A Reliable Travel Time Prediction System with Sparsely Distributed Detectors

Ph.D. Dissertation Defense

Nan Zou

May 10th, 2007

Outline

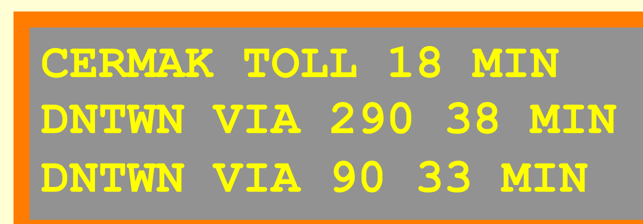
- Introduction
- Research Objectives
- Framework of the Travel Time Prediction System
- System Components
  - Travel Time Estimation Module
  - Travel Time Prediction Module
  - Missing Data Estimation Module
- Summary
- On-going Works

Introduction

- Travel times (completed and en-route trips) are crucial information for an Advanced Traveler Information System (ATIS)

Baltimore, MD

Houston, TX



Chicago, IL

Introduction (cont'd)

- Travel time prediction is a challenging task due to the impacts of
  - Geometric features
  - Traffic patterns
  - Availability of the detection system
  - Delay and/or missing of the real-time data, etc.

Introduction (cont'd)

- Issues Associated with Existing Models and Systems:
  - High system costs
    - Densely distributed detectors (i.e., 0.5-mile apart)
    - Accurate speed detection
    - Recurrent measurement on travel times

      Coifman et al. (2002, 2003), van Lint et al. (2003), Liu et al. (2006)
  - Reliability
    - Missing or delayed data
    - Nonrecurrent congestions (for example, incidents)

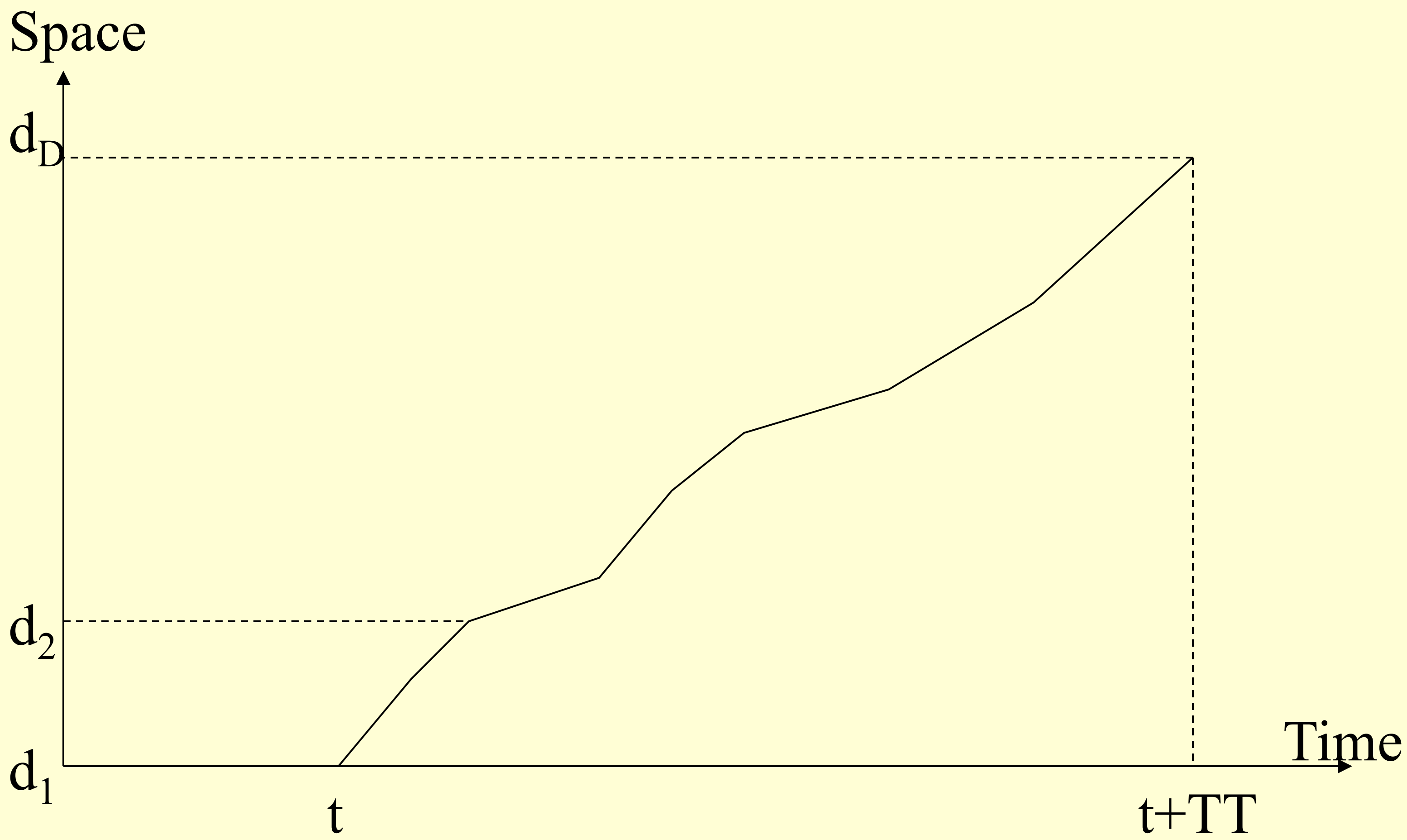
Features of A Cost-efficient and Reliable Travel Time Prediction System

- Required input variables should be obtainable from sparsely distributed traffic detectors
- Take advantage of some actual travel times from the field, but not rely on a large number of such data.
- Be capable of operating under normal and/or some data-missing scenarios and effectively dealing with related issues during real-time operations.
- Estimate the impact of the missing data and avoid potential large prediction errors.

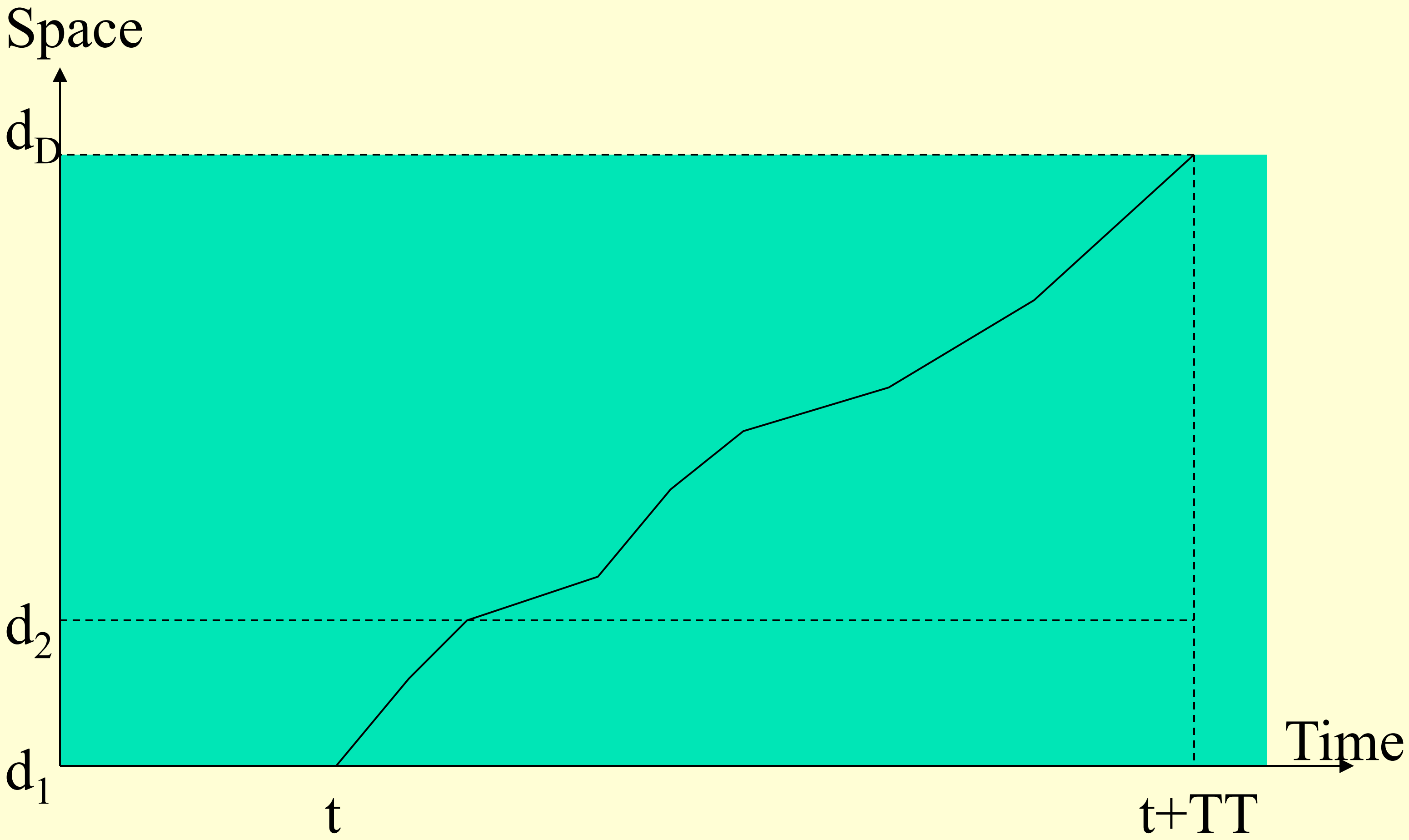
Research Objectives

- Develop a travel time estimation module
  - Reliable estimates of completed trips
  - Under all types of recurrent traffic patterns
  - With sparsely distributed traffic detectors
- Construct a travel time prediction module
  - For freeway segments
  - Large detector spacing
  - Historical travel times and traffic patterns
- Integrate a missing data estimation module
  - To deal with various missing data and delay patterns
  - Estimate the impact of the missing data

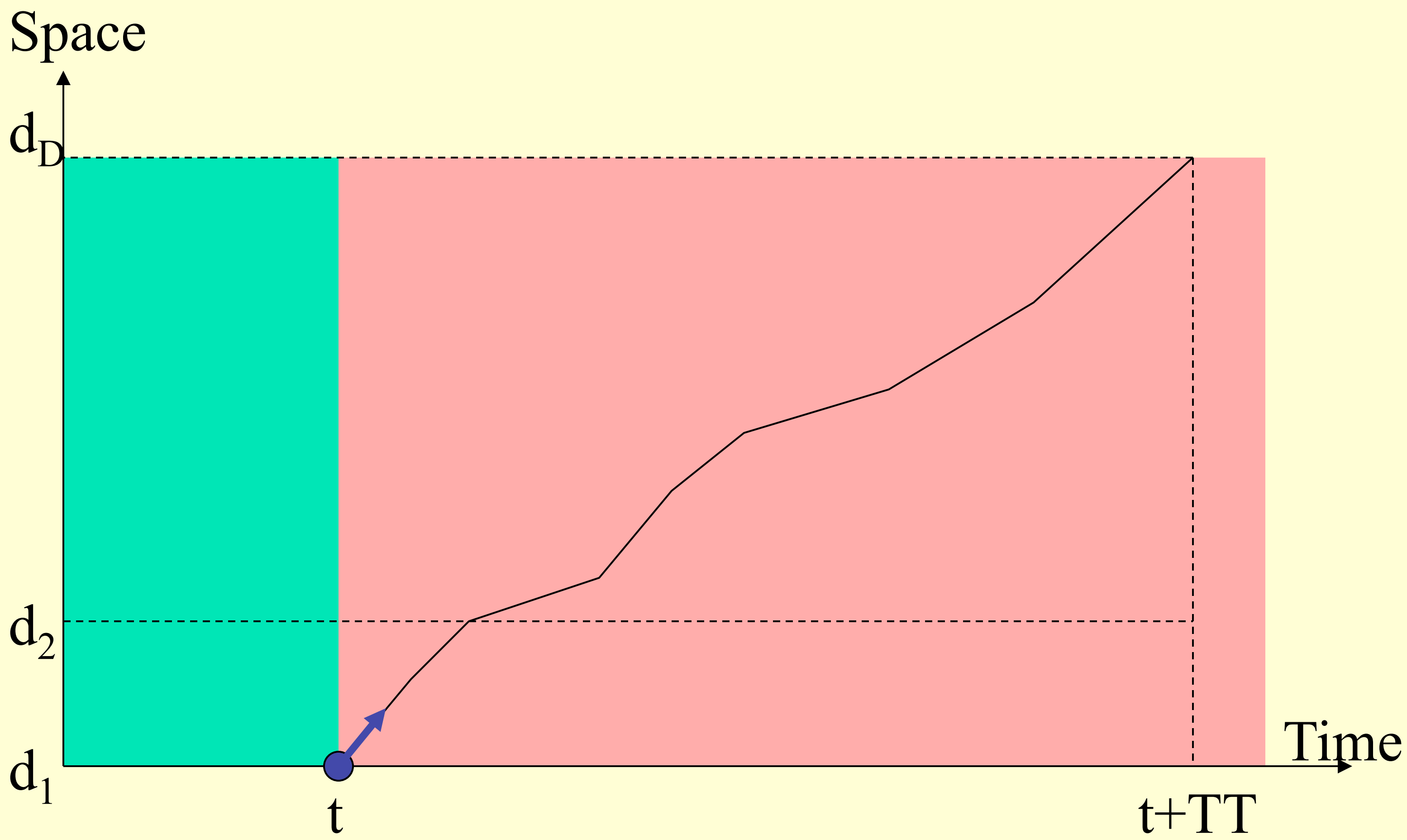
T.T. Estimation vs. Prediction



T.T. Estimation



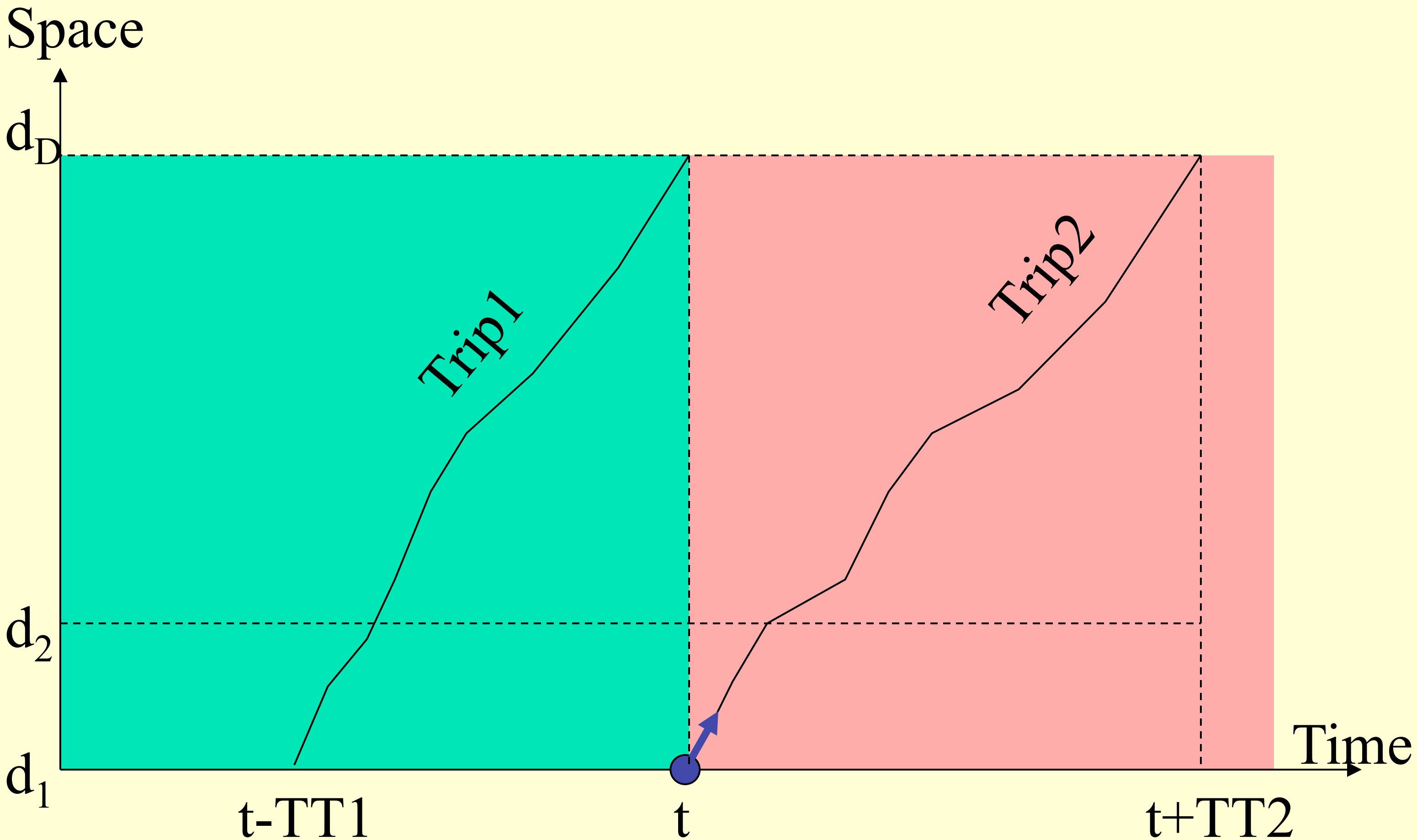
T.T. Prediction



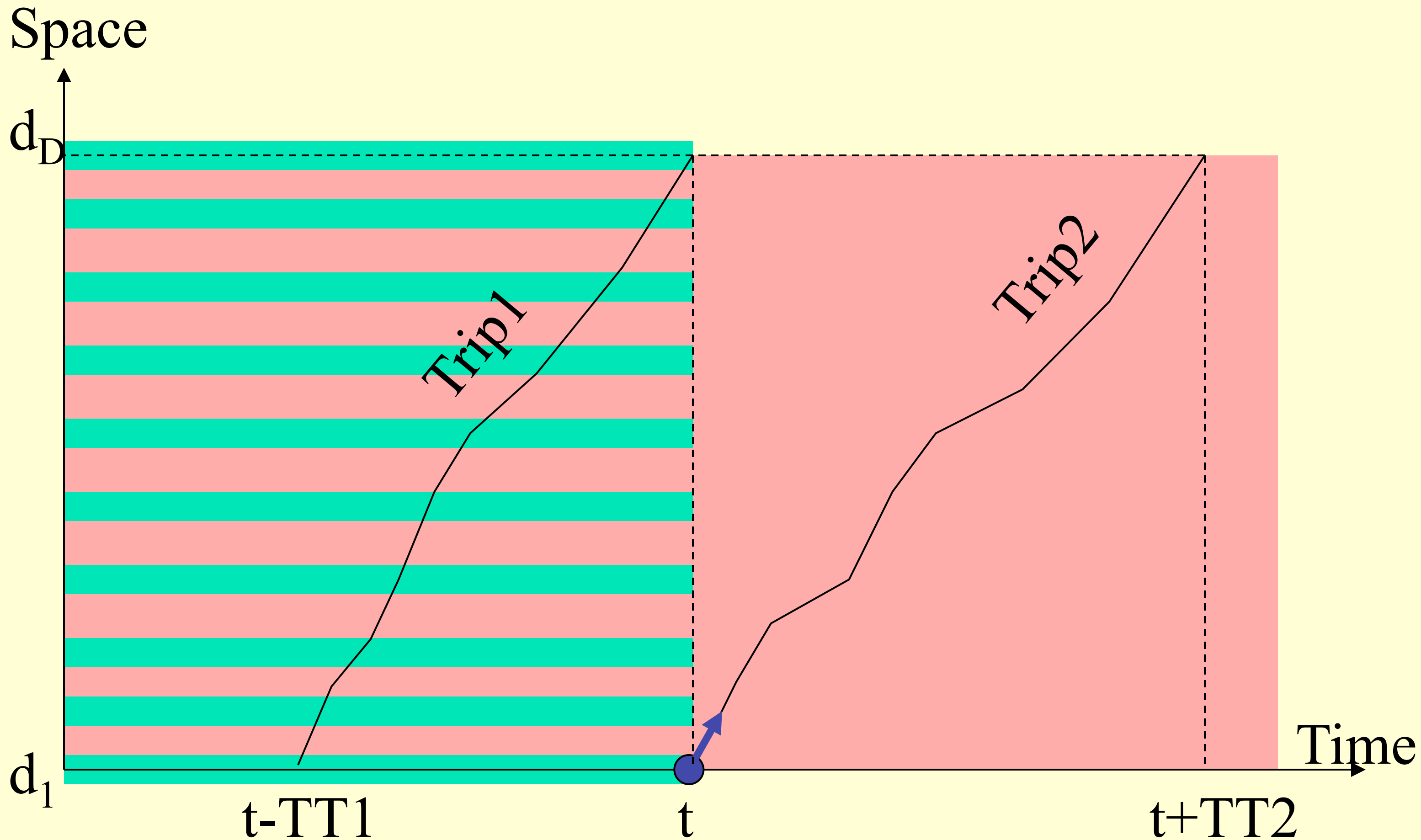
Existing Travel Time Prediction Systems

- Example systems
  - Houston, TX; Atlanta, GA; Chicago, IL; and Seattle, WA, etc.
- Almost all real-world systems use current detected traffic conditions as the prediction of the future
  - Completed trips instead of en-route trips
  - Big difference

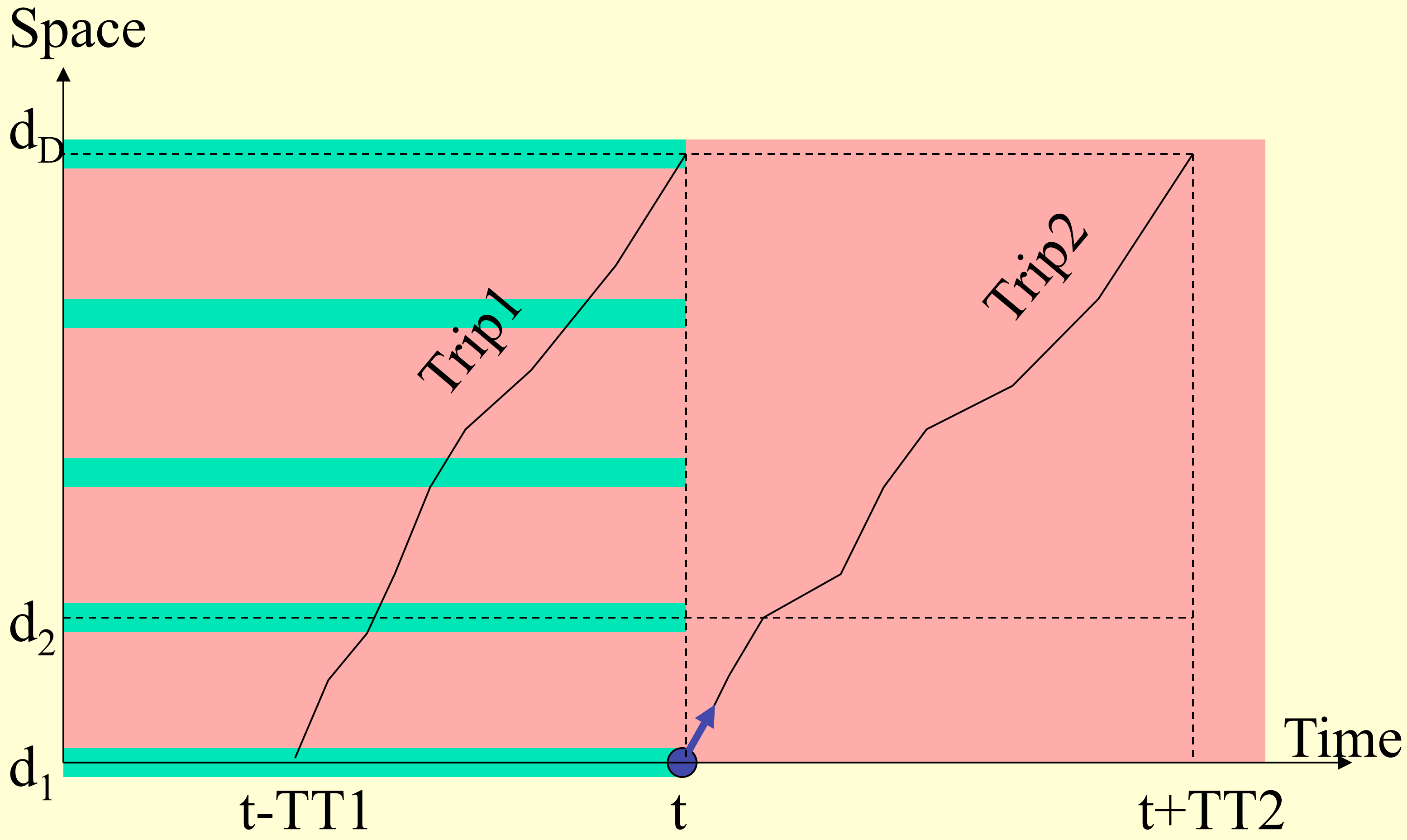
Completed Trips vs. En-route Trips



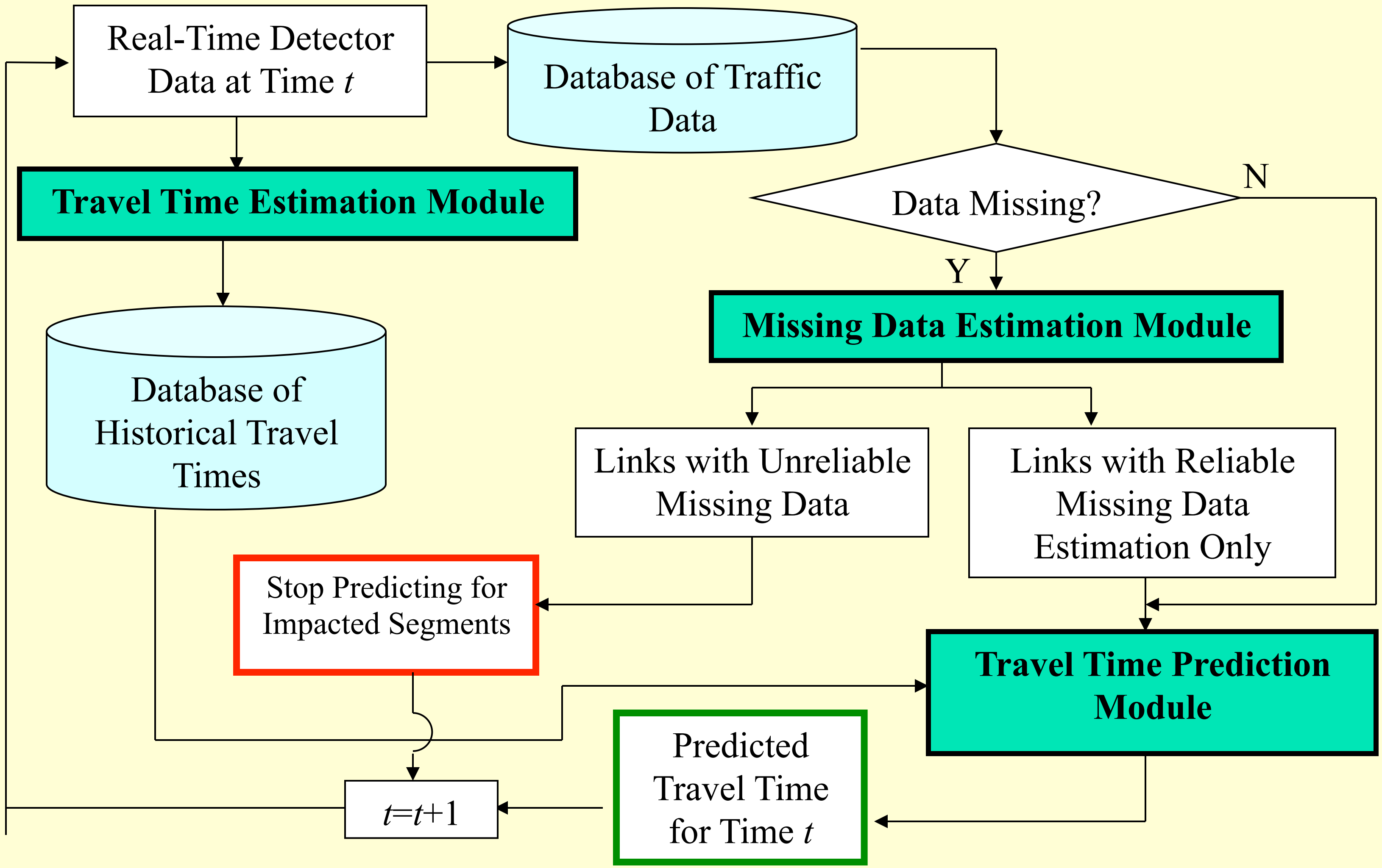
Completed Trips vs. En-route Trips



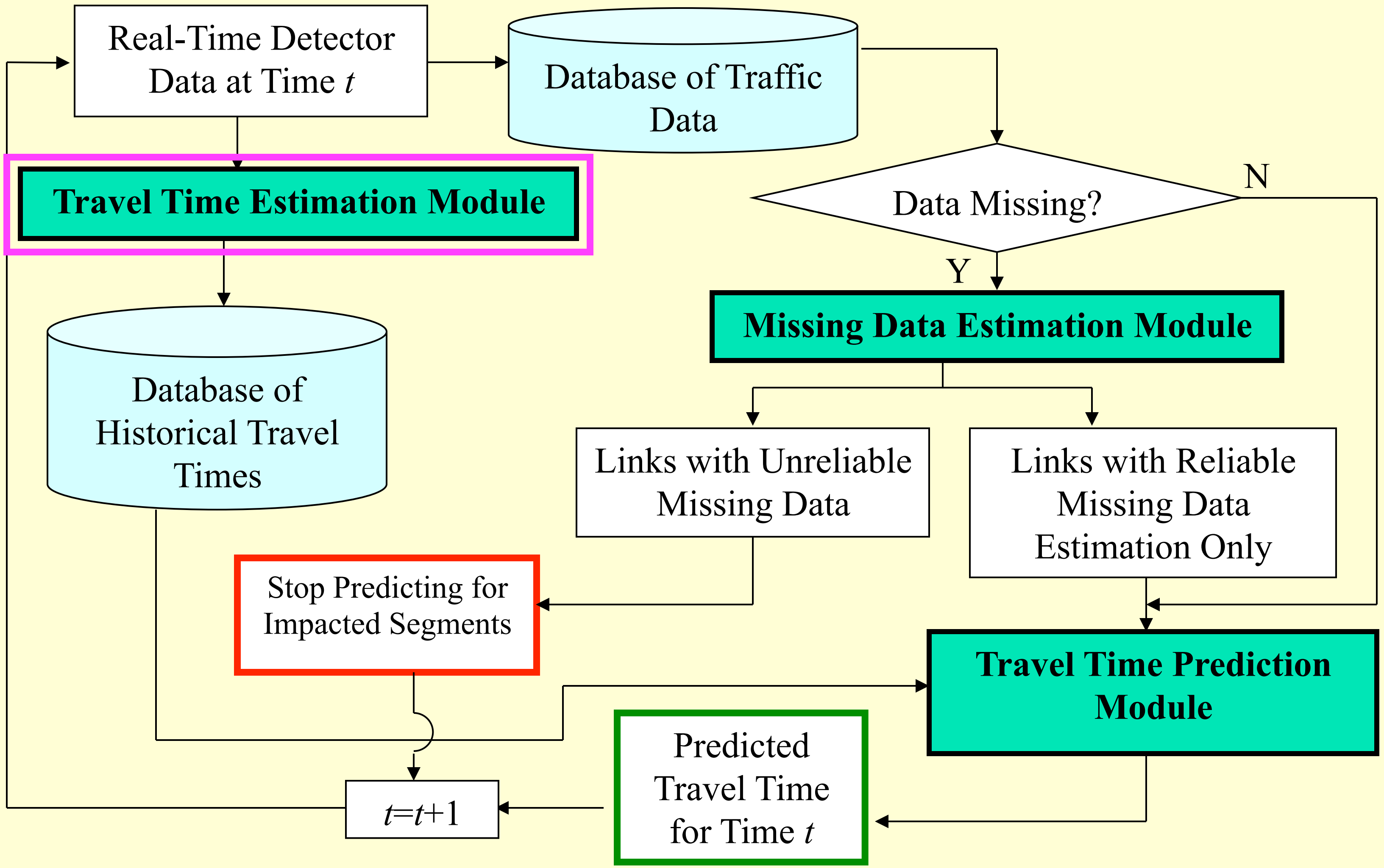
Completed Trips vs. En-route Trips



System Flowchart



System Flowchart



Literature Review

- Travel Time Estimation
  - Flow-based models
  - Vehicle identification approaches
  - Trajectory-based models

Limitations of Flow-based Models

- Reliability of detector data
  - Detection errors (volume drifting) vary over time and space
- Traffic patterns
  - Require uniformly distributed traffic across all lanes
- Geometric features
  - Cannot model ramp impact

Limitations of Vehicle Identification Approach

- Traffic patterns
  - Lane-based approach, therefore requires low lane changing rate
  - Requires uniform traffic conditions across lanes
- Geometric features
  - May not fit geometric changes, such as lane drop and lane addition
- System cost
  - High. Require new hardware or high bandwidth
- Reliability
  - Low detection resolution under high speed
  - Reduced accuracy under low light (video-based)

Limitations of Existing Trajectory-based Models

- Requires reliable speed measurement
  - Not available from most traffic detectors
- Assumes constant traffic-propagation speed
- May not perform well on long links
  - currently all studies are based on detectors less than 0.5-mile apart

A Hybrid Travel Time Estimation Model with Sparsely Distributed Detectors

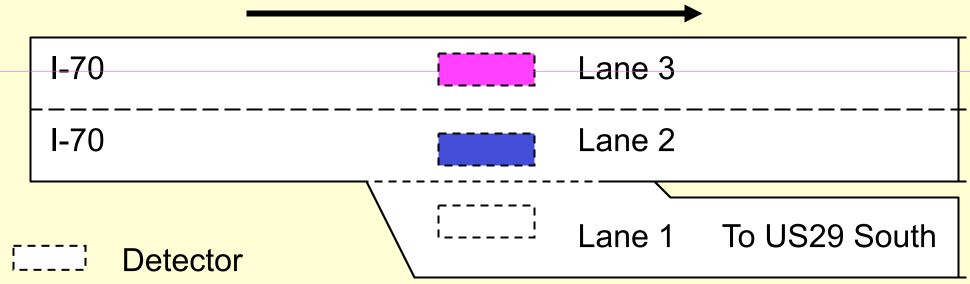
- A Clustered Linear Regression Model as the main model
  - For traffic scenarios that have sufficient field observations
- An Enhanced Trajectory-based Model as the supplemental model
  - For other scenarios

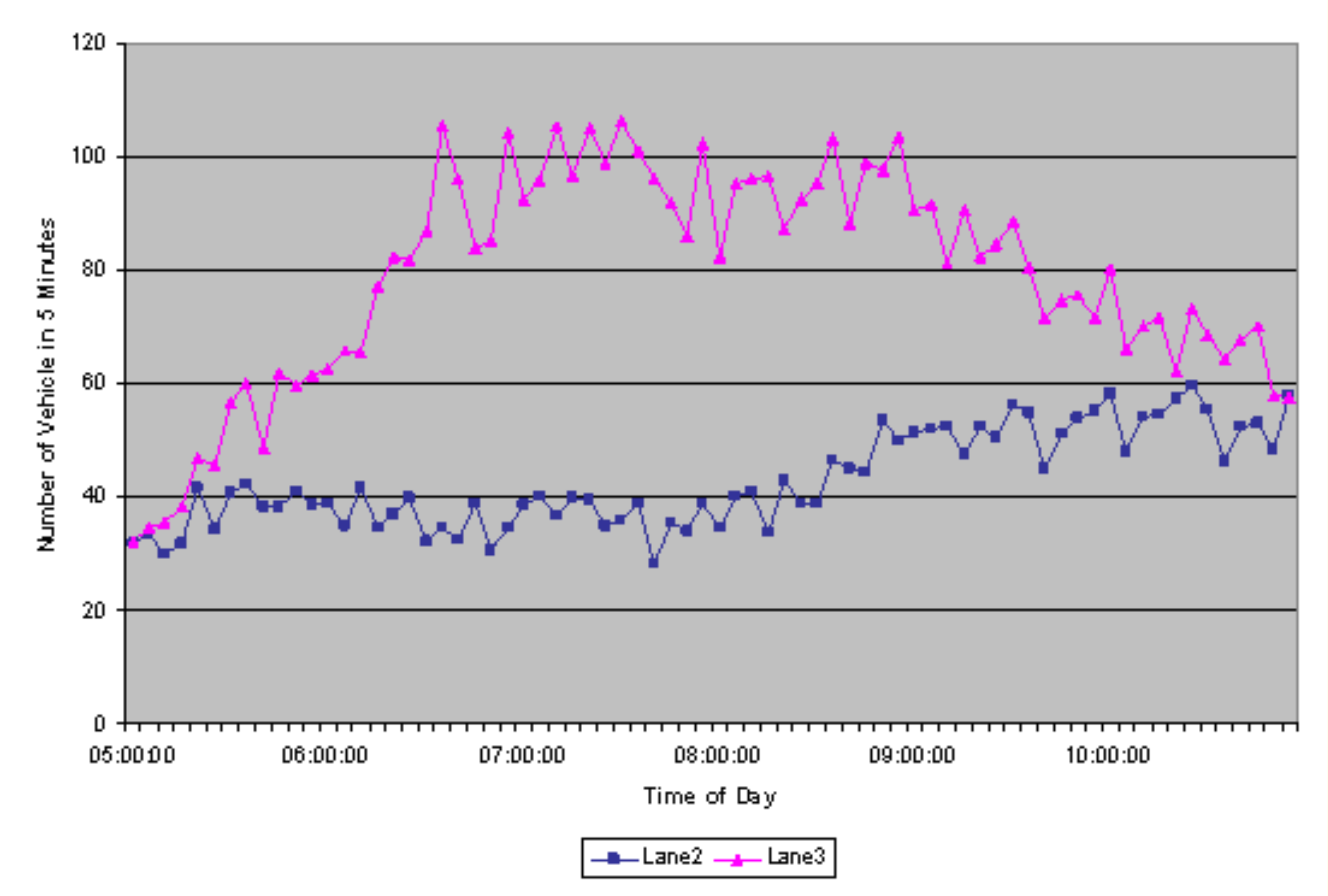
Clustered Linear Regression Model

- Travel times may be constrained in a range under one identified traffic scenario
  - For example, the travel time cannot be free-flow travel time when congestion is being observed at one detector
- Assume a linear relation between the travel time under one traffic scenario with traffic variables from pre-determined critical lanes

Critical Lanes

- Those lanes that directly contribute to estimate the average travel speed of through traffic
- May includes both mainline lanes and ramp lanes
- From both upstream and downstream detector locations


I-70
Lane 3
I-70
Lane 2
Lane 1
To US29 South
Detector


Number of Vehicle in 5 Minutes
120
100
80
60
40
20
0
05:00:00
06:00:00
07:00:00
08:00:00
09:00:00
10:00:00
Time of Day
Lane2
Lane3

Model Formulation of the Clustered Linear Regression Model

$$\tau_d(t) = \sum_{la \in \mathbf{CLT}^{\mathbf{d}}_{d,d+1}(p)} b^{T,p}_{d,la} \frac{o_{d,la}(t, \gamma^d_p \boldsymbol{\tau}^E_d(p))}{v_{d,la}(t, \gamma^d_p \boldsymbol{\tau}^E_d(p))} + \sum_{la \in \mathbf{CLR}^{\mathbf{d}}_{d,d+1}(p)} b^{R,p}_{d,la} \frac{o_{d,la}(t, \gamma^d_p \boldsymbol{\tau}^E_d(p))}{v_{d,la}(t, \gamma^d_p \boldsymbol{\tau}^E_d(p))}$$

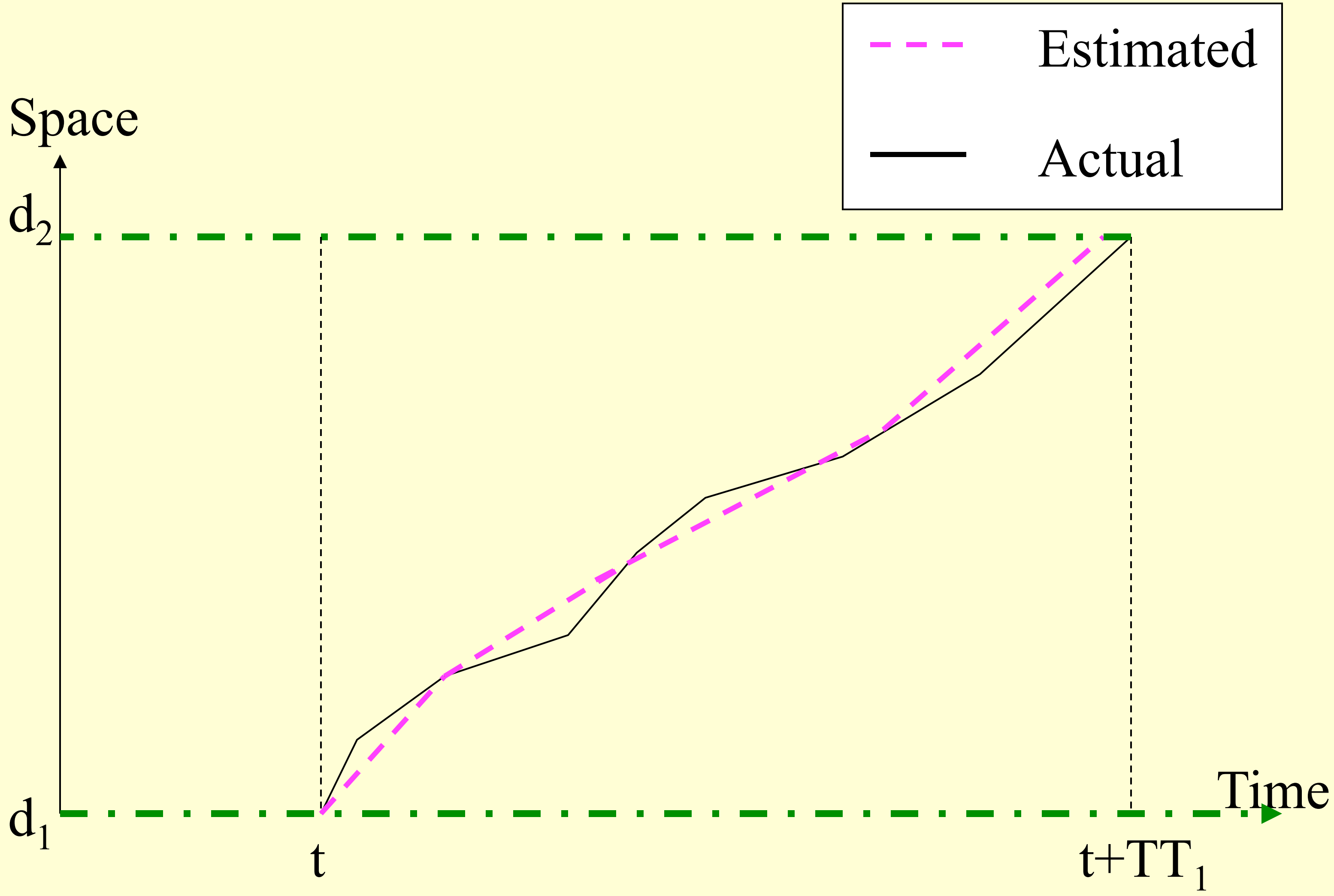
$$+ \sum_{la \in \mathbf{CLT}^{\mathbf{d+1}}_{d,d+1}(p)} b^{T,p}_{d+1,la} \frac{o_{d,la}(t + \gamma^d_p \boldsymbol{\tau}^E_d(p), (1-\gamma^d_p)\boldsymbol{\tau}^E_d(p))}{v_{d,la}(t + \gamma^d_p \boldsymbol{\tau}^E_d(p), (1-\gamma^d_p)\boldsymbol{\tau}^E_d(p))}$$

$$+ \sum_{la \in \mathbf{CLR}^{\mathbf{d+1}}_{d,d+1}(p)} b^{R,p}_{d+1,la} \frac{o_{d,la}(t + \gamma^d_p \boldsymbol{\tau}^E_d(p), (1-\gamma^d_p)\boldsymbol{\tau}^E_d(p))}{v_{d,la}(t + \gamma^d_p \boldsymbol{\tau}^E_d(p), (1-\gamma^d_p)\boldsymbol{\tau}^E_d(p))} + b^{0,p}_d$$

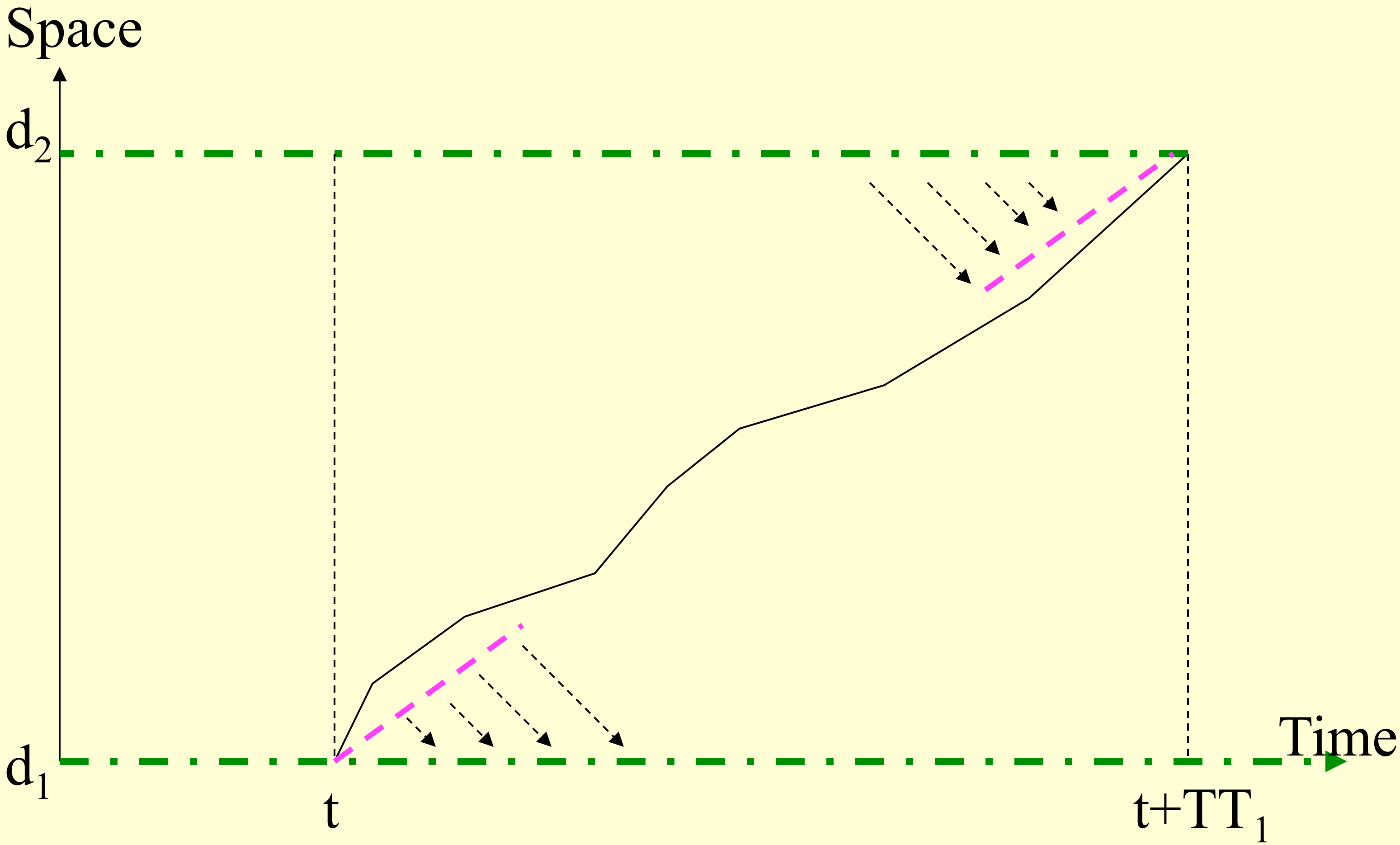
An Enhanced Trajectory-based Model

- Combines two types of trajectory estimation:
  - Traffic propagation relations when the vehicle is close to one detector
  - An enhanced piecewise linear-speed-based model when vehicle is far from both detector
- Does not require speed in input variables
  - Estimate the occupancy first, then use occupancy-flow-speed relation to estimate the vehicle’s speed

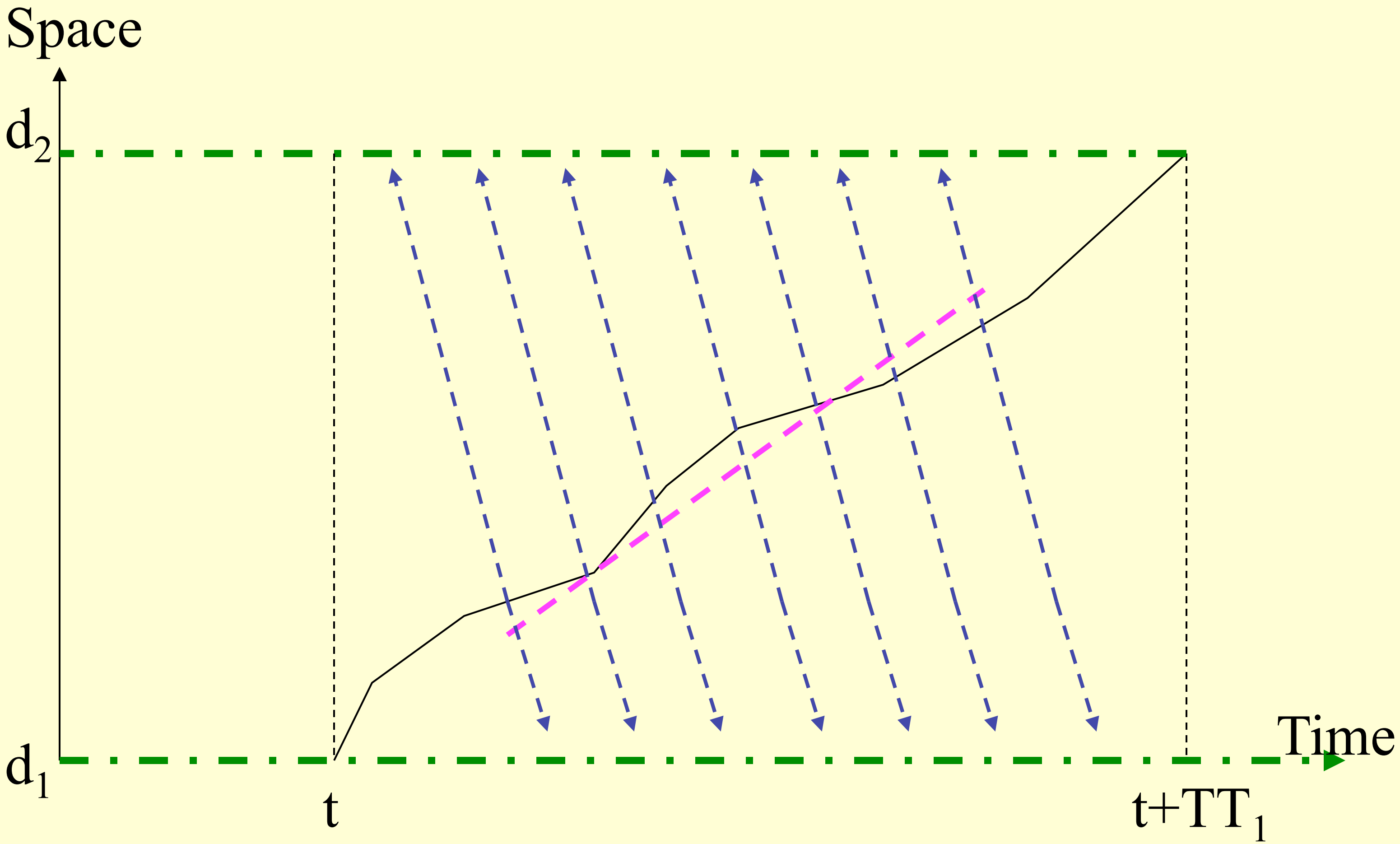
Trajectory-based Method



An Enhanced Trajectory-based Method



An Enhanced Trajectory-based Method



Model Formulation

$$
O(x,t) = \begin{cases}
o_d\left(t + \dfrac{x - x_d}{u_c^{\max}}, t + \dfrac{x - x_d}{u_c^{\min}}\right) & , \text{if } x - x_d < \hat{x} \\
o_{d+1}\left(t - \dfrac{x_{d+1} - x}{u_c^{\min}}, t - \dfrac{x_{d+1} - x}{u_c^{\max}}\right) & , \text{if } x_{d+1} - x < \hat{x} \\
o_d\left(t + \dfrac{\hat{x} - x_d}{u_c^{\max}}, t + \dfrac{\hat{x} - x_d}{u_c^{\min}}\right) & \\
+ \dfrac{(x - x_d - \hat{x})}{\hat{x}} & , \text{otherwise} \\
\times \left(o_{d+1}\left(t - \dfrac{x - (x_{d+1} - \hat{x})}{u_c^{\min}}, t - \dfrac{x - (x_{d+1} - \hat{x})}{u_c^{\max}}\right)\right. & \\
\left. - o_d\left(t + \dfrac{\hat{x}}{u_c^{\max}}, t + \dfrac{\hat{x}}{u_c^{\min}}\right)\right) &
\end{cases}
$$

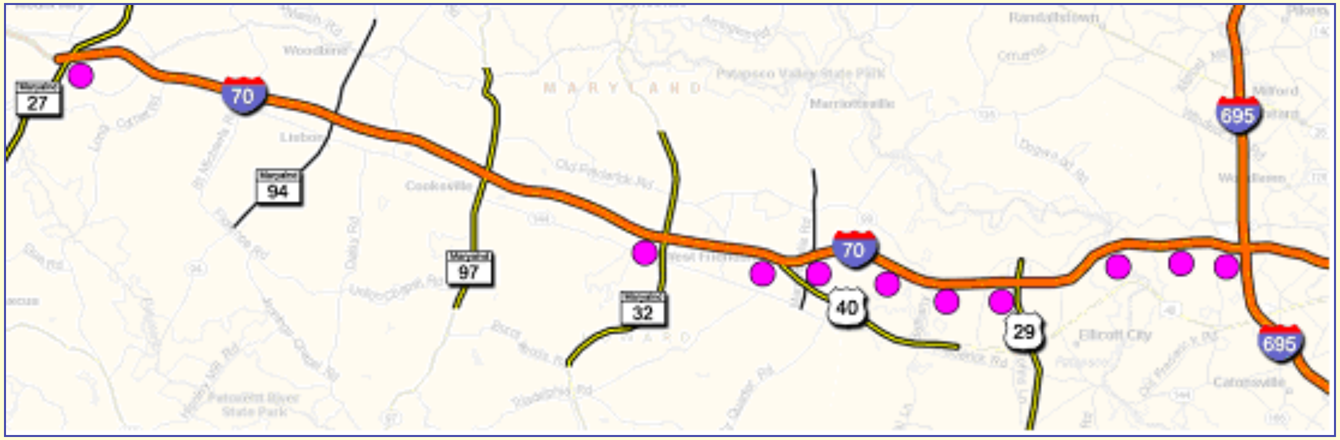
$$
\hat{x} = \begin{cases}
\min\left(\dfrac{l_d}{3}, \dfrac{1}{3}\text{mi}\right) & , \text{when } l_d \geq 1 \text{ mile} \\
\dfrac{l_d}{3} & , \text{otherwise}
\end{cases}
\qquad x_d \leq x \leq x_{d+1}
$$

$u_c^{\min}$ and $u_c^{\max}$ are the minimum and the maximum traffic propagation speeds.

Model Formation (cont'd)

$$u(x,t) = \begin{cases} u_{free} & , o(x,t) <= o_{free} \\ u_{cong} + (u_{free} - u_{cong})(1 - \dfrac{o(x,t) - o_{free}}{o_{cong} - o_{free}})^{m} & , o_{free} < o(x,t) <= o_{cong} \\ u_{min} + (u_{cong} - u_{min})(1 - \dfrac{o(x,t) - o_{cong}}{o_{max} - o_{cong}})^{n} & , o_{cong} < o(x,t) <= o_{\max} \\ u_{min} & , \text{otherwise} \end{cases}$$

Numeric Examples



- I-70 eastbound from MD27 to I-695
- 10 detectors on a 25-mile stretch
- Flow count and occupancy data
- 30-second intervals

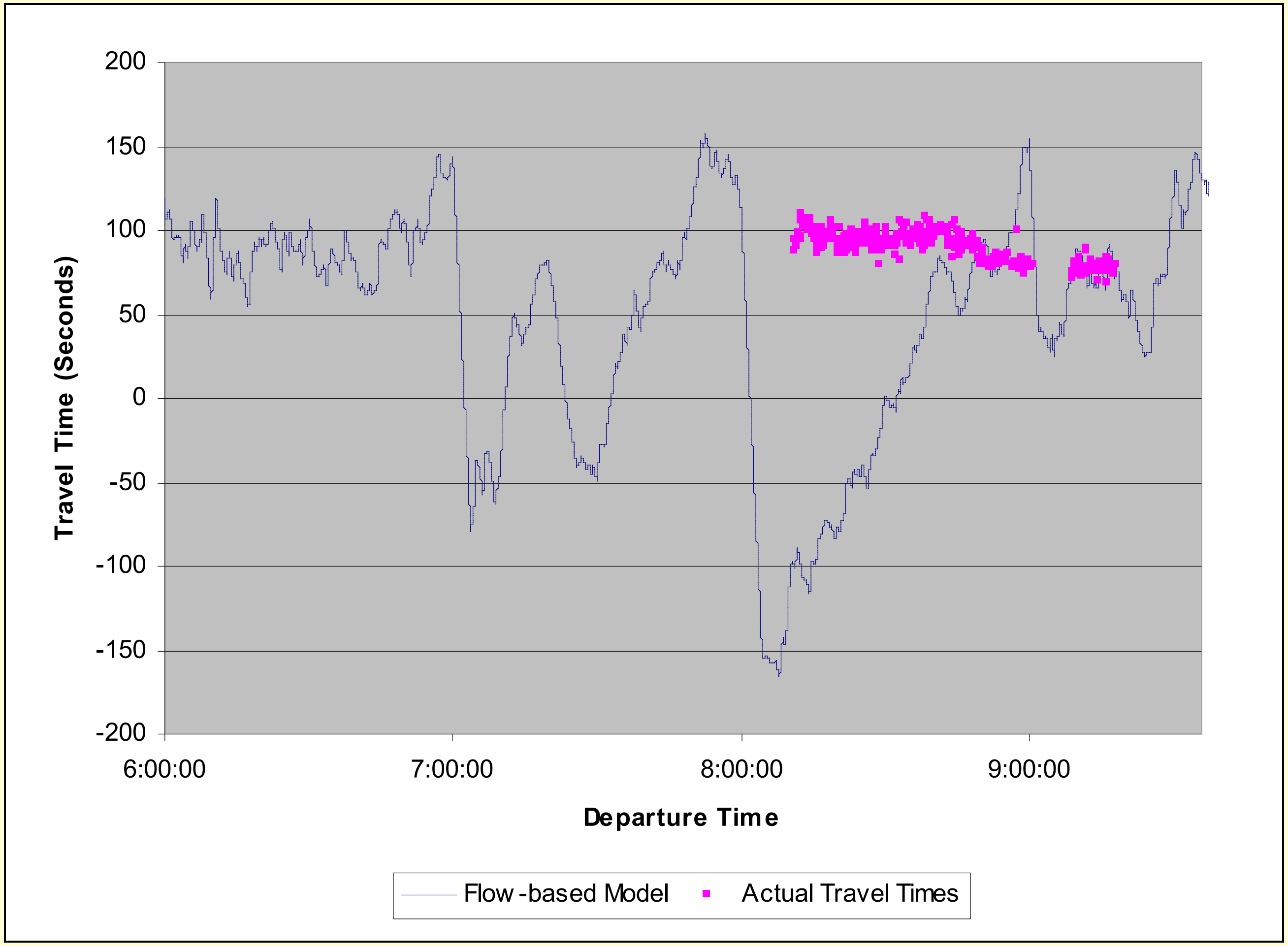
Methods for Comparison

- Proposed hybrid model
  - Clustered Linear Regression (CLR) model
  - Enhanced Trajectory-based (ETB) model
- Flow-based method (Nam and Drew, 1996)
- Piecewise Linear Speed-based (PLSB) method (Van Lint and van der Zijpp, 2003)

Volume Drifting Issue

	2006-06-27	2006-06-28	2006-06-29	2006-06-30	2006-07-01	2006-07-02
Daily Volume at Detector 4	37040	39121	41595	42707	35190	29891
Daily Volume at Detector 6	37903	39695	42373	43410	35117	29741
Difference	863	574	778	703	-73	-150
Relative Difference	**2.33%**	**1.47%**	**1.87%**	**1.65%**	**-0.21%**	**-0.50%**
Daily Volume at Detector 8	45332	49022	50160	50670	39469	34806
Daily Volume at Detector 9	44979	48945	49796	50449	39314	34784
Difference	-353	-77	-364	-221	-155	-22
Relative Difference	**-0.78%**	**-0.16%**	**-0.73%**	**-0.44%**	**-0.39%**	**-0.06%**

Flow-based Model



(481 actual travel time samples, January 19th, 2007)

Travel Time Surveys

Date and Time		Link								
		1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
12/1/2005	AM	Y	Y	Y	Y	Y	Y	Y	Y	Y
1/19/2006	AM					Y				
1/20/2006	AM						Y			
1/20/2006	PM									Y
2/1/2006	AM			Y						
2/2/2006	AM			Y						
2/7/2006	PM								Y	
2/28/2006	AM			Y	Y	Y	Y			
3/1/2006	PM							Y	Y	Y
3/7/2006	AM							Y	Y	Y
3/9/2006	PM							Y	Y	Y
4/6/2006	AM			Y						
4/20/2006	AM			Y						
6/13/2006	AM	Y		Y			Y	Y	Y	Y
6/15/2006	PM	Y		Y			Y	Y	Y	Y

Performance on Individual Links - Link (5,6)

- 446 samples on January 19th, 2006
- 411 samples on February 28th, 2006
- 4 identified scenarios

ID	Description of the Scenario	Detector 5		Detector 6	
		Occ. in Ln. 1	Occ. in Ln. 2	Occ. in Ln. 1	Occ. in Ln. 2
1	No congestion on the link	≤12	≤10	≤10	≤10
2	Congestion at Detector 5; no congestion at Detector 6	>12	>10	≤10	≤10
3	Congestion at both Detectors 5 and 6	>12	>10	>10	>10
4	Other	Other combinations			

Performance on Individual Links - Link (5,6) (cont'd)

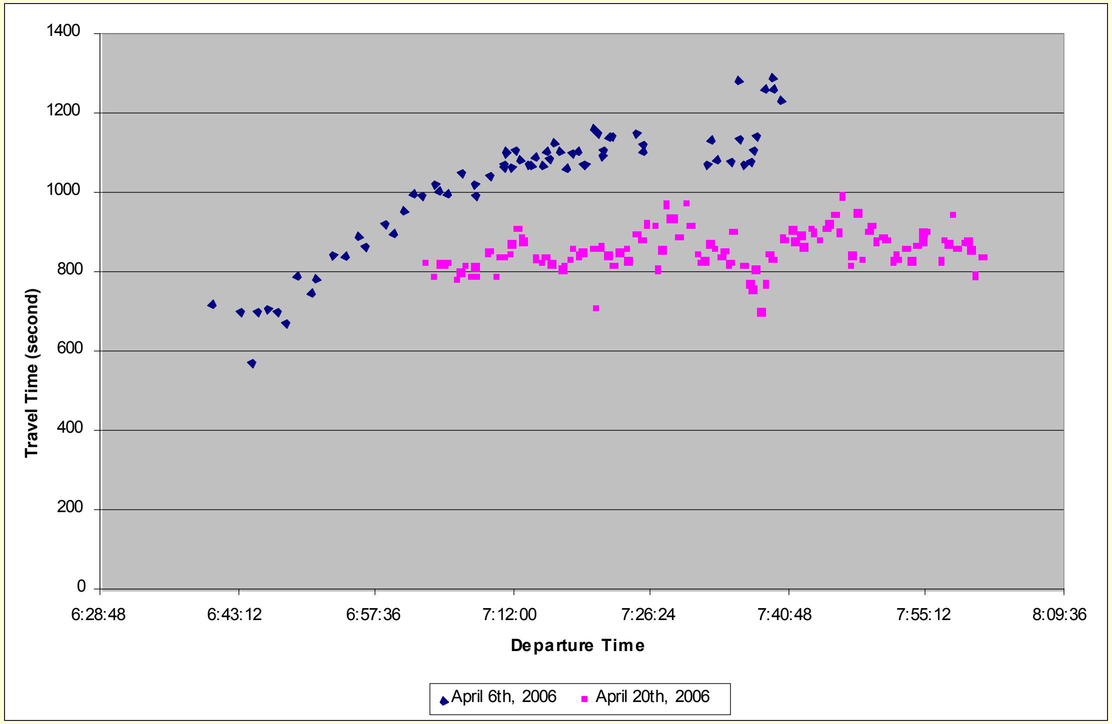
Scenario 2	All Samples (35 Observations)		Travel Times ≤ 95 sec. (20 Observations)		Travel Times > 95 sec. (15 Observations)	
	AAE (Sec.)	AARE (%)	AAE (Sec.)	AARE (%)	AAE (Sec.)	AARE (%)
CLR	5.63	6.57	6.16	8.31	4.46	4.26
ETB	**5.14**	**5.43**	3.71	4.19	7.20	7.08
PLSB	6.17	6.49	5.47	5.86	7.67	7.33

Scn. 3	All Samples (33 Observation)	
	AAE (Sec.)	AARE (%)
CLR	**6.60**	**4.79**
ETB	19.48	13.35
PLSB	26.33	17.65

	Scenario 1 (60 Observations)		**Scenario 4** (151 Observations)	
	AAE (Sec.)	AARE (%)	AAE (Sec.)	ARE<10% (%)
ETB	**2.67**	**3.33**	**7.22**	**6.54**
PLSB	2.92	3.67	8.69	7.66

Performance on Multiple Links

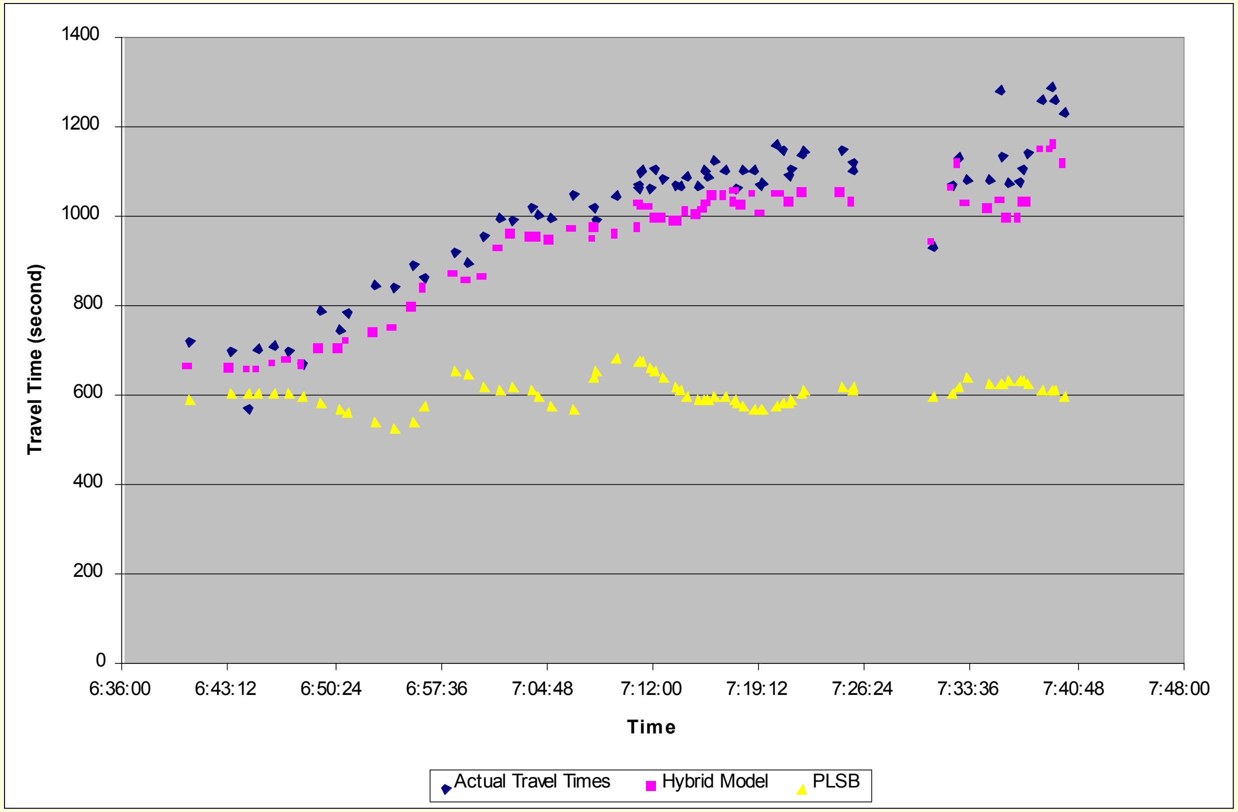
- Subsegment (3, 10)
- About 10 miles
- 71 samples on April 6th, 2006
- 114 samples on April 20th, 2006



Performance on Multiple Links (cont' d)

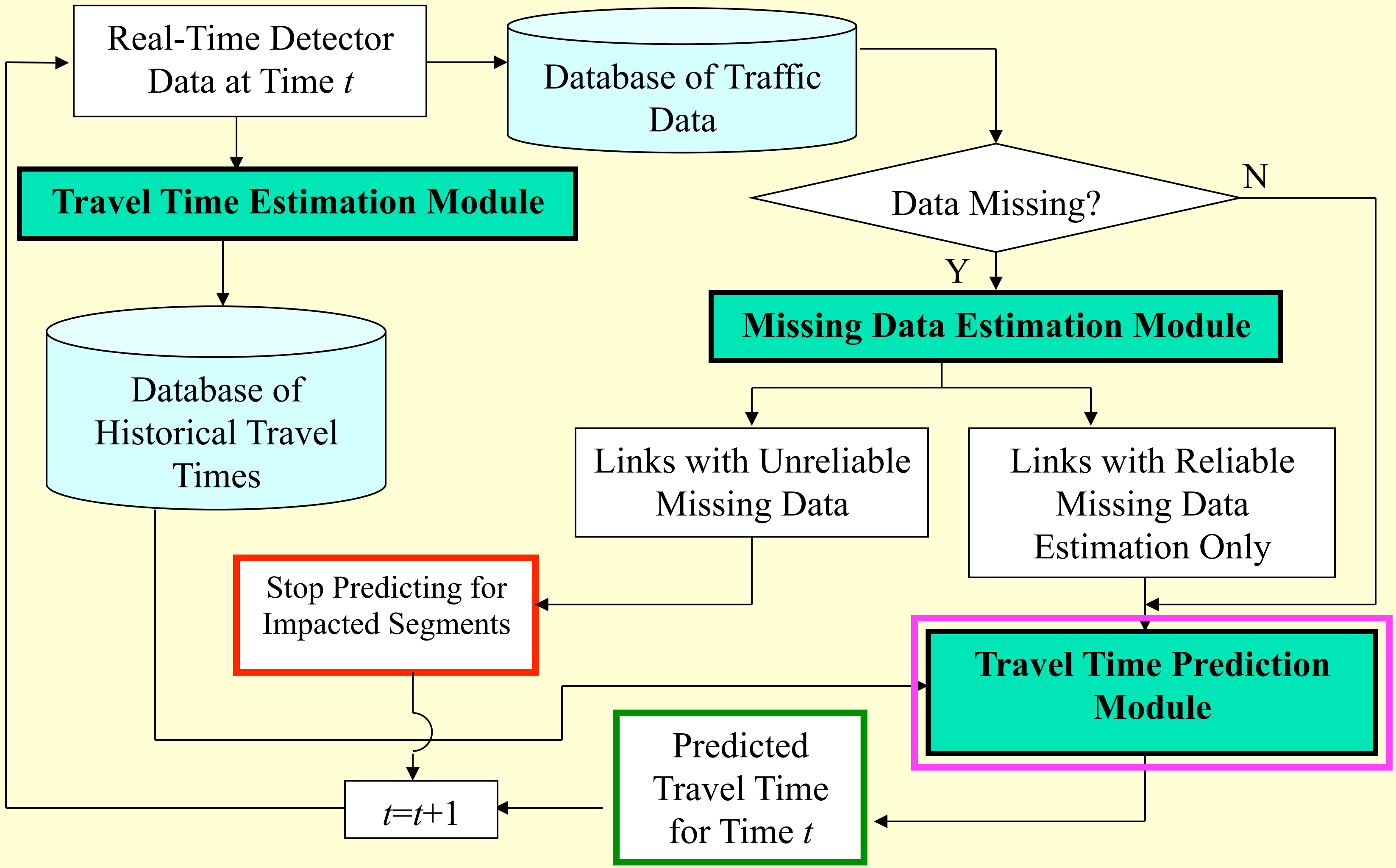
		Travel Time Range (sec)		
		520 to 800	800 to 1000	> 1000
Sample Size		23	113	49
Maximum Travel Time (sec)		796	998	1290
Average Travel Time (sec)		742.3	847.2	1109.1
Hybrid Model	AAE (sec)	59.4	54.1	83.6
	AARE (%)	8.1%	6.2%	7.5%
PLSB	AAE (sec)	139.8	266.6	493.8
	AARE (%)	18.4%	30.5%	44.3%

Performance on Multiple Links (cont' d)



April 6th, 2006

System Flowchart



Travel Time Prediction

- Parametric Models
  - Time series model
  - Linear regression model
  - Kalman Filter model
- Nonparametric models
  - Neural Network model
  - Nearest Neighbor model
  - Kernel model and local regression model

Autoregressive Integrated Moving Average (ARIMA)

- Advantages:
  - Ability to predict a time series data set
  - Good for predicting traffic data (volume, speed, or occupancy) at one detector
- Disadvantages:
  - Focus on the mean value, therefore cannot well predict scenarios that less frequently occur
  - It is hard to model multiple sets of time series data together (for example, multiple series of data from detectors)

Linear Regression Models

- One single linear regression model cannot predict well for all traffic scenarios, therefore multi-model structure is often used:
  - Layered/clustered linear regression model
  - Varying coefficient linear regression model

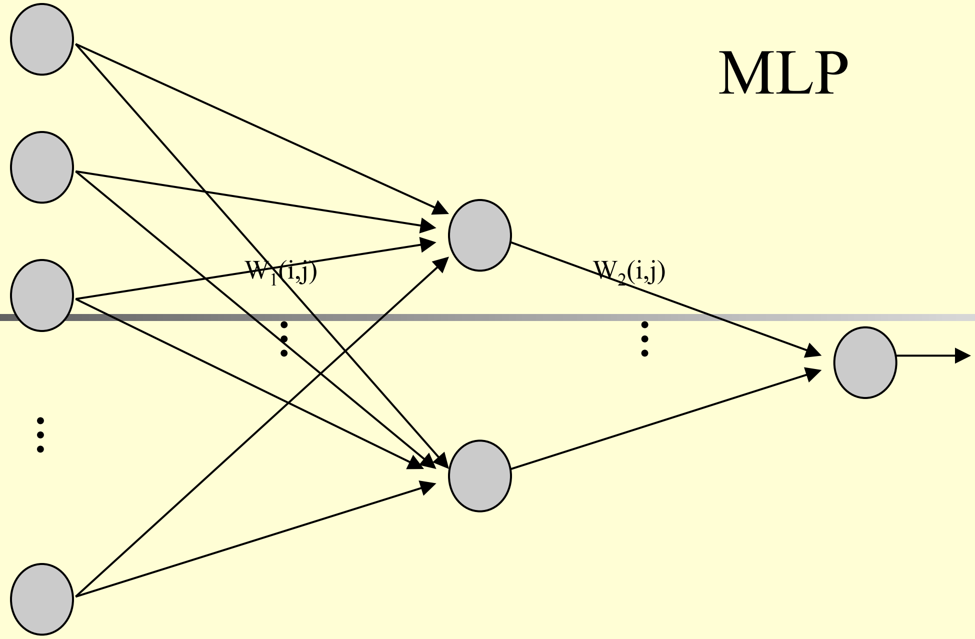
Kalman Filter Model

- Ability to auto-update parameters based on the evaluation of the prediction accuracy of the previous time interval
- Good performance when the true value can be obtained with a short delay (Chien et al., 2002 and 2003)
- May not work well for a prediction system with long travel times (long travel times = long delay for the update process)

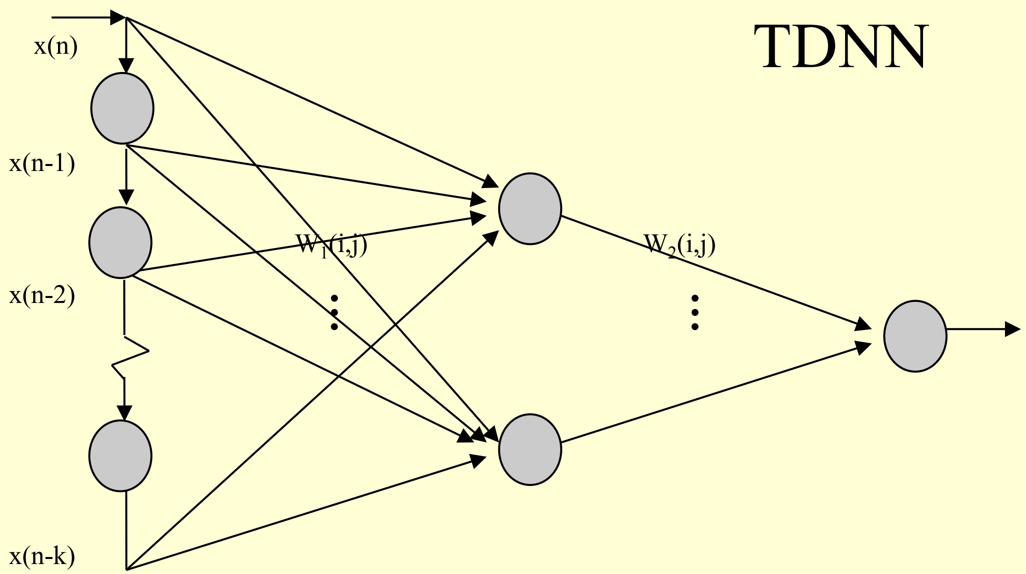
Neural Network Models

- Widely used to predict travel times
- Accurate and robust because of its good ability to recognize patterns
- Multi-layer Perceptron (MLP) and Time Delay Neural Network (TDNN) are mostly seen in the literature
- A large amount of training data

MLP



TDNN



k-Nearest Neighbor Model

- Looks for *k* most similar cases as the current condition from the historical database to come out a prediction
- Requires a fairly large historical database

$$dist_{EUC}(p,q) = \sqrt{\sum_{i=1}^{K}(p_i - q_i)^2}$$

$$dist_{NUW}(p,q) = \sqrt{\sum_{i=1}^{K}w_i(p_i - q_i)^2}$$

Other Nonparametric Models

- Share a common structure
  - A clustering function
  - A kernel function (linear, nonlinear and/or other form) for each cluster
- For example
  - Kernel regression
  - Layered linear regression
  - Time-varying coefficient linear regression

A Hybrid Travel Time Prediction Model

- A Multi-topology Neural Network model
  - A rule-based clustering function
  - Customized topologies for various traffic scenarios
- An Enhanced *k*-Nearest Neighbor Model
  - For cases with sufficient good matches in the historical data

A Multi-topology Neural Network Model

- Categorize congestion patterns, instead of time-of-day, with a rule-based clustering function
- Select only data in critical lanes as input variables
  - Geometric features
  - Traffic patterns
- Various topology to fit different traffic patterns

IF $t \geq TML_d^{wk}$ and $t \leq TMU_d^{wk}$ THEN

IF $\exists la, o_{d^*,la}(t-j) > OM_{d,la}$ for all j, where $la \in \mathbf{CLM}_{d,d+1}^{d^*}$ and $0 \leq j \leq THNN$, THEN

$p_d(t) = 1$ (morning congestion)

ELSE

IF $o_{d^*,la}(t-j) \leq OM_{d,la}$ for all la and j, where $la \in \mathbf{CLM}_{d,d+1}^{d^*}$ and $0 \leq j \leq THNN$, THEN

$p_d(t) = 0$ (off-peak period)

ELSE

$p_d(t) = p_d(t-1)$

END IF

END IF

ELSE

IF $t \geq TEL_d^{wk}$ and $t \leq TEU_d^{wk}$ THEN

IF $\exists la, o_{d^*,la}(t-j) > OE_{d,la}$ for all j, where $la \in \mathbf{CLE}_{d,d+1}^{d^*}$ and $0 \leq j \leq THNN$, THEN

$p_d(t) = -1$(evening congestion)

ELSE

IF $o_{d^*,la}(t-j) \leq OE_{d,la}$ for all la and j, where

$0 \leq j \leq THNN$, THEN

$p_d(t) = 0$ (off-peak period)

ELSE

$p_d(t) = p_d(t-1)$

END IF

END IF

ELSE

$p_d(t) = 0$ (off-peak period)

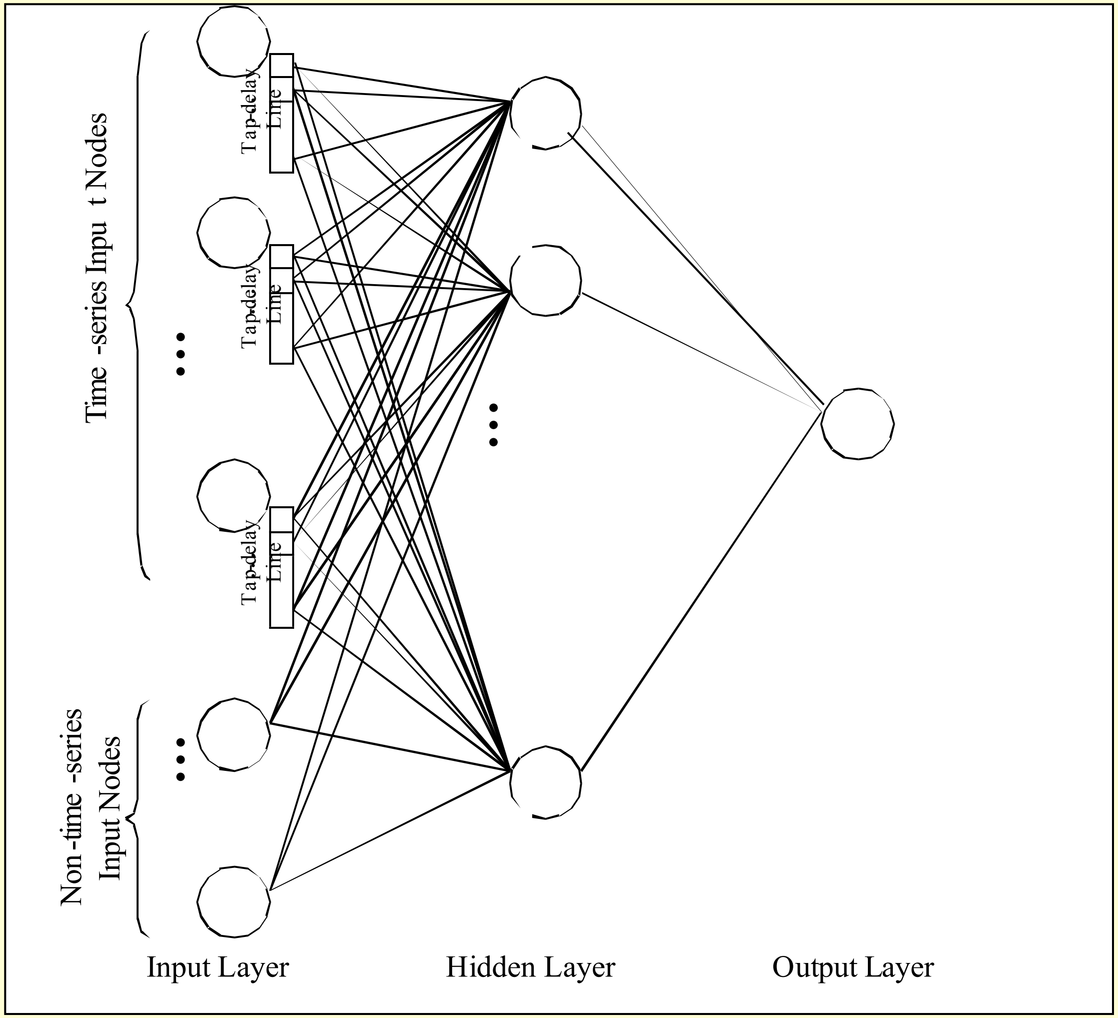
END IF

END IF

where, TML_d^{wk} and TMU_d^{wk} are the lower and upper time boundaries for morning peak hours in link (d, d+1) on weekday wk in the historical traffic patterns;
TEL_d^{wk} and TEU_d^{wk} are the lower and upper time boundaries for evening peak hours in link (d, d+1) on weekday wk in the historical traffic patterns;
$0:00 \leq TML_d^{wk} < TMU_d^{wk} \leq TEL_d^{wk} < TEU_d^{wk} < 24:00$;
$d^* = d$ or $d+1$;
$OM_{d,la}$ is the occupancy threshold at lane la at detector d in the morning;
$OE_{d,la}$ is the occupancy threshold at lane la at detector d in the evening;
$\mathbf{CLM}_{d,d+1}^{d^*}$ and $\mathbf{CLE}_{d,d+1}^{d^*}$ are sets of critical lanes at detector d^* in link (d, d+1) in the morning and in the evening respectively; and
$THNN$ is the required duration for the traffic condition to maintain congested or uncongested stably;

Enhanced Topology

- Combines time-series and non-time-series data



k-Nearest Neighbor Model for Travel Time Prediction

- An updated distance function
  - Based on three types of traffic state
- Geometric features
  - Take traffic data from critical lanes only
  - The time range of input data increases with the distance to the origin
- Daily and weekly traffic patterns
  - Varying search window based on historical traffic patterns

Modified Definition of the Distance

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (p_i^* - q_i^*)^2}$$

$$p_i^* = \begin{cases} p_i & \text{, when } TC_d^{l_a}(t, t+\Delta t) = 0 \\ OC_d^{l_a} & \text{, when } TC_d^{l_a}(t, t+\Delta t) = 1 \\ OF_d^{l_a} & \text{, when } TC_d^{l_a}(t, t+\Delta t) = -1 \end{cases}$$

$$q_i^* = \begin{cases} q_i & \text{, when } TC_d^{l_a}(t_h, t_h+\Delta t) = 0 \\ OC_d^{l_a} & \text{, when } TC_d^{l_a}(t_h, t_h+\Delta t) = 1 \\ OF_d^{l_a} & \text{, when } TC_d^{l_a}(t_h, t_h+\Delta t) = -1 \end{cases}$$

Consideration of Traffic Patterns

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (\hat{p}_i - q_i^*)^2}$$

Where

$$\hat{p}_i = \begin{cases} M & \text{, if } |t - t_h| > T_{th}(d,t) \\ p_i^* \times \hat{w} & \text{, otherwise} \end{cases}$$

$$\hat{w} = \begin{cases} 1 & \text{, if } \exists s, wk_h \in \boldsymbol{W}_s \text{ and } wk_c \in \boldsymbol{W}_s \ (1 \le s \le \mathrm{S}) \\ M & \text{, otherwise} \end{cases}$$

$$\bigcup_{s=1}^{S} \boldsymbol{W}_s = \{\text{all weekdays}\}$$

M is a very large number.
wk_c and wk_h are weekdays of the current case and the historical case respectively

Numerical Examples

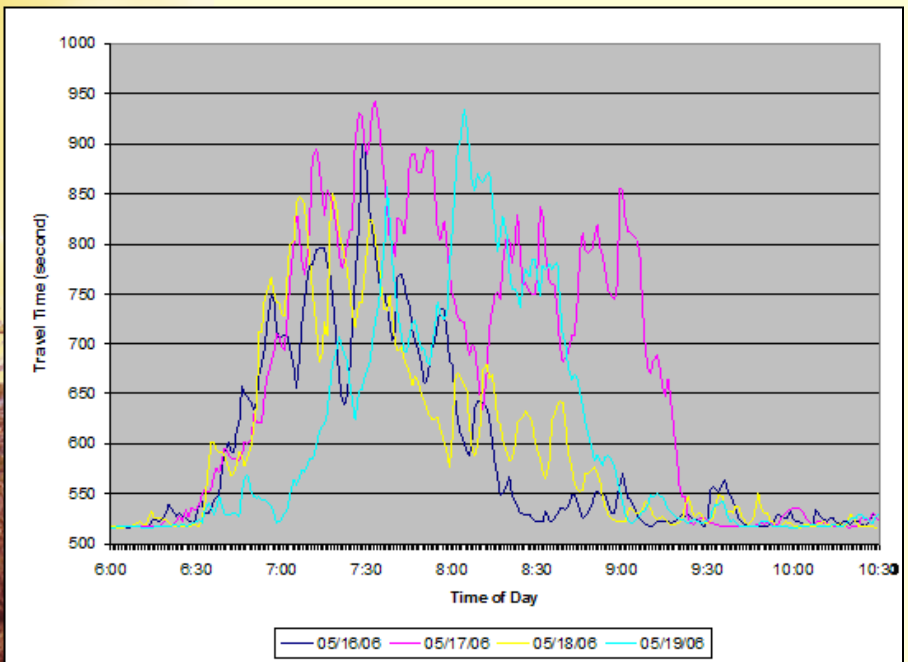
- Same dataset from I-70 eastbound
- Subsegment (3, 10)
  - About 10 mile
- Comparison 1:
  - Predicted travel times vs. estimated travel times
- Comparison 2:
  - Predicted travel times vs. actual travel times

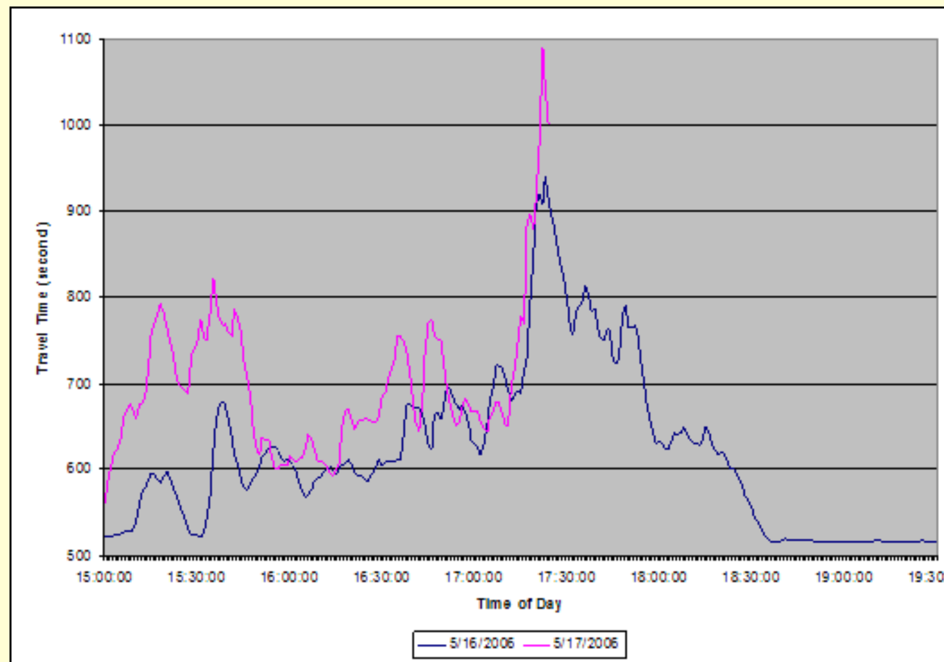
Models for Comparison

	4 Weeks of Training Data	10 Weeks of Training Data
Hybrid model developed in this study	HM4	HM10
Neural Network model in the developed hybrid model	NN4	NN10
k-Nearest Neighbors model in the developed hybrid model	kNN4	kNN10
Constant current speed-based model	CCSB	
Time-varying coefficient model	TVC4	TVC10

Predicted vs. Estimated

- 6:00 to 10:30 and 15:00 to 19:30
- AM: May 16th, 2006 to May 19th, 2006
- PM: May 16th, 2006 and May 17th, 2006





All Sample Days

Model	Average Absolute Error (second)	Average Absolute Relative Error (%)
CCSB	77.92	10.89
TVC4	173.99	28.10
TVC10	65.64	9.44
kNN4	64.38	9.04
kNN10	60.86	8.56
NN4	53.88	7.81
NN10	48.68	7.07
HM4	48.84	6.92
HM10	45.69	6.53

Each Peak Period

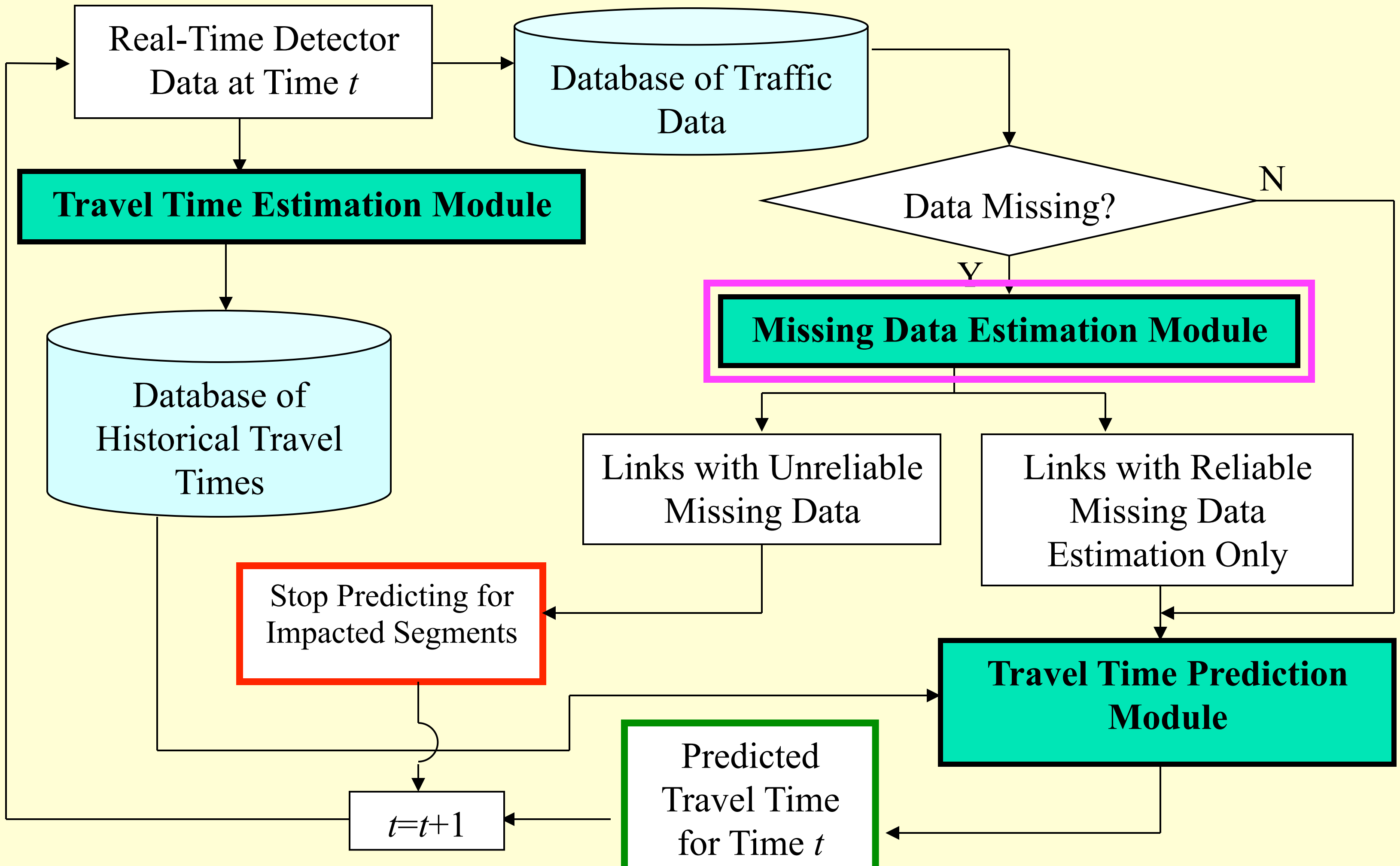
Average Absolute Error (seconds)	5/16 AM	5/16 PM	5/17 AM	5/17 PM	5/18 AM	5/19 AM
CCSB	56.37	73.93	106.62	106.84	63.95	71.97
TVC4	186.04	127.27	232.82	128.47	166.45	168.96
TVC10	39.64	105.05	83.84	121.86	41.38	36.99
kNN4	34.42	84.09	81.48	126.45	34.10	58.85
kNN10	31.71	71.08	79.46	127.66	33.68	54.25
NN4	31.81	68.18	64.77	93.47	32.80	53.39
NN10	30.70	65.38	55.64	75.96	36.83	43.92
HM4	29.44	54.00	58.37	87.17	28.95	49.82
HM10	29.09	52.10	53.75	75.96	35.26	36.69

Predicted vs. Actual

- 70 actual travel times collected by a third party company
- Same sample peak periods

	Average Travel Time (seconds)	HM4 AAE (seconds)	HM10 AAE (seconds)	Number of Samples
All samples	655.67	56.58	51.69	70
TT≤580	532.58	15.74	15.11	24
580<TT≤900	703.86	80.45	72.02	36
TT>900	949.67	113.43	95.29	10

System Flowchart



Missing Data Estimation

- Data discard
- Single imputation
- Multiple imputation

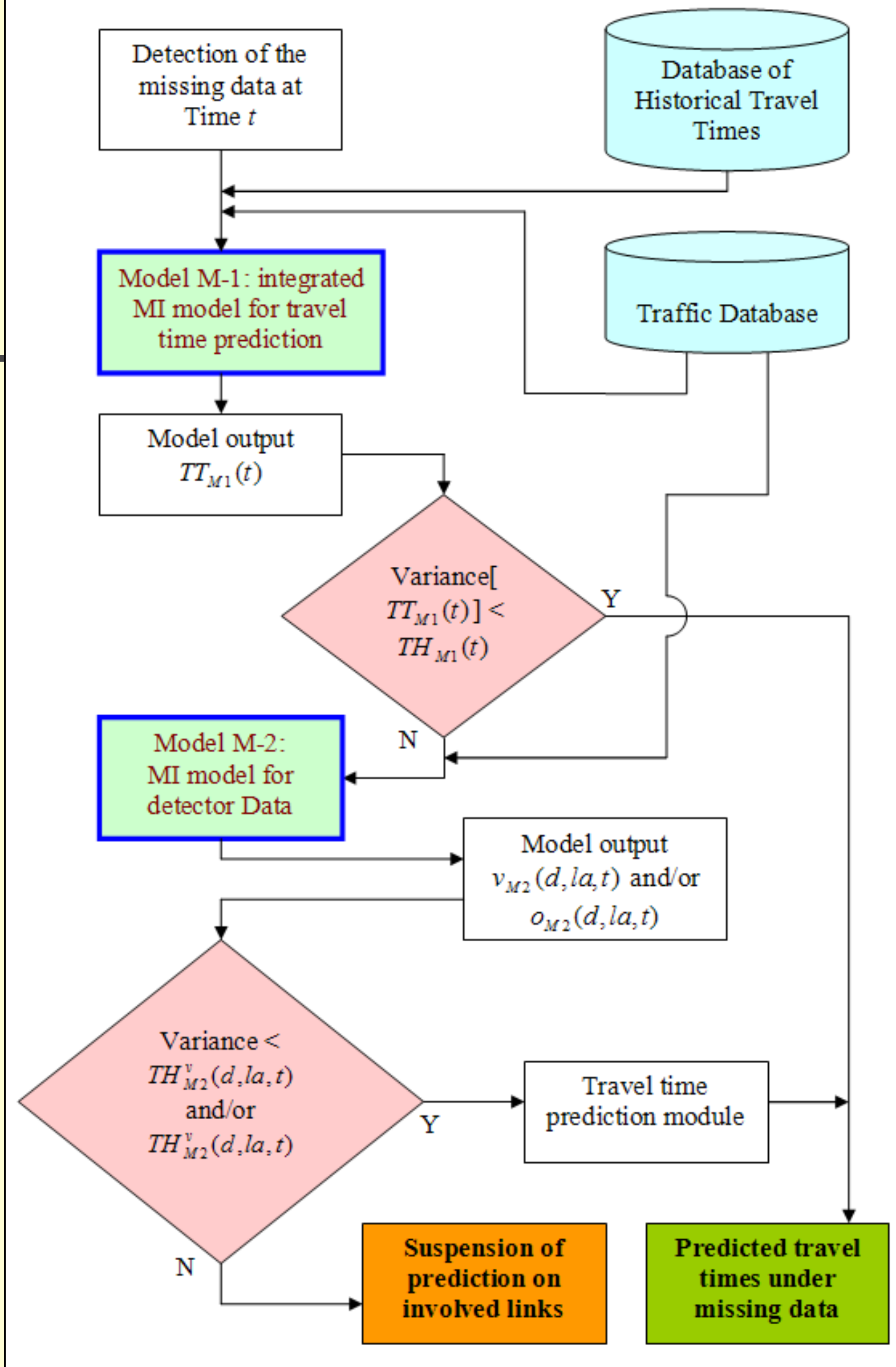
Multiple Imputation Technique

- Estimate the distribution of the missing values
- Randomly draw missing values until the distributions converge
- Repeat the imputation for *m* times

Proposed Models

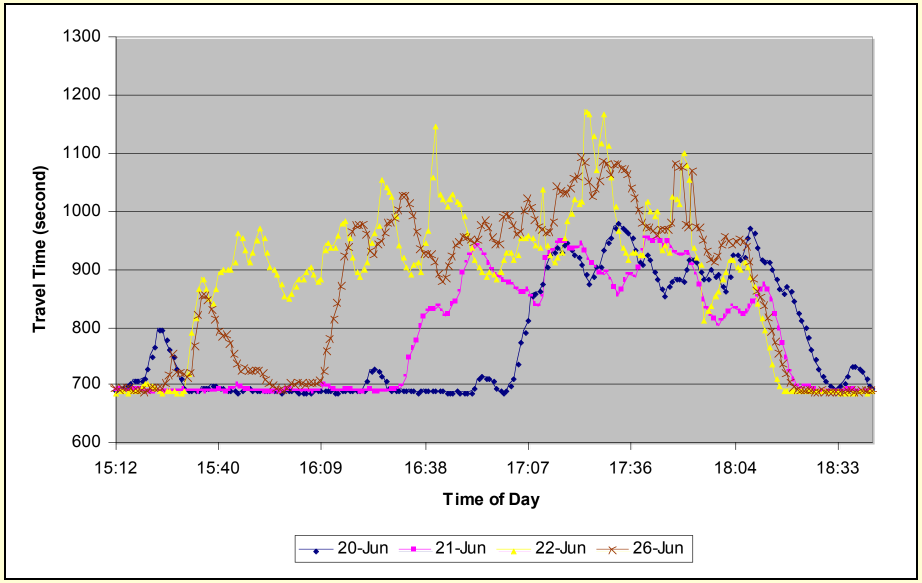
- Model M-1:
  - An integrated missing data imputation and travel time prediction model
  - Rely on data of the entire target segment
- Model M-2:
  - Multiple imputation model for missing values
  - Rely on data from predefined subsegment

Model Flowchart



Numerical Examples

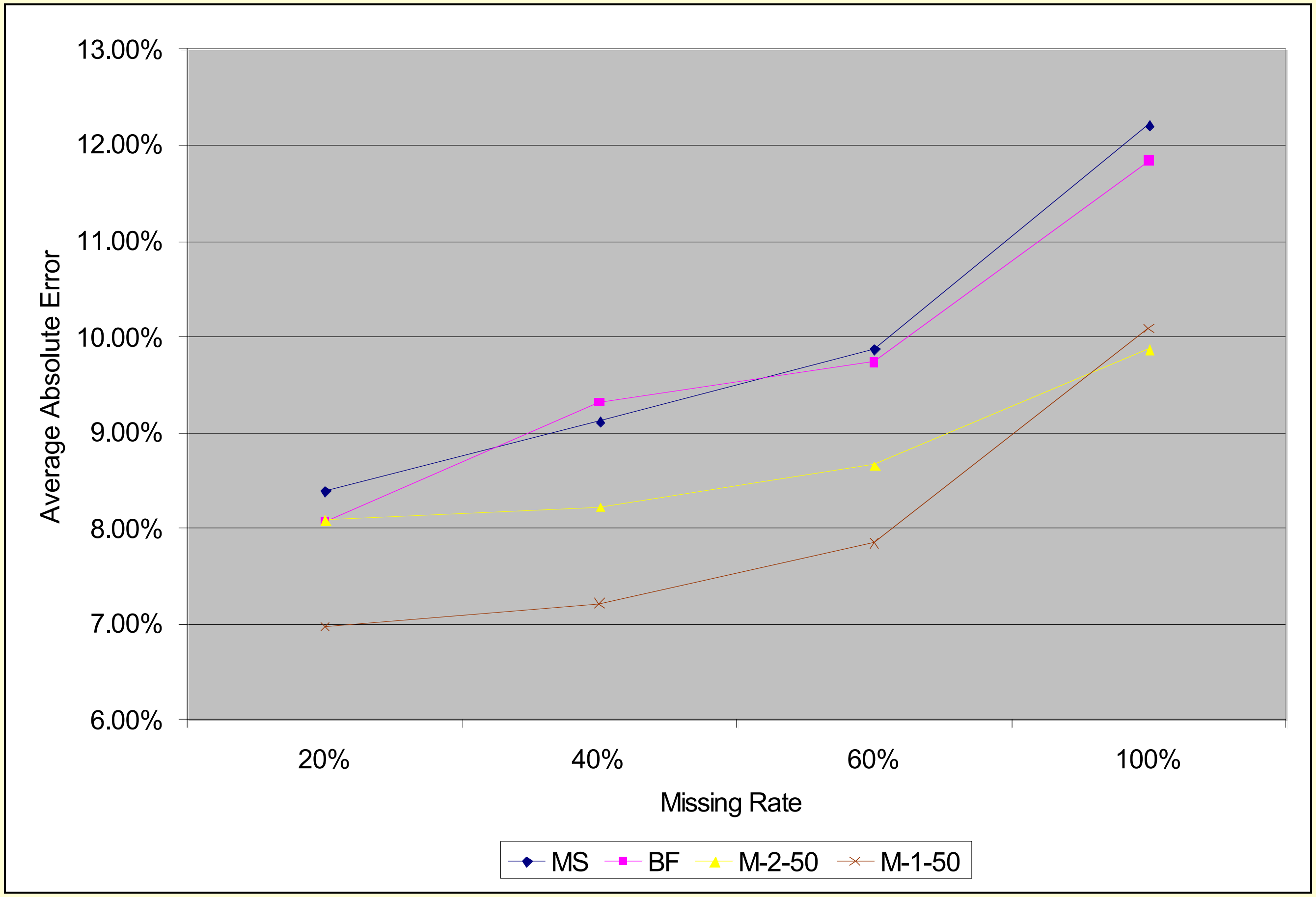
- Same dataset on I-70
- Four weekdays
  - June 20th (Tuesday) 21st (Wednesday), 22nd (Thursday) and 26th, 2006 (Monday)



- Comparison focuses on the impacts of:
  - The missing rate
  - The imputation models
    - Mean substitution (MS), Bayesian Forecast (BS)
  - The number of imputation

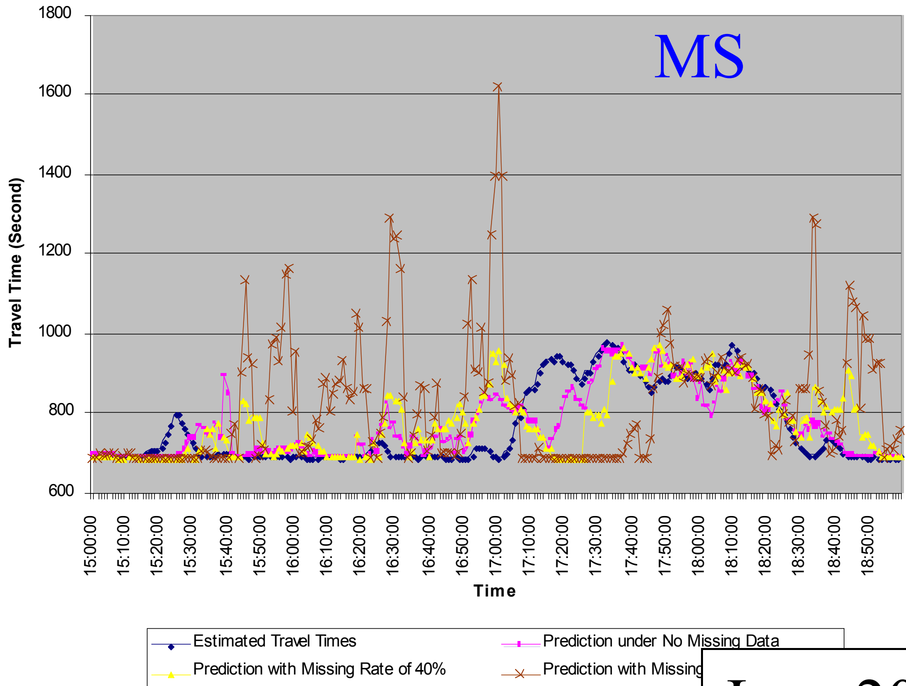
Performance Comparison (Travel Time)

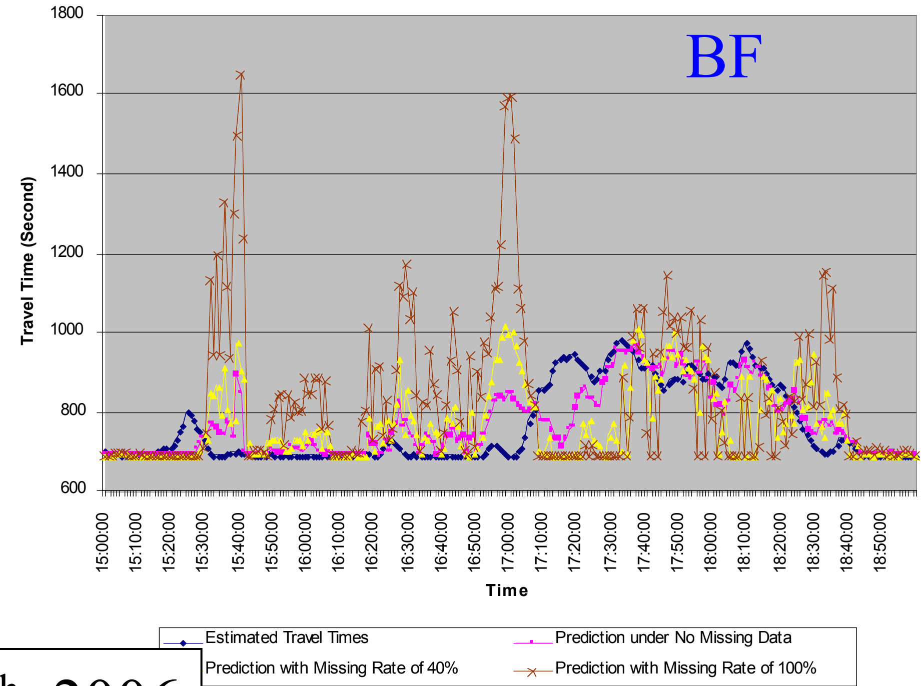
- One most critical detector has missing data



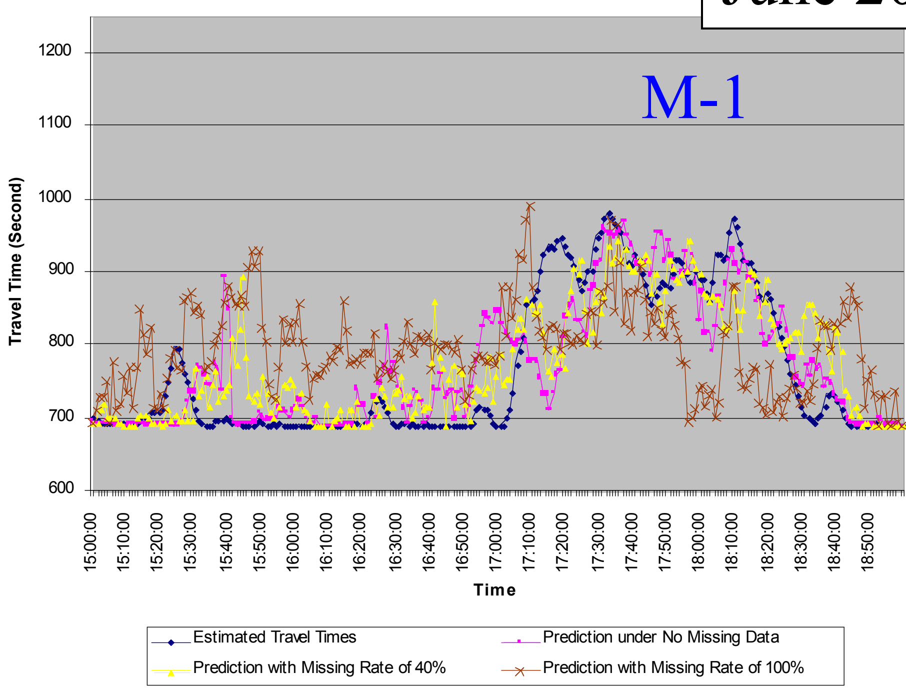
Different Congestion Levels

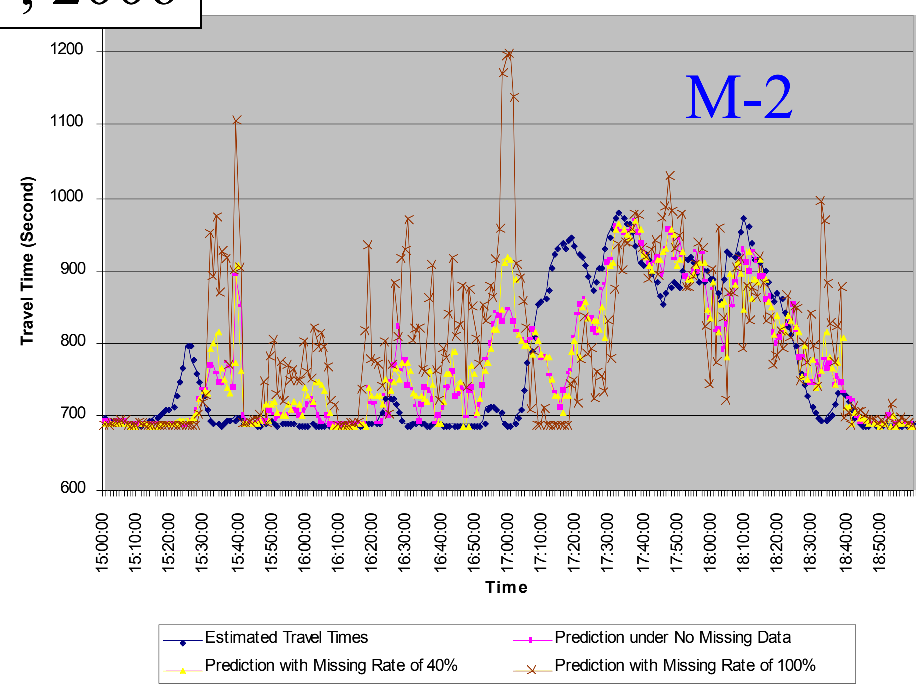
TT≤700		MS	BF	M-2-50	M-1-50
Missing Rate	20%	3.10%	2.78%	3.11%	2.54%
	40%	4.10%	3.63%	3.26%	2.80%
	60%	5.05%	4.47%	3.73%	3.07%
	100%	8.53%	7.47%	5.42%	6.37%
700<TT≤900					
Missing Rate	20%	8.23%	7.35%	7.43%	6.65%
	40%	8.76%	8.56%	7.29%	6.43%
	60%	9.33%	8.64%	7.71%	6.95%
	100%	10.48%	10.26%	8.58%	8.66%
TT>900					
Missing Rate	20%	13.46%	12.80%	12.36%	10.96%
	40%	13.82%	15.05%	12.57%	11.86%
	60%	14.29%	15.28%	12.99%	12.76%
	100%	16.12%	15.86%	13.55%	14.07%



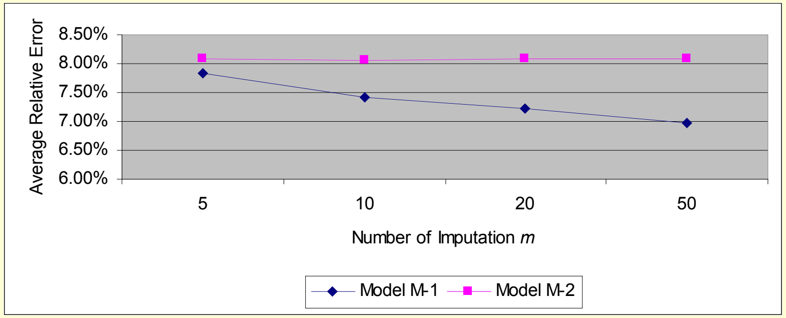


June 20th, 2006

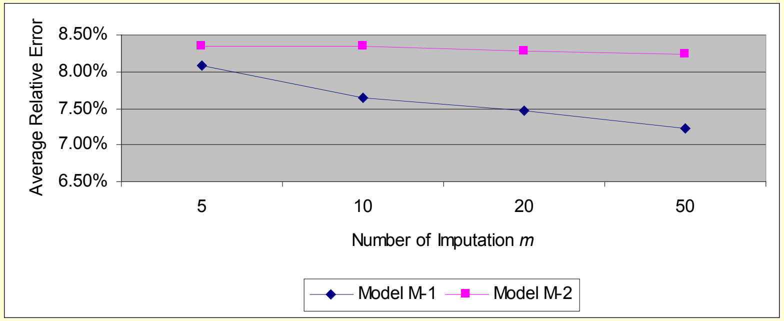




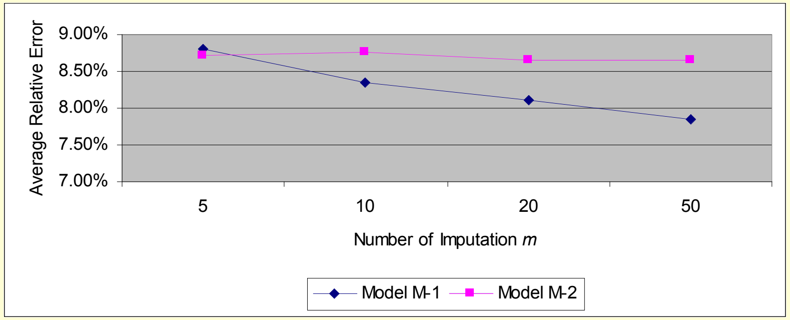
m=5, 10, 20, 50



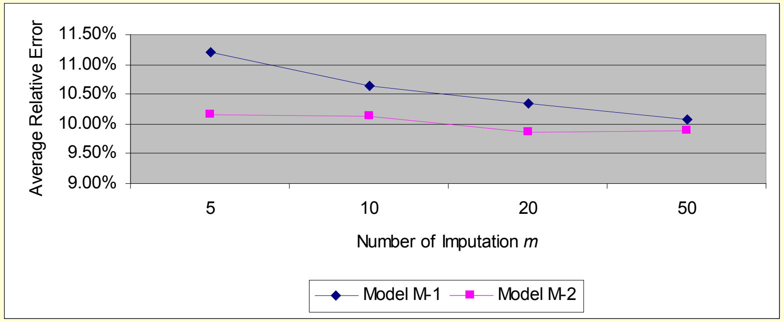
Missing rate: 20%



Missing rate: 40%



Missing rate: 60%



Missing rate: 100%

Summary

- Contributions
  - Perform an in-depth review of literature associated with travel time prediction
  - Develop a modeling framework for a travel time prediction system with widely spaced detectors on the freeway
    - Propose a hybrid model for estimating travel times
    - Develop a hybrid model for travel time prediction
    - Construct an integrated missing data estimation model for contending missing data issue

Future Research

- Determining optimal detector locations for better prediction performance
- Detecting incidents and other special events to minimize the potential prediction errors
- Monitoring change in traffic patterns and estimating potential impacts

Thank you!

Any questions?