Using Temporal Detrending to Observe the Spatial Correlation of Traffic

Alireza Ermagun (Corresponding Author)
Ph.D. Candidate
University of Minnesota, Department of Civil, Environmental, and Geo- Engineering
500 Pillsbury Drive SE, Minneapolis, MN 55455 USA
ermag001@umn.edu

Snigdhansu Chatterjee
Associate Professor
University of Minnesota, School of Statistics
224 Church Street SE, Minneapolis, MN 55455 USA
chatt019@umn.edu

David Levinson
Professor
University of Minnesota and University of Sydney
500 Pillsbury Drive SE, Minneapolis, MN 55455 USA
dlevinson@umn.edu

Paper submitted for:
Presentation at 96th Annual Transportation Research Board Meeting, January 2017
Standing Committee on Freeway Operations (AHB20)

4745 words + 5 figures + 1 tables
July 31, 2016
ABSTRACT

This empirical study sheds light on the correlation of traffic links under different traffic regimes. We mimic the behavior of real traffic by pinpointing the correlation between 140 freeway traffic links in a sub-network of the Minneapolis - St. Paul highway system with a grid-like network topology. This topology enables us to juxtapose positive correlation with negative correlation, which has been overlooked in short-term traffic forecasting models. To accurately and reliably measure the correlation between traffic links, we develop an algorithm that eliminates temporal trends in three dimensions: (1) hourly dimension, (2) weekly dimension, and (3) system dimension for each link. The correlation of traffic links exhibits a stronger negative correlation in rush hours, when congestion affects route choice. Although this correlation occurs mostly in parallel links, it is also observed upstream, where travelers receive information and are able to switch to substitute paths. Irrespective to the time-of-day and day-of-week, a strong positive correlation is witnessed between upstream and downstream links. This correlation is stronger in uncongested regimes, as traffic flow passes through consecutive links more quickly and there is no congestion effect to shift or stall traffic. The extracted correlation structure can augment the accuracy of short-term traffic forecasting models.

Keywords: Highway Network; Traffic Flow; Spatial Correlation; Data Detrending; Traffic Forecasting
INTRODUCTION

The rapid development of technology and availability of large amounts of data enhances the ability to monitor traffic data over time and space, and eases analyzing the correlation between traffic links. Traffic analysts have utilized the spatial dependency of road segments to solve three typical problems in a traffic network: (1) short-term traffic forecasting, (2) reliable path problem, and (3) missing data estimation, discussed in the next section.

Irrespective of the problem, studies have revealed the positive spatial correlation between road segments for two main reasons. First, the topology of studied networks typically consists of traffic links that are immediately upstream or downstream, and thereby they exhibit positive correlation in terms of traffic due to the physics of conservation of flow. Second, traffic rises and falls by time-of-day, day-of-week, and week-of-year across the network, more-or-less independent of spatial configuration. Failing to extract the exact temporal dependency of traffic characteristics throughout a network results in neglecting the negative correlation between traffic links.

The positive correlation stands on traffic flow theory, and more precisely on the time-space diagram. This positivity derives from vehicles observed upstream at one time slice being observed downstream in the same or at a later time slice. It is true when we assume changing the road costs and demands over time has no impact. In reality, however, traffic may shift from one road to another due to traveler responses to congestion or closure. This may result in negative correlation between two links which are in series, as well as negative correlation for links in parallel. Although recent contributions from network science emphasize the necessity for procuring the exact interdependency between road segments, little is known about existing negative and positive correlations in a complex traffic network.

We hypothesize that traffic links in a real network exhibit both negative and positive correlations after detrending. This empirical study sheds light on the correlation of traffic links under different traffic regimes by adopting an in-depth statistical analysis to pinpoint the correlation of traffic. We contribute to the literature by defining and measuring the correlation of traffic on a real-word network, and explore their causes. This correlation structure is capable to augment the accuracy of short-term traffic forecasting models. We restrict our attention to a sub-network of the Minneapolis - St. Paul freeway system composed of 140 loop detectors with a grid-like network topology. This topology enables us to juxtapose the negative correlation of competitive segments with the positive correlation of complementary segments.

The remainder of the paper is set out as follows. First, we review the literature discussing the correlation nature of traffic links in road networks. Next, we discuss the data and methodology used in this study in detail. We proceed to graphically display the empirical correlation of traffic links, and collate the results in individual traffic regimes. We then conclude the paper by broaching a number of recommendations for future research.

PREVIOUS STUDIES

Traffic analysts scrutinize the correlation of traffic links to augment the accuracy of short-term traffic forecasting, reliable path finding, and missing data estimation. The literature discussing these three branches of research is prolific, and a well-established body of literature reviews the methodologies used in these studies. In 2004, Vlahogianni et al. (4) reviewed objectives and methods used in short-term traffic forecasting. They examined the pros and cons of modeling frameworks under the umbrella of parametric and non-parametric techniques. In 2014, Vlahogianni et al. (5) examined the challenges of modeling in short-term traffic forecasting, and concluded there is an
uncertainty whether the accuracy of developed complex methods are better than researchers models developed 30 years ago. More recently, Ermagun and Levinson (6) systematically reviewed more than 130 papers using spatiotemporal models for traffic forecasting. They emphasized that a large gulf exists between the spatial dependence of traffic links on a real network and the networks studied in current literature, and drew attention to these three shortcomings: (1) looking at spatial dependency of either adjacent or distant upstream and downstream of study link, (2) prejudging the spatial dependence between traffic links in modeling, and (3) neglecting the negative correlation between traffic links in modeling.

One of the main difficulties in the literature is that it is plagued with multifarious complex forecasting methods, while representing a long but shallow comprehension of spatial dependency between traffic links. In this part, hence, we dig into the correlation analysis used in the literature, and emphasize the approach of capturing spatial dependence between traffic links.

Early researchers used information upstream and downstream of the study link, as there is a reasonable belief that they are highly and positively correlated with the study link. Okutani and Stephanedes (7) were the first to utilize the information of adjacent upstream links in predicting traffic flow in 1984, although they never pointed out the correlation between traffic links. This approach spread through the literature for two major reasons. First, it was simple. As alluded to previously, traffic network is a complex system and understanding the detailed interrelationship between all traffic links requires comprehensive knowledge and large computational efforts. Thus, considering only the immediate upstream and downstream of the study link eases the calculation. Second, it was effective. Research typically studied a corridor comprising a small number of traffic links, which narrows the neighboring links of the study link down to adjacent upstream and downstream links. Then, it is not surprising to achieve decent results.

Stathopoulos and Dimitriou (8) used spatial correlation between two loop detectors. Embedding the information of the immediate upstream link, they improved traffic forecasts. Although Chandra and Al-Deek (9) examined a significant correlation of the study link with both adjacent and far traffic links, they only utilized the information of immediate upstream and downstream links.

Despite the simplicity and effectiveness, this method ignores the effects of other traffic links, as correlation only between adjacent links was presented in the literature. Li et al. (3), for instance, narrowed their study area to three consecutive traffic links: “It had been shown that the correlation degrees among different points decreases significantly with respect to distances. So, [...] we only consider $m = 3$ in this paper. That is, only the upstream and downstream neighboring detecting points are studied.” This approach is incomplete, as it selects only a part of the network and neglects the correlation between other traffic links.

More recently studies have emerged to examine the effects of not just adjacent traffic links, and thereby embed more information to enhance the accuracy of forecasting methods. One class of studies prejudices the correlation between traffic links in different distance thresholds. This class is so-called “$l^{th}$-order neighbors,” where $l$ represents the ring of neighbors. For instance, the first-order neighbors are those links that adjoin the study link, while the second-order neighbors are indirectly joined to the study link, having the adjacent links in the middle. Studies falling into this class assign a similar correlation value to each neighbor. In 2003, for example, Kamarianakis and Prastacos (10) considered both the first- and second-order neighbors and equally weighted all first- and second-order neighbors.

The other class of studies benefits from the correlation coefficient analysis to determine the
correlated links with the study link. In 2005, Sun et al. (11) analyzed a grid network comprising 31 traffic links in Beijing, China. To capture all spatial and temporal correlation between traffic links, they adapted Pearson correlation coefficient analysis. The results indicated the traffic flows of links are positively correlated, and the correlation does not follow any distance pattern.

Using cross-correlation analysis, Yue and Yeh (12) quantitatively measured the correlation between seven traffic links in an urban corridor of Kowloon, Hong Kong. They illustrated that the consecutive traffic links are positively correlated, and this correlation decreases by distance. They also found a significant drop in the correlation coefficient of one upstream link, which was justified by the presence of an off-ramp before the upstream link to a large residential area. A recent study (13) scrutinized the correlation between 3,254 loop detectors installed on the Minneapolis - St. Paul freeway system. Their analysis underlined that positive correlations exist in hundreds of sensors distributed on the whole road network sparsely, not just the neighborhood around the study link. Although they were the first to reveal the sparse correlation between traffic links, they overlooked the negative correlation nature of traffic links.

In defiance of various approaches to capture spatial correlation between traffic links, the literature has come to a longstanding agreement that traffic links are positively correlated.

We argue, that after properly controlling for temporal demand effects (i.e. by detrending as described below), network segments are both positively and negatively correlated, as one would expect from an understanding of spatial network structure which has links in both series and parallel, and where travelers have choice of route and are sensitive to perceived travel time (14).

DATA
In 2007, the Minnesota Department of Transportation (MnDOT) developed Intelligent Roadway Information System (IRIS), an open-source advanced traffic management system to monitor and manage highway traffic. The system collects and reports traffic flow, speed, occupancy, volume, density, and headway from 7,246 loop and virtual detectors in 30 seconds increments. Detectors are located in five distinct places: (1) Mainline Detectors, which collect data from all traffic lanes of interstates and highways, (2) Entrance ramp detectors, which collect the data of on-ramps, (3) Exit ramp detectors, which collect the data of off-ramps, (4) Queue ramp detectors at the start of ramps, and (5) Passage ramp detectors near ramp meters.

For the purpose of this study, we extracted traffic flow of a major sub-network of the Minneapolis - St. Paul freeway system. The sub-network consists of major highways in the western suburbs, specifically I-494, I-94, I-394, US 169, TH 212, TH 100, and TH 62 for the East-West and South-North directions. They are the busiest major highways in the Minneapolis - St. Paul highway system, particularly TH 62 in south Minneapolis is a notorious hotspot for traffic congestion. This sample includes 687 detectors, 146 of which are entrance and exit ramps. In road segments, the number of detectors varies from 1 to 4 depending on the number of lanes. We aggregated the flow information of all traffic lanes on a road segment, which results in 149 stations. We excluded 9 stations and 91 ramps due to lack of data. We collected the traffic flow measurements for all Tuesdays of 2015 in three distinct times-of-day:

1. Morning rush hour: From 7:30-8:30 AM
2. Non-rush hour: From 10:00-11:00 AM
3. Evening rush hour: From 4:30-5:30 PM
We also extracted the same information for all Saturdays of 2015. This trajectory enables us to compare the variation of competitive and complementary nature of traffic links not only over congested and uncongested regimes, but also over weekdays and weekends. We smoothed the traffic flow over 1-minute, which is assumed reasonable for the purpose of this study. This results in 3,120 observations (52 × 60) for each detector for each time-of-day. The missing data are excluded from the analysis for each detector. We cannot easily graph a 140 × 140 matrix for the analysis purpose, so we select some illustrative examples. Four stations were targeted in a stratified sampling method. They are stations 719, 340, 933, and 762, which are located in I-494, I-394, TH 100, and US 169, respectively. The characteristics of traffic flow for these four stations for all weeks are summarized in Table 1. As shown, the maximum traffic flow belongs to link 340 for Tuesday evening rush hour. The minimum traffic flow was observed on Saturday between 7:30 AM and 8:30 AM in link 719.

**TABLE 1**: Traffic flow characteristics of selected stations over week-of-year

<table>
<thead>
<tr>
<th>Link</th>
<th>Time</th>
<th>Average</th>
<th>St. Dev.</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>719</td>
<td>Tuesday 7:30-8:30</td>
<td>6082.59</td>
<td>849.71</td>
<td>7166.00</td>
<td>3788.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 10:00-11:00</td>
<td>3399.45</td>
<td>608.30</td>
<td>4426.00</td>
<td>2138.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 16:30-17:30</td>
<td>6227.45</td>
<td>1248.70</td>
<td>8036.00</td>
<td>3273.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 7:30-8:30</td>
<td>1921.12</td>
<td>880.07</td>
<td>3152.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 10:00-11:00</td>
<td>3690.43</td>
<td>1069.95</td>
<td>4826.00</td>
<td>18.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 16:30-17:30</td>
<td>4223.35</td>
<td>1167.78</td>
<td>5468.00</td>
<td>8.00</td>
</tr>
<tr>
<td>340</td>
<td>Tuesday 7:30-8:30</td>
<td>5470.54</td>
<td>411.86</td>
<td>5864.00</td>
<td>3818.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 10:00-11:00</td>
<td>2679.08</td>
<td>248.57</td>
<td>3151.00</td>
<td>1942.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 16:30-17:30</td>
<td>7768.58</td>
<td>643.83</td>
<td>8634.00</td>
<td>5966.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 7:30-8:30</td>
<td>1544.63</td>
<td>161.93</td>
<td>1870.00</td>
<td>1266.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 10:00-11:00</td>
<td>3182.50</td>
<td>202.86</td>
<td>3594.00</td>
<td>2834.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 16:30-17:30</td>
<td>3826.38</td>
<td>406.47</td>
<td>4673.00</td>
<td>2915.00</td>
</tr>
<tr>
<td>933</td>
<td>Tuesday 7:30-8:30</td>
<td>3672.42</td>
<td>299.76</td>
<td>4188.00</td>
<td>2576.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 10:00-11:00</td>
<td>2292.58</td>
<td>199.11</td>
<td>2653.00</td>
<td>1714.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 16:30-17:30</td>
<td>6687.75</td>
<td>706.48</td>
<td>7425.00</td>
<td>3341.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 7:30-8:30</td>
<td>1148.41</td>
<td>202.08</td>
<td>1450.00</td>
<td>703.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 10:00-11:00</td>
<td>2172.04</td>
<td>315.30</td>
<td>2639.00</td>
<td>1379.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 16:30-17:30</td>
<td>2810.31</td>
<td>406.34</td>
<td>3824.00</td>
<td>1672.00</td>
</tr>
<tr>
<td>762</td>
<td>Tuesday 7:30-8:30</td>
<td>4608.48</td>
<td>620.92</td>
<td>5517.00</td>
<td>2281.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 10:00-11:00</td>
<td>3314.12</td>
<td>838.73</td>
<td>5839.00</td>
<td>1691.00</td>
</tr>
<tr>
<td></td>
<td>Tuesday 16:30-17:30</td>
<td>4832.98</td>
<td>764.05</td>
<td>7368.00</td>
<td>3290.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 7:30-8:30</td>
<td>2017.15</td>
<td>589.23</td>
<td>3527.00</td>
<td>849.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 10:00-11:00</td>
<td>3602.87</td>
<td>744.23</td>
<td>5058.00</td>
<td>2023.00</td>
</tr>
<tr>
<td></td>
<td>Saturday 16:30-17:30</td>
<td>3970.42</td>
<td>753.70</td>
<td>5985.00</td>
<td>2157.00</td>
</tr>
</tbody>
</table>

To portray the traffic oscillation during day-of-week and day-of-weekend, we plotted the traffic flow of the selected links in Figure 2 for February 24th and 28th, 2015. As we expected, the traffic flow pattern of Tuesday is markedly different from Saturday. On Tuesday, traffic flow has two major peaks. One is happened in morning between 7:30 and 8:30, and the other is observed for
a longer period of time in evening between 15:00 and 18:00. Comparing the traffic flow of morning
rush hour with evening rush hour, we observe evening rush hour is generally more congested than
the mornings due to more personal trips. On Saturday, we witness on major traffic peak, which
starts about 10:00 AM. However, the traffic volume on Saturday is entirely lower than Tuesday.

METHODOLOGICAL FRAMEWORK
Three-dimensional Data Detrending
Traffic flow exhibits time trends in time-of-day, day-of-week, and week-of-year. These trends are
witnessed not only at the link level, but also at the system level, which is the total system travel by
time-of-day. Eliminating these time variations is fundamental to capture more accurate and reliable
spatial correlation between traffic links. As discussed in the preceding section, we extracted traffic
flow from three different one-hour time threshold in both Tuesdays and Saturdays of 2015. For
the purpose of temporal detrending and observing spatial correlation, we detrend the data in three
dimensions:

1. Hourly Dimension: In this step, we remove the trend in each one-hour time threshold
   from each traffic link. For example, we eliminate the trend from time threshold of 7:30-
   8:30 AM of the first Tuesday of 2015 from link 719. We repeat this step for all Tuesdays
   and for all links.

2. Weekly Dimension: The hourly detrended data of each traffic link has a weekly trend of
   52 weeks of the year. In this step, we eliminate this trend from the data.

3. System Dimension: Although removing the trend in two aforementioned directions is
   prevalent in the traffic literature, this dimension is typically overlooked in the traffic
data analysis. Unlike the previous two dimensions that focus on removing time trend,
   this dimension emphasizes on extracting the total system travel by time-of-day. Traffic
   flow of each link in a specific time span during a day displays a remarkable correlation
   with total flow of all traffic links. Deriving this trend is fundamental to observe the
   competitive nature of traffic links.

In the remainder of this section, we unpack the statistical steps behind the three-dimensional
data detrending. We utilize an algorithm to remove time-of-day and day-of-week trend for each
link, and the total system travel by time-of-day trend.

Data Detrending Algorithm
Loop detectors may not be functional for different periods of time during any given day owing
to malfunction or other technical issues, and in some cases they may not be functional for longer
stretches of time due to construction work or other longer term issues. There are various ways in
which such lack of data from a detector are indicated.

After reading the data and parsing it correctly to account for malfunctions, we concentrate
on the specific day of the week and duration of time that is of interest for our analysis. To verify
algorithmic robustness, we tested the algorithm for all days of the week, at various start and end
time points, and different levels of data aggregation. This yields a vector of volume of traffic data
for each traffic link and each day. Consider \( m \) the total number of aggregated data points, then the
vector of observations is represented by \( Y(s,t) = (Y_{s,t,1}, \ldots, Y_{s,t,m}) \) for each station \( s \) and each day.
FIGURE 1: Traffic Flow of Selected Sections for February 24th and 28th, 2015
$t$. The notations $s \in \{s_1, \ldots, s_5\}$ and $t = 1, 2, \ldots, T$ stand for study stations and days, respectively.

In our present data analysis $T = 52$. We fit a robust location estimator to the data vectors $Y(s, t)$ for each station $s \in \{s_1, \ldots, s_5\}$. This is captured by obtaining the minimizer $\hat{\mu}(s) \in \mathbb{R}^n$ as per Equation 1.

$$\sum_{t=1}^{T} ||Y(s, t) - \hat{\mu}(s)||_1,$$

where $|| \cdot ||_1$ is the $L_1$-norm of a vector. This yields the vector of medians for each location.

This step removes the secular trend for each coordinate of the vector obtained from the previous step. To this detrended data, we fit an autoregression model of appropriate order, to model the temporal dependencies between successive time aggregation intervals. This step involves a model selection, and we select the best available autoregression up to and including lags of order 0 to 5. A lag zero model implies no temporal dependency.

In order to do this, we fit under penalization the following model using the assumption that for each $s$ and $t$, the sequence $\{\varepsilon(s, t, k)\}$ is a mean zero, finite variance white noise sequence.

$$Y(s, t, k) - \hat{\mu}(s, k) = \sum_{j=0}^{J} \phi_{s,j}(Y(s, t, k - j) - \hat{\mu}(s, k - j)) + \varepsilon(s, t, k),$$

Figure ?? represents the selected models for Link 719 in different time thresholds.

We assume second order stationarity for the above model fitting. We obtain the residuals after this autoregression model fitting. As a result, we derive Equation 3

$$R(s, t, k) = Y(s, t, k) - \hat{\mu}(s, k) - \sum_{j=1}^{J} \phi_{s,j}(Y(s, t, k - j) - \hat{\mu}(s, k - j)).$$

Following the aforementioned steps to remove the trend and temporal dependencies, we embark on steps to obtain the spatial dependency patterns using the $R(s, t, k)$ values. The first step is to elicit the neighborhood dependency relations. For this, we obtain serial correlations across each pair of station $s_1$ and $s_2$ for each time $t$. This results in:

$$C(s_1, s_2, t) = Cor(R(s_1, t, k), R(s_2, t, k)).$$

We construct a robust yearly summary of these by taking the median $\hat{C}_1(s_1, s_2)$ of

$\{C(s_1, s_2, 1), \ldots, C(s_1, s_2, T)\}$. If $\hat{C}_1(s_1, s_2)$ is above a threshold $c_1$, we consider the stations $s_1$ and $s_2$ to be spatially correlated. We adopt $c_1 = 0.10$ for the present study.

After obtaining and identifying correlation structures in the above manner, we study longer range of complementary relations between stations. To achieve this, we first compute the proportion of trend and temporal dependency adjusted residuals for each day $t$ and each station $s$, which represents the proportion of traffic flowing through station $s$ on day $t$ at each time aggregation step.

Let these proportional residuals be $\tilde{R}(s, t, k)$. We use the same measure of association, namely the correlation, using these. That is, across each pair of station $s_1$ and $s_2$ for each time $t$, we obtain:

$$\tilde{C}(s_1, s_2, t) = Cor(\tilde{R}(s_1, t, k), \tilde{R}(s_2, t, k)).$$
FIGURE 2: Fitted autoregressive model to Link 719 in different time thresholds
As in the previous step, we construct a robust yearly summary of these by taking the median \( \hat{C}_2(s_1, s_2) \) of \( \{ \tilde{C}(s_1, s_2, 1), \ldots, \tilde{C}(s_1, s_2, T) \} \), and obtain a negative or positive relation between stations \( s_1 \) and \( s_2 \) if \( \hat{C}_2(s_1, s_2) < -c_2 \) for a chosen threshold \( c_2 \). In the present study, we used \( c_2 = 0.10 \).

We have cross checked our computations with other choices of thresholds and other tuning parameters of our algorithm, and the overall pattern of the results we obtain are quite stable. In the following section, we depict the extracted correlation for selected links and discuss the correlation of traffic.

**GRAPHICAL DISCUSSION**

After temporal detrending the data, we represent and discuss the results of spatial correlation of selected traffic links in this section. To give the reader a sense of how the value of correlation fluctuates between traffic links and time-of-day, we plotted the box and whisker diagram of four traffic links in Figure 3. Looking at the plots, the dots above and below each box show only a few number of links are highly correlated with the study link. The positive correlation is stronger than negative correlation, although they are competitive in the number. In general, the negative correlation is more prevalence on Tuesday morning and evening rush hours than other time-of-day. It is justified by the congestion during the rush hour, which brings to light the competitive role of parallel traffic links in the network. A weak negative correlation is observed during non-rush hour and weekends, due to the low level of traffic congestion.

To examine the relationship of negative and positive correlations with the network structure, and more precisely the parallel and series links, we mapped the correlation results of the selected links for Tuesday morning rush hour in Figure 4. In this figure, the color spectrum of negative correlation changes from light pink to violet, and for positive correlation it varies from light blue to dark blue. The study link is shown by a black star. The correlation magnitude greater than \( |10.0| \) represents a strong significant correlation at the 90% confidence interval.
The correlation results of station 719 highlight that, after detrending, there is a network structure effect. Both negative and positive correlation exist between flows at this station and others. This correlation ranges from -48.3 to 70.0 for station 719. A strong positive correlation belongs to the immediate upstream and downstream links of station 719. It is in line with our hypothesis and previous studies. The strength of positive correlation declines with distance. The positive correlation stretch upstream turns negative at station 515, which is located before an off-ramp. We posit traffic congestion propagation on station 719 results in some upstream traffic switching to a substitute path, and thereby more traffic on station 719 reduces traffic upstream as travelers seek substitutes. A strong stretch of negative correlation is also observed in the links parallel to station 719. This supports our hypothesis about competitive links. US 169 and TH 100 are two main competitive paths for I-494. Thereupon, it is not surprising that traffic flow passes through the substitute paths, when traffic congestion has a strong effect on the network.

Likewise, there is a strong positive correlation between station 340 and its immediate upstream and downstream. This correlation is weakened by distance station station 340 and is transformed into the negative correlation upstream. There is a strong negative correlation between station 340 and its competitive links in TH 62 and I-494. Looking at the correlation analysis of station 933, we observe a strong positive correlation between station 933 and its two immediate upstream links, but not its downstream link. As shown, the downstream station 935 stands in a
FIGURE 4: Correlation of Four Selected Sections for Tuesday between 7:30 AM and 8:30 AM
significant distance from station 933, which results in a weak positive correlation. Stations 755, 756, and 724 that are strong substitutes with station 933 exhibits a strong negative correlation.

Noteworthy is that spurious correlation appears in correlation analysis of all links. Although it includes fewer than 10% of the correlation results, it should be kept in mind that it stems from the nature of using real-world data and a significant number of missing data in loop detector data samples. For example, we do not have any physical justification to support why there is a strong negative correlation between stations 762 and 1769 or stations 340 and 935. Instead we believe it is a spurious correlation.

Traffic flow varies between weekdays and weekends. This variation results in a different correlation structure between traffic links. For example, we do not expect a strong negative correlation between traffic links during non-rush hour, as there is little congestion causing traffic flow to switch to the competitive paths. However, we still expect a strong positive correlation between the study link and its immediate upstream and downstream links. We also expect the evening rush hour and morning rush hour are alike in the correlation structure. To test these hypotheses, we present the correlation analysis of station 719 for different times of day in Tuesday and Saturday in Figure 5.

First cut analysis shows a significant difference between rush hour and non-rush hour, and between weekdays and weekends. In Tuesday non-rush hour, we observe a positive correlation between upstream and downstream of the study link. Not only does a strong correlation exist between the immediate links, but also in a second-order upstream link. Traffic flow passes through links faster in the uncongested traffic condition than congested traffic condition. As a consequence, the traffic observed in the upstream links are observed in the study link in a shorter time slice, and thereby they show a stronger positive correlation. A strong point of emphasis is the strength of this correlation in comparison with morning rush hour. The correlation between upstream and downstream in non-rush hour is stronger than rush hour, as fewer travelers divert to alternative routes. As we expected, there is no significant negative correlation in non-rush hour. Comparing the evening rush hour with morning rush hour, we detect a similar correlation not only in pattern, but in the magnitude as well. The results indicate dissimilarities between correlation patterns for Saturday and Tuesday between 7:30 AM and 8:30 AM. The correlation pattern of station 719 for Saturday between 7:30 AM and 8:30 AM is fairly similar to Tuesday between 10:00 AM and 11:00 AM. It is not surprising as there is no congestion on Saturday early morning, and thereby there is no negative correlation effect. Interestingly, the negative correlations show up between 10:00 AM and 11:00 AM on Saturday.

CLOSING REMARKS

Okutani and Stephanedes (7) directed attention to spatial correlation of traffic links. They did not recommend incorporating the information of correlated links in traffic forecasting models, but rather the immediate upstream link. This school of thought has spread through the literature of short-term traffic forecasting. Using the spatial correlation between links has grown in popularity, not just because it is a way to augment short-term traffic forecasting models, but also because it is a way to cope with missing data and path selection. However, the literature provides little empirical evidence for the correlation of traffic in a real-word network, and is limited to correlation analysis of links in a series corridor encompassing consecutive links. The literature is comprehensive in the sense that it deals with positive correlation among the study links and its immediate upstream and downstream links. However, it is not generic in that it sets broad principles for complementary
Correlation Analysis of Link 719 for Tuesday Non Rush Hour

Correlation Analysis of Link 719 for Tuesday Evening Rush Hour

Correlation Analysis of Link 719 for Saturday Early Morning

Correlation Analysis of Link 719 for Saturday Before Noon

FIGURE 5: Comparison of Correlation for Different Times and Days
nature of traffic links, and leaves the correlation analysis of competitive traffic links for later.

This empirical study instead applies a three-dimensional data detrending algorithm and tests it on a grid-like network topology consisting of both competitive and complementary traffic links. This methodological approach enabled us to shed more light on the understanding of the traffic phenomena. We added to the body of knowledge on short-term traffic forecasting problem by capturing the realistic spatial correlation between traffic links. The key findings from correlation analysis of 140 traffic links and 54 ramps in the Minneapolis - St. Paul network are as follows:

- In a network comprising links in parallel and series, both negative and positive correlation shows up between links.
- The strength of correlation varies by time-of-day and day-of-week.
- The strong negative correlation is observed in rush hours, when congestion affects travel behavior. This correlation occurs mostly in parallel links, and in far upstream links where travelers receive information about congestion (for instance from media, variable message signs, or personal observation of propagating shockwaves) and are able to switch to substitute paths.
- Irrespective of time-of-day and day-of-week, a strong positive correlation is observed between upstream and downstream sections. This correlation is stronger in uncongested regimes, as traffic flow passes through the consecutive links in a shorter time and there is no congestion effect to shift or stall traffic.

The sub-network used in this study includes a significant number of missing data pertaining to both traffic links and time-of-day. To extract more accurate correlation between traffic links, we need data that represents all traffic demands in the network for a specific time slice. We argue that accuracy, robustness, and adaptivity are fundamental for successful implementation of short-term traffic prediction models in advanced traveler information service. The proposed algorithm is practical for deployment in any traffic network to achieve persistent and accurate correlation between traffic links. Spelling out the details of how to integrate these correlation effects into short-term traffic forecasting models remains a research challenge.

REFERENCES


