Integrating Spatial-Temporal Traffic Data on an Urban Network for Change Detection

Zhengyuan Zhu
Department of Statistics
Iowa State University

Aug. 17, 2017

Acknowledgement

- This talk is based on joint work with PhD students Zhonglei Wang and Weicheng Zhu, and Tina Greenfield from Iowa DOT.
- The authors want to thank the Iowa Department of Transportation (DOT), Midwest Transportation Center, U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology, and Federal Highway Administration (FHWA) Aurora Transportation Pooled Fund (TPF-5(290)) partners for sponsoring this research.

Background and Motivation

- Winter weather in Iowa is often unpredictable and can have an adverse impact on traffic flow.
- The Iowa DOT uses various proactive maintenance operations to lessen the impact of winter weather events on traffic speeds, and want to assess the performance of these maintenance operations.
- In previous research, we developed models for estimating expected speed reductions for performance evaluation based on weather variables and normal maintenance schedules for a single site.
- How about doing this on an urban network?

Objectives

- Develop spatial temporal models to integrate multiple sources of traffic data from an urban network;
- Early detection of traffic speed and volume change in space and time;
- Forecasting traffic speed and volume change during winter weather for performance evaluation.

Data sources

- Wavetronix:
  - 2013 and 2014 data. Resolution: 5-minute.
  - Location: about 20 locations covering sections of I-35, I-80, US-65 and IA-5 in the metropolitan area of Des Moines.
  - It provides accurate readings for traffic speed and volume.
- INRIX:
  - 2013 and 2014 data. Resolution: 2-minute.
  - Location: dense in main roads of Des Moines area.
  - It only provide readings for traffic speed
  - Its reading not consistent with those from Wavetronix
  - Kim et al., 2014: "efficiently calculating the speeds over a longer period".
  - The traffic speed reading appears to be self-consistent.

Data sources (Cont'd)

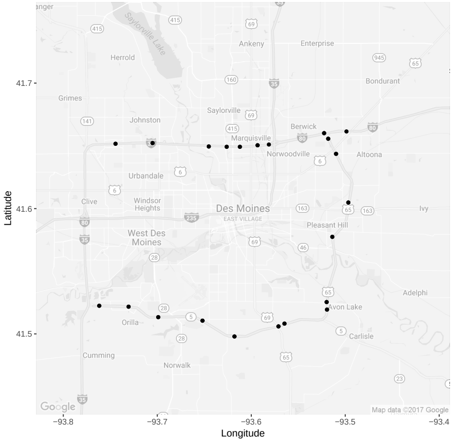


Figure: Locations of Wavetronix sensors in Des Moines metro-area.

Data sources (Cont'd)

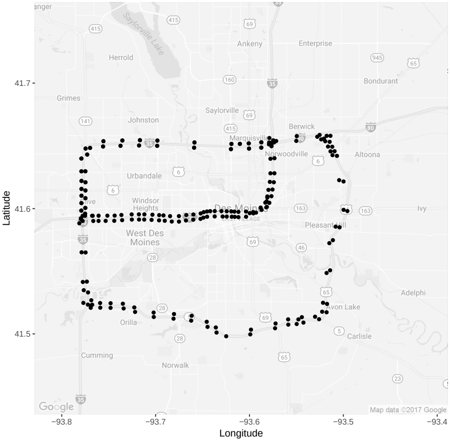
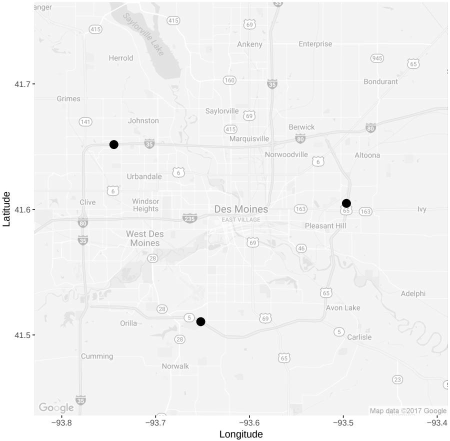


Figure: Locations of INRIX in Des Moines metro-area with proper jitter.

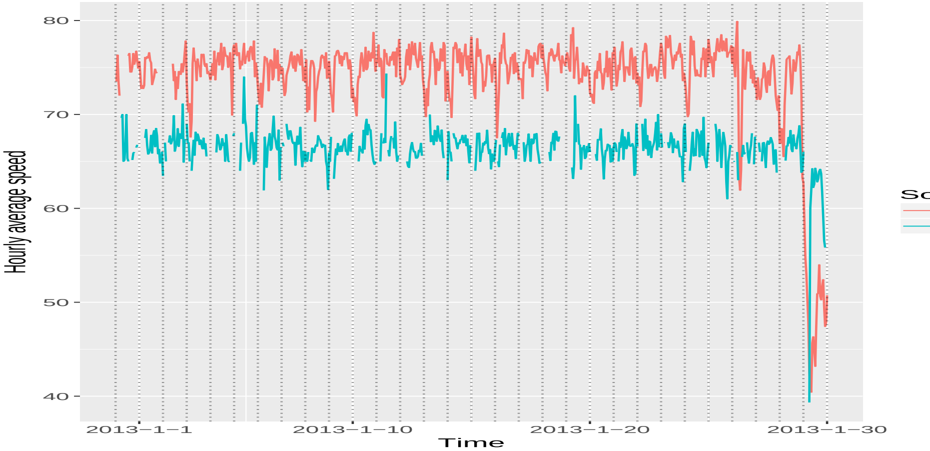
Data sources (Cont'd)

- Three locations are used for the comparison of the hourly mean speeds from two sources.



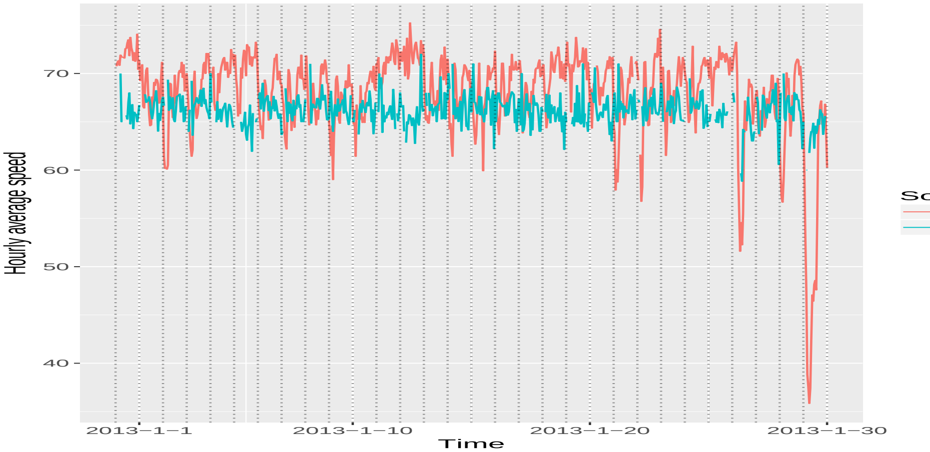
Data sources (Cont'd)

The lower one.



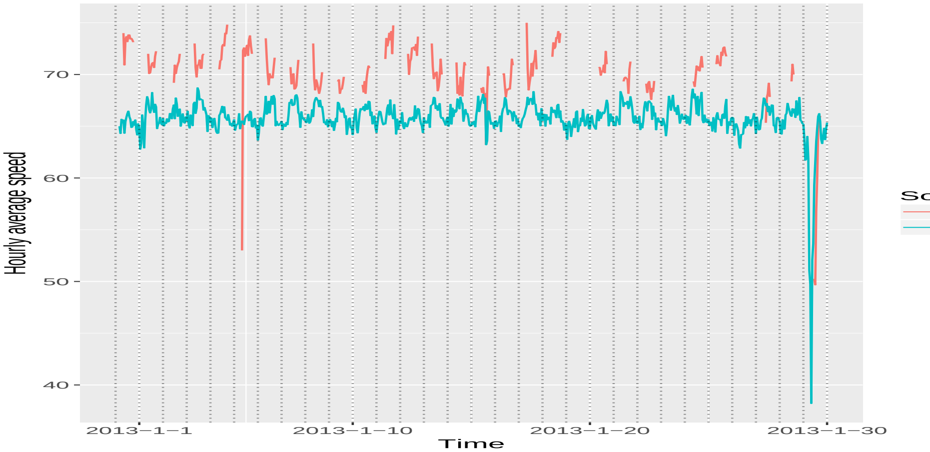
Data sources (Cont'd)

The middle right one.



Data sources (Cont'd)

The upper one.



Integrating the two data sources: the Curve Kriging approach

- Wavetronix data have higher quality but are sparse in space.
- INRIX data have lower quality but are dense in space.
- How to integrate these two sources? Key idea:
  - Use the INRIX data to estimate the spatial functional variogram (the dependence structure),
  - Use the Curve Kriging approach (Giraldo et al., 2011) to predict the Wavetronix traffic speed/volume data at any location using functional variagram estimated from INRIX data.

Curve Kriging approach: Technical details

- The spatial dependence is modeled by the functional variogram

$$\gamma_{t_0,t_0'}(\|\boldsymbol{h}\|;\boldsymbol{\theta}) = \frac{1}{2}V[Y_{\boldsymbol{s}}(t_0) - Y_{\boldsymbol{s}+\boldsymbol{h}}(t_0')].$$

- For location $\boldsymbol{s}_0$, the best linear unbiased predictor (BLUP) for $\boldsymbol{Y}_{\boldsymbol{s}_0}$ is

$$\hat{\boldsymbol{Y}}_{\boldsymbol{s}_0} = \sum_{i=1}^{n} \lambda_i^{(\boldsymbol{s}_0)} \boldsymbol{Y}_{\boldsymbol{s}_i},$$

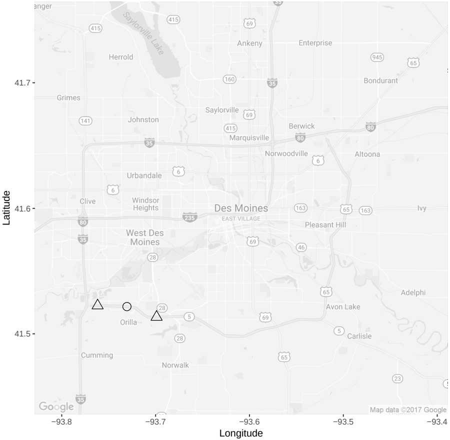
  where $\{\boldsymbol{s}_i : i = 1, 2, \ldots, n\}$ is the Wavetronix location, and the coefficients $\{\lambda_i^{(\boldsymbol{s}_0)} : i = 1, 2, \ldots, n\}$ are obtained by minimizing

$$E\left[\int_T \left(\hat{Y}_{\boldsymbol{s}_0}(t) - Y_{\boldsymbol{s}_0}(t)\right)^2 \mathrm{d}t,\right]$$

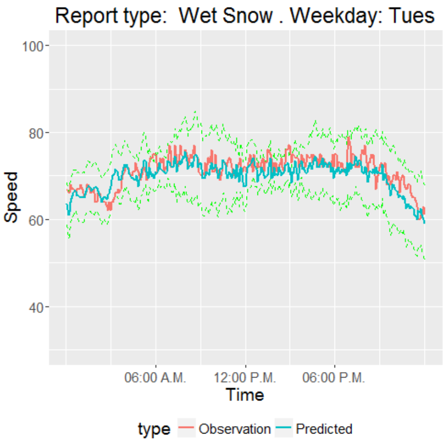
  subject to $\sum_{i=1}^{n} \lambda_i^{(\boldsymbol{s}_0)} = 1$.
- $\lambda_i^{(\boldsymbol{s}_0)}$ is a function of $\gamma_{t_0,t_0'}(\|\boldsymbol{h}\|;\boldsymbol{\theta})$.

Example

As an example, we predict the traffic speed at the Wavetronix location "IA-5 EB to SW CONNECTOR-EB" on 2013/1/29 using the observations at its two nearest neighbors.



Example (Cont'd)



Robust functional change detection to detect traffic anomaly

- Functional approach takes care of temporal dependence
- Estimation of normal traffic variability: functional quantile approach for robustness.
  - Abdous and Theodorescu (1992) extended the definition of quantile to multivariate case.
  - Functional depth (Liu, 1990; Zuo and Serfling, 2000; Lopez-Pintado and Romo, 2009) can also be used.

Multivariate quantile estimation: Technical details

- Denote $\boldsymbol{x} \in \mathbb{R}^d$ is a d dimensional vector with the l_p norm $\|\cdot\|_p$.
- For fixed $\alpha \in (0, 1)$, define $\|\cdot\|_{p,\alpha} : \mathbb{R}^d \to \mathbb{R}$ as

$$\|\boldsymbol{x}\|_{p,\alpha} = \|(x_1, \ldots, x_d)\|_{p,\alpha} = \left\| \left(\frac{|x_1| + (2\alpha - 1)x_1}{2}, \ldots, \frac{|x_d| + (2\alpha - 1)x_d}{2} \right) \right\|_p .$$

- Assume $\boldsymbol{X} \in \mathbb{R}^d$ to be a random variable, and $\mu(\cdot)$ is the corresponding induced probability measure. Define

$$H_{p,\alpha}(\boldsymbol{\theta}) = E\left\{ \|\boldsymbol{X} - \boldsymbol{\theta}\|_{p,\alpha} - \|\boldsymbol{X}\|_{p,\alpha} \right\} = \int_{\mathbb{R}^d} \left\{ \|\boldsymbol{x} - \boldsymbol{\theta}\|_{p,\alpha} - \|\boldsymbol{x}\|_{p,\alpha} \right\} \mu(\mathrm{d}\boldsymbol{x}).$$

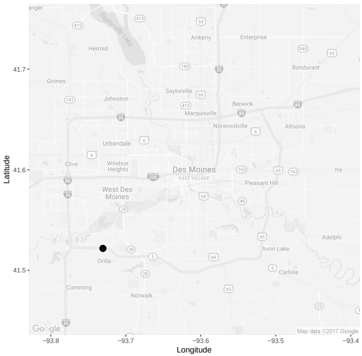
- The α-quantile of $\boldsymbol{X}$, say $\boldsymbol{\theta}_{p,\alpha}(\boldsymbol{X})$, is obtained by

$$\boldsymbol{\theta}_{p,\alpha}(\boldsymbol{X}) = \arg \inf_{\boldsymbol{\theta} \in \mathbb{R}^d} H_{p,\alpha}(\boldsymbol{\theta}).$$

- Newton-Rahpson method can be used to estimate $\boldsymbol{\theta}_{p,\alpha}(\boldsymbol{X})$.

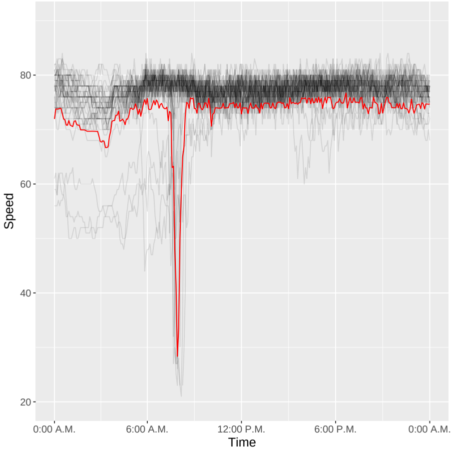
Multivariate quantile estimation (Cont'd)

- For each Wavetronix station, a 10%-th quantile is estimated for each day of a week based on the observations of 2013.
- One Wavetronix station is used as an example.



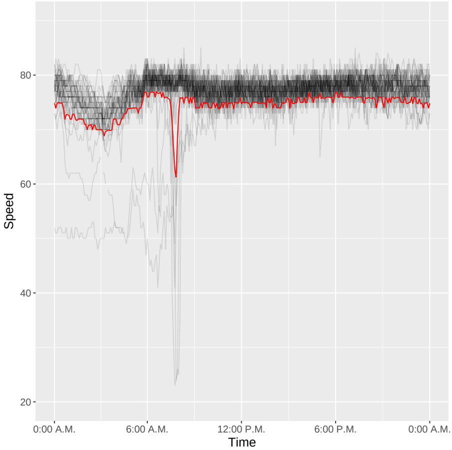
Example

Estimation result for Monday.



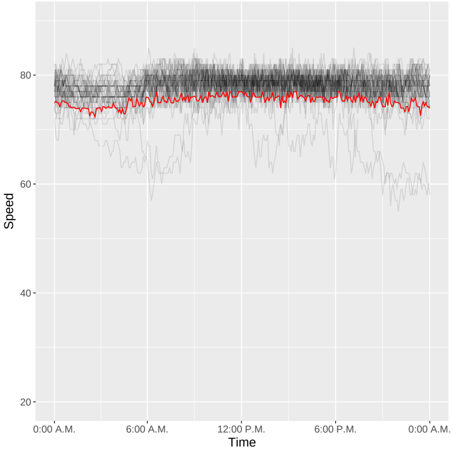
Example (Cont'd)

Estimation result for Friday.



Example (Cont'd)

Estimation result for Sunday.



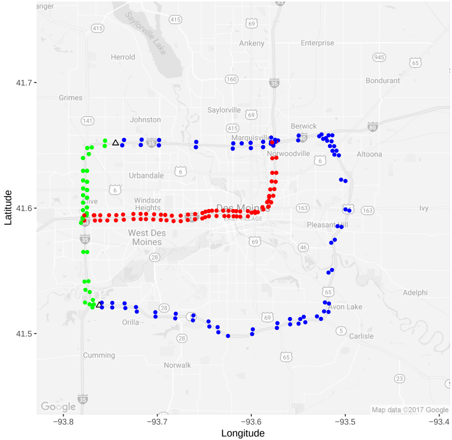
Detection of traffic change in space and time

- We first integrate the two data sources, then detect change on the integrated data.
- The proposed procedure:
  1. Use the multivariate quantile estimation method to obtain the 10%-th quantile of normal traffic for each Wavetronix station for each day of the week.
  2. Use the curve Kriging approach to get the date-based spatial structure.
  3. For each target location, use the curve Kriging method to estimate both the current speed and the 10%-th quantile curve of normal traffic.
  4. For each target location, on an hourly bases compare the estimated speed to the corresponding 10%-th quantile to detect change.

Detection of traffic change (Cont'd)

- More details about the second step.
  - We divide the target locations into three groups based on the Wavetronix observation locations.
  - For a target location in each group, we specify its two surrounding Wavetronix stations with observations.
  - The curve Kriging approach is conducted based on the observations from its two neighbors.
- Imputation is used for missing observations.
  - The estimated median curve is used for the missing segment that starts with the first observation or ends with the last observation.
  - Otherwise, the weighted average is used.
- We use 2013-1-1 as an example.

Example



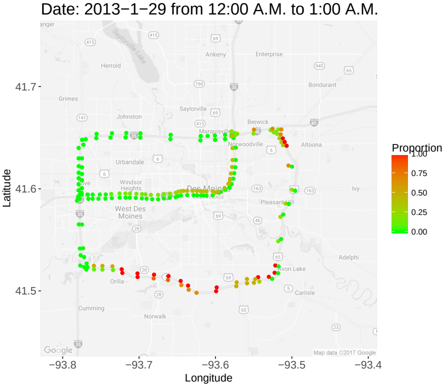
- Triangle: the most left Wavetronix locations with observations in the I-35 and I-5 segments.

Detection of traffic change (Cont'd)

- Blue group: find the two surrounding Wavetronix location directly.
- Green group: use the two triangle locations as the nearest neighbors.
- Red group: Use the predicted value at both ends for kriging.

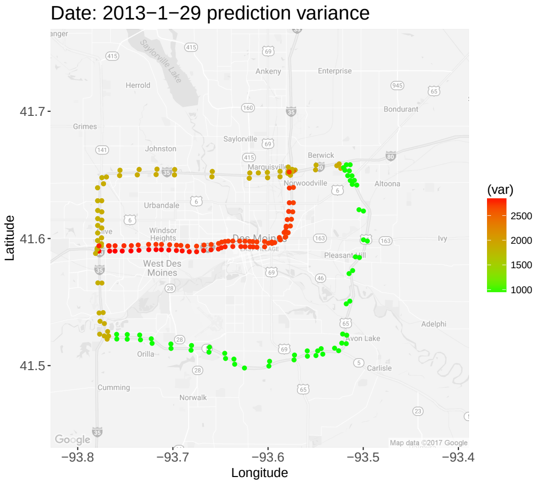
Detection results

The detection results from 2013-1-29 to 2013-1-31.



Uncertainty estimation

The variance results for 2013-1-29



Improve the integrate

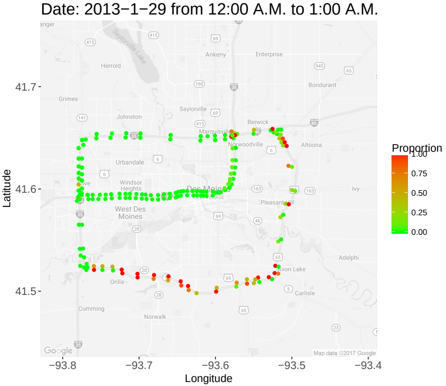
- The aggregated INRIX data over a longer period is believed to be reasonable and may be used to improve the integrated data.
- Idea: build a regression model between the hourly means of the two sources to correct for bias.
- For a specific hour,
  - Let w_i and x_i be the hourly mean of Wavetronix and INRIX data.
  - Consider the following linear model

$$E(w_i \mid x_i) = \beta_0 + \beta_1 x_i.$$

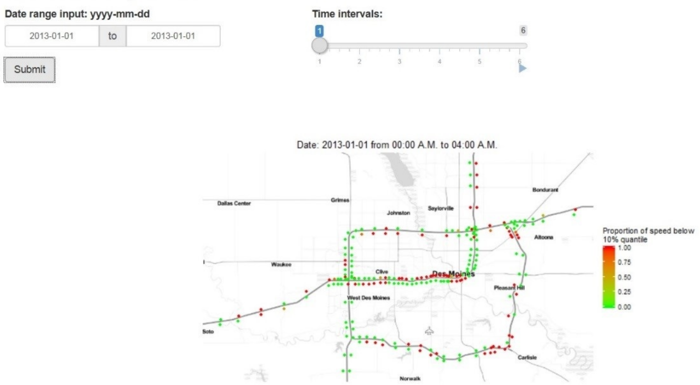
  - Let $\hat{\beta}_0$ and $\hat{\beta}_1$ to be the estimated parameters.
  - Conduct the curve Kriging approach based on the mean-adjusted Wavetronix data.

Detection results

The detection results from 2013-1-29 to 2013-1-31 are shown.



We have developed an online APP to interactively present the results. The computation is fast and can be done in real time.



Future Work

- Spatial dependence can be explored for early detection.
- Spatial-temporal interaction can be explored for prediction.
- Integrating other sources of data to build spatial-temporal data base for traffic forecasting, anomaly detection, etc. In particular, tools are available to convert video camera data to speed data, which can be integrated.
- Network design: the model can provide kriging uncertainty, which can guide the selection of new sensor locations.

Reference

Abdous, B.,& Theodorescu, R. (1992). Note on the Spatial Quantile of a Random Vector. *Statistics & Probability Letters* **13**, 333–336.

Giraldo, R., Delicado, P. & Mateu, J. (2011). Ordinary Kriging for Function-valued Spatial Data. *Environmental and Ecological Statistics* **18**, 411–426.

Kim, S. & Coifman, B. (2014). Comparing INRIX speed data against concurrent loop detector stations over several months. *Transportation Research Part C: Emerging Technologies* **49**, 59–72.

Liu, R. Y. (1990). On a Notion of Data Depth based on Random Simplices. *The Annals of Statistics* **18**, 405–414.

López-Pintado, S. & Romo, J. (2009). On the Concept of Depth for Functional Data *Journal of the American Statistical Association* **104**, 718–734.

Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics* **28**, 461–482.