AUGUST 9, 2019

# DIFFUSION CONVOLUTION RECURRENT NEURAL NETWORK FOR TRAFFIC FORECASTING

## SCIENCE USE CASE 2

Tanwi Mallick and Prasanna Balaprakash

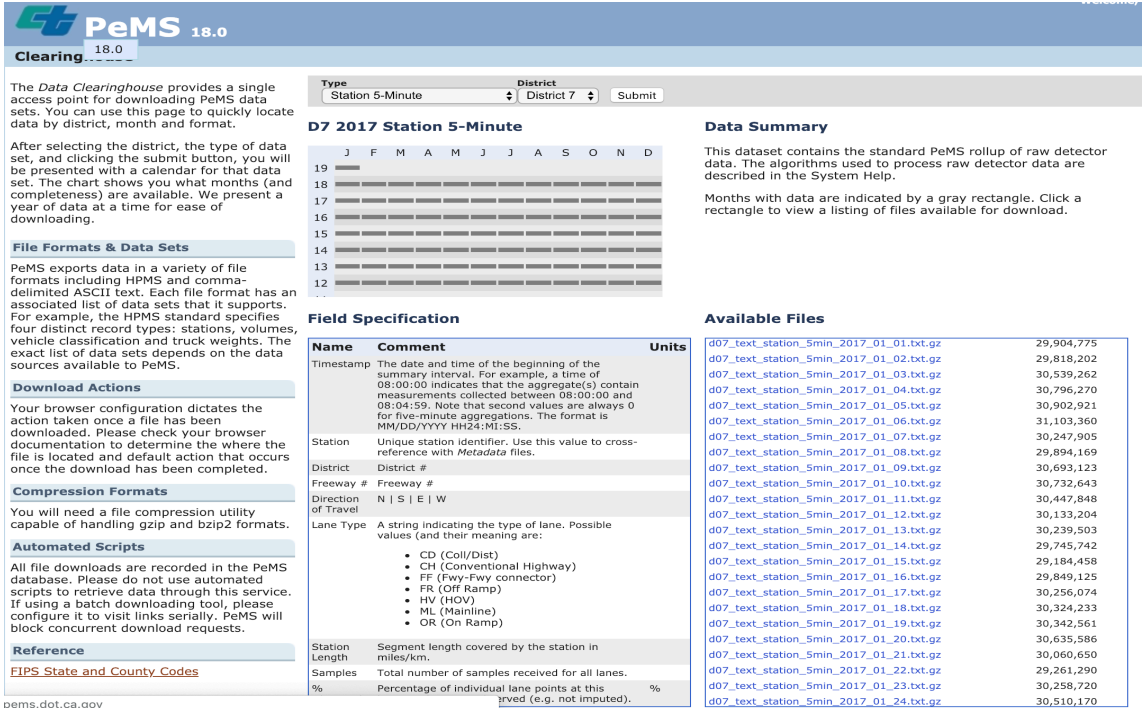
ATPESC 2019

# Traffic forecasting

- Given
  - Historic traffic metrics {speed, flow, etc}
  - Road network distance and connectivity
- Predict
  - Future traffic metrics

Large-scale machine (supervised) learning problem!

# Traffic data from Caltrans Performance Measurement System (PeMS)



PeMS 18.0

Clearing 18.0

The *Data Clearinghouse* provides a single access point for downloading PeMS data sets. You can use this page to quickly locate data by district, month and format.

After selecting the district, the type of data set, and clicking the submit button, you will be presented with a calendar for that data set. The chart shows you what months (and completeness) are available. We present a year of data at a time for ease of downloading.

**File Formats & Data Sets**

PeMS exports data in a variety of file formats including HPMS and comma-delimited ASCII text. Each file format has an associated list of data sets that it supports. For example, the HPMS standard specifies four distinct record types: stations, volumes, vehicle classification and truck weights. The exact list of data sets depends on the data sources available to PeMS.

**Download Actions**

Your browser configuration dictates the action taken once a file has been downloaded. Please check your browser documentation to determine the where the file is located and default action that occurs once the download has been completed.

**Compression Formats**

You will need a file compression utility capable of handling gzip and bzip2 formats.

**Automated Scripts**

All file downloads are recorded in the PeMS database. Please do not use automated scripts to retrieve data through this service. If using a batch downloading tool, please configure it to visit links serially. PeMS will block concurrent download requests.

**Reference**

FIPS State and County Codes

pems.dot.ca.gov

Type: Station 5-Minute | District: District 7 | Submit

**D7 2017 Station 5-Minute**

J F M A M J J A S O N D

19 18 17 16 15 14 13 12

**Data Summary**

This dataset contains the standard PeMS rollup of raw detector data. The algorithms used to process raw detector data are described in the System Help.

Months with data are indicated by a gray rectangle. Click a rectangle to view a listing of files available for download.

**Field Specification**

| Name | Comment | Units |
|---|---|---|
| Timestamp | The date and time of the beginning of the summary interval. For example, a time of 08:00:00 indicates that the aggregate(s) contain measurements collected between 08:00:00 and 08:04:59. Note that second values are always 0 for five-minute aggregations. The format is MM/DD/YYYY HH24:MI:SS. | |
| Station | Unique station identifier. Use this value to cross-reference with *Metadata* files. | |
| District | District # | |
| Freeway # | Freeway # | |
| Direction of Travel | N \| S \| E \| W | |
| Lane Type | A string indicating the type of lane. Possible values (and their meaning are: • CD (Coll/Dist) • CH (Conventional Highway) • FF (Fwy-Fwy connector) • FR (Off Ramp) • HV (HOV) • ML (Mainline) • OR (On Ramp) | |
| Station Length | Segment length covered by the station in miles/km. | |
| Samples | Total number of samples received for all lanes. | |
| % | Percentage of individual lane points at this rved (e.g. not imputed). | % |

**Available Files**

| | |
|---|---|
| d07_text_station_5min_2017_01_01.txt.gz | 29,904,775 |
| d07_text_station_5min_2017_01_02.txt.gz | 29,818,202 |
| d07_text_station_5min_2017_01_03.txt.gz | 30,539,262 |
| d07_text_station_5min_2017_01_04.txt.gz | 30,796,270 |
| d07_text_station_5min_2017_01_05.txt.gz | 30,902,921 |
| d07_text_station_5min_2017_01_06.txt.gz | 31,103,360 |
| d07_text_station_5min_2017_01_07.txt.gz | 30,247,905 |
| d07_text_station_5min_2017_01_08.txt.gz | 29,894,169 |
| d07_text_station_5min_2017_01_09.txt.gz | 30,693,123 |
| d07_text_station_5min_2017_01_10.txt.gz | 30,732,643 |
| d07_text_station_5min_2017_01_11.txt.gz | 30,447,848 |
| d07_text_station_5min_2017_01_12.txt.gz | 30,133,204 |
| d07_text_station_5min_2017_01_13.txt.gz | 30,239,503 |
| d07_text_station_5min_2017_01_14.txt.gz | 29,745,742 |
| d07_text_station_5min_2017_01_15.txt.gz | 29,184,458 |
| d07_text_station_5min_2017_01_16.txt.gz | 29,849,125 |
| d07_text_station_5min_2017_01_17.txt.gz | 30,256,074 |
| d07_text_station_5min_2017_01_18.txt.gz | 30,324,233 |
| d07_text_station_5min_2017_01_19.txt.gz | 30,342,561 |
| d07_text_station_5min_2017_01_20.txt.gz | 30,635,586 |
| d07_text_station_5min_2017_01_21.txt.gz | 30,060,650 |
| d07_text_station_5min_2017_01_22.txt.gz | 29,261,290 |
| d07_text_station_5min_2017_01_23.txt.gz | 30,258,720 |
| d07_text_station_5min_2017_01_24.txt.gz | 30,510,170 |

The data includes:

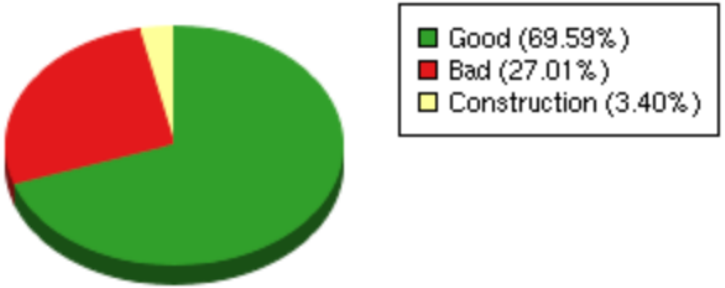
- Timestamp
- Loop IDs
- District
- Freeway name
- Freeway direction
- Total flow
- Average speed

# Location data of Loop IDs

| Fwy | District | County | City | CA PM | Abs PM | Length | ID | Name | Lanes | Type | Sensor Type | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SR2-E | 7 | Los Angeles | Los Angeles | 15.2 | 7.762 | 2.65 | 767494 | E-O ROUTE | 3 | Mainline | loops | 34.10388 | -118.25 |
| SR2-E | 7 | Los Angeles | Los Angeles | 15.5 | 8.062 | 0.45 | 767509 | FLETCHER D | 4 | Mainline | loops | 34.10825 | -118.25 |
| SR2-E | 7 | Los Angeles | Los Angeles | 16.1 | 8.662 | 0.45 | 767523 | SAN FERNA | 4 | Mainline | loops | 34.11428 | -118.242 |
| SR2-E | 7 | Los Angeles | Los Angeles | 16.4 | 8.962 | 0.45 | 767541 | 34TH AVE | 4 | Mainline | loops | 34.11635 | -118.238 |
| SR2-E | 7 | Los Angeles | Los Angeles | R17 | 9.562 | 0.55 | 767554 | EAGLE ROC | 4 | Mainline | loops | 34.12151 | -118.229 |
| SR2-E | 7 | Los Angeles | Los Angeles | R17.5 | 10.062 | 0.5 | 767572 | YORK BLVD | 5 | Mainline | loops | 34.1286 | -118.229 |
| SR2-E | 7 | Los Angeles | Los Angeles | R18 | 10.562 | 0.65 | 767620 | N-O YORK B | 4 | Mainline | loops | 34.1359 | -118.229 |
| SR2-E | 7 | Los Angeles | Los Angeles | R18.8 | 11.362 | 0.45 | 767455 | CONN ROU | 4 | Mainline | loops | 34.14712 | -118.226 |
| SR2-E | 7 | Los Angeles | Los Angeles | R18.9 | 11.462 | 0.335 | 768202 | HOLLY DR | 4 | Mainline | loops | 34.14847 | -118.226 |
| SR2-E | 7 | Los Angeles | Los Angeles | R19.47 | 12.032 | 0.65 | 767470 | N-O CHEVY | 6 | Mainline | loops | 34.15554 | -118.224 |
| SR2-E | 7 | Los Angeles | Glendale | R20.2 | 12.762 | 0.665 | 767585 | MOUNTAIN | 5 | Mainline | loops | 34.1669 | -118.224 |
| SR2-E | 7 | Los Angeles | Glendale | R20.8 | 13.362 | 0.65 | 767597 | N-O MOUN | 5 | Mainline | loops | 34.17536 | -118.221 |
| SR2-E | 7 | Los Angeles | Glendale | R21.5 | 14.062 | 0.913 | 767609 | FERN LANE | 5 | Mainline | loops | 34.185 | -118.217 |
| SR2-E | 7 | Los Angeles | Glendale | R22.626 | 15.188 | 3.063 | 768238 | VERDUGO E | 2 | Mainline | loops | 34.2005 | -118.219 |
| SR2-W | 7 | Los Angeles | Los Angeles | 15.2 | 7.762 | 2.863 | 767495 | E-O ROUTE | 3 | Mainline | loops | 34.1039 | -118.25 |
| SR2-W | 7 | Los Angeles | Los Angeles | 15.925 | 8.487 | 0.599 | 718064 | SAN FERNA | 4 | Mainline | loops | 34.11266 | -118.245 |
| SR2-W | 7 | Los Angeles | Los Angeles | 16.4 | 8.962 | 0.561 | 767542 | 34TH AVE | 4 | Mainline | loops | 34.11651 | -118.238 |
| SR2-W | 7 | Los Angeles | Los Angeles | R17.048 | 9.61 | 0.55 | 718066 | VERDUGO F | 4 | Mainline | loops | 34.12161 | -118.23 |
| SR2-W | 7 | Los Angeles | Los Angeles | R17.5 | 10.062 | 0.476 | 767573 | YORK BLVD | 4 | Mainline | loops | 34.1286 | -118.229 |
| SR2-W | 7 | Los Angeles | Los Angeles | R18 | 10.562 | 0.368 | 767621 | N-O YORK B | 4 | Mainline | loops | 34.13588 | -118.23 |
| SR2-W | 7 | Los Angeles | Los Angeles | R18.237 | 10.799 | 0.4 | 762329 | COLORADO | 4 | Mainline | loops | 34.13879 | -118.229 |

# Available data set for Caltrans

| District | Total Stations |
|---|---|
| District 3 | 1247 |
| District 4 | 3880 |
| District 5 | 382 |
| District 6 | 624 |
| District 7 | 4856 |
| District 8 | 2115 |
| District 10 | 1189 |
| District 11 | 1502 |
| District 12 | 2539 |
| Total | 18334 |




http://pems.dot.ca.gov/

# Available data set for Caltrans

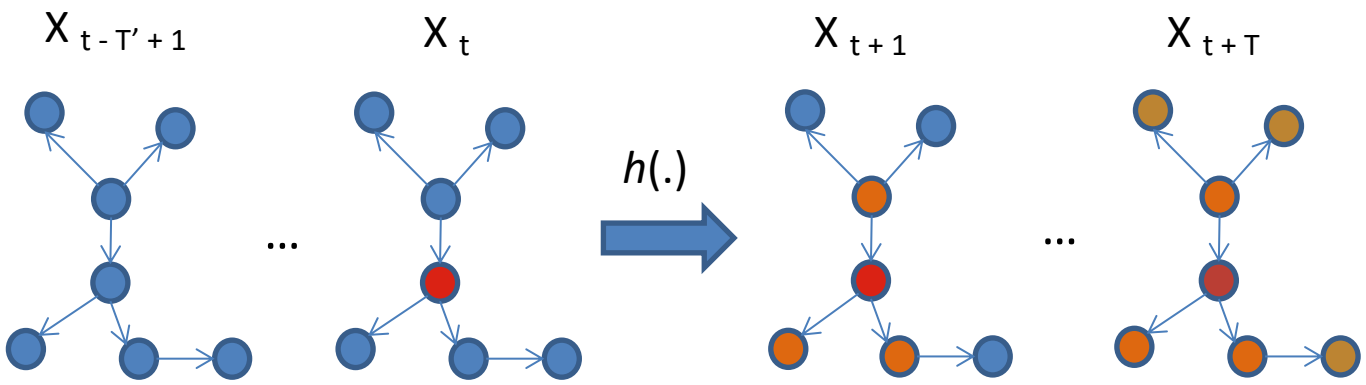
Total : 11, 160 detector stations

Duration: 1st Jan 2018 to 31st Dec 2018

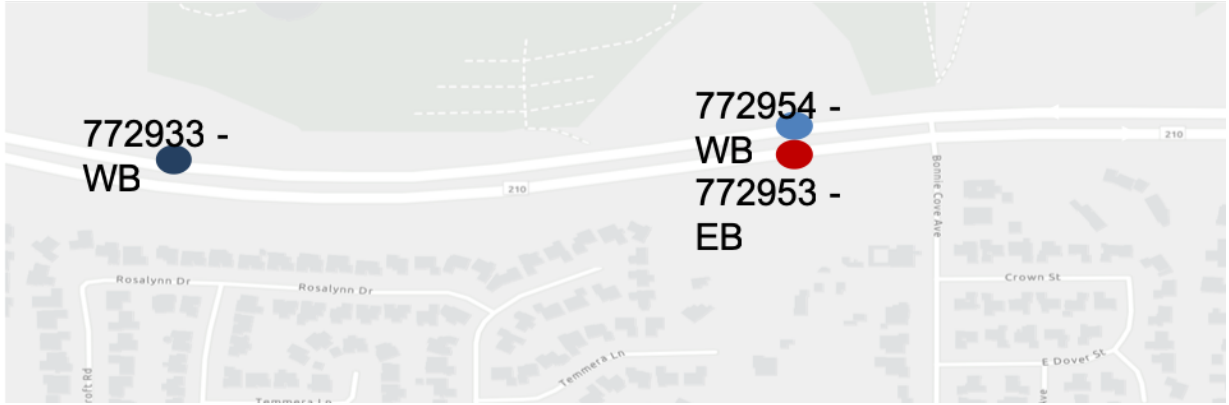
# Network distance computation

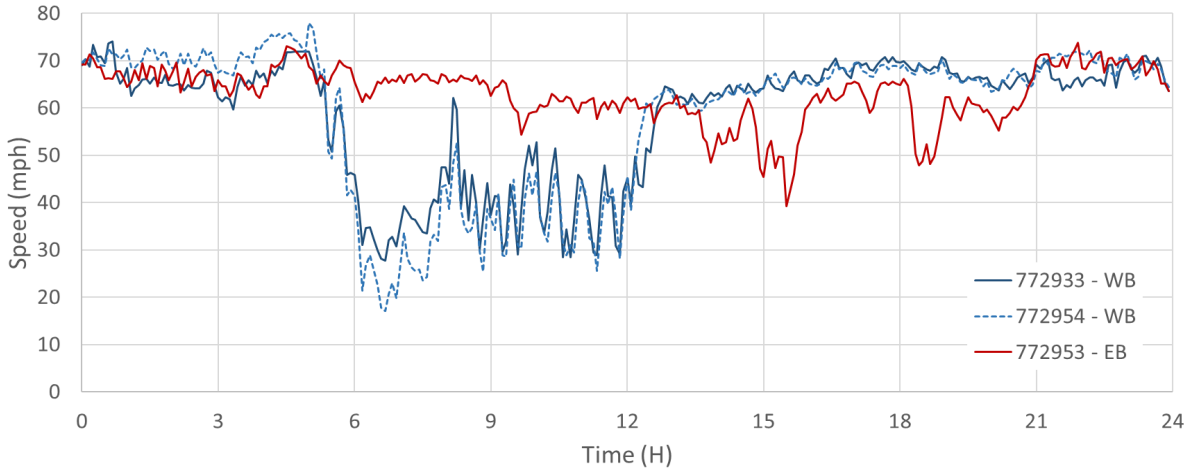
- OSRM local server (unlimited query)
- Find the driving distance between two nearest loop IDs by querying in OSRM server
  - Nearest neighbor using Euclidean distance (to speedup)
- OSRM gives shortest driving distance between two latitude and longitude

# Problem statement: Traffic prediction



# Challenges



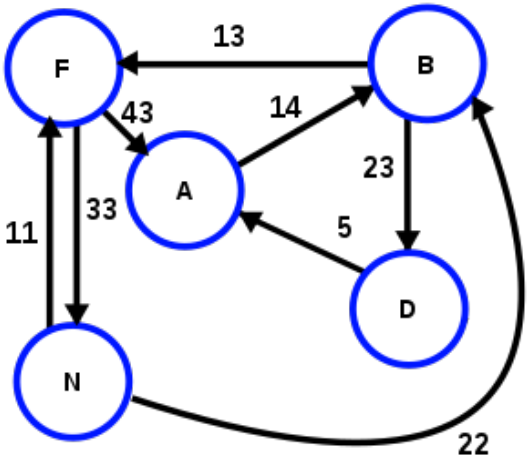




- Complex spatial dependency
- Non-stationary temporal dynamics
- Non-Euclidean spatial geometry
- Model each sensor independently fail to capture spatial correlation

# Forecasting for multiple loop detector

## Capturing the road network in terms of graphs

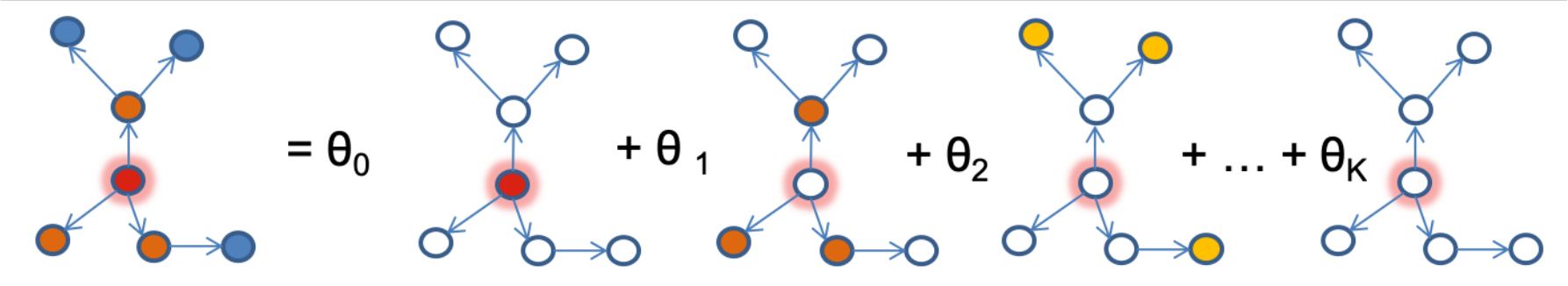
- Transportation network as graph
  - V = Vertices (sensors)
  - E = Edges (roads)
  - A = Weighted adjacency matrix
    (A function of the road network distance)




$$A_{ij} = \exp\left(-\frac{\mathrm{dist}_{\mathrm{net}}(v_i, v_j)^2}{\sigma^2}\right) \text{ if } \mathrm{dist}_{\mathrm{net}}(v_i, v_j) \leq \kappa$$

$\mathrm{dist}_{\mathrm{net}}(v_i, v_j)$: road network distance from $v_i$ to $v_j$,
$\kappa$: threshold to ensure sparsity, $\sigma^2$ variance of all pairwise road network distances

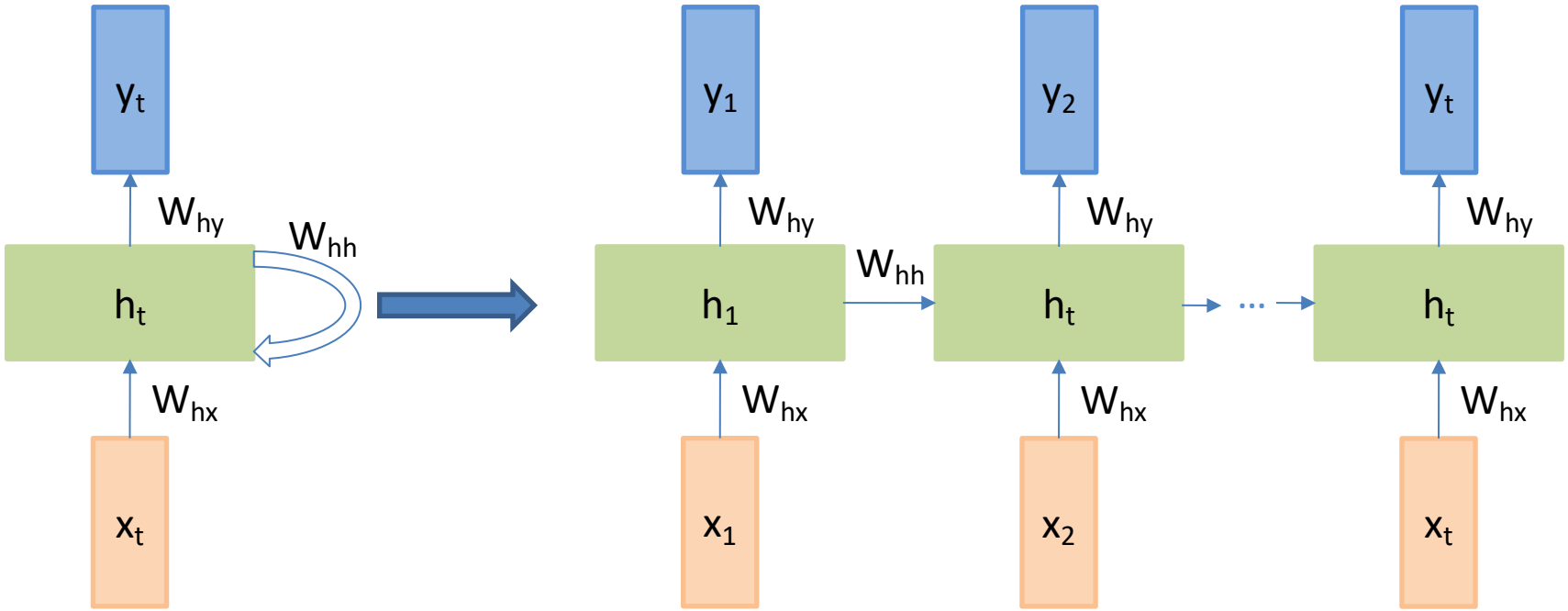
# Diffusion convolution



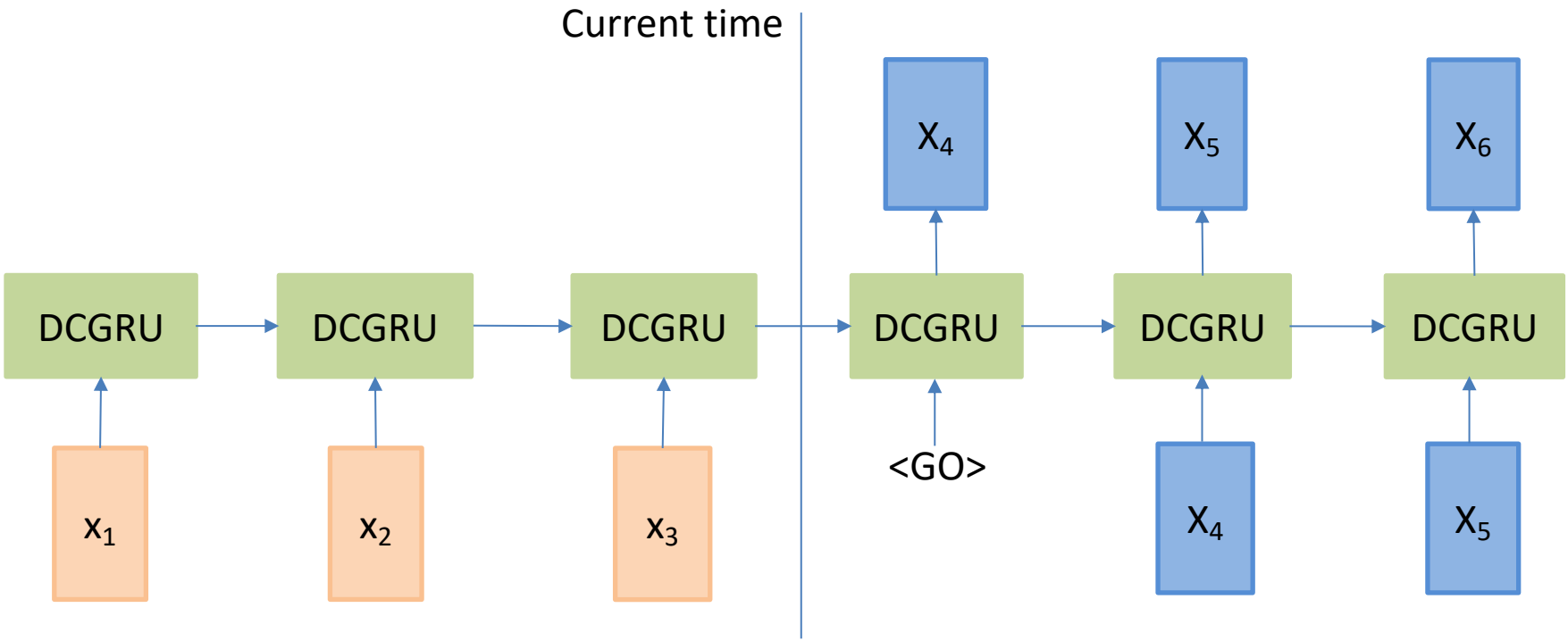

Out-degree In-degree

$$\boldsymbol{X}_{:,p} \star_{\mathcal{G}} f_\theta = \sum_{k=0}^{K-1} \left( \theta_{k,1} \left( \boldsymbol{D_O^{-1}} \boldsymbol{A} \right)^k + \theta_{k,2} \left( \boldsymbol{D_I^{-1}} \boldsymbol{A}^\intercal \right)^k \right) \boldsymbol{X}_{:,p}$$

# Recurrent Neural Network



# Encoder decoder Framework of DCRNN



# GRU cell

$$u_t = \sigma(w_u[h_{t-1}, x_t])$$

$$r_t = \sigma(w_r[h_{t-1}, x_t])$$

$$c_t = tanh(w_c[r_t * h_{t-1}, x_t])$$

$$h_t = (1 - u_t) * h_{t-1} + u_t * c_t$$

# DCRNN cell

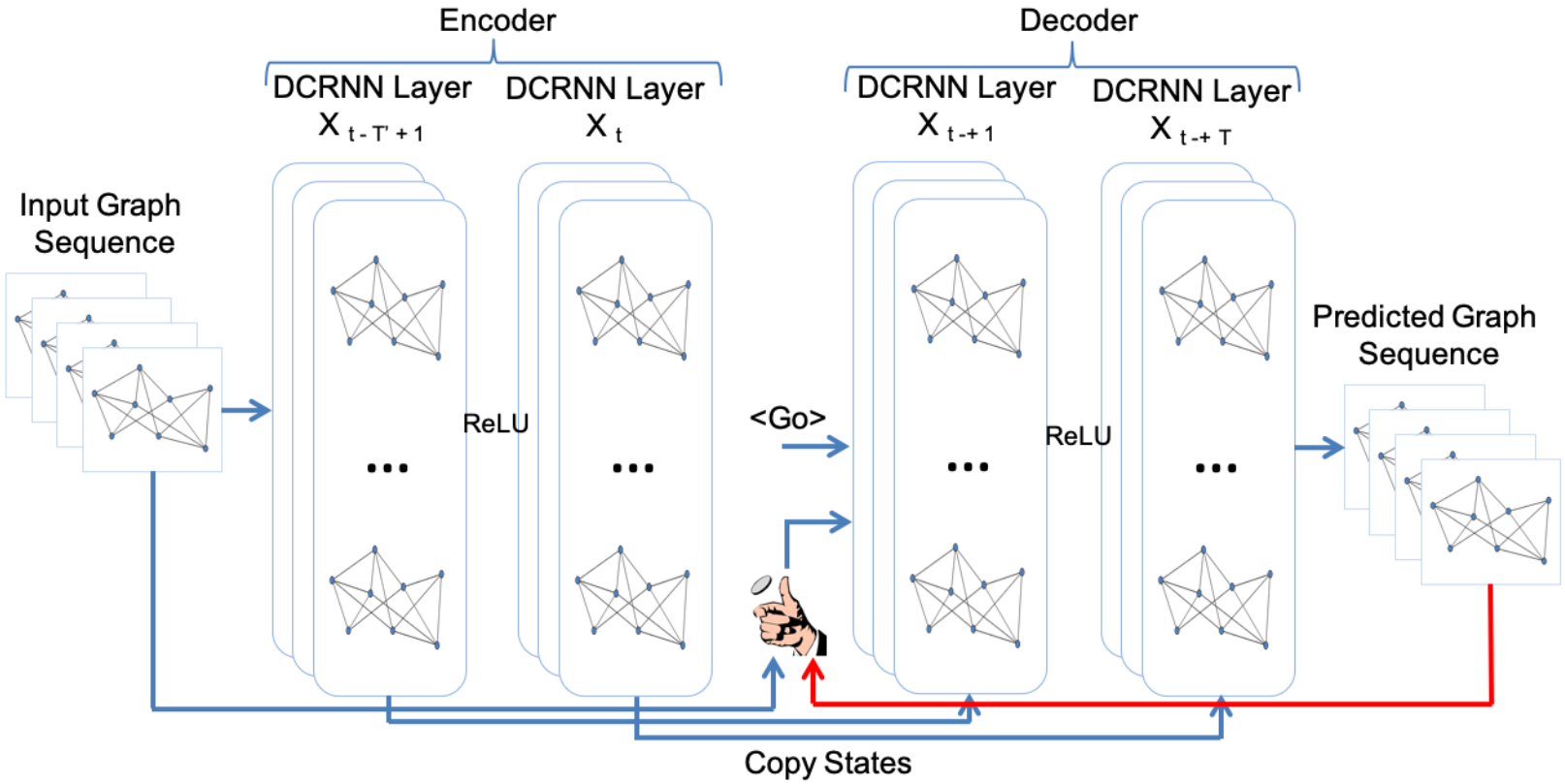
$$u_t = \sigma(w_{u*G}[h_{t-1}, x_t])$$

$$r_t = \sigma(w_{r*G}[h_{t-1}, x_t])$$

$$c_t = tanh(w_{c*G}[r_t * h_{t-1}, x_t])$$

$$h_t = (1 - u_t) * h_{t-1} + u_t * c_t$$

# Traffic forecasting using DCRNN

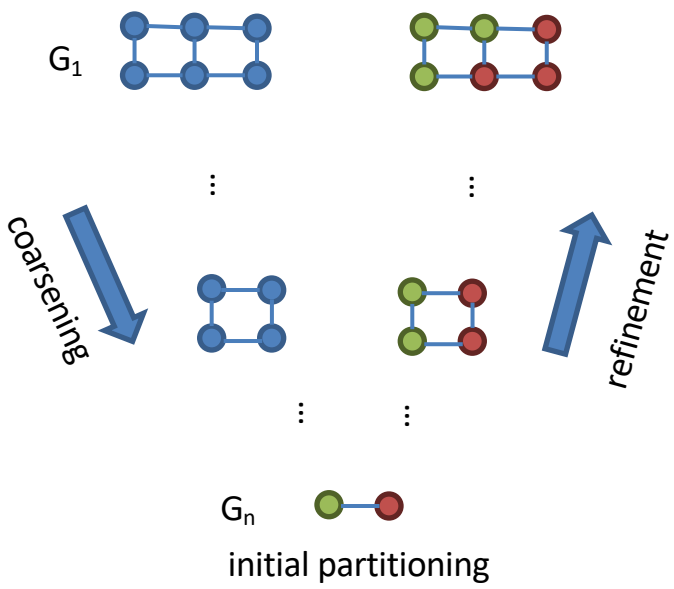


# Scaling DCRNN on HPC systems

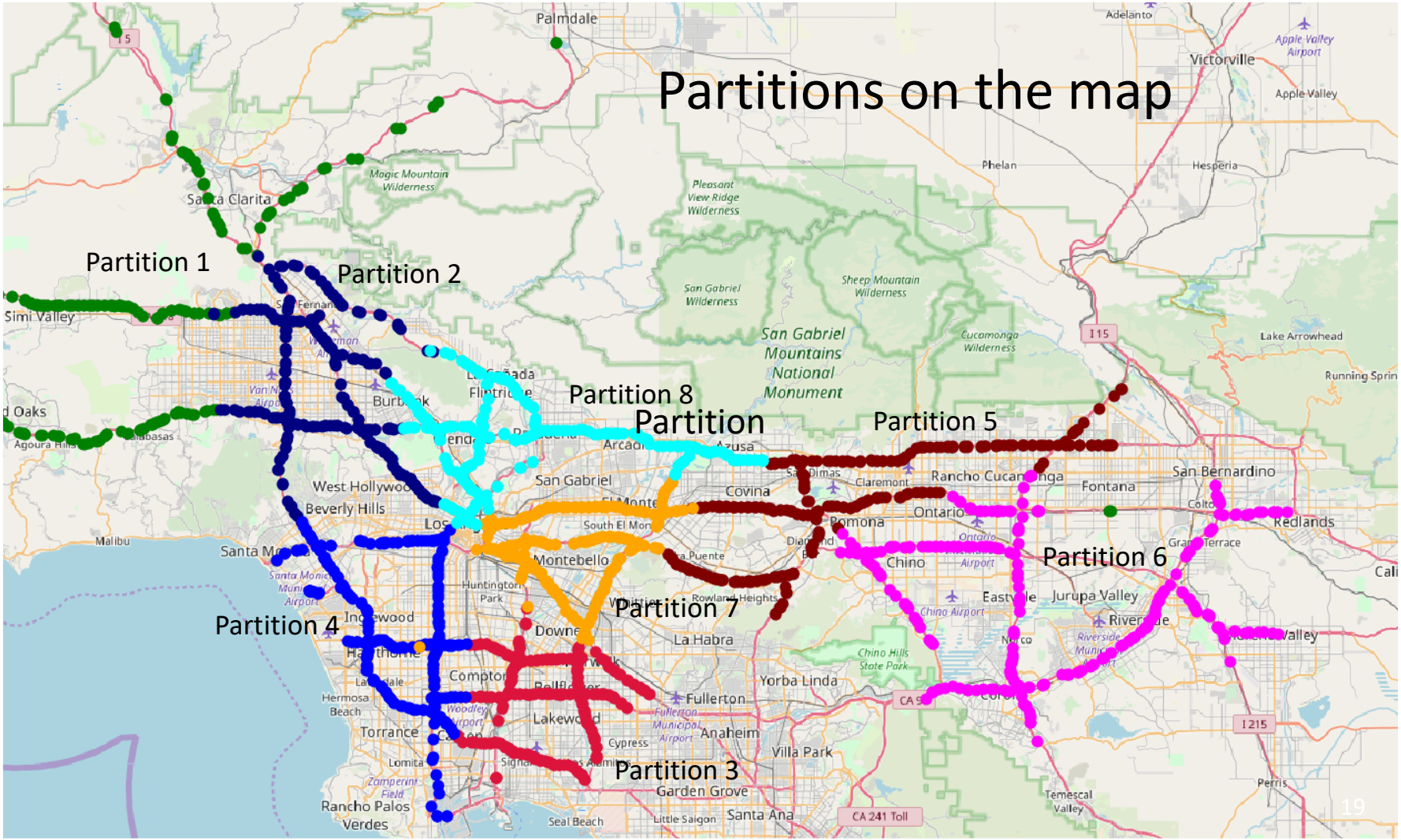
- Domain decomposition for large networks
  - Partition large graph into sub-graph
  - Run DCRNN for each sub-graph on a compute node
  - Combine the results and forecast traffic
- Simultaneous execution of DCRNN

# Graph partitioning using Metis

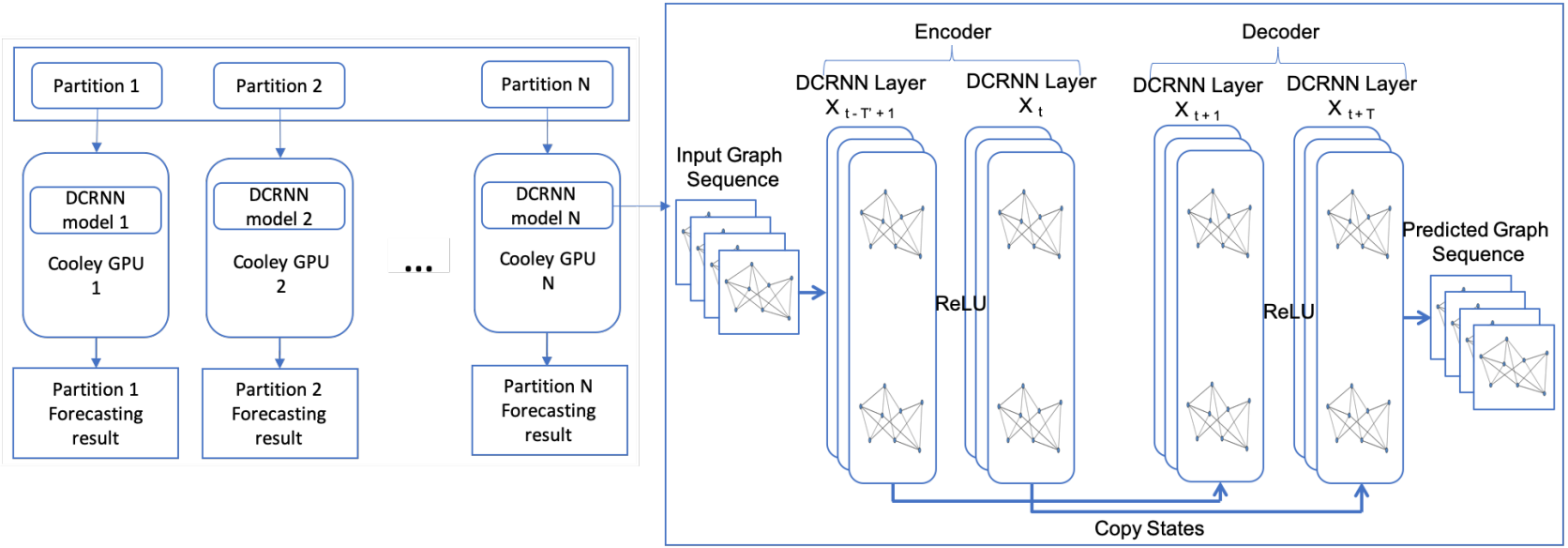
- 3 major phases
  - coarsening
  - initial partitioning
  - refinement
- Key advantage
  - millions of vertices in few seconds



# Partitions on the map



# Simultaneous execution of DCRNN

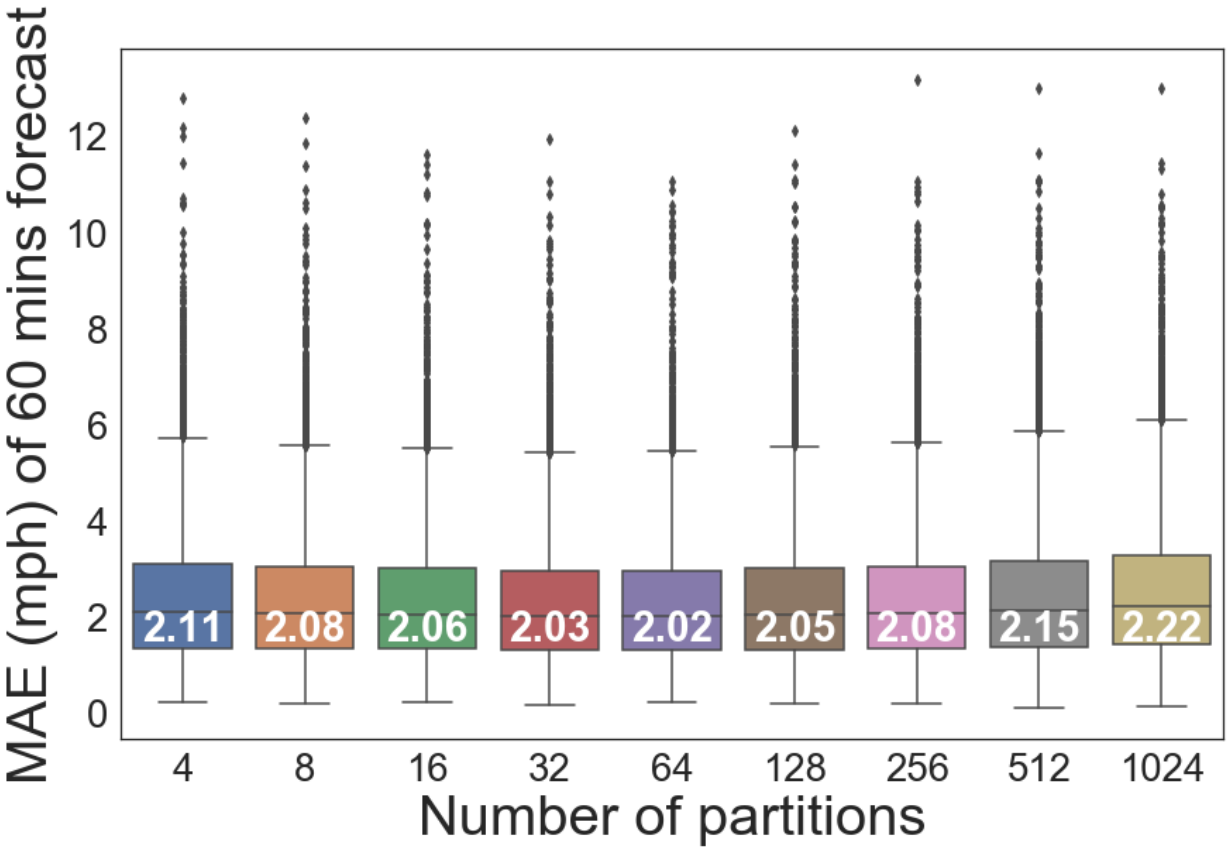


# Data set

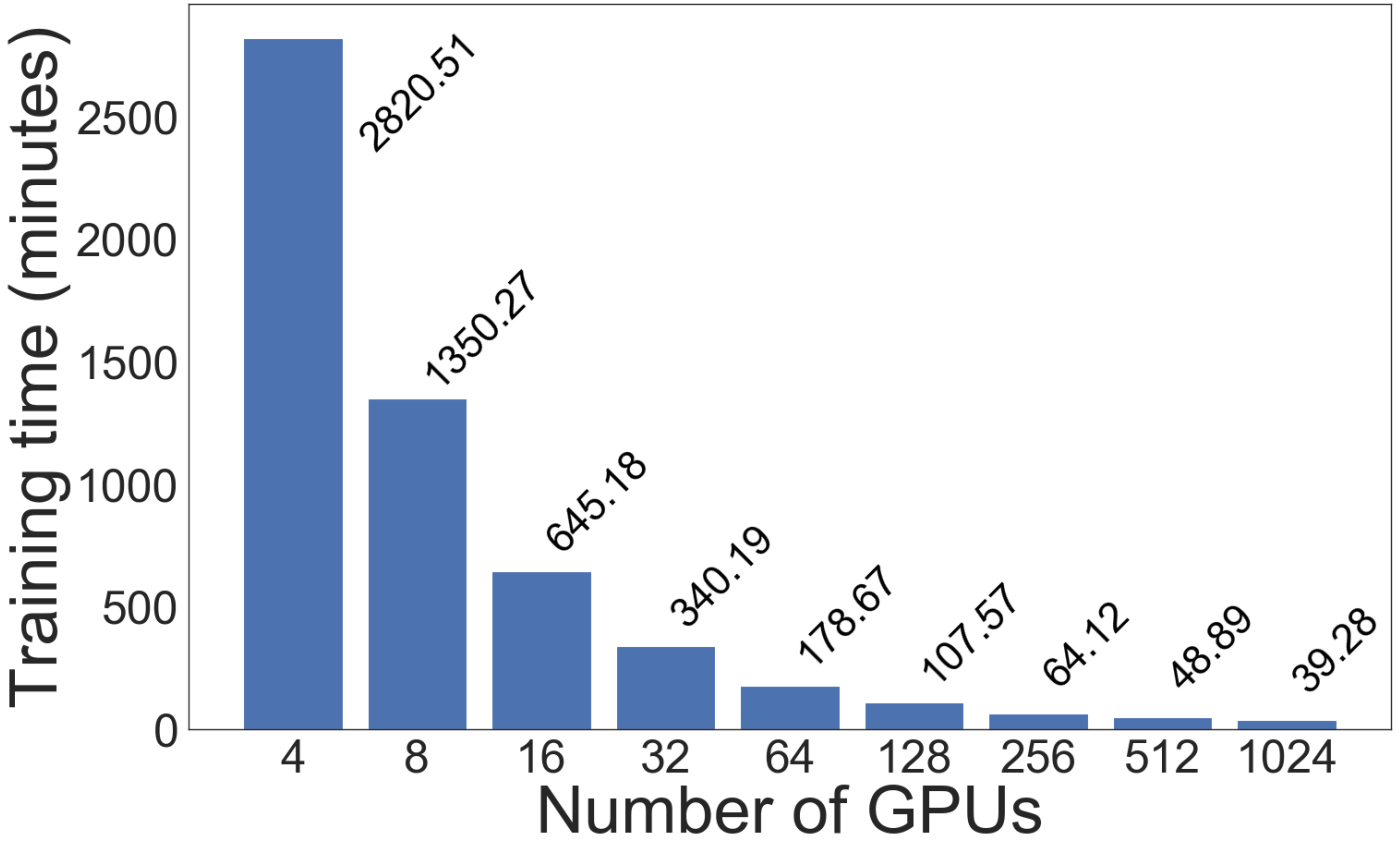
| Number of partitions | Stations per partition (aprox) |
|---|---|
| 2 | 5580 |
| 4 | 2790 |
| 8 | 1395 |
| 16 | 697 |
| 32 | 348 |
| 64 | 174 |
| 128 | 87 |
| 256 | 43 |
| 512 | 21 |
| 1024 | 11 |

- Total stations: 11160
- Divide the whole data set in different number of partitions

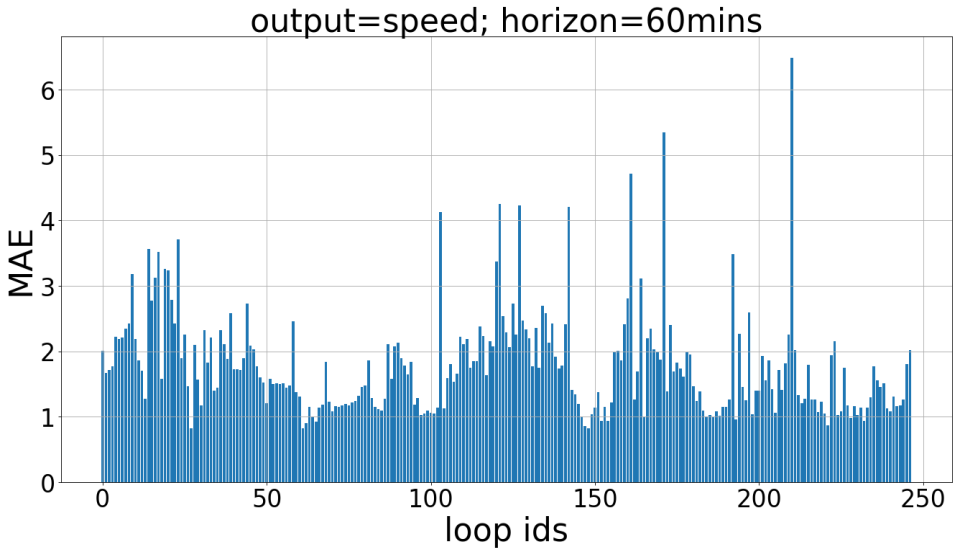
# Forecasting accuracy



# Execution time

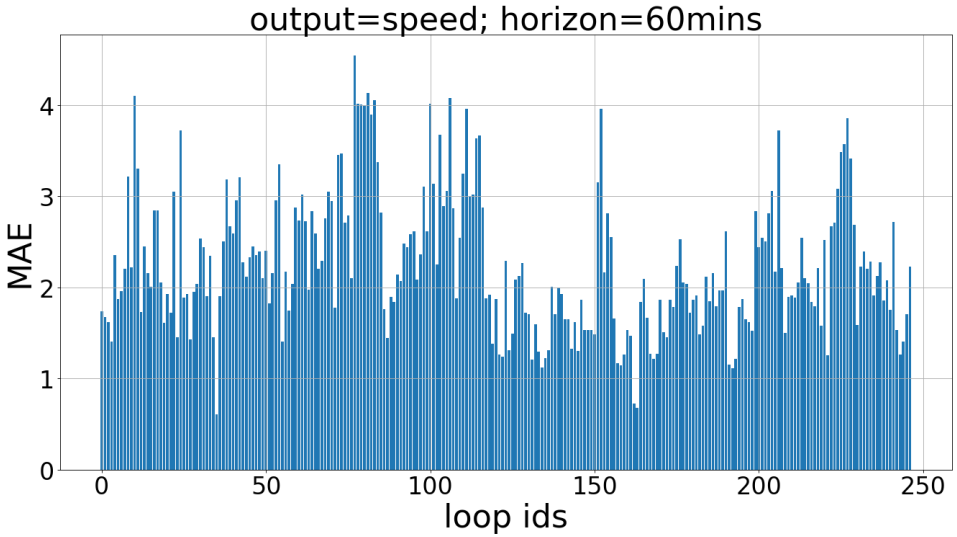


# DCRNN results




Group: 1

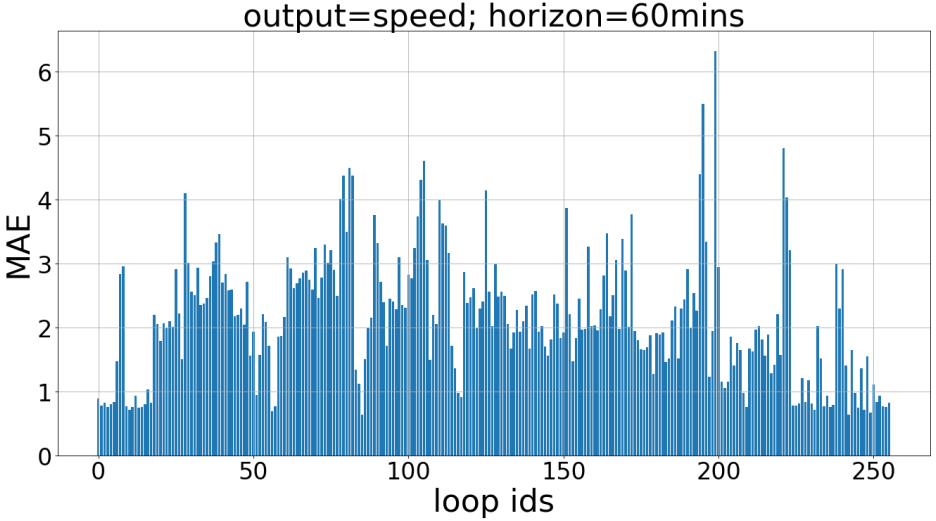
Node: 247



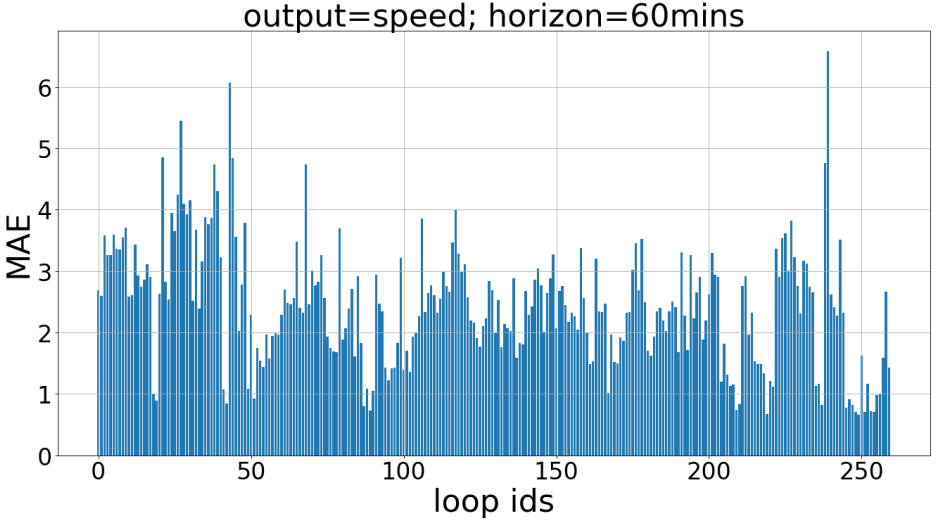

Group: 2

Node: 247

# DCRNN results

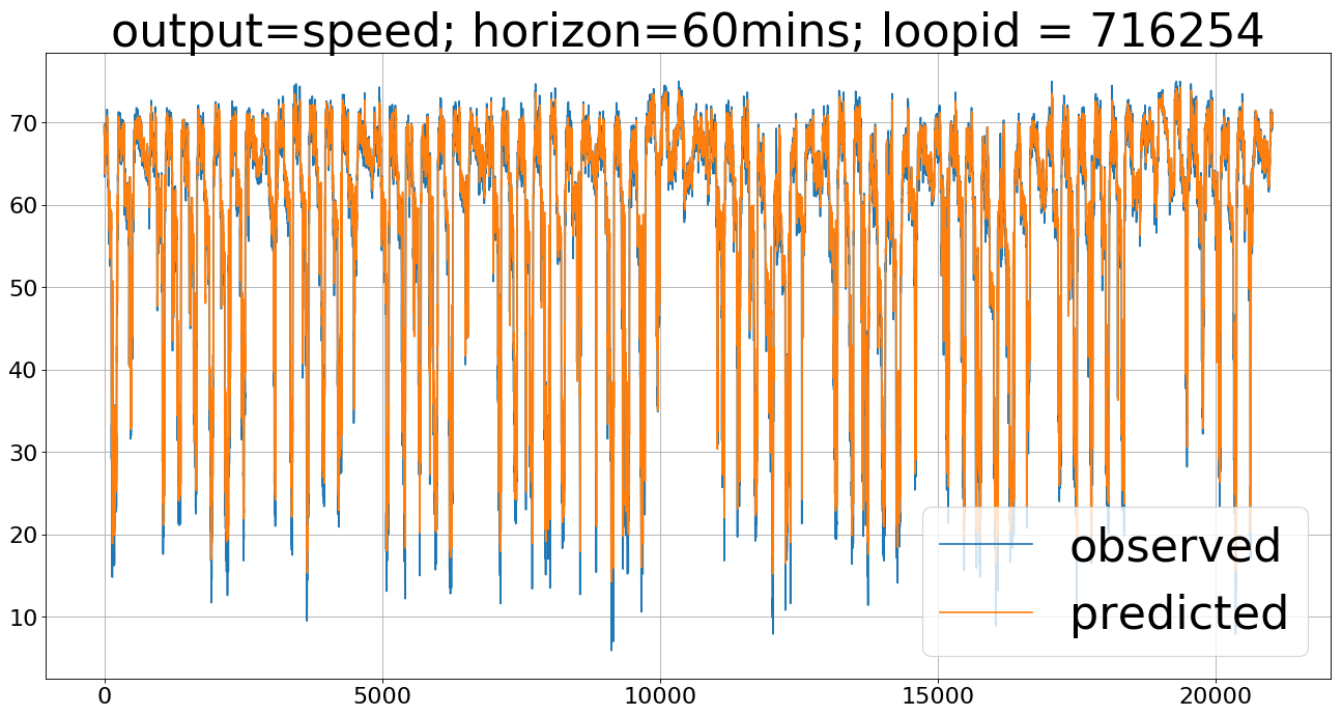



Group: 3
Node: 256



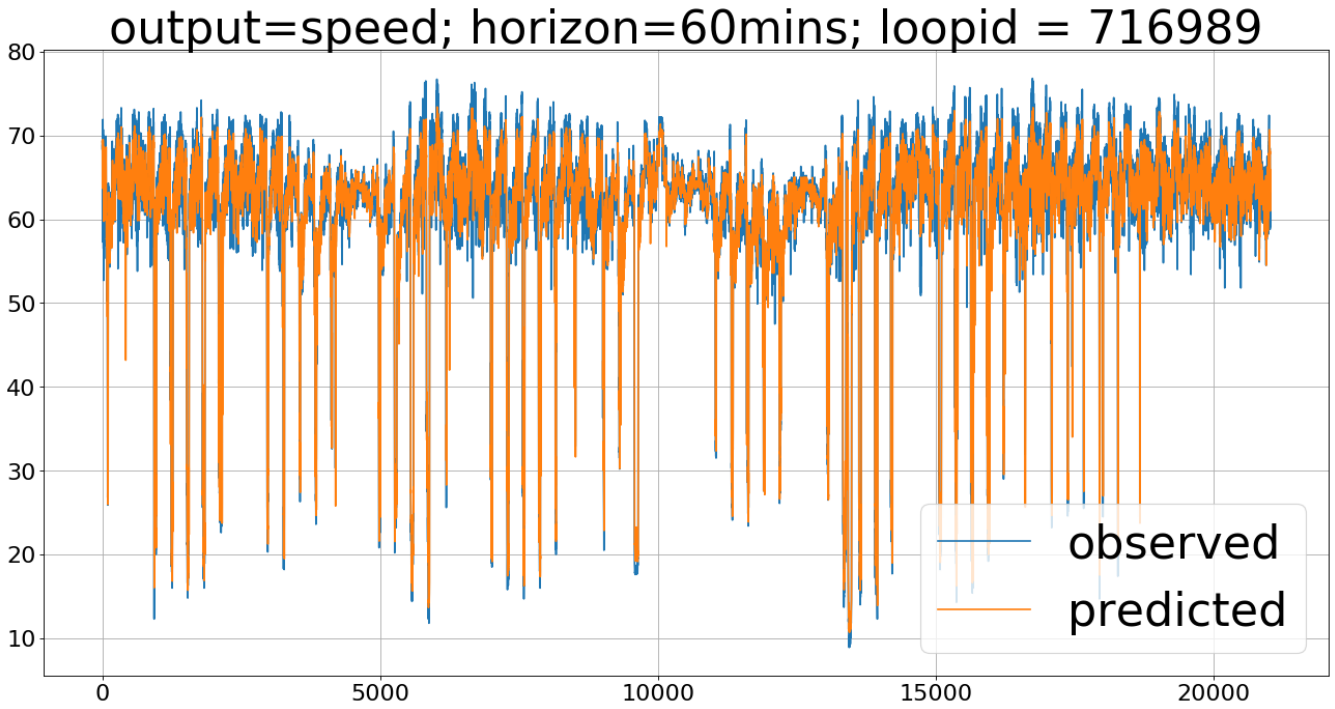

Group: 7
Node: 260

# Speed prediction



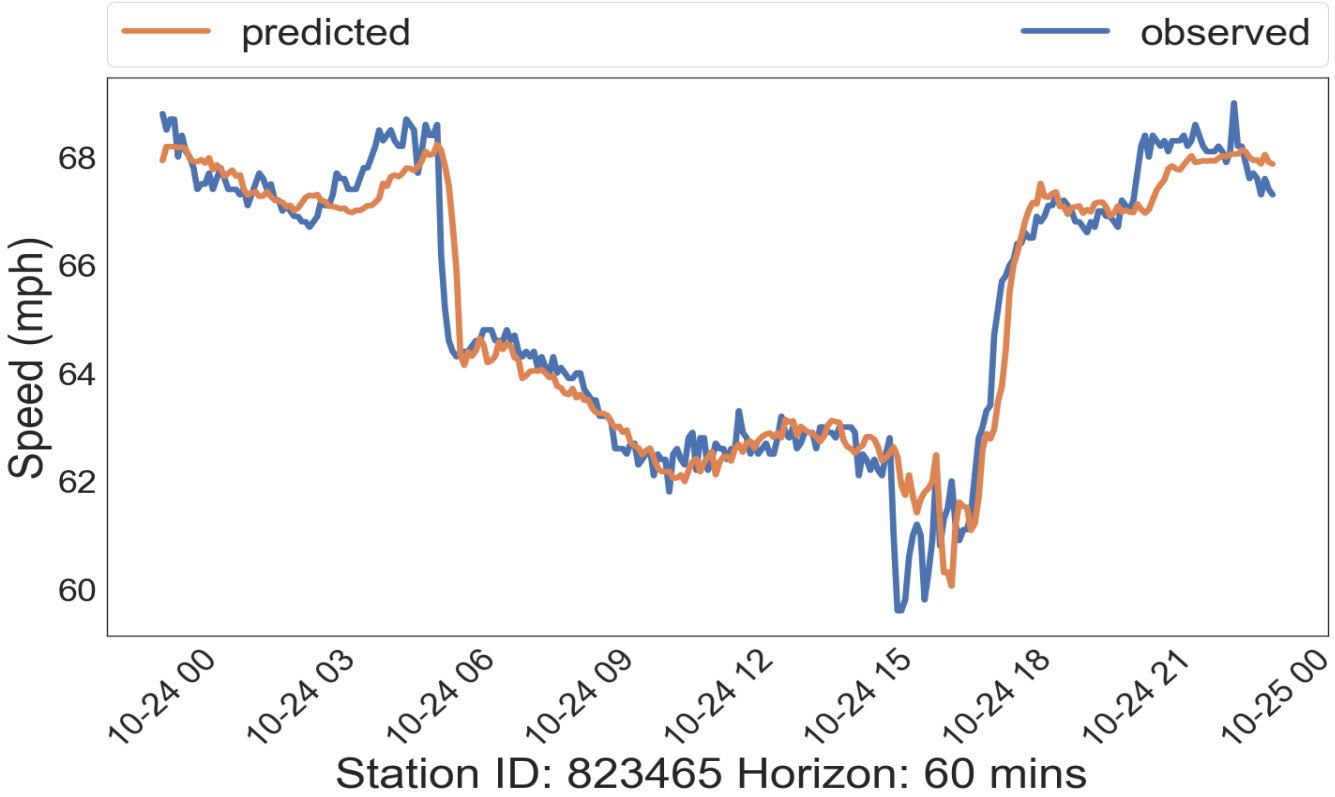

Group: 5

# Speed prediction

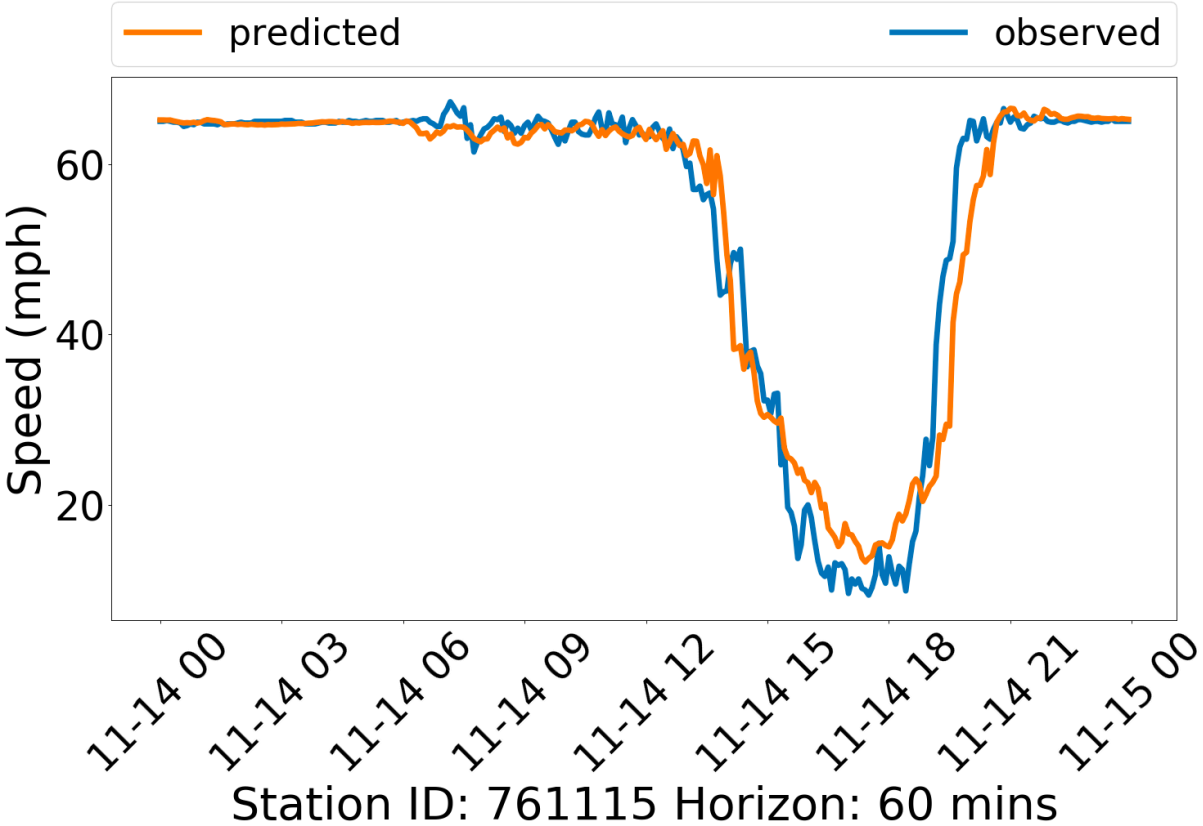



Group: 2

# Speed prediction



# Speed prediction



# Summary

- DCRNN captures the spatial as well as temporal dynamics well
- DCRNN is scalable for large network without performance degradation