

Improve traffic forecasting accuracy using spatial-temporal dynamics

Jianjiang Yang¹

Lee D. Han¹

Phillip B. Freeze²

University of Tennessee, Knoxville¹

Tennessee Department of Transportation²

BACKGROUND

- How upstream and downstream speed correlated?
- How to quantify the correlations?
- How to use neighboring information to improve traffic forecast accuracy?

OBJECTIVE OF THE STUDY

- ❑ To investigate spatial temporal relationships between speed series from consecutive segments under different conditions;
- ❑ To quantify the effect of upstream and downstream traffic information;
- ❑ To propose an algorithm on properly selecting upstream and downstream traffic information for short-term traffic forecasting.

DATA AND STUDY LOCATIONS

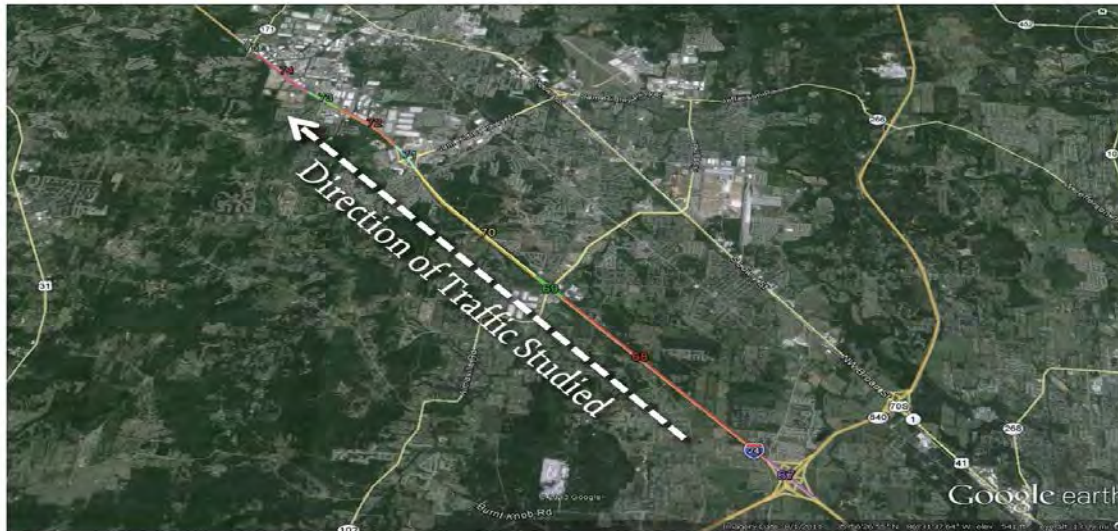


Table 1 Description of Study Segments

Segment No.	TMC	Location	Direction	Length (miles)	Type
67	121P04467	TN-840/EXIT 74 @ I-24	WESTBOUND	0.86	Internal
68	121+04468	ALMAVILLE RD/EXIT 70 @ I-24	WESTBOUND	3.94	External
69	121P04468	ALMAVILLE RD/EXIT 70 @ I-24	WESTBOUND	0.58	Internal
70	121+04469	SAM RIDLEY PKWY/EXIT 66 @ I-24	WESTBOUND	2.94	External
71	121P04469	SAM RIDLEY PKWY/EXIT 66 @ I-24	WESTBOUND	0.63	Internal
72	121+04470	WALDRON RD/EXIT 64 @ I-24	WESTBOUND	1.34	External
73	121P04470	WALDRON RD/EXIT 64 @ I-24	WESTBOUND	0.68	Internal
74	121+04215	OLD HICKORY BLVD/EXIT 62 @ I-24	WESTBOUND	1.32	External
75	121P04215	OLD HICKORY BLVD/EXIT 62 @ I-24	WESTBOUND	0.51	Internal

- INRIX link speed data collected for weekdays from April 07, 2013 to May 03, 2013;
- 9 segments (12.80 miles) from Exit 71 to Exit 62 (westbound) along I-24 in Nashville, TN;
- Segments are ordered from 65 to 75

METHODOLOGY

❖ Cross-correlation function (CCF)

Speed series is defined as $V_{s,t}$ with s as segment ID and t as time. The CCF for upstream and downstream speed series is defined as follows:

$$\rho_{s,s+1}^k = \frac{E[(V_{s,t+k} - \mu_s)(V_{s+1,t+k} - \mu_{s+1})]}{\sigma_s \sigma_{s+1}}, \quad k = 0, \pm 1, \pm 2, \dots$$

Where μ_s and μ_{s+1} are the means of $V_{s,t}$ and $V_{s+1,t}$, σ_s and σ_{s+1} are the standard deviation of $V_{s,t}$ and $V_{s+1,t}$, and k is the time lag between two series.

METHODOLOGY (CONT.)

❖ Pre-whitened Correlation method

To deal with complicating effects of autocorrelation and correctly indentify patterns in CCF, a pre-whitened cross-correlation method was proposed:

Fit model to $V_{s,t}$ to get an autoregressive model

Pre-whiten $V_{s,t}$ and $V_{s+1,t}$ based on AR model

Compute CCF for pre-whitened series α_t and β_t

Select significant CCFs for lagged regression model

METHODOLOGY (CONT.)

❖ Lagged regression

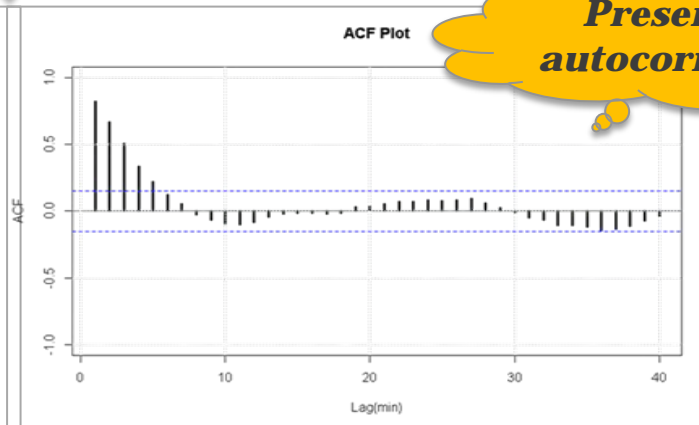
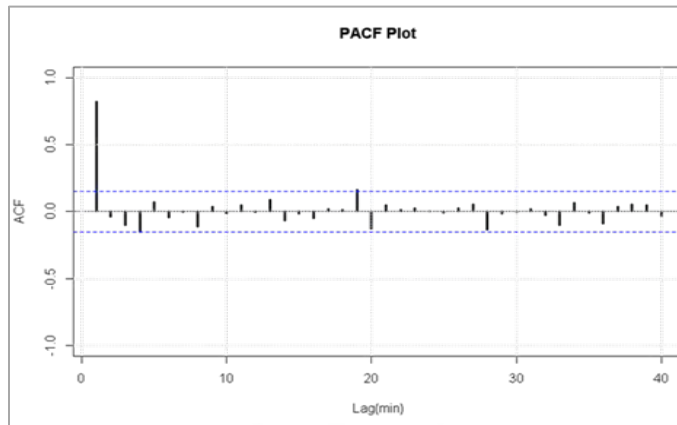
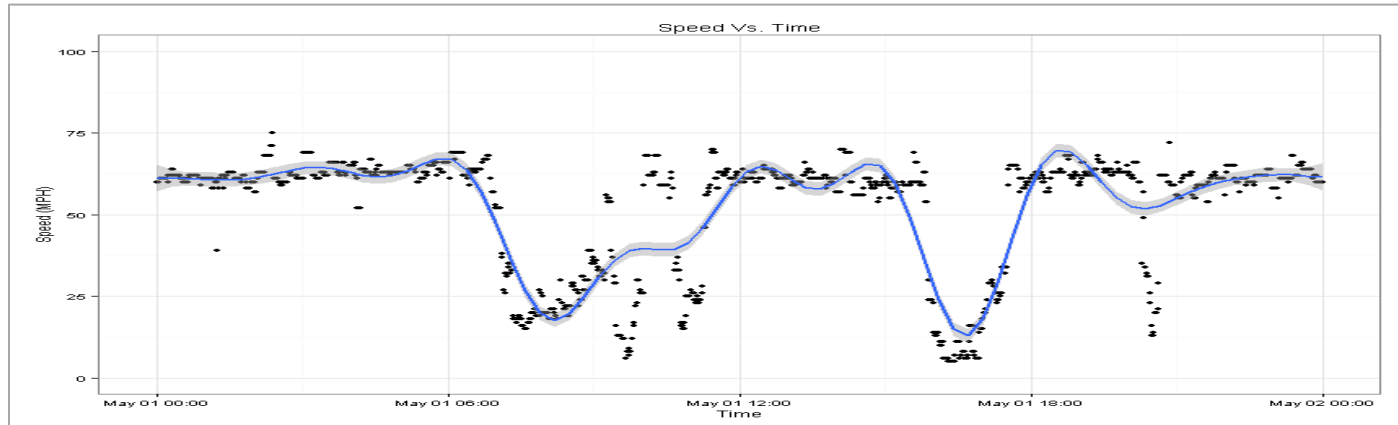
Suppose target segment ID is 71, lagged regression model in this study can be expressed as follows:

$$V_{71,t} = \sum \phi_{s,k} V_{s,t-k} + \varepsilon_t, \quad k \geq 0$$

Where ε_t is zero-mean uncorrelated error term.

PRELIMINARY ANALYSIS

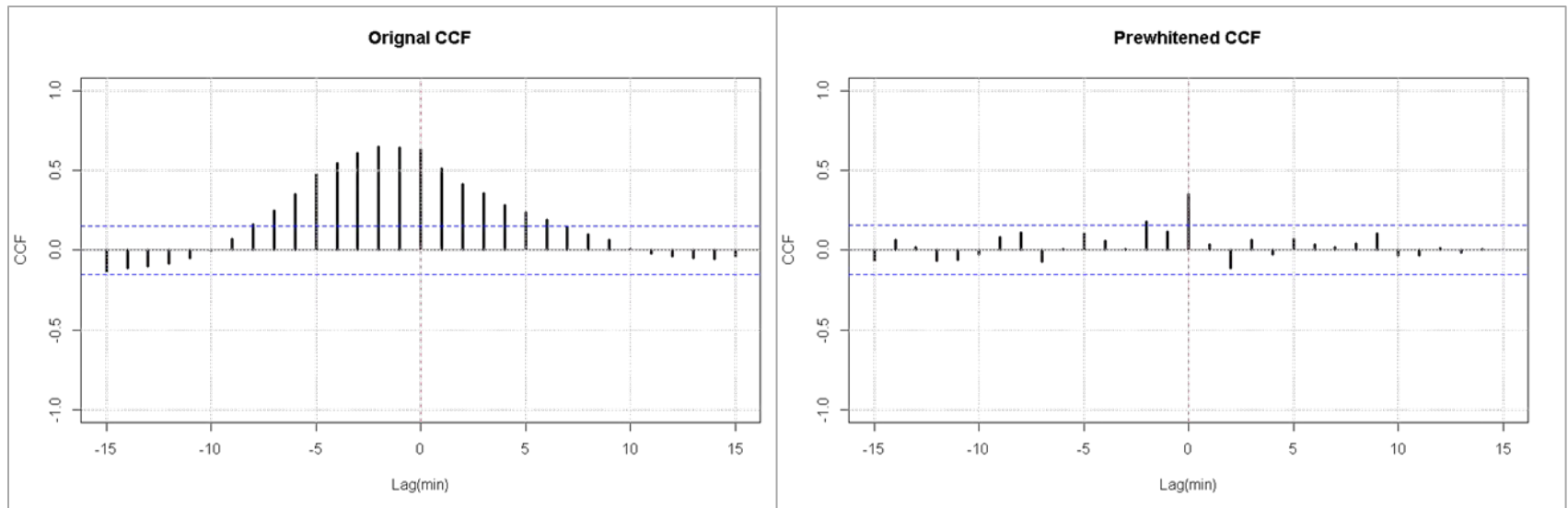
❖ ACF & PACF Analysis



Presence of autocorrelation

PRELIMINARY ANALYSIS (CONT.)

❖ Original CCF vs. Pre-whitened CCF

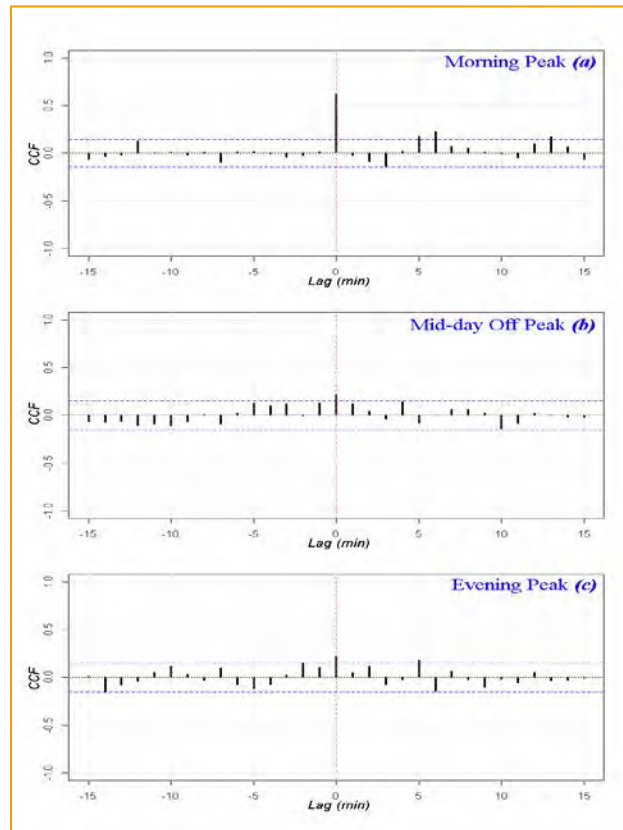


Original CCF: Real lagged relation between the two series is masked by the autocorrelation in the individual series ;

Pre-whitened CCF: clear spikes at lags 0 and -2 from the pre-whitened CCF plot, which means $V_{s,t}$ and $V_{s,t-2}$ can be used as predictor of $V_{s+1,t}$

ANALYSIS RESULTS

❖ Effect of traffic conditions ($V_{70,t+k}$, $V_{71,t}$)



Past values of speed at downstream can be used to predict future speed at upstream locations

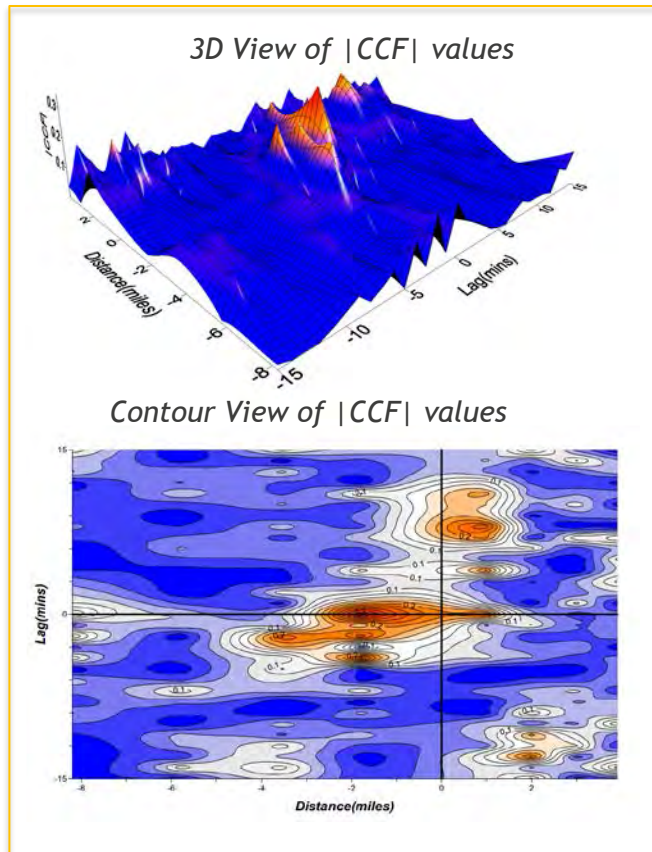
Effect of downstream traffic decreases; inter-influence exists

Effect of downstream traffic is not as strong as during morning peak

Cross-correlations between upstream (70) and downstream (71)

ANALYSIS RESULTS (CONT.)

❖ Effect of distance



From the distance perspective, the effective influence range is $[- 3.55, 1.99]$ based 99% significant confidence level;

From time lag perspective, most time lags are within 5 minutes

Spatial-temporal correlations regarding to segment 71 (morning Peak)

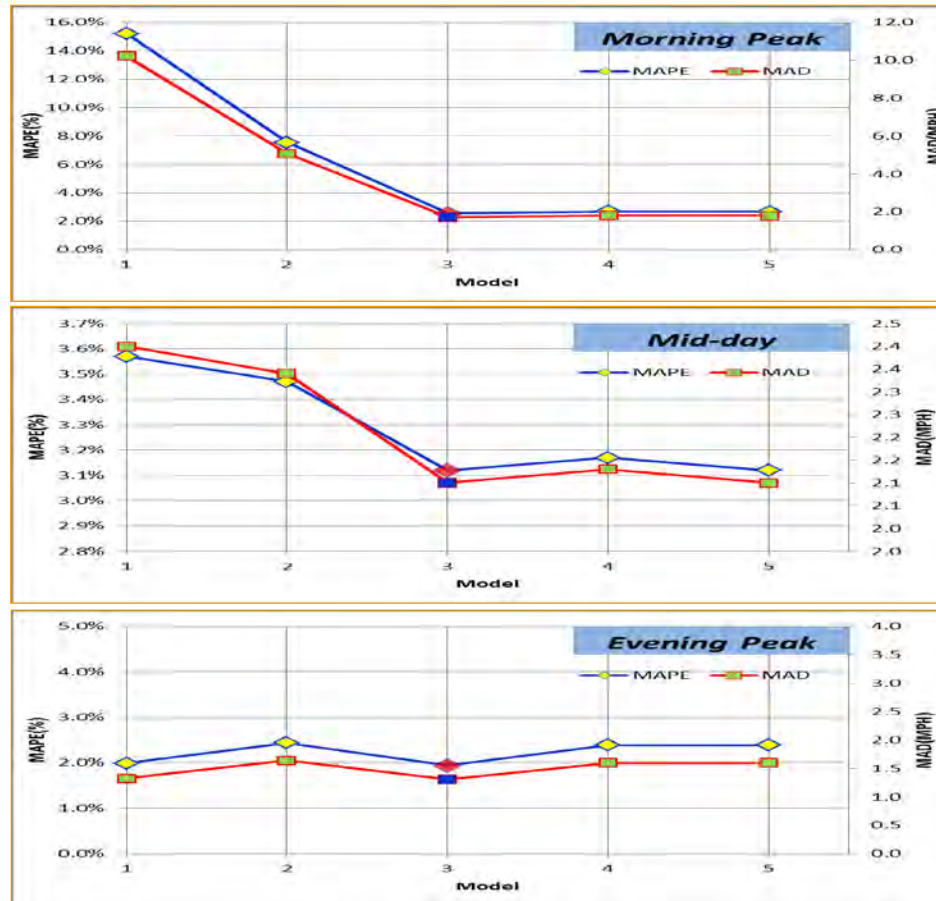
ANALYSIS RESULTS (CONT.)

❖ Models fitted and comparison results

Period	Model	Segments for Prediction		Holdout Performance	
		Upstream	Downstream	MAPE	MAD
AM Peak	1	70	-	15.21%	10.22
	2	69,70	-	7.56%	5.08
	3*	69,70	73	2.56%	1.72
	4	69,70	72,73	2.70%	1.81
	5	68, 69, 70	72,73	2.67%	1.80
Mid-day	1	70	-	3.57%	2.40
	2	69,70	-	3.47%	2.34
	3*	68,70	72	3.12%	2.10
	4	69,70	72,73	3.17%	2.13
	5	68, 69, 70	72,73	3.12%	2.10
PM Peak	1	70	-	1.99%	1.33
	2	69,70	-	2.44%	1.64
	3*	70	72	1.94%	1.31
	4	69,70	72,73	2.39%	1.60
	5	68, 69, 70	72,73	2.39%	1.60

ANALYSIS RESULTS (CONT.)

❖ Models fitted and comparison results



Conclusions

- ✓ The correlations for consecutive segments are highly related to traffic conditions and effect of downstream traffic increases with severity of congestion;
- ✓ Influences of up- and down-stream locations on current traffic are not symmetric with distance and upstream traffic tends to have greater effect;
- ✓ Properly using surrounding traffic information is important for forecasting accuracy.

