Time Series Based Traffic Flow Prediction

——Bill (Lin Yuchen)
2016.5.25

Outline

- **Introduction**
- **Problem Definition**
- **Solutions**
- **Performance Evaluation**
- **Conclusion**

Introduction

- Traffic Congestion wastes time and energy.
- **Traffic Flow Prediction** is a HOT topic in many disciplines:
  - Transportation Science
  - Civil Engineering
  - Policy Planning
  - Operation Research

Introduction

- **3 main previous methods:**
  - Mathematical models, Simulation studies, Field survey
- **Now we have a gold mine of real-time traffic data.**
  - CCTV cameras
  - GPS devices
  - Other Traffic sensors: under-pavement loop detectors
- **We can implement Data Mining Techniques on them.**

Problem Definition

- Consider a set of **n** road segments comprising **n** traffic sensors (e.g., loop detectors).
- We assume that at given time interval **t** (e.g., every 5 mins), each sensor provides a traffic data reading, e.g., speed v[t].
- Definition 1:
  - {**i**} denotes **sensors** and {**j**} denotes a series of continuous **time**
  - Given a set of observed speed readings $V = \{ v_i(j),\ i = 1,\ldots,n;\ j = 1,\ldots,t \}$
  - The prediction problem is to find the set $V = \{ v_i(j),\ j = t+1,\ t+2,\ldots t+h \}$ for each sensor i, where **h denotes the prediction horizon.**

Problem Definition

- Example:

Time Interval=5min
h=3

sensor \time	8:05	8:10	8:15	8:20	8:25	8:30
V1	10.1	13.5	14.9	?	?	?
V2	9.2	7.5	8.2	?	?	?
V3	9.2	9.6	9.4	?	?	?

- Definition 2:
  - When h = 1, it refers to a short-term prediction
  - When h > 1, it refers to a long-term prediction

1. Introduction
2. Problem Definition
3. Solutions
    1. ARIMA
    2. HAM
    3. H-ARIMA
4. Performance Evaluation
5. Conclusion

ARIMA Model

- Auto-Regressive Integrated Moving Average (ARIMA)
- $\{Y_t\}$ refers to a time series data (e.g., the sequence of speed readings).

$$Y_{t+1} = \sum_{i=1}^{p} \alpha_i Y_{t-i+1} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i+1} + \varepsilon_{t+1}$$

Auto-Regressive component：
a linear weighted combination of previous data is calculated, where **p** refers to the order of this model and $\boldsymbol{\alpha_i}$ refers to the weight of (t – i + 1)-th speed

Noise Items From Moving Average Model

the sum of weighted noise from the moving average model is calculated, where **q refers to its order** ,**ε denotes the noise**, and $\boldsymbol{\beta_i}$ **represents the weight of (t – i + 1)-th noise.**

ARIMA Model

$$Y_{t+1} = \sum_{i=1}^{p} \alpha_i Y_{t-i+1} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i+1} + \varepsilon_{t+1}$$

- The predicted value mainly relies on the linear combination of the data that occurred before time t.
- This model can be directly used to predict the traffic speed data, when prediction horizon h=1.
- When h >1, we can iterate the prediction process h times by <u>using the predicted value as the input</u> to predict the next value.
- An obvious disadvantage: When h is increasing, the accuracy is decreasing.

HAM

- Historical Average Model

Speed of A certain Sensor

	Monday Of Week3	Monday Of Week4	Monday Of Week5	Monday Of Week6
8:30	10.1	10.3	10.2	?
8:40	10.5	10.4	10.6	?
8:50	9.9	9.8	9.8	?
9:00	7.8	7.7	7.6	?

$$v(t_{d,w} + h) = \frac{1}{|V(d,w)|} \sum_{s \in V(d,w)} v(s)$$

V (d, w) refers to the subset of past observations that happened at the same time **d** on the same day **w**.

d captures the **daily** effects
w captures the **weekly** effects

(10.1+10.3+10.2) / 3 = 10.2

Hypothesis1：

The prediction horizon**(h)** has no noticeable effect on **HAM**. However, as the h increases, the accuracy of **ARIMA** decreases.

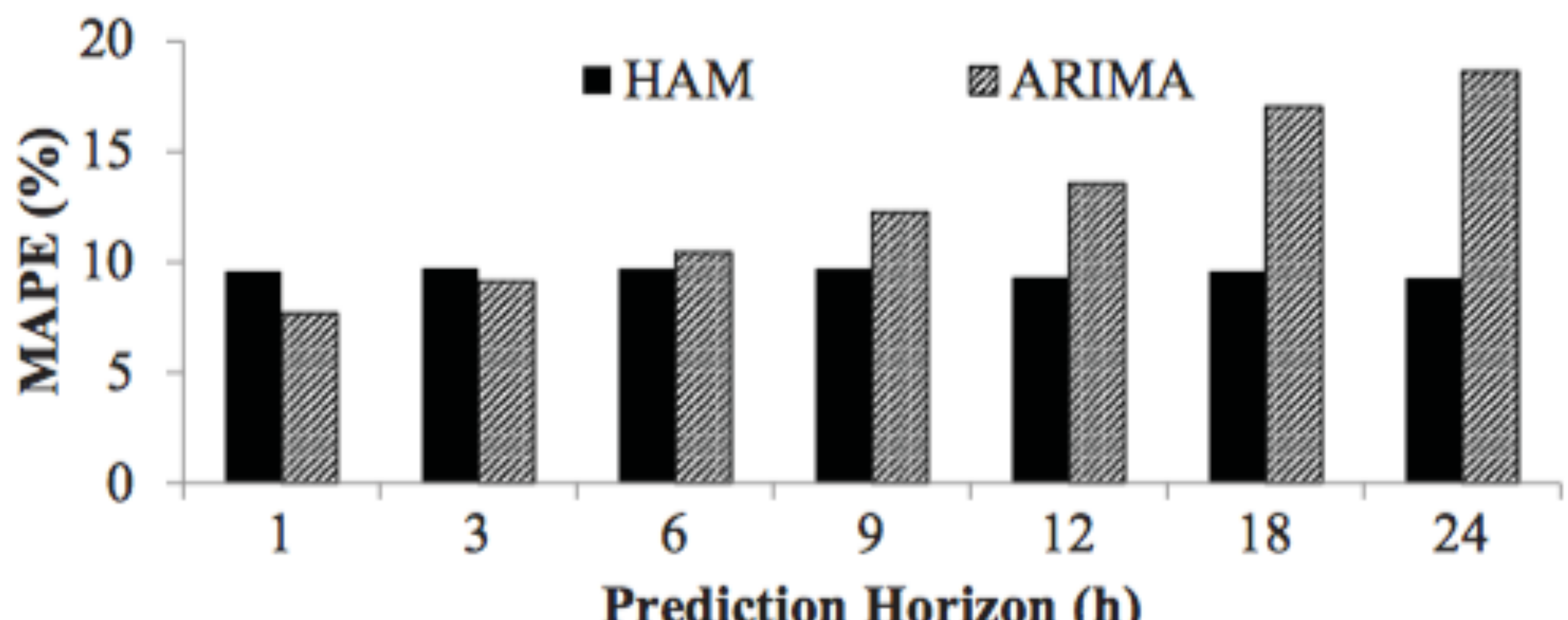


Figure 1. Effect of prediction horizon (*h*)

The dataset is from LA County road network

Hypothesis2：

HAM can effectively predict the **sudden speed changes** at the boundaries (i.e., beginning and end) of rush hours.

ARIMA has a delayed reaction on the boundaries

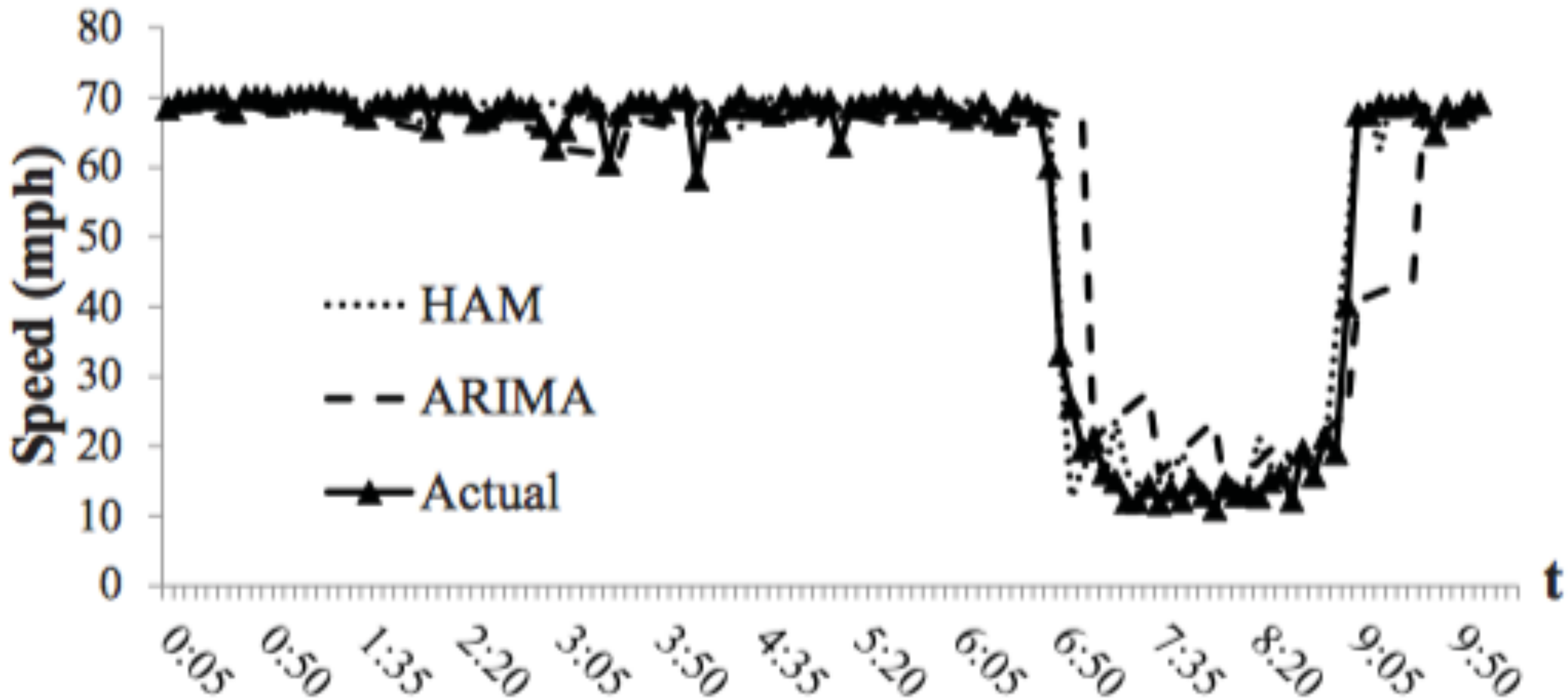


Figure 2. Effect of rush-hour boundaries

1. Introduction
2. Problem Definition
3. Solutions
    1. ARIMA
    2. HAM
    3. H-ARIMA
4. Performance Evaluation
5. Conclusion

H-ARIMA

Historical ARIMA (H-ARIMA) **selects** in real- time between ARIMA or HAM based on their **accuracy**.

We train a **decision-tree model**.
(λ) = the decision parameter
(ϕ) = threshold
For each **t**, we choose between ARIMA and HAM based on the trained value of $\boldsymbol{\lambda_t}$. If $\boldsymbol{\lambda_t \leq \phi}$, we choose ARIMA, otherwise, we choose HAM.

H-ARIMA

dataset

4:00PM, Monday

Algorithm 1 Get $\lambda(\{v(j)\}$, d, w)

Output: λ

Let S = $\{V(\{\text{v(j)}\}, d, w)\}$

subset of {v(j)} for HAM

Let $Err_{\text{ARIMA}} = 0$; $Err_{\text{HAM}} = 0$

init two Error

Initialize ARIMA model with training dataset $\{v(j)\}$

Get ARIMA Model

v_{HAM}= Average(V{d,w});

Get prediction **vHAM** from HAM

for all $v_i \in S$ **do**

Get prediction **vARIMA** from ARIMA model

  v_{ARIMA} = ARIMA(i);

  $Err_{\text{ARIMA}} = Err_{\text{ARIMA}}$+ RMSE($v_i, v_{\text{ARIMA}}$);

  $Err_{\text{HAM}} = Err_{\text{HAM}}$+ RMSE($v_i, v_{\text{HAM}}$);

accumulate the two error

end for

$\lambda = Err_{\text{ARIMA}}$ / (Err_{ARIMA}+Err_{HAM})

Return λ.

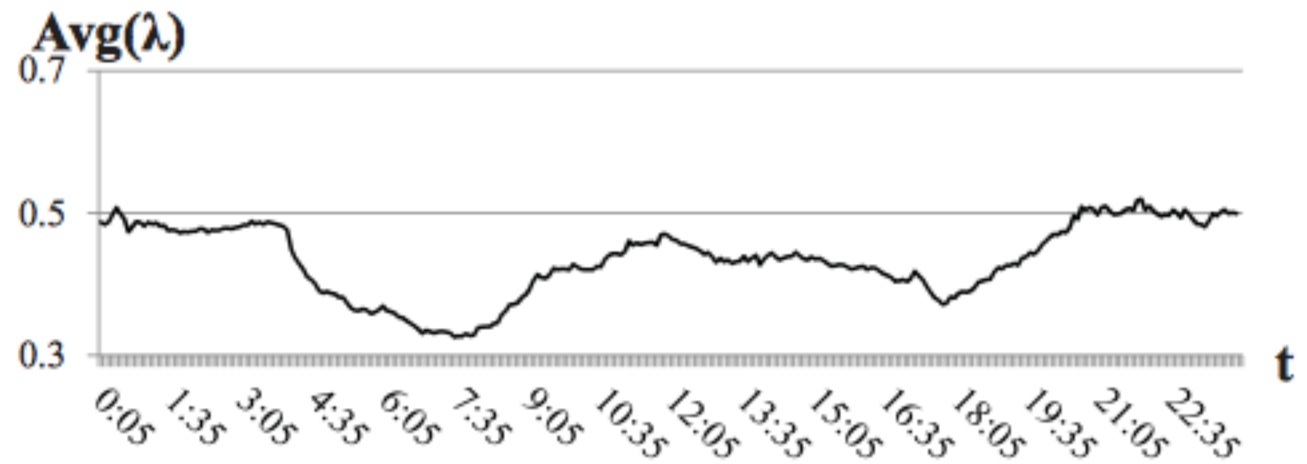
set **λ** the ratio of **ErrARIMA**

if λ <= 0.5, ARIMA is better
if λ > 0.5, HAM is better
so threshold ϕ = 0.5

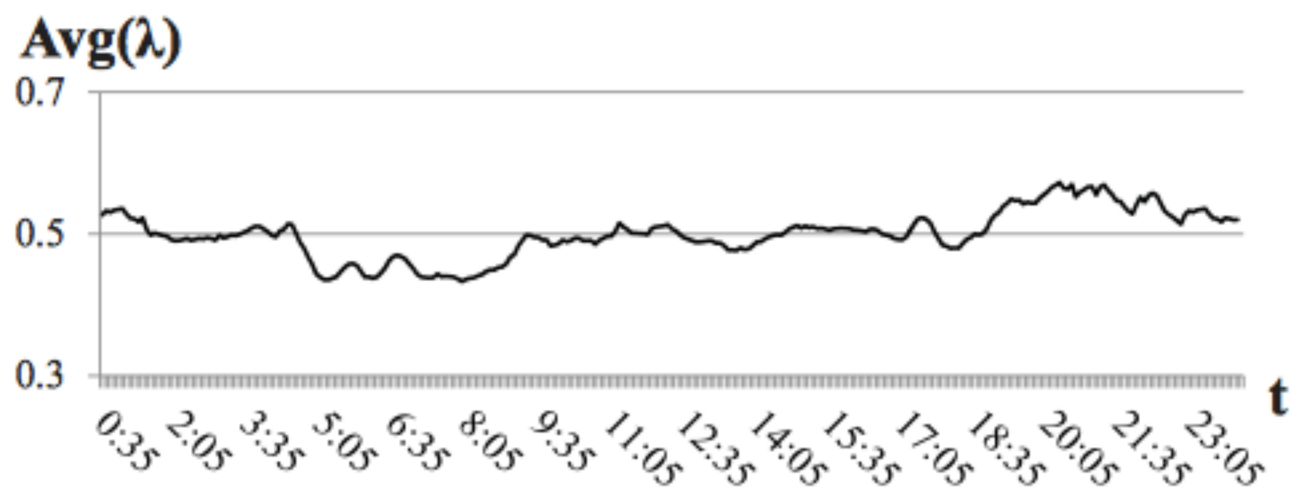
1. Introduction
2. Problem Definition
3. Solutions
    1. ARIMA
    2. HAM
    3. H-ARIMA
4. Performance Evaluation
5. Conclusion

H-ARIMA

The effect of d on λ.



(a) h=1 (5-min in advance prediction)



(b) h=6 (30-min in advance prediction)

H-ARIMA

The effect of rush hours.

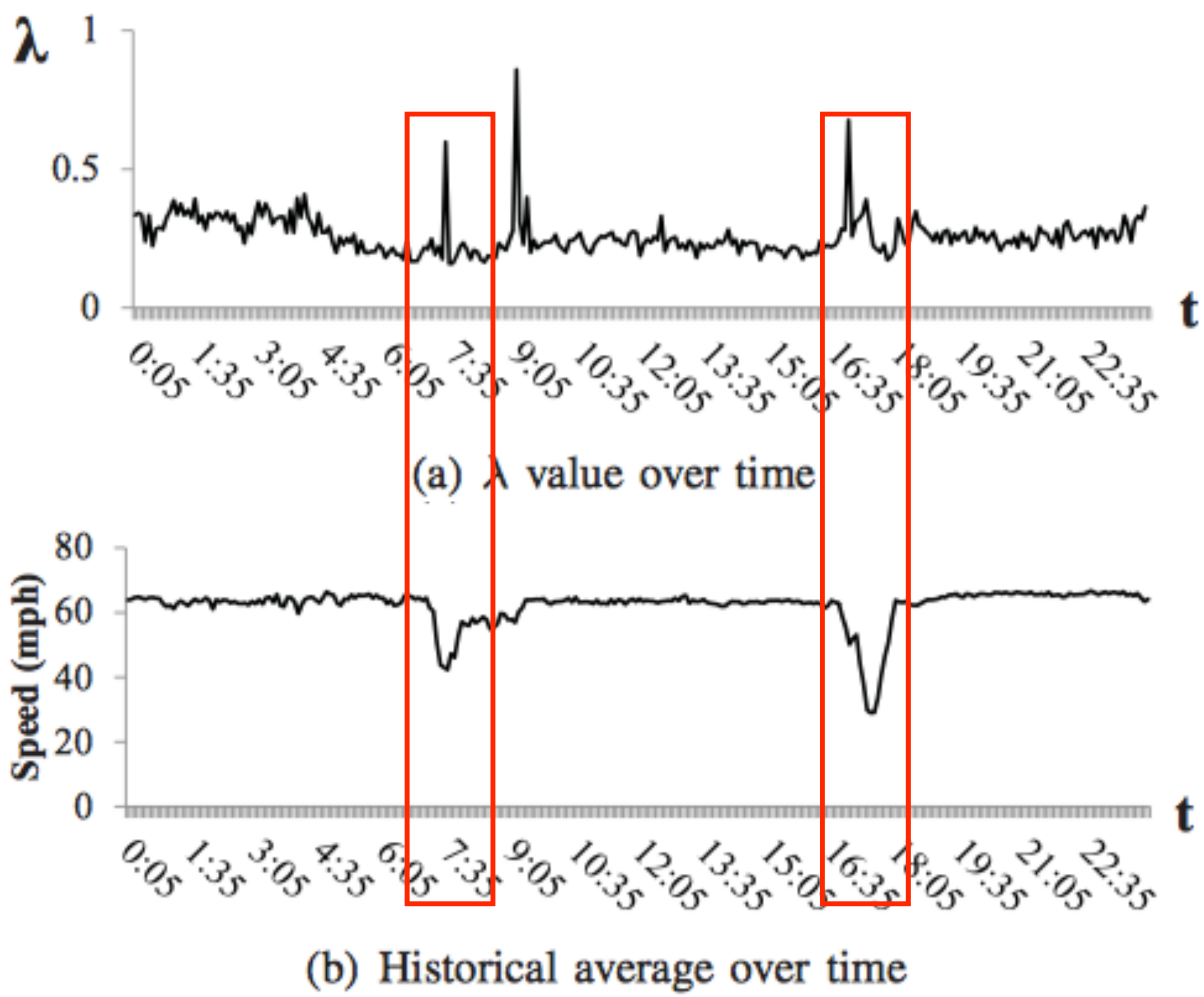


Figure 4. Effects of rush-hour boundaries over λ

Fitness Measurements

(MAPE) Mean Absolute Percent Error
(RMSE) Root Mean Square Error

$$\text{MAPE} = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \widehat{y}_i|}{y_i}\right) \times 100$$

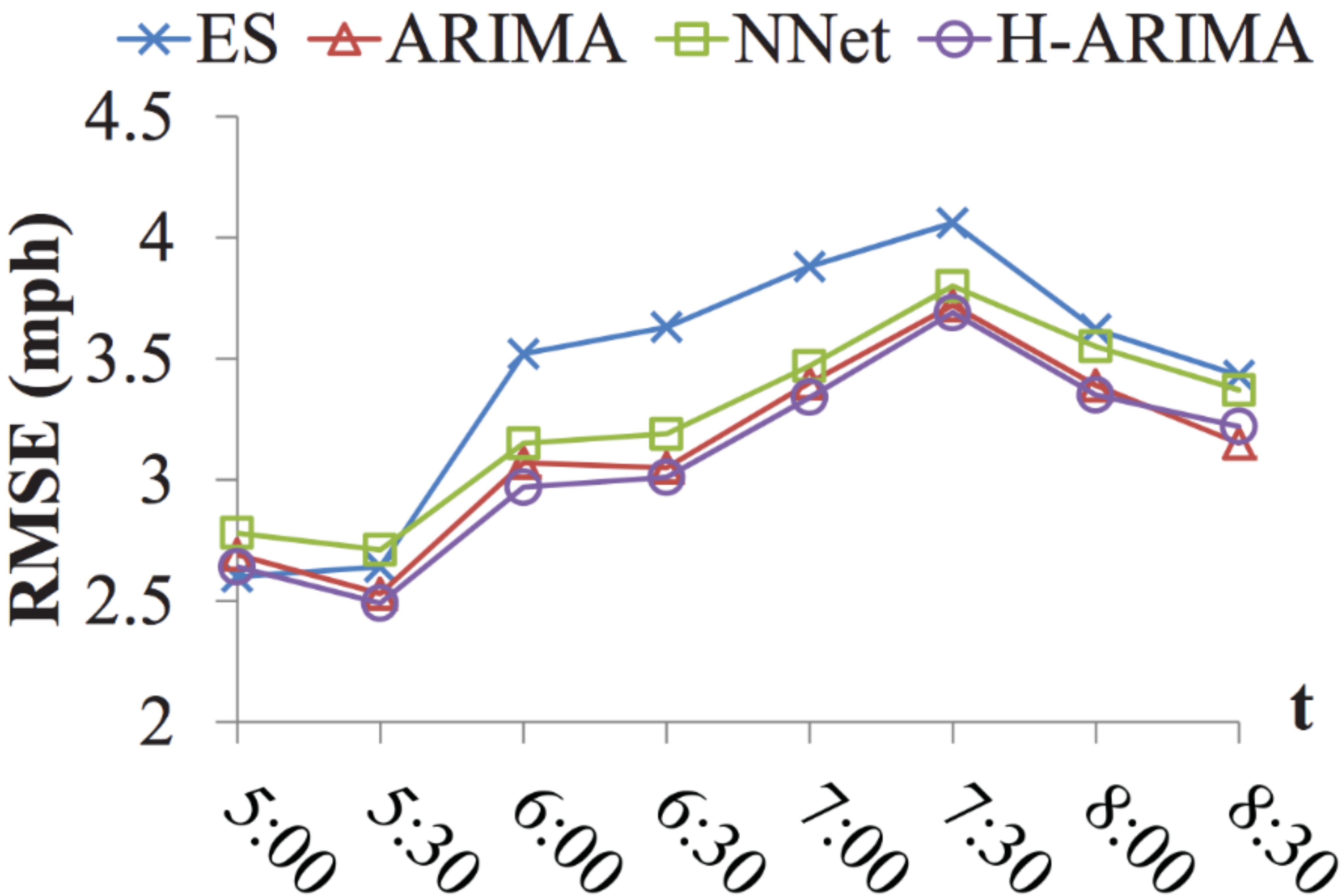
$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}$$

actual

predicted

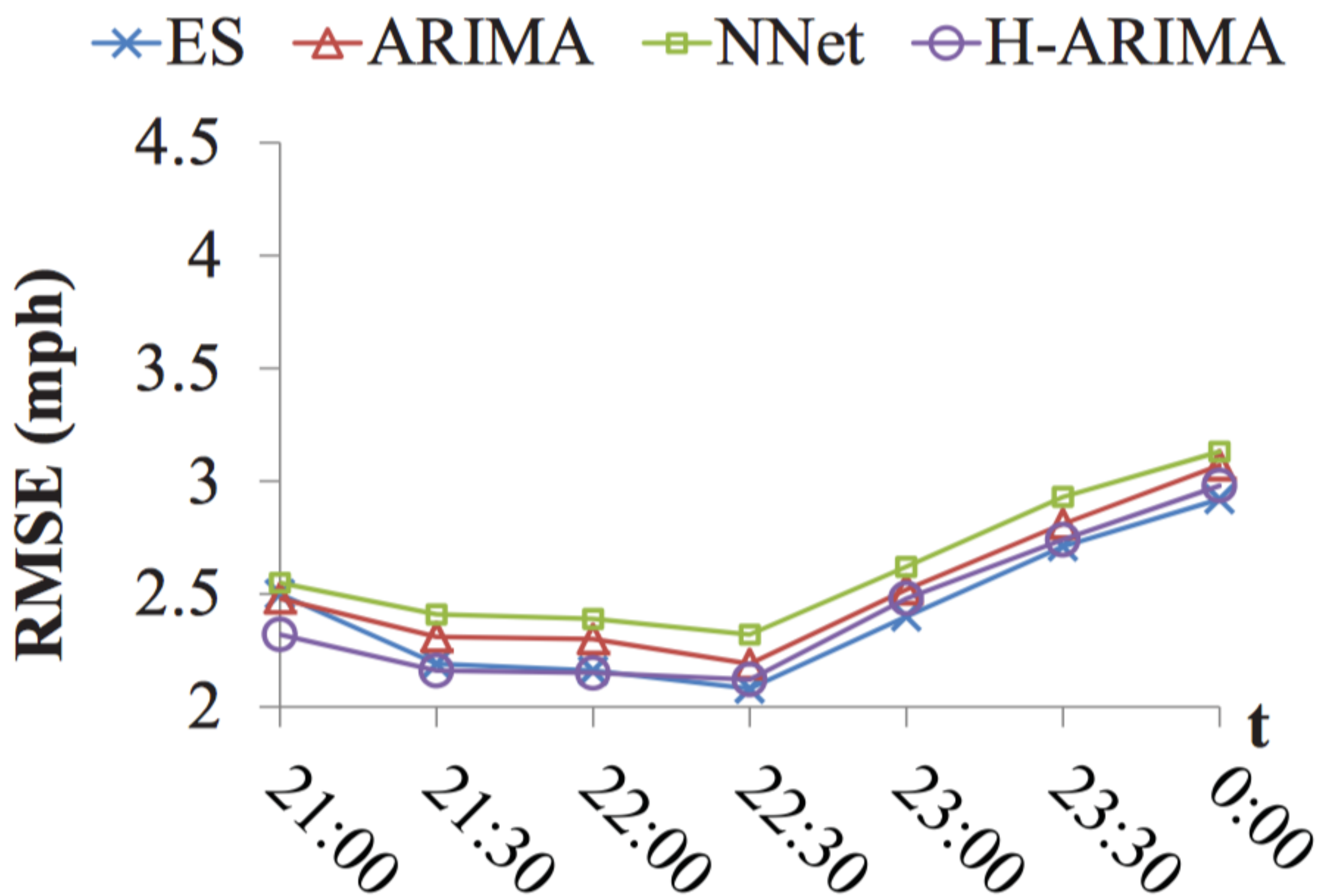
the number of predictions

Short-term Prediction(h=1)



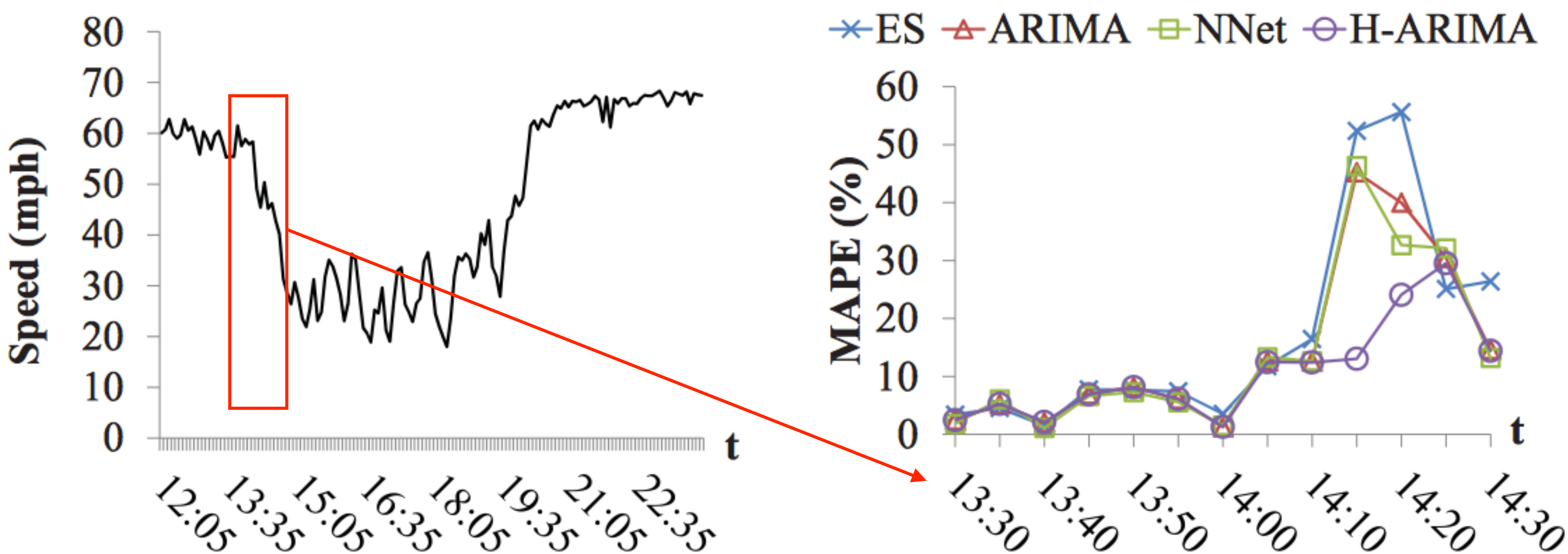
(a) Rush hour

Short-term Prediction(h=1)



(b) Non-rush hour

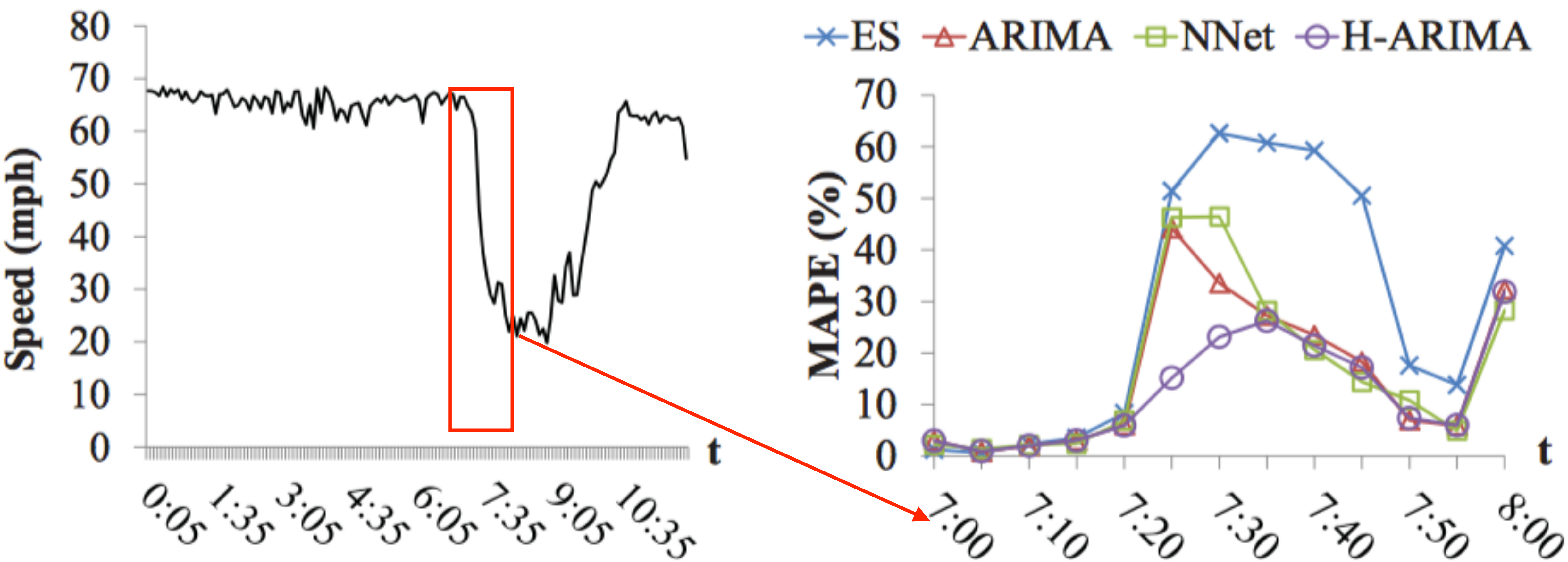
Short-term Prediction(h=1)



(a) Actual speed (b) MAPE of the road

Figure 8. Case study on I-5 S. segment from Downtown

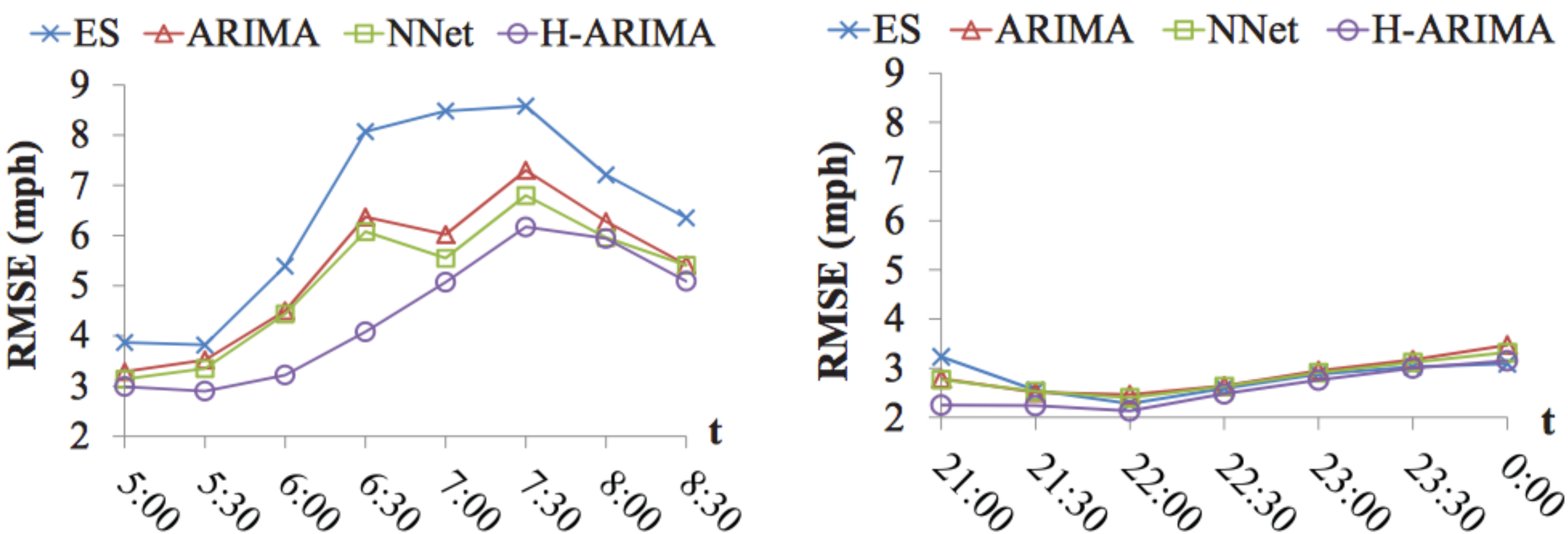
Short-term Prediction(h=1)



(a) Actual speed (b) MAPE of the road

Figure 9. Case study on I-10 W. segment to West-LA

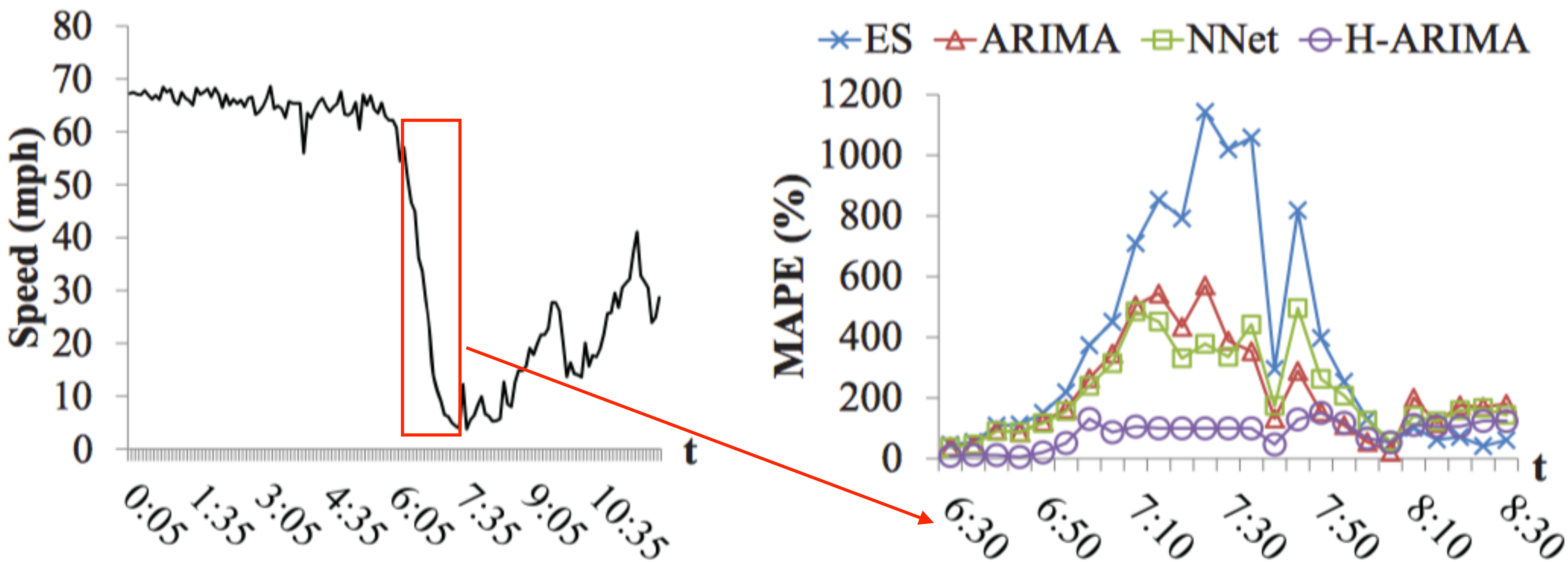
Long-term Prediction(h=6)



(a) Rush hour (b) Non-rush hour

Figure 10. Overall RMSE (h=6)

Long-term Prediction(h=6)



(a) Actual speed (b) MAPE of the road

Figure 11. Case study on I-10 E. segment to Downtown

Conclusion

1.The traditional prediction approaches that treat traffic data streams as <u>generic time series</u> fail to forecast traffic during traffic <u>peak hours</u>.

2.H-ARIMA significantly improves the prediction accuracy of existing approaches(HAM、ARIMA) by incorporating the historical traffic data into the prediction model.

3. The accuracy of H-ARIMA is 67% in short-term and 78% in long-term predictions.

4.Event information (e.g. traffic accidents) and location features(e.g. there is a super market nearby), Thank you for listening.