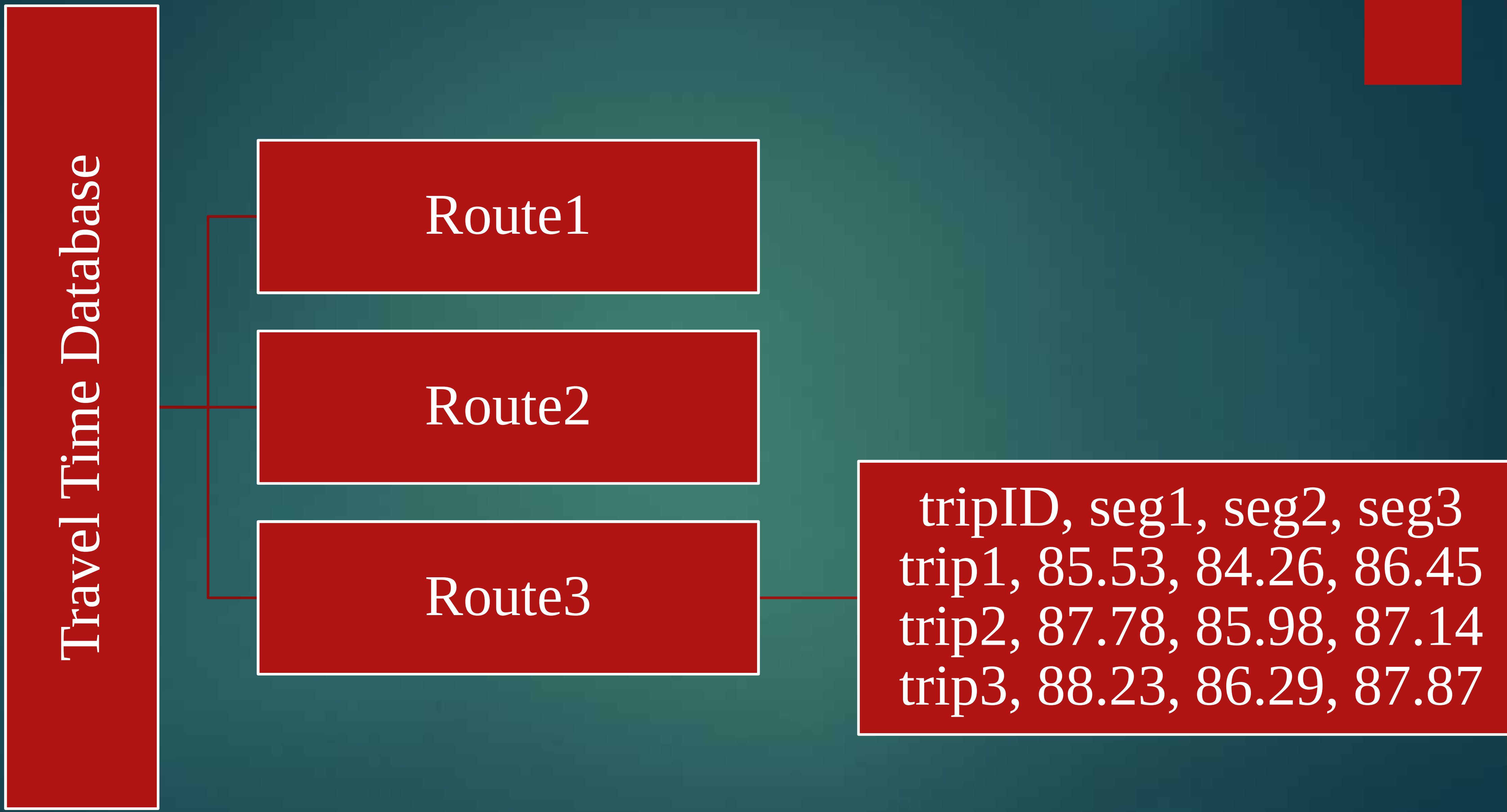
Figure 1. GPS dat two routes plotted Google Earth

Data Preprocessing

- The data were transformed into traffic data e.g. travel time, distance, speed.
- Distances were calculated using haversine formula.
- The data were split into sets belonging to individual trips.
- Historical travel times on each 500 m segment of each route were calculated.

Processed Data

Timestamp (dd-mm-yyyy HH:MM:SS)	Longitude (degrees)	Latitude (degrees)	TimeDiff (s)	Dist (m)	CumulDist (m)	Spee (m/s
-09-2013 09:30:22	80.12769318	12.92300034	10	43.11	294.16	4.31
-09-2013 09:30:32	80.12789917	12.92298985	10	22.38	316.54	2.24
-09-2013 09:30:43	80.12844849	12.92292023	11	60.11	376.65	5.46
-09-2013 09:30:53	80.12899781	12.92286968	10	59.87	436.52	5.99
-09-2013 09:31:03	80.12966156	12.92282963	10	72.17	508.69	7.22


Travel Time Database
Route1
Route2
Route3
tripID, seg1, seg2, seg3
trip1, 85.53, 84.26, 86.45
trip2, 87.78, 85.98, 87.14
trip3, 88.23, 86.29, 87.87

Outline

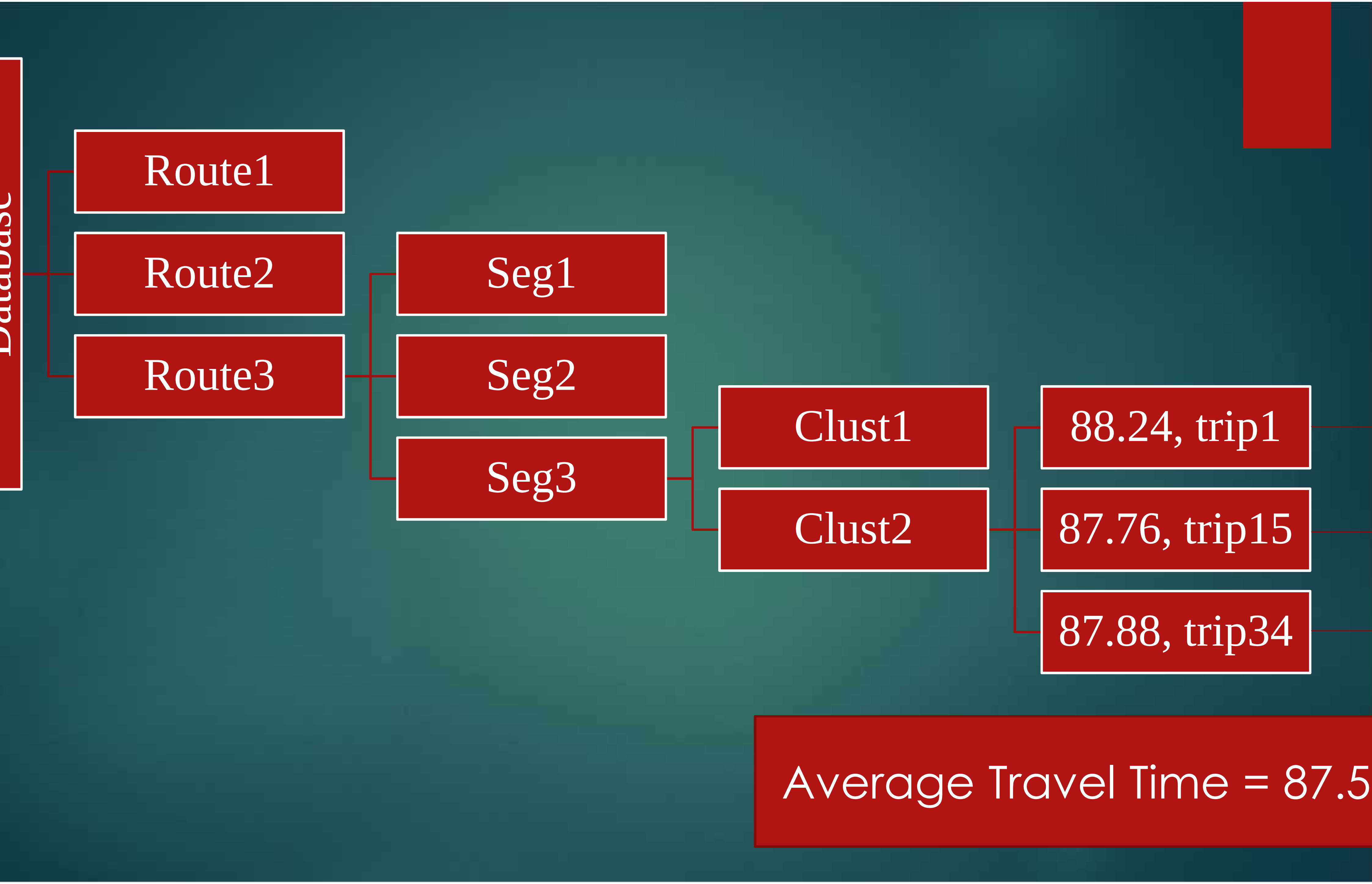
- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion

Clustering

- Historical travel times on segments were clustered using the V-clustering algorithm (Lee et al. 2012).
- Brief explanation of V – clustering:
  - For a segment, historical travel times are sorted.
  - The list is split at an element that minimizes weighted average variance.
  - The child lists are split recursively until a minimum of 25 travel times in each list.

$$WAV^i = \left|\frac{L_1^i}{L}\right| * Var\left(L_1^i\right) + \left|\frac{L_2^i}{L}\right| * Var\left(L_2^i\right)$$

- This led to the reduction of search space for prediction algorithm.


Route1
Route2
Route3
Seg1
Seg2
Seg3
Clust1
Clust2
88.24, trip1
87.76, trip15
87.88, trip34
Average Travel Time = 87.5

Travel time prediction

- Prediction of travel time on upcoming segments:
  1. Past segment travel times are calculated.
  2. The cluster to which these travel times should belong to is searched for.
  3. The cluster found suggests historical trajectories similar to the current one.
  4. Travel times on upcoming segments are calculated by averaging corresponding values from similar historical trajectories.

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Application
- Conclusion

Application

- Bus arrival information system
  - Predicted travel times on the segments are added up to obtain arrival times at bus stops.
  - Predictions update every one minute as the bus moves from one segment to another.
- Trip planner
  - The rider inputs the origin and destination.
  - The algorithm suggests the quickest route by finding the minimum weight (time) path from origin to destination.

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Application
- Conclusion

Conclusion

- Classification of historical trips and clustering of travel times led to faster real time predictions.
- Use of optimal number of previous segments led to more accurate predictions.
- Smaller interval between updates provides the rider with the latest information.
- The promising results indicate that the approach can be implemented under Indian traffic scenarios.

Contact

- Rakesh Behera
  - Email: rakesh8248@gmail.com
  - Phone: +919884995394
- Devarsh Kumar
  - Email: devarsh10iitm@gmail.com
  - Phone: +919884992880
- Lelitha Vanajakshi
  - Email: lelitha@iitm.ac.in

Thank you

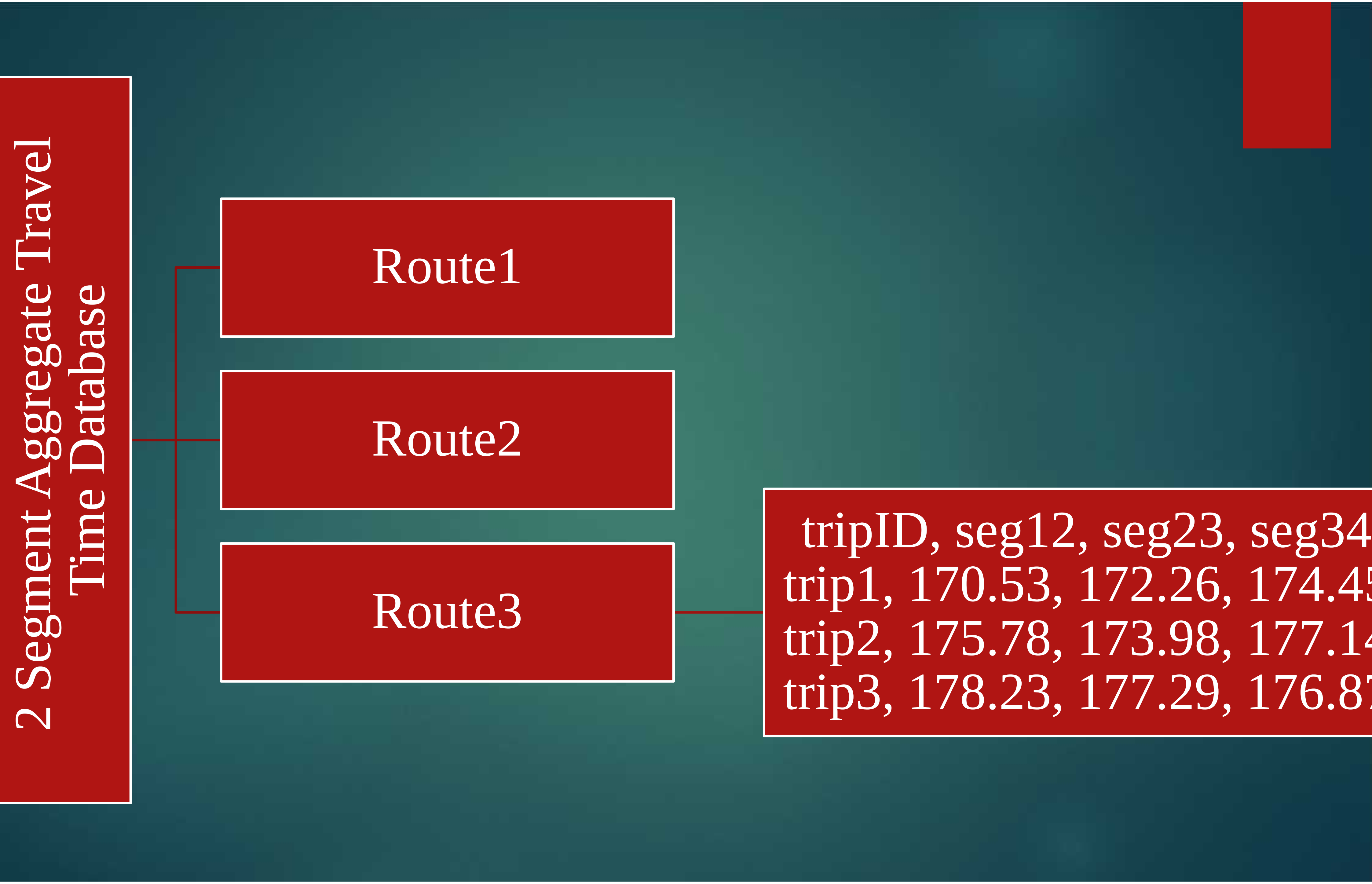
Questions

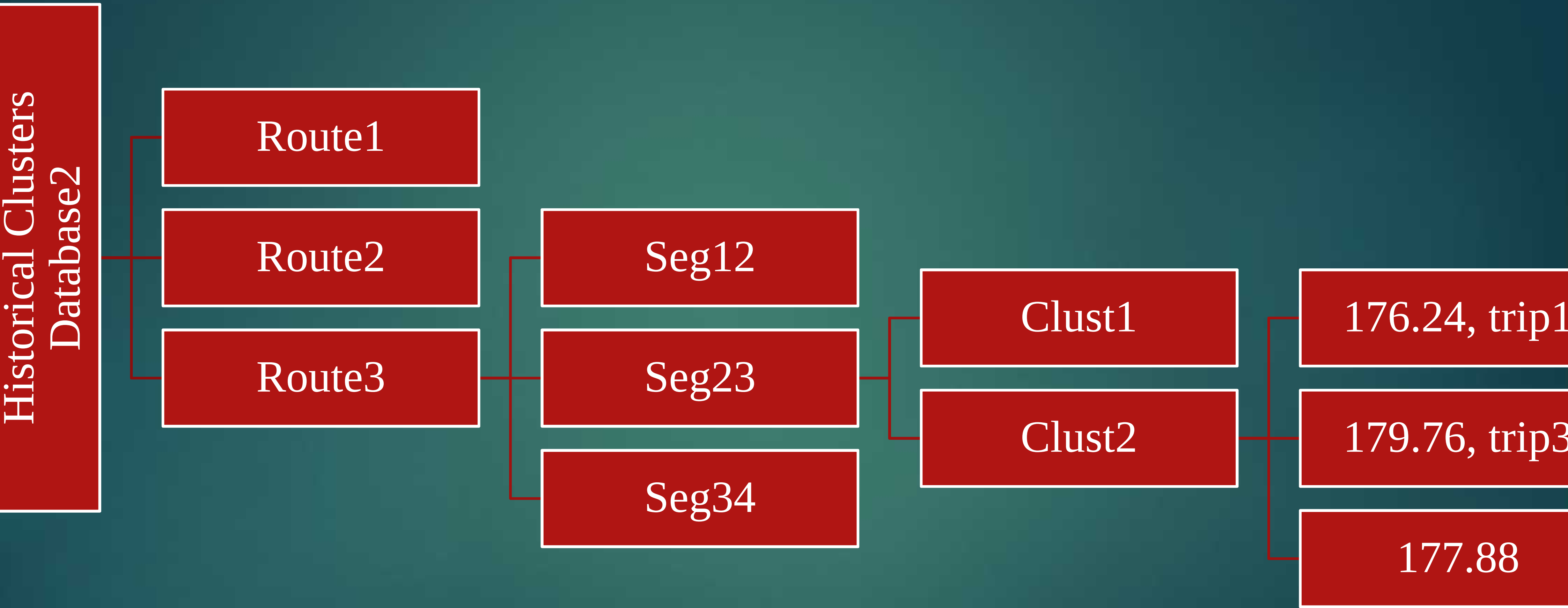
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Results
- Application
- Conclusion

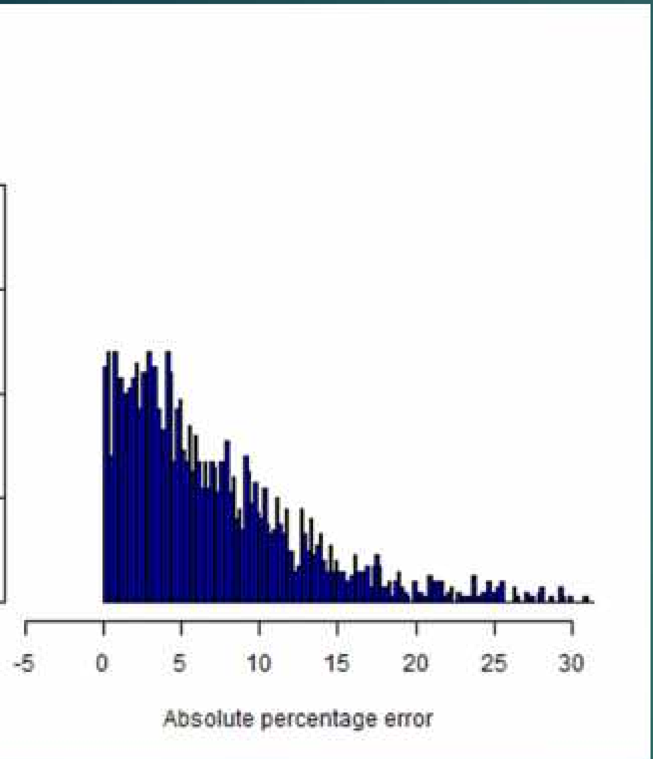
Results

- Measures of accuracy
  - MAPE, MPE, RMSE
- Histograms of prediction errors were plotted.
- Effect of considering varying number of previous segments was analyzed using error distributions.


2 Segment Aggregate Travel Time Database
Route1
Route2
Route3
tripID, seg12, seg23, seg34
trip1, 170.53, 172.26, 174.4
trip2, 175.78, 173.98, 177.1
trip3, 178.23, 177.29, 176.8


Historical Clusters Database2
Route1
Route2
Route3
Seg12
Seg23
Seg34
Clust1
Clust2
176.24, trip1
179.76, trip3
177.88

Error distributions



re 5. Absolute percentage
rs by considering 1 previous
nent.

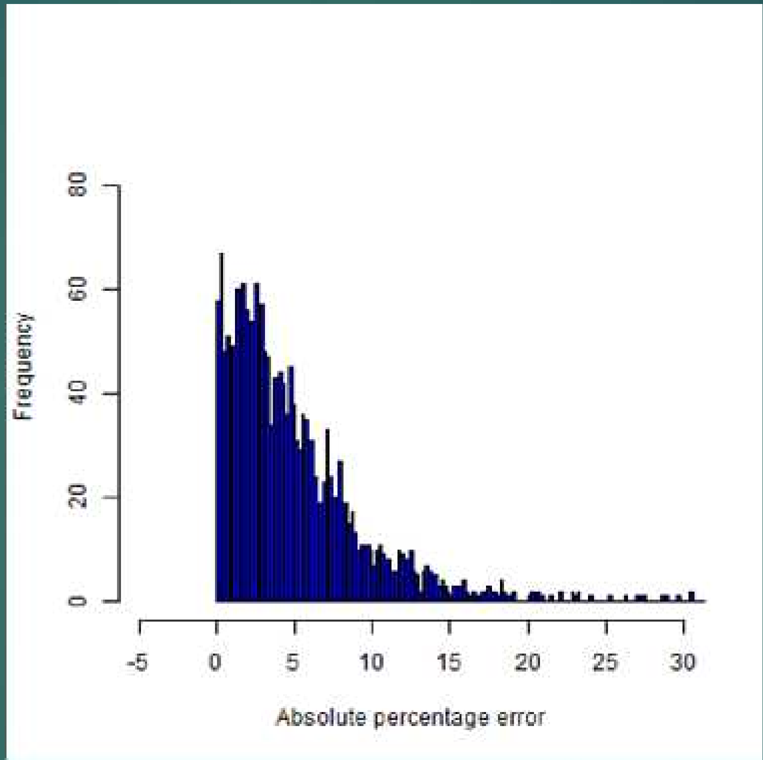


Figure 6. Absolute percentage errors by considering 2 previous segments.

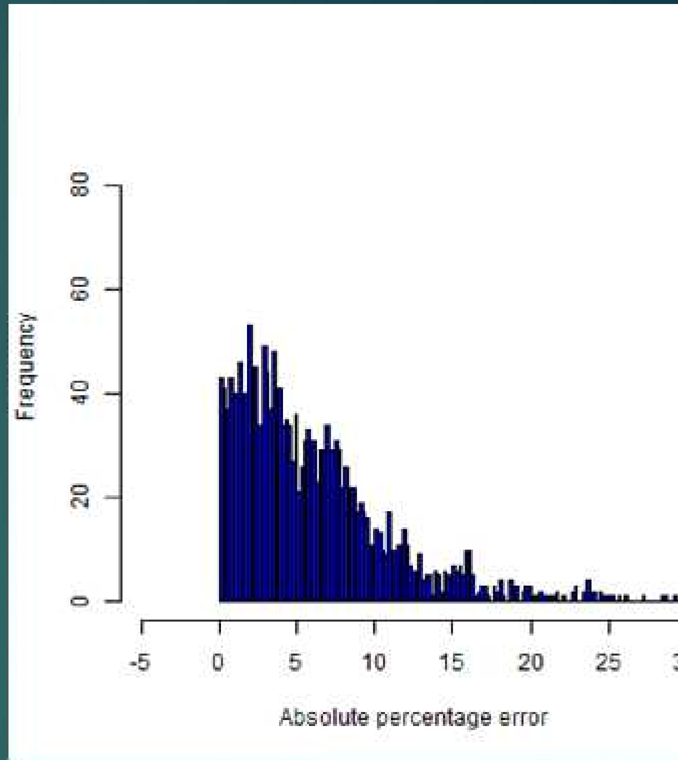


Figure 7. Absolute percentage
errors by considering 3 previou
segments.

Results

MSL	MPE (%)	MAPE (%)	RMSE (s)
1	2.80	12.02	9.22
2	**0. 75**	**8. 72**	**7.20**
3	1.78	9.19	8.08
4	1.89	9.38	8.26
5	1.96	9.76	8.46
6	1.93	10.10	8.68
7	1.99	10.05	8.40

Measures of Accuracy

- MAPE: Mean Absolute Percentage Error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{T}_i - T_i}{T_i}\right| * 100$$

- MPE: Mean Percentage Error

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{T}_i - T_i}{T_i} * 100$$

- RMSE: Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{T}_i - T_i\right)^2}{T_i}}$$

Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Results
- Application
- Conclusion

Data Analysis

- Used for classifying historical trips into groups of similar characteristics.
- Analyses
  - Within – day travel time variation
  - Weekdays versus weekends
  - Monthly variation

Within – day variation

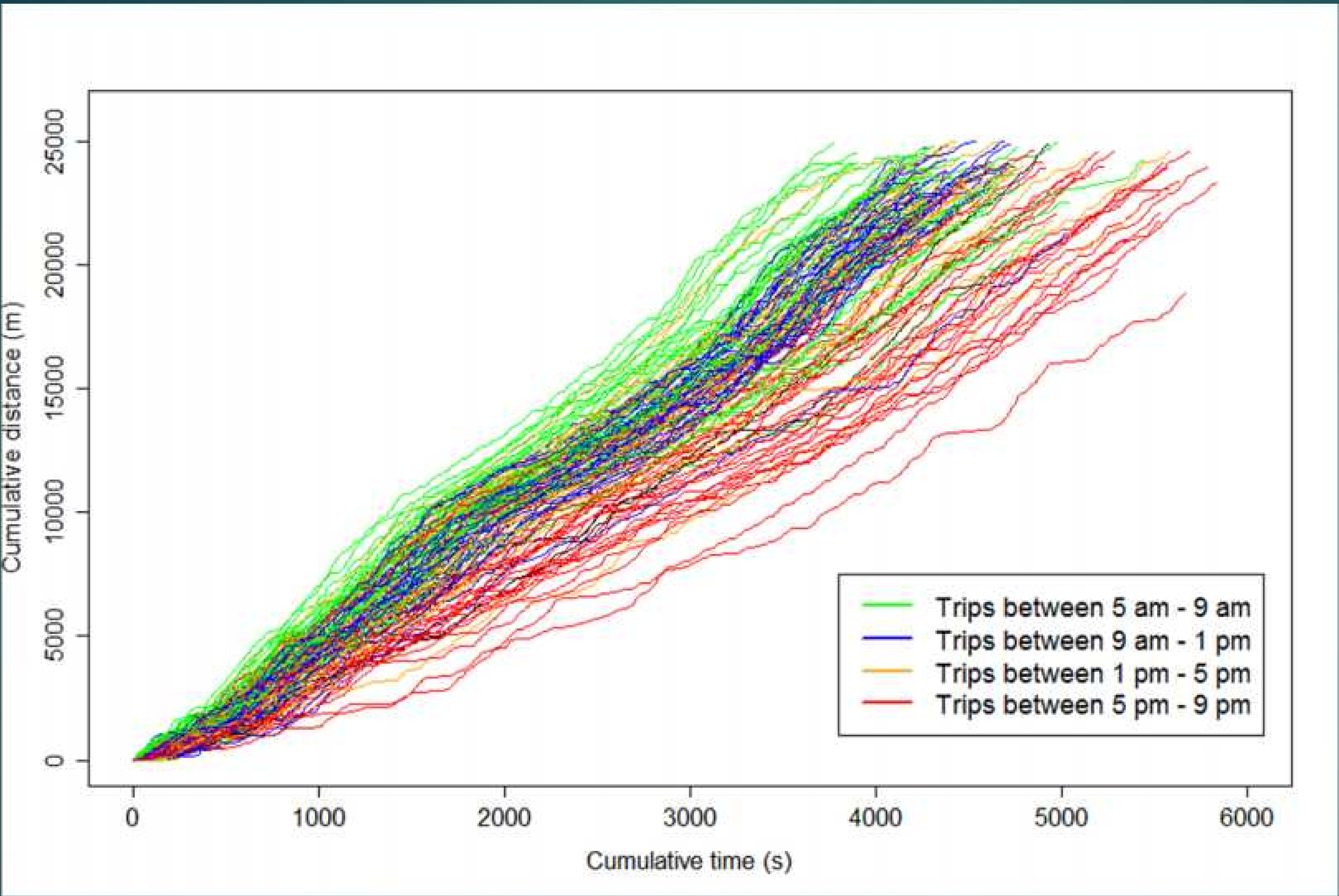


Figure 2. Space-time trajectories for south bound trips on route B51 in Chennai showing within – day variation

Weekdays versus weekends

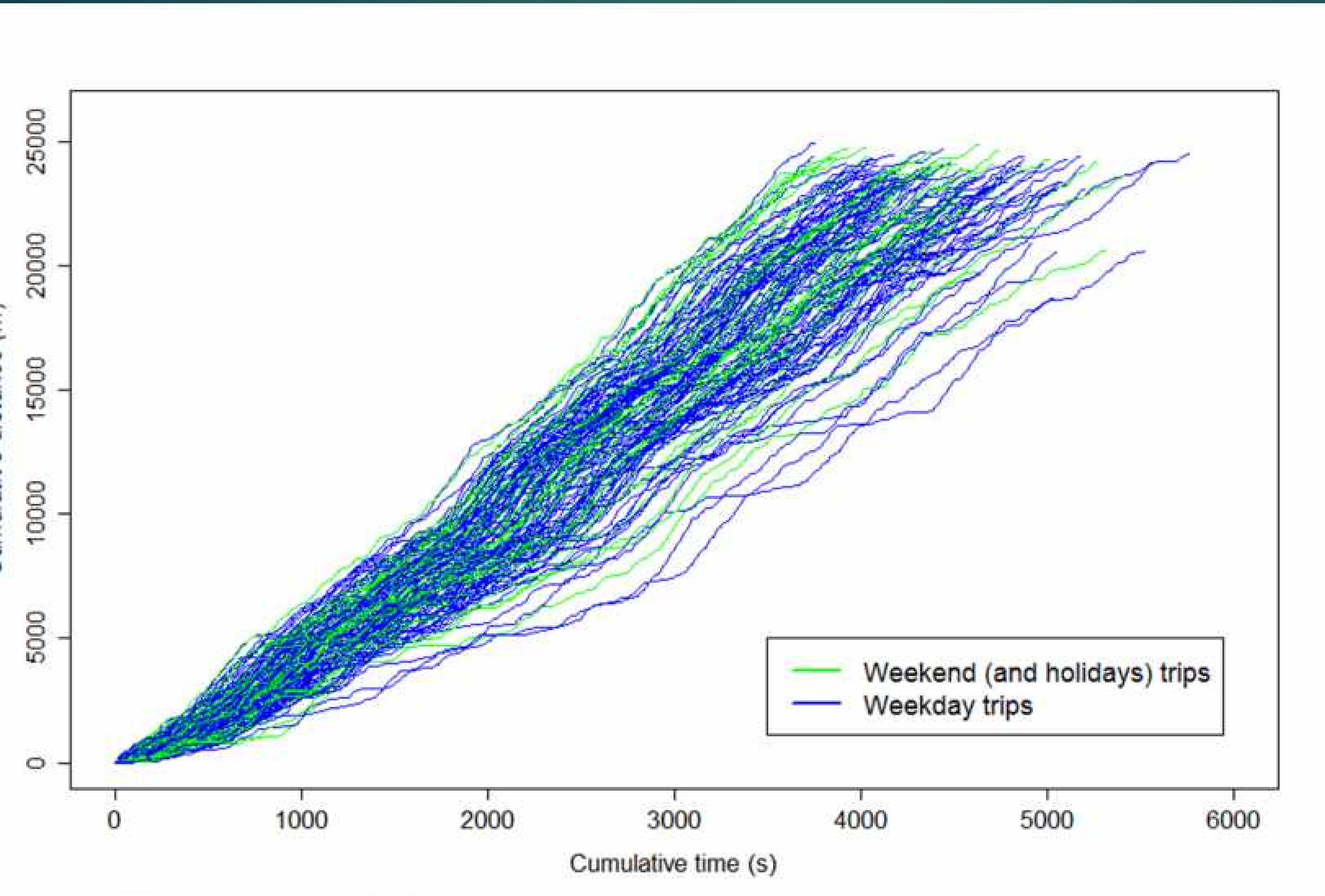


Figure 3. Space-time trajectories for south bound trips on route V51 in Chennai comparing weekend ar weekday trips.

Monthly variation

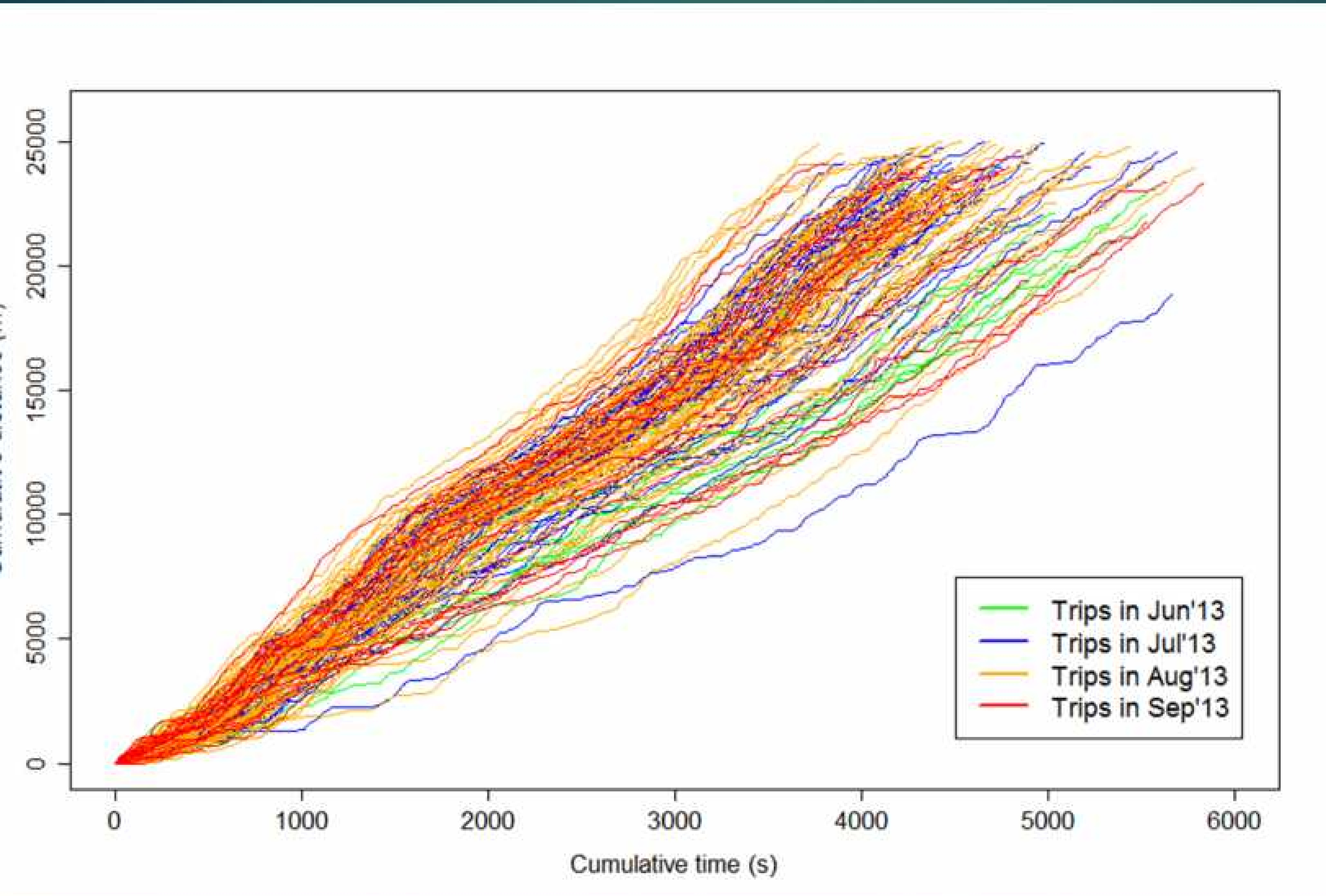


Figure 4. Space-time trajectories for south bound trips on route B51 in Chennai show monthly variation.

Haversine Formula

$$D = R * \cos^{-1}(a + b)$$

where

D = distance between two GPS points

R = mean radius of Earth = 6371000 m

$$a = \cos\left(\frac{\circledR}{2} - lat_1\right)\cos\left(\frac{\circledR}{2} - lat_2\right)$$

$$b = \sin\left(\frac{\circledR}{2} - lat_1\right)\sin\left(\frac{\circledR}{2} - lat_2\right)\cos(long_1 - long_2)$$

lat_1 = latitude of the first GPS point

lon_1 = longitude of the first GPS point

lat_2 = latitude of the second GPS point

lon_2 = longitude of the second GPS point