Aurora

working to advance road weather information systems technology

Improving Estimates of Real-Time Traffic Speeds During Weather for Winter Maintenance Performance Measurements

Project Sponsors
Iowa Department of Transportation
Aurora - Transportation Pooled Fund
Midwest Transportation Center
U.S. Department of Transportation
Office of the Assistant Secretary
for Research and Technology

Aurora Projects 2013-03 and 2015-03
Final Report • April 2017

MIDWEST TRANSPORTATION CENTER
MTC

IOWA STATE UNIVERSITY
Institute for Transportation

About Aurora

Aurora is an international program of collaborative research, development, and deployment in the field of road and weather information systems (RWIS), serving the interests and needs of public agencies. The Aurora vision is to deploy RWIS to integrate state-of-the-art road and weather forecasting technologies with coordinated, multi-agency weather monitoring infrastructures. It is hoped this will facilitate advanced road condition and weather monitoring and forecasting capabilities for efficient highway maintenance and real-time information to travelers.

About MTC

The Midwest Transportation Center (MTC) is a regional University Transportation Center (UTC) sponsored by the U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology (USDOT/OST-R). The mission of the UTC program is to advance U.S. technology and expertise in the many disciplines comprising transportation through the mechanisms of education, research, and technology transfer at university-based centers of excellence. Iowa State University, through its Institute for Transportation (InTrans), is the MTC lead institution.

About InTrans

The mission of the Institute for Transportation (InTrans) at Iowa State University is to develop and implement innovative methods, materials, and technologies for improving transportation efficiency, safety, reliability, and sustainability while improving the learning environment of students, faculty, and staff in transportation-related fields.

ISU Non-Discrimination Statement

Iowa State University does not discriminate on the basis of race, color, age, ethnicity, religion, national origin, pregnancy, sexual orientation, gender identity, genetic information, sex, marital status, disability, or status as a U.S. veteran. Inquiries regarding non-discrimination policies may be directed to Office of Equal Opportunity, 3410 Beardshear Hall, 515 Morrill Road, Ames, Iowa 50011, Tel. 515-294-7612, Hotline: 515-294-1222, email eooffice@ iastate.edu.

Notice

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The opinions, findings and conclusions expressed in this publication are those of the authors and not necessarily those of the sponsors.

This document is disseminated under the sponsorship of the U.S. DOT UTC program in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document. This report does not constitute a standard, specification, or regulation.

The U.S. Government does not endorse products or manufacturers. If trademarks or manufacturers' names appear in this report, it is only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. The FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Iowa DOT Statements

Federal and state laws prohibit employment and/or public accommodation discrimination on the basis of age, color, creed, disability, gender identity, national origin, pregnancy, race, religion, sex, sexual orientation or veteran's status. If you believe you have been discriminated against, please contact the Iowa Civil Rights Commission at 800-457-4416 or the Iowa Department of Transportation affirmative action officer. If you need accommodations because of a disability to access the Iowa Department of Transportation's services, contact the agency's affirmative action officer at 800-262-0003.

The preparation of this report was financed in part through funds provided by the Iowa Department of Transportation through its "Second Revised Agreement for the Management of Research Conducted by Iowa State University for the Iowa Department of Transportation" and its amendments.

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Iowa Department of Transportation or the U.S. Department of Transportation Federal Highway Administration.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
InTrans Project 13-485		

4. Title and Subtitle	5. Report Date
Improving Estimates of Real-Time Traffic Speeds During Weather Events for Winter Maintenance Performance Measurement	April 2017
	6. Performing Organization Code
7. Author(s)	**8. Performing Organization Report No.**
Vianey Leos Barajas (orcid.org/0000-0001-8016-773X), Zhonglei Wang (orcid.org/0000-0001-6712-0750), Mark Kaiser (orcid.org/0000-0003-0449-0388), and Zhengyuan Zhu (orcid.org/0000-0002-2266-0646)	InTrans Projects 13-485 and 16-156
9. Performing Organization Name and Address	**10. Work Unit No. (TRAIS)**
Institute for Transportation Iowa State University 2711 South Loop Drive, Suite 4700 Ames, IA 50010-8664	**11. Contract or Grant No.** Part of DTRT13-G-UTC37
12. Sponsoring Organization Name and Address	**13. Type of Report and Period Covered**
Midwest Transportation Center 2711 S. Loop Drive, Suite 4700 Ames, IA 50010-8664 Iowa Department of Transportation and Aurora Transportation Pooled Fund 800 Lincoln Way Ames, IA 50010 U.S. Department of Transportation Federal Highway Administration and Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE Washington, DC 20590	Final Report **14. Sponsoring Agency Code** Part of TPF-5(290) and SPR-3(042) Aurora Projects 2013-03 and 2015-03

15. Supplementary Notes

Visit www.intrans.iastate.edu and www.aurora-program.org for color pdfs of this and other research reports.

16. Abstract

This report describes two related projects, the second of which builds on the first.

In Part I, a model was developed to relate weather variables to traffic flow changes at a local level. Weather station data and maintenance crew reports were used to develop an empirical adaptive stochastic model using a Bayesian formulation. Data from early time segments provide a prior quantification of the expected effects of weather variables on traffic speed over subsequent time segments. Data in the next time segment are then used to adjust these quantifications to reflect observed traffic speeds during that period. Thus, rather than explicitly determining numerous temporally dependent interactions, the main effects associated with key factors are allowed to undergo small shifts over time to fit the data. The model incorporates an autoregressive error structure to reflect temporal dependencies in observations that occur at reasonably high frequencies.

In Part II, INRIX and Wavetronix traffic data and limited weather information were used to develop models for detecting abnormal traffic patterns and predicting traffic speed and volume at any location on a network. Multivariate quantiles were estimated for the INRIX observations, and the INRIX data were compared with the estimated quantiles to identify abnormal traffic patterns in both space and time. An online interactive app was developed to visualize the results and inform decisions about winter maintenance. A dynamic Bayesian model was implemented at two Wavetronix sensor locations where weather information was available, with the corresponding median curve as the baseline. The fitting results were satisfactory. The INRIX data's spatial structure was explored, and curve Kriging was used to predict traffic speed and volume at any location. The prediction method worked well.

17. Key Words	18. Distribution Statement
highway maintenance operations—performance measurement—real-time traffic speed predictions—resource utilization—rural traffic speeds—urban traffic speeds—winter highway maintenance	No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	57	NA

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Improving Estimates of Real-Time Traffic Speeds During Weather Events for Winter Maintenance Performance Measurements

Final Report
April 2017

Principal Investigator
Zhengyuan Zhu
Associate Professor, Department of Statistics, Iowa State University

Co-Principal Investigator
Mark Kaiser
Professor, Department of Statistics, Iowa State University

Research Assistants
Vianey Leos Barajas and Zhonglei Wang

Authors
Vianey Leos Barajas, Zhonglei Wang, Mark Kaiser, and Zhengyuan Zhu

Sponsored by
Iowa Department of Transportation,
Federal Highway Administration Aurora - Transportation Pooled Fund,
Midwest Transportation Center, and
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology

Preparation of this report was financed in part
through funds provided by the Iowa Department of Transportation
through its Research Management Agreement with the
Institute for Transportation
(InTrans Projects 13-485 and 16-156)

A report from
Institute for Transportation
Iowa State University
2711 South Loop Drive, Suite 4700
Ames, IA 50010-8664
Phone: 515-294-8103 / Fax: 515-294-0467
www.intrans.iastate.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

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.

The authors would also like to thank Tina Greenfield with the Iowa DOT for her valuable input and advice throughout these projects.

EXECUTIVE SUMMARY

This report describes two related projects, the second of which builds on the first.

The Iowa Department of Transportation (DOT) and Federal Highway Administration (FHWA) Aurora Transportation Pooled Fund TPF-5(290) partners sponsored both projects, and the Institute for Transportation Center's Midwest Transportation Center at Iowa State University and U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology co-sponsored Part II.

Part I: Improving Estimations of Real-Time Traffic Speeds During Weather for Winter Performance Measurement

Winter weather in Iowa is often unpredictable and can have an adverse impact on traffic flow. The Iowa DOT attempts to lessen the impact of winter weather events on traffic speeds using various proactive maintenance operations. In order to assess the performance of these maintenance operations, a model was developed for estimating expected speed reductions based on weather variables and normal maintenance schedules. Such a model would allow the Iowa DOT to identify situations in which speed reductions are much greater or less than expected for a given set of storm conditions and modify maintenance operations to improve efficiency and effectiveness.

For a previous Iowa DOT project, the authors of this report developed a sequential Bayesian dynamic model based on a non-stochastic model proposed in Iowa Highway Research Board Project TR-491. The sequential Bayesian dynamic model was designed to predict speed changes relative to baseline speeds under normal conditions. The model assumes nominal maintenance schedules and incorporates winter weather covariates (snow type, temperature, and wind speed) as measured by roadside weather stations. A limitation of this model is that it is not flexible enough to accommodate temporal heterogeneity. The speed change predictions are inconsistent across different locations and events, and measures of uncertainty in the model do not account for model uncertainty and are therefore overly optimistic.

The objective of the work performed for the first part of this project was to improve that model to achieve real-time prediction of traffic speed changes with realistic uncertainty measures. Automated systems that record information on traffic speed and road conditions at particular points on a road system offer a way to improve the information available for building empirical models that relate weather variables to changes in traffic flow at a local level. For this application, we used data from two sources: road weather information systems (RWIS) and automated weather observing systems (AWOS). We further used maintenance crew reports to identify winter weather events throughout the year when reliable information on precipitation, such as snow intensity and type, was not available. Atmospheric variables, along with variables reflecting physical pavement condition, may contain a large number of complex interactions upon which realized changes in traffic speed depend, and these interactions vary with time.

To accommodate these types of interactions and temporal dynamics, we developed an empirical adaptive stochastic model. Our approach made use of a Bayesian model formulation in which the effects of weather variables are allowed to adapt over four-hour time segments. Data from prior four-hour time segments provide a prior quantification of the effects that variables such as lane condition, temperature, and wind are expected to have on changes in traffic speed over the next four hours. Additional data in the next time segment are then used to adjust these quantifications to better reflect observed traffic speeds during that period. This model would allow, for example, the effects of wind speed to change somewhat (but not radically) over the course of a longer storm event to reflect the fact that low or high wind speeds have different effects on traffic speeds.

The effect of this modeling approach is to circumvent the impossible task of explicitly determining a plethora of temporally dependent interactions. Under this approach, interactions are not explicitly identified and modeled. Rather, the problem is approached by allowing the main effects associated with key factors to undergo small shifts over time. The model also incorporates an autoregressive error structure to reflect temporal dependencies in observations that occur at a reasonably high frequency, such as every few minutes.

Part II: Data-Driven Urban Traffic Prediction for Winter Performance Measurements

The model developed in Part I is for data from a single site, which is useful for predicting traffic speed changes in rural areas. In an urban setting, multiple sites collect traffic data from a network of roads, and the data are typically correlated in both time and space. Modeling data from multiple sites jointly can help in detecting abnormal traffic patterns earlier and in more accurately predicting traffic speed changes.

The objective of Part II of this report was to use traffic data and limited weather information to develop models for detecting abnormal traffic patterns and predicting traffic speed and volume at any location on a network.

We introduced several models for detecting abnormal traffic patterns based on two sources of traffic data, INRIX and Wavetronix, and limited weather information. Because the INRIX data included more locations than the Wavetronix data and the corresponding observations were self-consistent, we used the INRIX data to detect abnormal traffic areas. We developed a method to estimate the multivariate quantiles for the INRIX observations, and the INRIX data were compared with the estimated quantiles to identify abnormal traffic patterns in both space and time. An online interactive app was developed to visualize the results and help the Iowa DOT make informed decisions about winter weather maintenance.

A dynamic Bayesian model was implemented at two Wavetronix sensor locations where weather information was available, with the corresponding median curve as the baseline. The fitting results were satisfactory. Furthermore, we also explored the spatial structure of the traffic data using the INRIX dataset and used curve Kriging to predict traffic speed and volume at any location. The prediction method was tested at the Wavetronix locations and was found to work well.

PART I: IMPROVING ESTIMATIONS OF REAL-TIME TRAFFIC SPEEDS DURING WEATHER FOR WINTER PERFORMANCE MEASUREMENT

1.1 Introduction and Overview of Previous Work

Winter weather in Iowa is often unpredictable and can have a large impact on traffic flow. The Iowa Department of Transportation (DOT) attempts to lessen the impact of winter weather events on traffic speeds using various maintenance operations. In order to assess the performance of these maintenance operations, a model was developed for estimating expected speed reductions based on different winter weather variables and normal maintenance schedules. Such a model would allow the Iowa DOT to identify situations in which speed reductions are much greater than expected for a given set of storm conditions.

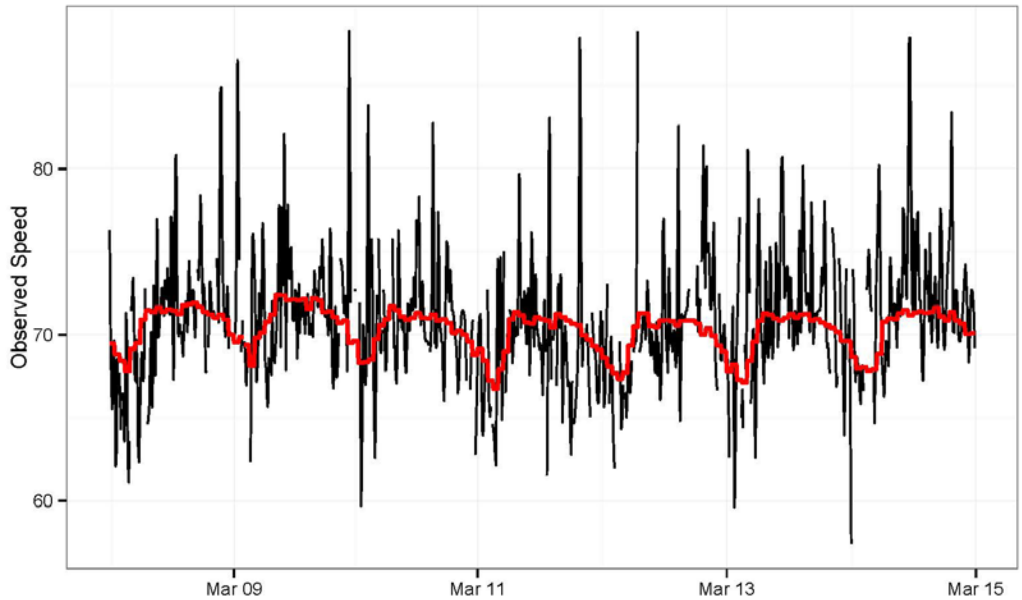
Overview of Previous Work

In 2009, Qiu and Nixon (2009) developed a model to predict speed reduction. This model was then modified in a study by Greenfield et al. (2012), which reviewed previously developed models and discussed modifications to introduce more variability into the model inputs. Although these modifications were an improvement, they failed to account for variability in the model structure itself. In Greenfield et al. (2012), an exploratory analysis of sensor data was used to help formulate a hierarchical model for estimating the effects of storm events at individual locations. This model is in the form of a dynamic linear regression and takes into account winter weather variables as well as plow operation schedules, which makes it different from previous models. In addition, the researchers use deviation from “normal,” or baseline, traffic speeds as the response variable instead of deviation from the posted speed limit. Temperature in the model is represented by the absolute deviation from freezing, as opposed to the raw temperature value. The model does a good job predicting the expected speed reduction and captures some of the variability inherent in the data.

Model for Speed Reduction

Traffic speeds are likely to be impacted by various factors, such as time of day or day of the week. In order to understand the impact of a winter storm on traffic speeds, we developed a baseline measure of traffic speed for each hour of each day. The baseline speeds represented the expected traffic speed for each hour of each day when conditions were favorable, i.e., when no winter events were recorded at the time speed was measured or in the previous 24 hours. We included three years of speed data to estimate the baseline speeds.

Figure 1.1 shows the observed traffic speeds for the week of March 8 through March 14, 2013, when conditions were favorable, along with the estimates of the baseline speeds.



Black lines = Observed traffic speeds; Red line = Estimated baseline speeds

Figure 1.1. Traffic speeds with favorable conditions for March 8 through March 14, 2013 versus estimated baseline speeds during the respective hours of each day

Using the baseline speeds, we defined the response variable as the difference between the observed speed and the baseline speed, $y_t = observed_t - baseline_t$. When the response variable is negative, the observed traffic speed is lower than the baseline speed, and vice versa. In the event of a winter storm, we typically observe traffic speeds consistently in the negative range.

Consider one storm event. Let $\{Y_t: t = 1,2, \ldots, T\}$ denote the recorded speed reduction from the baseline at time (t) during a given 10-minute time interval. The model for Y_t is as follows:

$$Y_t = h(\beta, x_t) + w_t$$

$$w_t = \gamma w_{t-1} + v_t$$

where $w_0 = 0$ and v_t are independent, identically distributed, random variables following a normal distribution with a mean of 0 and a variance of τ^2.

Let $X_{t,NoSnow}, X_{t,ModSnow}$and $X_{t,HeavySnow}$ be indicator variables that take on values of 0 or 1 for provided values of snow type, e.g., $X_{t,NoSnow} = 1$ if no snow is observed during a given time period and 0 if conditions are otherwise. We define

$$h^*(\beta, x_t) = \beta_o X_{t,NoSnow} + \beta_1 X_{t,|temp-32|} + \beta_2 X_{t,LightSnow} + \beta_3 X_{t,ModSnow} + \beta_4 X_{t,HeavySnow} + \beta_5 X_{t,wind}$$

and

$$h(\beta, x_t) = \begin{cases} h^*(\beta, x_t), & t \leq 4 \\ h^*(\beta, x_t) + \lambda(4 - t), & 4 < t \leq 10 \\ h^*(\beta, x_t) + \lambda(10 - t), & 10 < t \leq 16 \\ h^*(\beta, x_t) + \lambda(16 - t), & 16 < t \leq 22 \\ \vdots & \vdots \end{cases}$$

$h(\beta, x_t)$ is comprised of two components. The first component, $h^*(\beta, x_t)$, is a linear combination of weather covariates. The second component takes into account the impact of snowplows on speed reduction, represented here by the parameter λ. The values of 4, 10, 16, etc., were chosen to reflect the time points at which the snowplow would pass.

A Bayesian analysis was conducted using Markov chain Monte Carlo methods to simulate values from the joint posterior distributions. Prior distributions were placed on β_i ($i = 1,2, \ldots ,5$), λ, γ, and τ^2. The following diffuse priors were used:

$$\beta_i \sim N\left(0, \sigma_{\beta_i}^2 = 100\right)$$

$$\lambda \sim N\left(0, \sigma_{\lambda}^2 = 100\right)$$

$$\gamma \sim Unif(-1,1)\tau^2 \sim IG(0.01, 0.01),$$

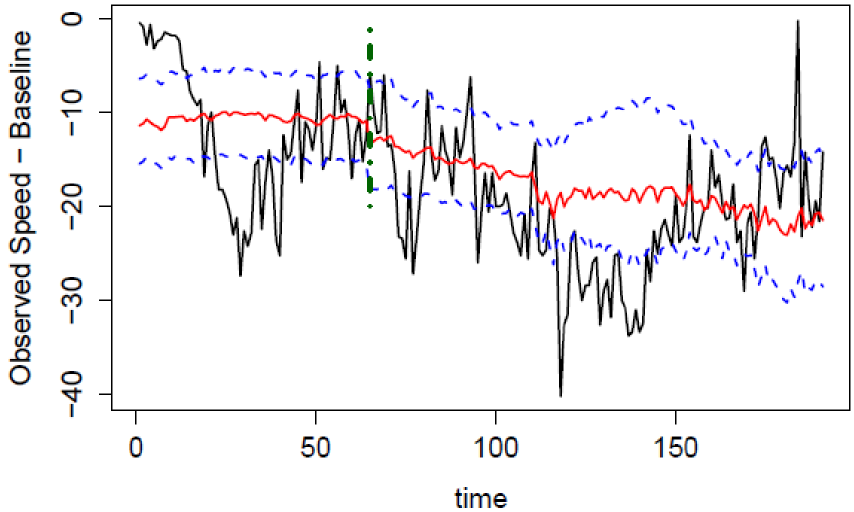
where $N(\mu, \sigma^2)$ denotes a normal distribution with parameters μ and σ^2, $Unif(a, b)$ denotes a uniform distribution on the interval (a, b), and $IG(\alpha, \beta)$ denotes an inverse Gaussian distribution with parameters α and β.

1.2 Evaluation of the Hierarchical Model

We applied the hierarchical model to traffic data from Newton/Colfax, Iowa, gathered during a number of winter events. We show three examples here to demonstrate the strengths and weaknesses of the model formulation.

Event 74 with Periods of Heavy and Medium Snow

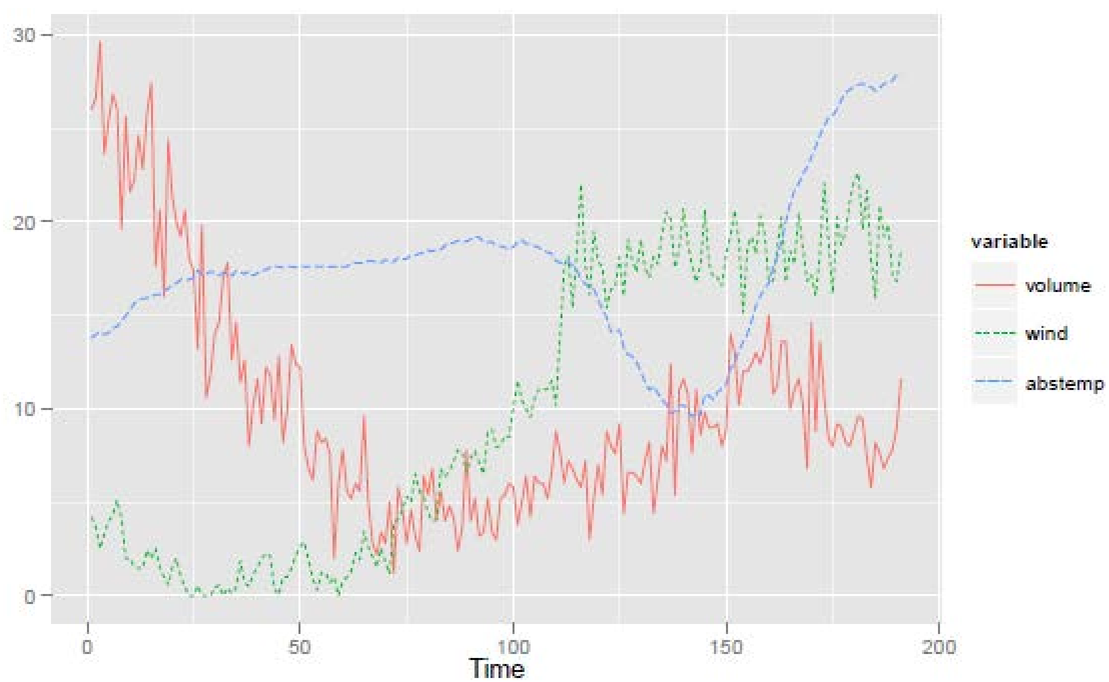
Event 74 (as designated in the data provided by the Iowa DOT) was characterized by periods of heavy snow and medium snow. Most of the observations are negative, indicating a visible reduction in traffic speed during this time. There is a sharp drop in speed during the initial period of the winter weather event, with further general fluctuations occurring over this time period, i.e., increases in traffic speeds followed by decreases, and so on (Figure 1.2).



Solid black line = Observed changes in traffic speeds from the baseline speeds;
Solid red line = Estimated changes in traffic speeds from baseline speeds;
Dashed blue lines = 90% credible intervals for the expected changes in speeds
Green vertical dashed-dotted line = Iowa DOT-designated change from heavy to medium snow

Figure 1.2. Changes in traffic speed compared to baseline speed for Event 74 with periods of heavy and medium snow

After fitting the model, the only variable that seems to significantly describe the change in traffic speed from the baseline is wind speed (Figure 1.3).



Orange/red solid line = Traffic volume; Green dotted line = Wind speeds;
Blue dashed line = Absolute temperature

Figure 1.3. Traffic volume, wind, and absolute temperature (deviations from 32°F) for Event 74 with periods of heavy and medium snow

No other variables have significant effects on traffic speeds, including the maintenance variable (Table 1.1).

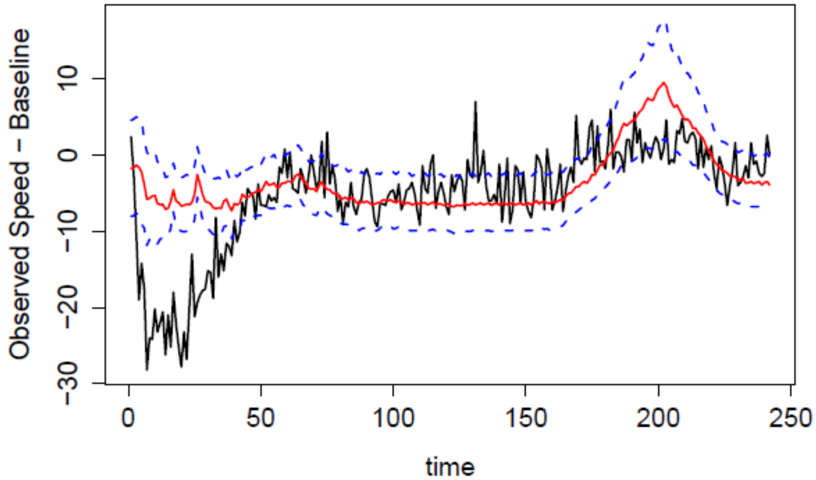
Table 1.1. Parameter estimates and 90% intervals for Event 74

Parameter	Estimate	90% Interval
β_0 (No Snow)	-0.02	(-16.40, 16.47)
β_1 (|temp - 32|)	-0.16	(-0.81, 0.37)
β_2 (Light Snow)	-0.08	(-16.5, 16.46)
β_3 (Moderate Snow)	-9.19	(-19.90, 4.20)
β_4 (Heavy Snow)	-7.40	(-16.68, 3.97)
β_5 (wind)	-0.43	(-0.73, -0.10)
λ	0.05	(-0.43, 0.57)
τ^2	23.25	(19.27, 27.93)
γ	0.78	(0.65, 0.91)

However, the fitted model does not capture the multiple increases and decreases in traffic speeds occurring over time, but rather seems to follow a somewhat decreasing linear trend overall.

Event 78 with Freezing Rain

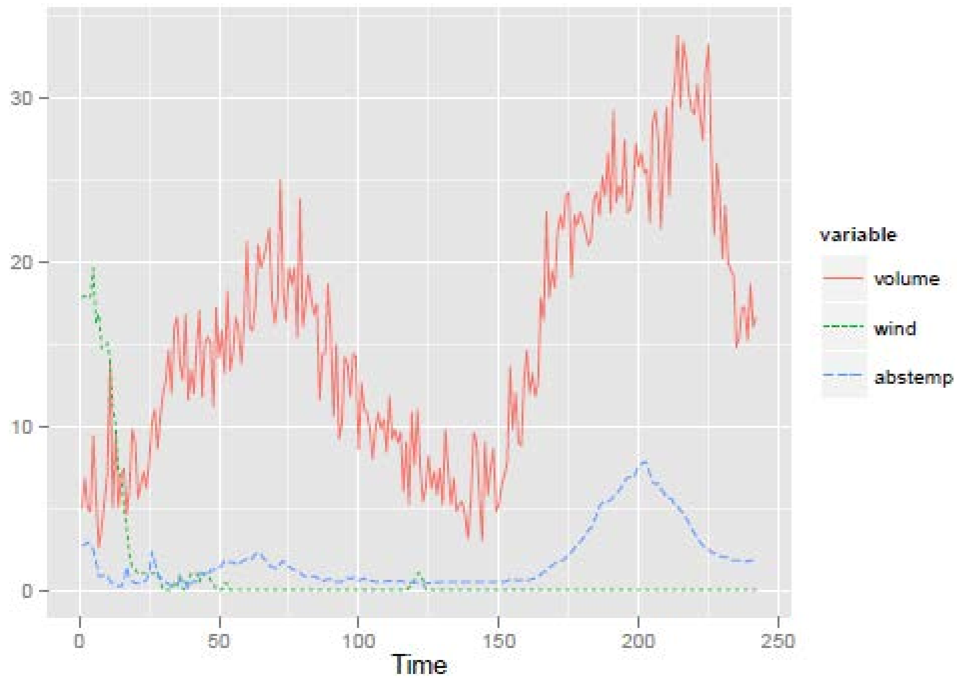
Event 78 was categorized by freezing rain. We observe a dramatic decrease in traffic speeds at the beginning of the winter weather event, followed by a sharp increase, and then little overall change in the subsequent traffic speeds observed throughout the remainder of the time period (240 10-minute intervals or 40 hours) (Figure 1.4).



Solid black line = Observed change in traffic speeds from the baseline speeds;
Solid red line = Estimated change in traffic speeds from baseline speeds;
Dashed blue lines = 90% credible intervals for the expected change in speeds

Figure 1.4. Changes in traffic speed compared to baseline speed for Event 78 with freezing rain

The only variables that seem to significantly describe the changes in speeds from the baseline levels are absolute deviation in temperature from 32°F and the indicator variable for no snow (Figure 1.5).



Orange/red solid line = Traffic volume; Green dotted line = Wind speeds;
Blue dashed line = Absolute temperature

Figure 1.5. Traffic volume, wind speed, and absolute temperature (deviations from 32°F) for Event 78 with freezing rain

The maintenance parameter is not significant for this event (Table 1.2).

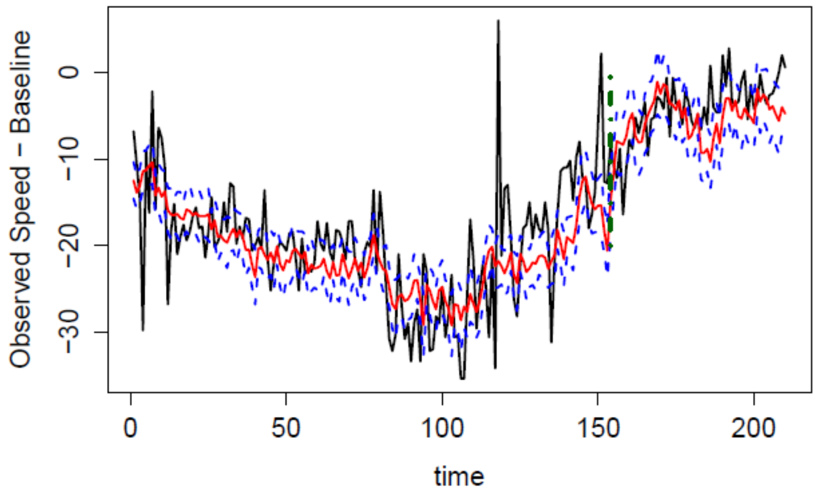
Table 1.2. Parameter estimates and 90% intervals for Event 78

Parameter	Estimate	90% Interval
β_0 (No Snow)	-7.75	(-11.40, -3.79)
β_1 (|temp - 32|)	2.16	(1.05, 3.33)
β_2 (Light Snow)	-0.01	(-16.48, 16.51)
β_3 (Moderate Snow)	0.08	(-16.40, 16.59)
β_4 (Heavy Snow)	0.00	(-16.45, 16.43)
β_5 (wind)	0.01	(-0.35, 0.38)
λ	0.13	(-0.17, 0.44)
τ^2	12.93	(11.02, 15.18)
γ	0.85	(0.78, 0.93)

The model for this event fits better than the model for Event 74, although we can distinctly see that this model picks up the behavior of the temperature variable and therefore overestimates certain areas while underestimating others.

Event 114 with Periods of Heavy and Blowing Snow

Event 114 was categorized by periods of heavy and blowing snow. Reductions in observed speeds from the baseline slowly increase at the beginning of the event followed by a gradual decrease to zero (Figure 1.6).



Solid black line = Observed change in traffic speeds from the baseline speeds;
Solid red line = Estimated change in traffic speeds from the baseline speeds;
Dashed blue lines = 90% credible intervals for the expected change in speeds

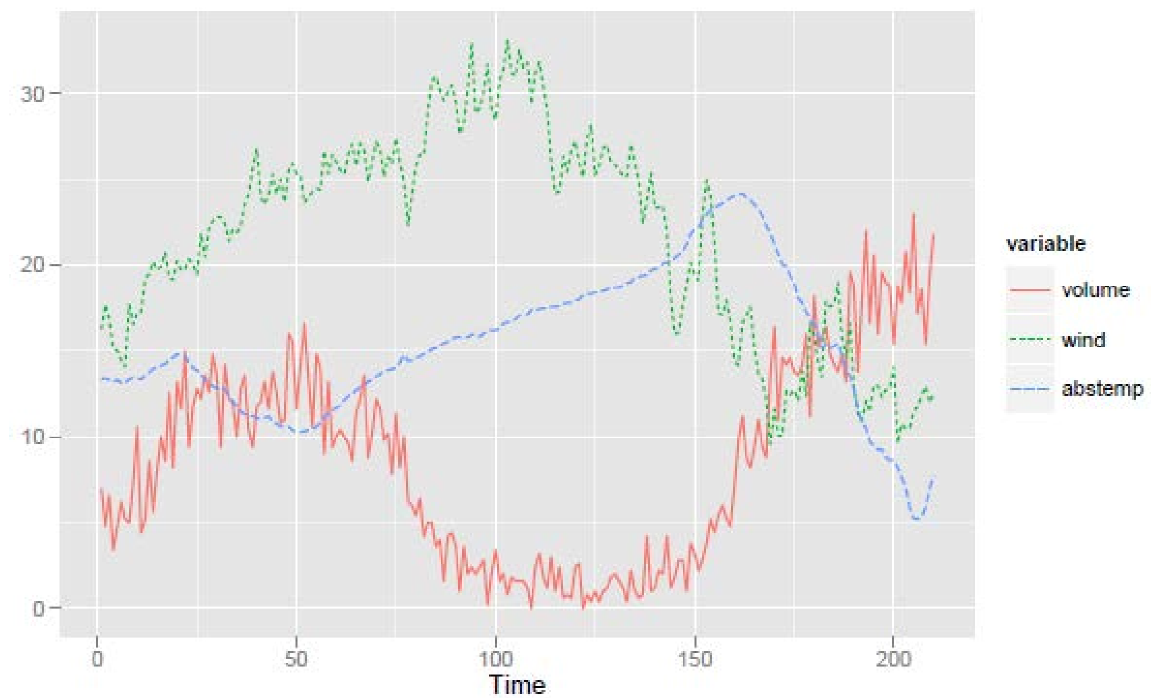
Figure 1.6. Changes in traffic speed compared to baseline speed for Event 114 with periods of heavy and blowing snow

We do not observe many fluctuations similar to the number of general increases and decreases that are present in Events 74 and 78. The model adequately describes the trend in changes in speed from the baseline levels in terms of capturing the mean deviations from the baseline. The intercept term and the parameter for wind are significantly different from zero in this model (Table 1.3).

Table 1.3. Parameter estimates and 90% intervals for Event 114

Parameter	Estimate	90% Interval
β_0 (No Snow)	7.21	(2.91, 11.39)
β_1 (|temp - 32|)	0.08	(-0.12, 0.30)
β_2 (Light Snow)	0.01	(-16.34, 16.37)
β_3 (Moderate Snow)	0.02	(-16.53, 16.59)
β_4 (Heavy Snow)	2.01	(-3.37, 7.36)
β_5 (wind)	-0.96	(-1.16, -0.75)
λ	-0.29	(-0.85, 0.28)
τ^2	25.53	(20.86, 31.66)
γ	0.34	(0.20, 0.53)

In particular, wind seems to be driving the pattern we observe in the deviations in traffic speeds from the baseline levels (Figure 1.7).



Orange/red solid line = Traffic volume; Green dotted line = Wind speeds;
Blue dashed line = Absolute temperature

Figure 1.7. Traffic volume, wind speed, and absolute temperature (deviations from 32°F) for Event 114 with periods of heavy and blowing snow

Comparing Event 74, Event 78, and Event 114, we see that the model can work well, but its performance is not consistent. That is, while the model as a function of type of snow variable, wind speed, and temperature may adequately describe the observed deviations in traffic speeds from the baseline for some event, it is not general or flexible enough to capture all patterns observed.

Issues with the Type of Snow Variable

During the application of the model to traffic data, there seemed to be a problem with the formulation for the "type of snow" variable, where "type of snow" was provided as either no snow, medium snow, or heavy snow:

- Events tended to be categorized as having only one type of snow. This formulation is the same as only having one intercept. For events that had more than one type of snow reported, there were very few (one or two) changes between snow type categories. This did not seem to truly capture the behavior of the storm.

- The lack of a more informative "type of snow" variable led to inconsistent model fitting performance because the other variables, wind and temperature, were not sufficient to explain the overall pattern.

The general formulation of the hierarchical model also included a maintenance parameter. The snowplow was scheduled to plow the roads every 30 minutes. The assumption was that speeds improve when the roads are recently plowed and decrease until the next time the roads are plowed. The maintenance parameter was included as a series of indicator functions, where the maintenance parameter had a linear effect for each 30-minute span. This formulation was used to capture the so-called sawtooth effect that is expected to be seen in traffic speed data.

The maintenance parameter was not significant in any of the fitted models. We may not have had enough time to truly see the effects of the plow, or it could be unrealistic to see such a dramatic pattern within a 30-minute span. Additionally, the maintenance parameter was included using the assumption that the plow would pass by exactly every 30 minutes. If the plows did not adhere to this schedule, the maintenance parameter would not have adequately described the deviations in traffic speeds from the baseline. However, because we did not observe a sawtooth pattern, we suspect that the maintenance effect did not manifest itself as expected.

After fitting the model to multiple datasets, we found that both the manner in which the "type of snow" variable was provided and the formulation of the maintenance parameter were inadequate.

1.3 Modification of the Model

In modifying the model, we removed the maintenance effect parameter and type of snow formulation. In order to denote winter weather events, we used maintenance crew reports to obtain the dates and times when there was a winter weather event. Although we did not explicitly include a term in the model for type of snow, we at least knew that there should be some effect of winter weather events on traffic speeds.

We received data from road weather information systems (RWIS) and obtained visibility data from automated weather observing systems (AWOS). While most of the new information was provided in 10-minute intervals, there were some inconsistencies in the time duration between observations. We used a linear imputation approach to provide the missing data in order to have an observation every 10 minutes, unless the gap between observations was more than 30 minutes. With the new data, we developed five candidate models to supplement the previous hierarchical model. These new models resulted from redefinitions of the function $h(\beta, x_t)$ used in the model described in the previous section. We were able to obtain data for two new variables: precipitation type (snow, rain, none) and lane condition (wet, ice, dry). While precipitation type and lane condition initially contained more than three categories each, we combined the information into the three categories listed above in order to have enough data for each type of event.

Using this information about the new models, the covariates are defined as follows:

- $X_{t,|temp-32|}$ is the absolute deviation of the temperature from 32°F
- $X_{t,wind}$ is the observed wind speed
- $X_{t,visib}$ is the visibility (in miles) as reported by the nearest AWOS station in Newton, Iowa
- $X_{t,rain}$, $X_{t,snow}$, and $X_{t,none}$ are weather condition indicator variables taking on the value of 1 if the weather condition was observed during the given time period and 0 if otherwise
- $X_{t,wet}$, $X_{t,ice}$, and $X_{t,dry}$ are pavement condition indictor variables taking on the value of 1 if the pavement was observed to have that condition and 0 if otherwise

Using these covariates, the following forms of $h(\boldsymbol{\beta}, \boldsymbol{x_t})$ were considered:

$$h(\beta, x_t) = \beta_1 X_{t,|temp-32|} + \beta_2 X_{t,wind} + \beta_3 X_{t,visib}$$

$$h(\beta, x_t) = \beta_1 X_{t,|temp-32|} + \beta_2 X_{t,wind} + \beta_4 X_{t,snow} + \beta_5 X_{t,rain} + \beta_6 X_{t,none}$$

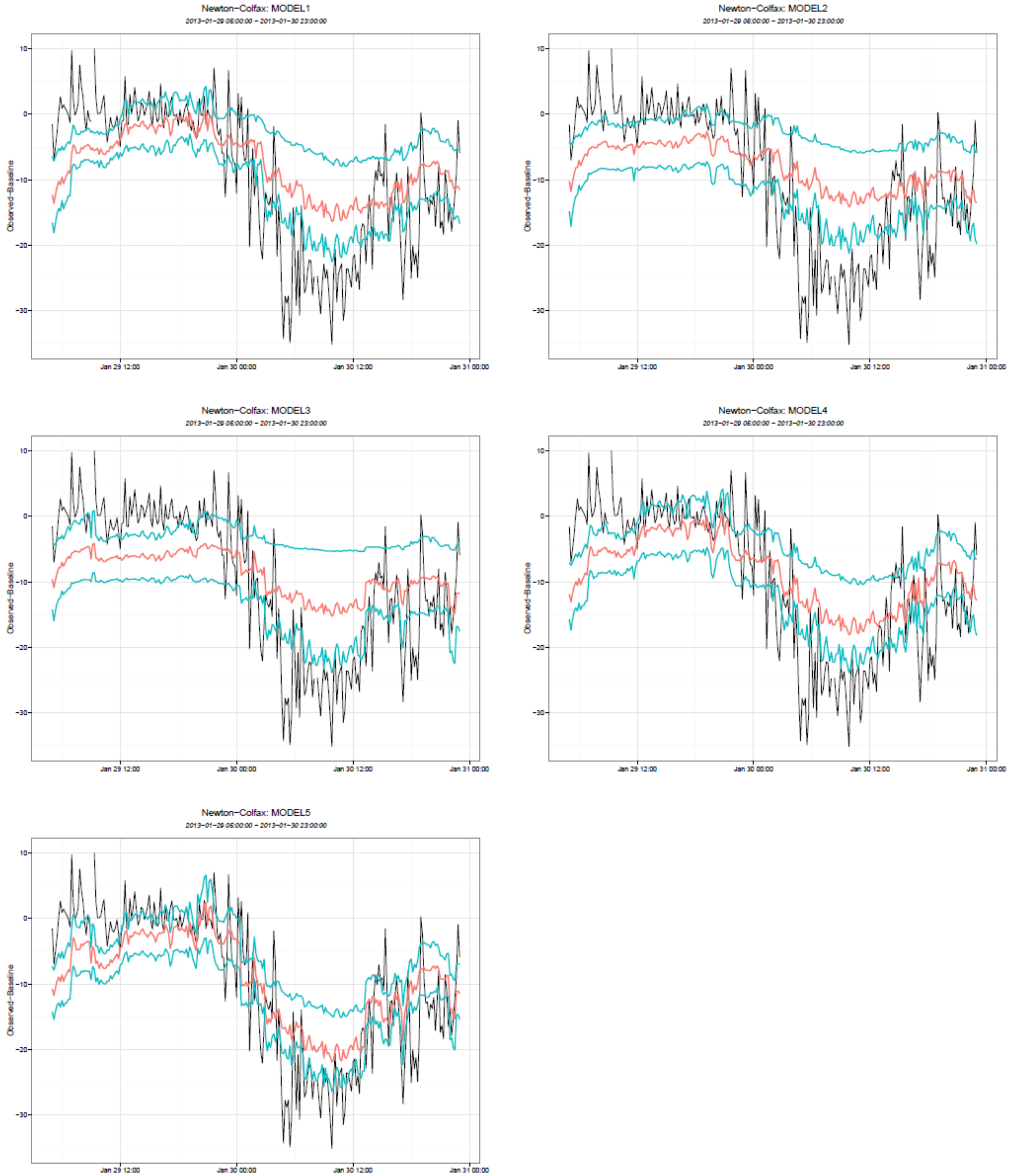
$$h(\beta, x_t) = \beta_1 X_{t,|temp-32|} + \beta_2 X_{t,wind} + \beta_7 X_{t,wet} + \beta_8 X_{t,ice} + \beta_9 X_{t,dry}$$

$$h(\beta, x_t) = \beta_1 X_{t,|temp-32|} + \beta_2 X_{t,wind} + \beta_3 X_{visib} + \beta_4 X_{t,snow} + \beta_5 X_{t,rain} + \beta_6 X_{t,none}$$

$$h(\beta, x_t) = \beta_1 X_{t,|temp-32|} + \beta_2 X_{t,wind} + \beta_3 X_{visib} + \beta_7 X_{t,wet} + \beta_8 X_{t,ice} + \beta_9 X_{t,dry}$$

Using the same diffuse priors described above, we can simulate values from the joint posterior distributions for each of these five models using Markov chain Monte Carlo methods. We fit the models to multiple events and compared the performance of each model for each event.

Figure 1.8 presents the results of fitting the five models to one storm event.



Black = Observed changes in traffic speeds; Orange = Estimated expected changes in speeds;
Teal/aqua = 90% credible intervals for expected changes in speeds

Figure 1.8. Hierarchical model under the five candidate mean structures

If the five models listed above are numbered 1 through 5, the first row of Figure 1.8 corresponds to Models 1 (left) and 2 (right), the second row to Models 3 (left) and 4 (right), and the final row to Model 5. The best fit is provided by Model 5, which contains covariates of temperature (departure from freezing), wind, visibility, and the three lane conditions (wet, ice, and dry). Model 5 was generally the best fit to various storm events at a number of locations, so the results depicted in Figure 1.8 are fairly general as an overall indication of the model's performance.

The model developed in this section incorporating temperature, wind, visibility, and lane condition improved the fit of the models compared to the initial hierarchical model, but there were still inconsistencies in performance. The effects of the covariates included in the improved models do not seem to affect traffic speeds in exactly the same manner over the entire course of a storm event. For instance, during a given 24-hour period defined by blowing snow, wind speed can be quite variable and snow may fall at different rates, and the effects that these variables have on traffic speed can differ depending on the overall conditions. In order to account for the possible heterogeneous effects of the covariates on observed traffic speeds, we modified the hierarchical model as described in the following section.

1.4 Adaptive Bayesian Model Formulation

One strength of the Bayesian approach to model development that we employed in this project is that it lends itself to sequential updating. If data are accumulated over successive time periods, the posterior from one period (reflecting what we believe after observing the data from that period) becomes the prior for the next period (reflecting what we believe before seeing the data from the new period). A sequential model fitting process thus consists of a sequence of prior-to-posterior transitions in which the posterior for one transition becomes the prior for the next. To put this concept into practice in modeling traffic speeds, we partitioned the data into four-hour periods. In this manner, we used sequential modeling to account for the fact that covariates may impact traffic speeds differently throughout the course of the day (i.e., the model can vary across four-hour periods). The modeling assumption is that the effects of the covariates remain constant within each four-hour period, but those effects may vary across time periods. The selection of a four-hour period was somewhat arbitrary and was driven in part by the desire to ensure that sufficient data were available within each time period for estimating the number of parameters included in the model.

Posterior to Prior Discounting

In the process of turning the posterior from one period into the prior for the next, it is beneficial to allow the model to reflect an increased level of uncertainty at the start of the new data period. This is accomplished by the process of discounting, which increases the variance of a posterior before it is used as the next prior. Discounting typically involves increasing the variance using a fixed discounting factor, which, although arbitrary, is treated as a tuning parameter in the model and subject to sensitivity analysis. We used a discounting factor of 0.01 to update the variances for the prior distributions of β across all segments of the data to fit the model. We assessed the fit under various values and selected 0.01 because this value provided flexibility while still allowing the prior means and variances to inform the estimates in a manner we deemed suitable.

Although this value is small, there is considerable difference between using a small value of 0.01 and estimating each segment of the data using diffuse priors.

Pavement Sensor Data Thresholding

The manner in which the traffic sensors record lane condition makes it possible to assign multiple categories in a short span of time. For example, there may be intermittent periods of dry recordings during a snow storm for various reasons. However, it is likely that the roads are not, in fact, completely dry each time a sensor records a dry value, particularly when those values are embedded within a larger string of wet or icy values.

In order to manage the effect of a few data points on the updating of priors, as detailed in the following section, we chose to impose a threshold on the number of data points required for each lane condition category to update the prior distribution. That is, if only a very small number of observations were connected with one or two of the wet, icy, or dry categories, we did not update the prior distribution. Conversely, if a large majority of the lane conditions reported fell into only one category, we adjusted this period to only reflect that particular category. The goal was to not allow a small number of observations for wet, icy, or dry conditions to impact the estimates drastically, which would thus improve the stability of the estimation in terms of the effect of lane conditions on changes in traffic speed.

For the data thresholding, we tested whether to exclude changes in lane condition if there were only one, two, or three successive observations in a category. We decided to disallow changes in lane condition if a run of the same values had a length less than or equal to three observations across all events and all periods, which reflected a minimum of 12.5% of the data for each period.

Each four-hour period had the same exact structure as the hierarchical model given so far. The one exception involved the prior distributions for the β parameters. We had initially selected diffuse normal prior distributions, whereas after the thresholding we only used diffuse prior distributions for the first four-hour period. We subsequently updated the prior distributions for the second through last four-hour periods for β. This updating allowed the covariates to have different effects on traffic speeds throughout the day(s), yet we ensured that the changes were not drastic, and evolved gradually, through the use of the prior distributions.

The thresholding algorithm was applied as follows:

1. Segment the dataset into four-hour periods, each with a minimum length imposed.

2. For each segment, if there are less than three observations for dry, wet, or icy conditions, change the values of all covariates to zero. Alternatively, if more than 87.5% of the observations for lane condition fall into one category, adjust all covariate values to one for that category.

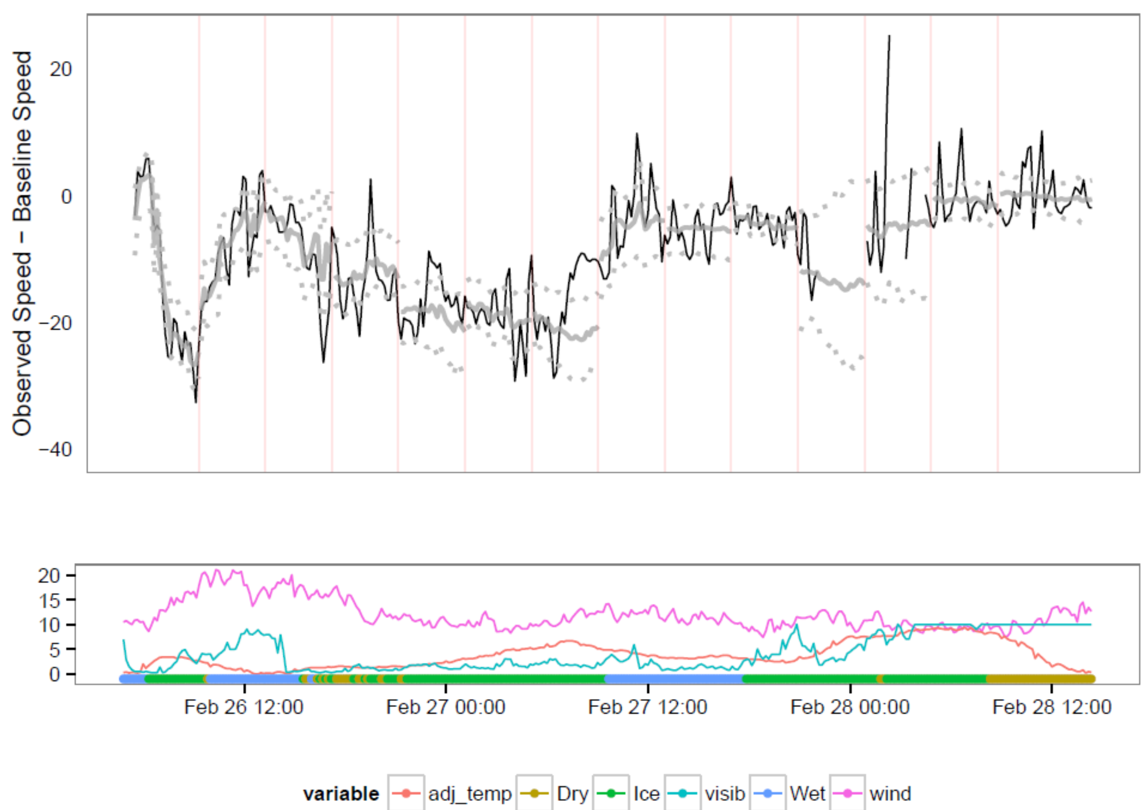
3. For the first time period, use diffuse priors for all parameters. Update prior means and variances for β for all subsequent periods, including a discount factor for the variance term.

4. Use the posterior draws from the n^{th} segment to update the prior means and variances of the $(n+1)^{th}$ segment, until n = the final four-hour period.

We illustrate the use of this adaptive version of the model with two winter storm events, one occurring in 2013 and the other in 2012.

Event 1 with Wet Snow and Blowing Snow

We selected a winter storm that occurred in Newton, Iowa in February 2013. Based on a crew report from a member of the winter storm maintenance crew active during this event, this particular event began on February 26, 2013 at 04:50 and ended on February 28, 2013 at 14:20. The event was categorized by periods of wet snow and blowing snow. The changes in speed for this event are shown in Figure 1.9, along with the fitted model results (gray curve) for each four-hour segment and 90% credible intervals (dotted curves) for the expected change in baseline traffic speeds.



Top: Black lines = Changes in speed; Gray lines = Fitted model results; Gray dots = 90% credible intervals

Figure 1.9. Fitted model for Event 1 with wet snow and blowing snow, along with covariate values observed over the course of the event

The lower panel of Figure 1.9 displays the observed wind speeds, visibility, deviation in temperature from 32°F, and lane condition. Traffic speeds are clearly affected by the winter storm, as evidenced by the large, consistently negative values in the deviations from the baseline levels (which are represented by 0 on the vertical axis of Figure 1.9). The largest decreases in speeds occurred a couple of hours after the storm began and from about 13:00 on February 26 to about 12:00 on February 27. Until February 28, there was wet snow combined with blowing snow, which perhaps more greatly impacted traffic speeds. Throughout February 28, blowing snow was observed, and speeds seemingly returned to normal. However, these classifications are based purely on observations by crew members and can vary slightly from person to person.

There is one particular four-hour period for this event in the first hours of February 28 in which there is a very large decrease in estimated traffic speeds and a large discontinuity between the previous and subsequent estimates of the expected changes in baseline levels. However, there is quite a bit of missing information in this interval, and the resulting uncertainty is reflected in the wide 90% credible intervals. In addition, the large decrease in the expected deviation from the baseline for this one time period did not affect the performance of the model in the subsequent period when speeds increased.

Overall, Figure 1.9 illustrates that the adaptive nature of the model depends on the amount and consistency of data across time periods. In the earlier time period of this event, shown in the left-hand portion of the plot, transitions between time periods caused mostly minor adjustments to the estimated expected speed changes. One would infer that the effects of the covariates included in the model were fairly stable over this period (and that the storm event was meteorologically consistent over this stretch of time). From somewhere around the midpoint onwards there are a number of sharper change points in the estimated expected values, reflecting either relatively sudden changes in speeds (e.g., around midday on February 27) or discontinuities in the data record, as described above.

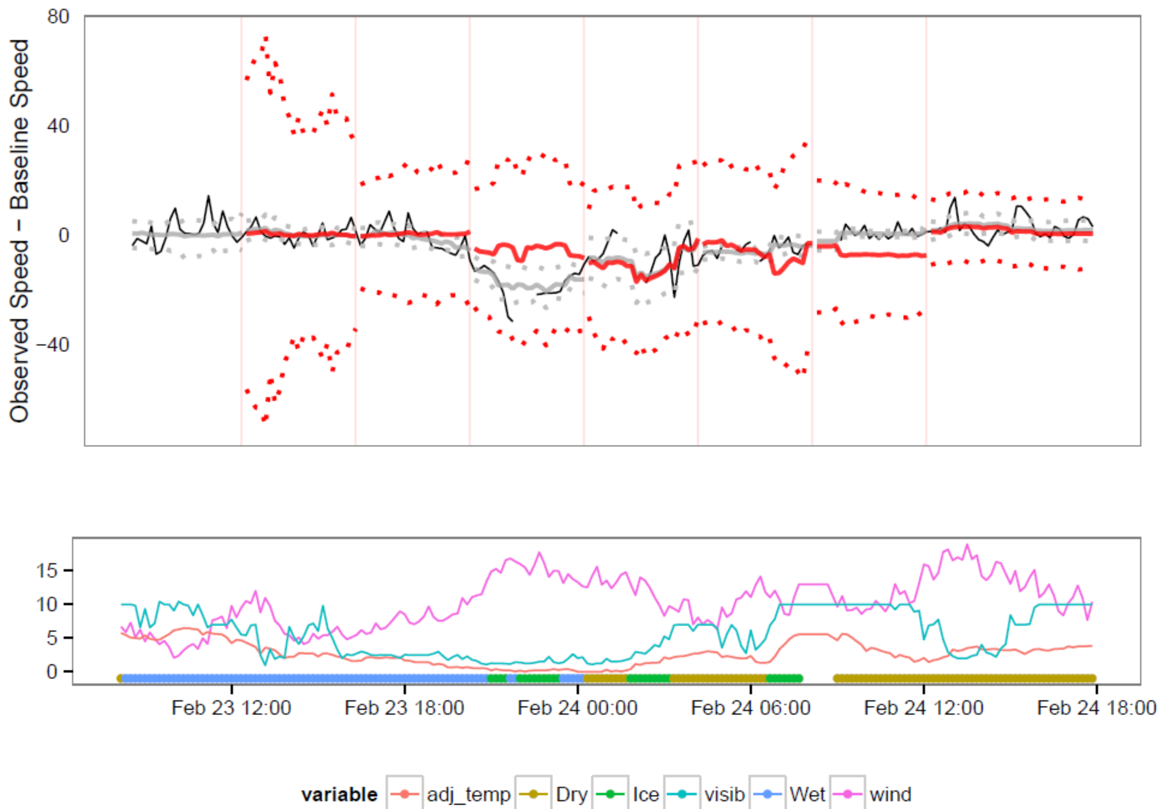
Short-Term Forecasting

As we developed a model that could adequately describe the changes in speeds from a baseline level, we used this model structure to forecast the behavior of traffic for each segment, given the values of the covariates, beginning with the second four-hour segment. We fit the model to the first segment and used the estimates to update the priors for the second segment and produced forecast estimates and prediction intervals using the covariate values from the second segment. Once we observed the speeds from the second segment, we fit the model again as described above and produced forecast estimates and prediction intervals for the third segment. We continued the pattern of forecast and fitting the model until we had produced forecasts for each segment.

We used the forecasts to evaluate changes in the covariates’ effects from segment to segment and demonstrate the model’s potential use in assessing highway maintenance during winter storms. Each forecast represents what is expected to happen if the traffic speed model does not change from the previous period. If the forecast is higher than what is observed, then there is evidence that a change has occurred in highway conditions to impact traffic speeds negatively. If the forecast estimates are lower than what is observed, then there is evidence that conditions have improved and traffic speeds have increased more than expected. As shown by the second storm event, described in the following section, this discrepancy between the estimated and expected speeds could indicate that the intensity of the storm subsided or that highway conditions improved due to better maintenance.

Event 2 with Periods of Rain, Wet Snow, and Blowing Snow

To further illustrate the use of the model, we selected a second winter weather event, which began on February 23, 2012 at 08:10 and ended on February 24, 2012 at 17:50. This storm featured periods of rain, wet snow, and blowing snow. This event was shorter than Event 1 and included a transition period from rain to snow, which is particularly troublesome for highway maintenance because of the potential for the rain to wash away salt and other road treatments. The changes in speeds for this event are shown in Figure 1.10, along with the fitted model results for each four-hour time period, 90% credible intervals for the expected change in baseline traffic speeds, and the estimates and intervals for the forecasts.



Top: Black lines = Changes in speed; Gray lines = Fitted model results; Gray dots = 90% credible intervals;
Red line = Forecast speed change; Red dots = Forecast intervals

Figure 1.10. Fitted model for Event 2 with periods of rain, wet snow, and blowing snow, along with covariate values observed over the course of the event

Fitted (i.e., estimated) expected values in Figure 1.10 are shown as the solid gray curve, 90% credible intervals are indicated by gray dots, forecast speed changes are shown as the solid red curve, and forecast intervals are indicated by red dots.

In the fourth segment of the dataset, the fitted model reflects a large decrease in speeds during this time period, but the forecast shows that this is not in line with what was expected to occur based on information from the previous segments. The fourth and fifth time periods include missing values in addition to the large decrease in speeds. These missing values affect the uncertainty of the data, which is reflected in wider forecast intervals in subsequent intervals. However, because the sixth and seventh segments do not have any missing data and the observed values are not highly variable, the forecast intervals are much smaller by the final segment.

Overall, the fitted model estimated the expected change from the baseline well. The forecast estimates closely matched the estimates of the expected change from the baseline in most cases, with the exception of two segments. In the first exception, the segment saw a drastic change in traffic speeds that was not in line with the previous segment's behavior. In the other exception, the forecast estimated lower traffic speeds than what was observed, providing evidence that highway conditions improved more than expected during this segment. This improvement could be connected to a change in weather conditions and/or improvements in road conditions due to highway maintenance.

1.5 Conclusions

In this part of the report, we have provided an overview of models developed to describe and potentially forecast changes in traffic speed during winter weather events in Iowa. The final adaptive model structure shows promise as a general approach for location-specific modeling. The key results of this research include the following:

- Typical traffic speeds depend on location, day of week, and time of day. Assessing traffic speed patterns on a location-specific basis, rather than attempting to model speed itself, is important for assessing the effect of storm events on changes in speed from a typical or baseline measure.

- The effects of winter weather variables and existing road conditions on changes in traffic speed are variable in both space and time. Important factors appear to be a mix of global effects, such as temperature and visibility, and highly localized effects, such as the condition of the pavement at sensor locations.

- It is important to allow mathematical models to adapt to varying conditions over time. We have presented one reasonably straightforward way to achieve this goal that is appropriate for the data sources and frequencies available to us. The key to our approach is to model the dynamic main effects for a small set of primary covariates rather than to model the fixed interactions among a larger set of both primary and secondary covariates. It is our opinion that this latter approach is not possible given existing data collection programs.

- The goal of running models, such as those developed here, in real time for the purposes of obtaining instantaneous feedback and forecasts is potentially achievable. Reaching this goal will require additional technological improvements in data collection, transfer, and processing, as well as the development of more efficient computational algorithms than those used in this project.

PART II: DATA-DRIVEN URBAN TRAFFIC PREDICTION FOR WINTER MAINTENANCE PERFORMANCE MEASUREMENTS

2.1 Introduction

The model developed in Part I is for data from a single site, which is useful for predicting traffic speed changes in rural areas. In urban settings, multiple sites collect traffic data from a network of roads, and the data are typically correlated in both time and space. Modeling data from multiple sites jointly can help in detecting abnormal traffic patterns earlier and in more accurately predicting traffic speed changes.

The objective of Part II of this report was to use traffic data and limited weather information to develop models for detecting abnormal traffic patterns and predicting traffic speed and volume at any location.

We received two sets of data for this project: Wavetronix data and INRIX data. Both datasets were collected in 2013 and 2014 on I-35, I-80, US 65, and IA 5 in Des Moines, Iowa.

The analysis of the data was based on each location and day of the week, and a multivariate quantile estimator was used for extreme value detection and baseline construction. The estimation of multivariate quantiles (Babu and Rao 1988, Liu 1990, Abdous and Theodorescu 1992, Chaudhuri 1996, López-Pintado and Romo 2009) is nontrivial because there is no unique way to order multivariate data. The method proposed by Chaudhuri (1996) requires a directional vector, which is difficult to determine. The computation of data depth discussed by Liu (1990) and López-Pintado and Romo (2009) is expensive when the sample size is large.

To circumvent such difficulties, we used the method by Abdous and Theodorescu (1992) in this project; the asymptotic properties of this method could be derived easily. Using this method, we estimated the median curve for traffic speed and volume for each day of the week and for each location, and the median curve provided us with a baseline of normal traffic conditions during various times of day and days of the week.

We also used the 10th quantile curve to detect extreme deviations in traffic speed during four-hour periods based on the INRIX data; an online app was developed for this task and is described in this report. During winter weather events, this method of extreme value detection gives an idea of the extent of the storm's impact across the Des Moines area and provides a method to identity areas that are affected more than others (as measured by traffic speed and volume). Note that the Wavetronix locations are sparse, so we used the INRIX data to detect the areas affected by the weather events.

We then fit a hierarchical Bayesian model to the deviations in Wavetronix speed data from the median speed curve at two locations where weather information was available. Furthermore, we demonstrated how a curve kriging (Giraldo et al. 2011) approach can be used to predict the speed and volume at any location in between by combining the two datasets.

This part of the report is organized as follows. Section 2.2 describes the data sources and the preliminary analysis of the data. In Section 2.3, the multivariate quantile estimation method is introduced, and Section 2.4 discusses the implementation of the method to detect extreme traffic conditions. By using the estimated median as the baseline, a dynamic Bayesian model is applied in Section 2.5 to the Wavetronix data from two Des Moines locations. Section 2.6 briefly introduces the curve Kriging method. The conclusions of this project and a discussion are given in Section 2.7.

2.2 Data Sources and Preliminary Analysis of the Data

Data Sources

We received two sets of data for this project: Wavetronix data and INRIX data. Both datasets were collected in 2013 and 2014 on I-35 and I-80, US 65, and IA 5 in Des Moines, Iowa. The data collection locations are shown in Figure 2.1.

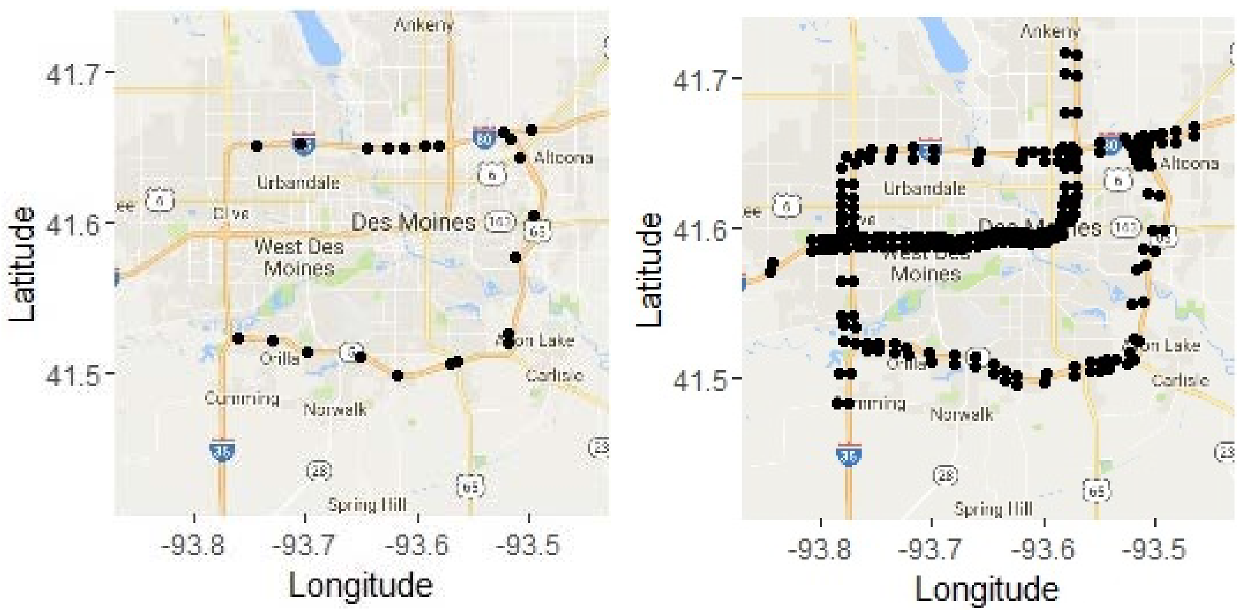


Figure 2.1. Des Moines metropolitan area with black points showing Wavetronix (left) and INRIX (right) data collection locations (with the INRIX data collection locations jittered)

The Wavetronix dataset included traffic speed and volume data collected at 5-minute intervals across 22 locations. The INRIX dataset was much denser spatially, and the corresponding temporal resolution was 2 minutes. Each location provided data from two lanes in either direction, including data collected from on/off-ramps. While the Wavetronix dataset was from Iowa DOT sensors, which provide high-quality readings for both speed and volume, the locations were sparse, and spatial dependence due to the long distances between sites was not evident; more details are provided below.

The INRIX speed dataset was not as consistently accurate as the Wavetronix dataset, but it was reasonable to assume that the INRIX dataset was self-consistent. The INRIX dataset did not include information about traffic volume.

Preliminary Analysis of the Data

We began the preliminary analysis by first examining the spatial correlation of the traffic speeds and volumes from the Wavetronix dataset across all available locations. To do this, we first selected a location and compared the observed traffic speeds at time n to the traffic speeds at time $n-1$, i.e., the observed traffic speeds were compared with those observed 5 minutes prior at every other location. A similar comparison was made for traffic volume. We expected that locations closer together would be more alike than those further apart. However, we did not find this. In general, there was no clear pattern in the data. This could have resulted from a number of factors, but we believe that the two main issues were the sparsity of the locations around the metropolitan area (in particular, the lack of data for certain highways) and the mismatch between the frequency at which the data were collected and the spatial distances among the sites.

The frequency at which the Wavetronix data were provided was every 5 minutes. While this may be considered a fine-scale resolution for some applications, it may not be the correct time scale to study spatial dependence, for which the distance that a vehicle can travel during this time period in different directions needs to be taken into account. During a 5-minute period, it is possible that a vehicle may pass multiple locations at which there is a sensor to record traffic speed and volume. This possibility may have affected the spatial dependence among locations at various distances from each other in the Wavetronix data.

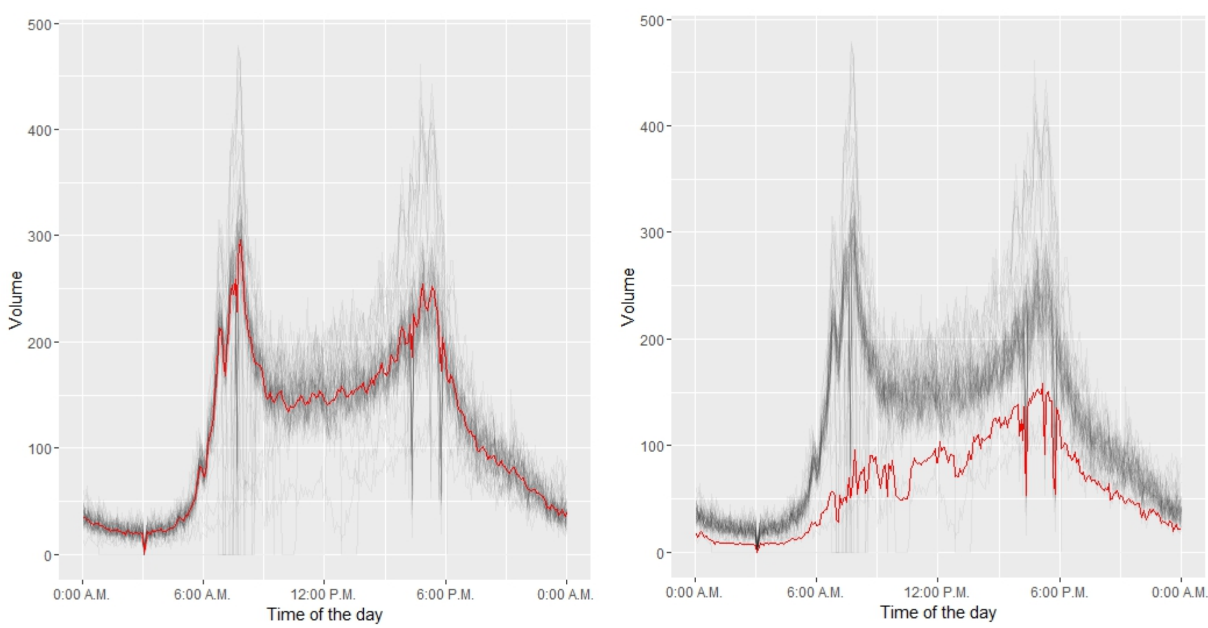
If the time scale at which data were collected had corresponded well to the distances between the traffic sensor locations, the data would exhibit strong spatial dependence, and it would be possible to make accurate predictions regarding future traffic speeds or volumes utilizing this dependence structure. Because this was not the case for this dataset, it was difficult to extract the necessary information from the Wavetronix data to understand how a slow-down or speed-up event at one location affected traffic speeds in the next few minutes at a nearby location.

We also thoroughly examined the INRIX data provided. The INRIX data were much denser in space, with one reading for each 1-mile segment, and the corresponding temporal resolution was 2 minutes. We first compared the Wavetronix data, which can be considered as the gold standard in terms of traffic speed accuracy, with the INRIX data at each location. The INRIX traffic speed data had high variability and did not always match the Wavetronix data, and frequently the recorded INRIX data were simply the posted speed limit in the area. The INRIX data included a score for the quality of data that was intended to be used to assess the data quality, although it is not clear how it was defined exactly. In general, the mismatch between the Wavetronix and INRIX data was too great for us to have confidence that the INRIX data at the point level provided actual traffic speeds.

However, the INRIX data were self-consistent and were collected at a finer time scale (every 2 minutes) and at many more locations than the Wavetronix data. The INRIX data therefore provided insight into the spatial structure of traffic speeds in the Des Moines area, which could be used in conjunction with the Wavetronix data for prediction; more details are provided in the section below on curve Kriging. While the traffic speeds reported were not always accurate at the point level, the data did reflect the general traffic pattern. For example, we saw a general pattern of lower traffic speeds reported by INRIX during winter storms, as we would expect. Using both sources of information, we could obtain a better picture of the impact of winter storms on traffic speeds across the Des Moines metropolitan area. In particular, we could identify when traffic speeds were greatly reduced and see which general locations were most affected. We discuss this further below in the section on extreme value detection.

2.3 Multivariate Quantile Estimation

Traffic speeds and volumes are likely to be impacted by numerous factors, such as time of day or day of the week. The morning rush, for instance, produces a spike in traffic volume, and speeds may be slower when traffic volume is too high. In general, for each location there tends to be similar traffic pattern on each day of the week, as seen in Figure 2.2 for traffic volume.



Red=Estimation of the corresponding quantile curves; Gray=2013 Thursday observations

Figure 2.2. Multivariate quantiles: median (left) and 10% (right) for Wavetronix volume data across all Thursdays in 2013 for location I-35/I-80 EB to Merle Hay Road WB

With this consistency in traffic patterns in mind, we constructed a baseline traffic pattern for both speed and volume for each location by day of the week. With a collection of traffic speeds and volume curves for a particular day of the week across the entire year, we estimated the median curve and used it as the baseline. Using the median rather than the mean provided a robust metric that was not influenced by outliers in traffic conditions, which was critical because we expected to observe when a winter weather event occurred.

To obtain the quantile curves, we used a methodology discussed by Abdous and Theodorescu (1992) that allowed us to compute the quantile curve of interest. This method generalizes the definition of the α^{th} quantile in the univariate case, and a Newton-Raphson procedure is implemented to estimate the quantile curve in the multivariate case. It can be shown (Kemperman 1987, Abdous and Theodorescu 1992) that, under mild conditions, there exists one unique theoretical quantile curve to which the estimated one converges.

We constructed the baseline (median) curves for both the Wavetronix (speed and volume) and INRIX (speed) data, as well as multiple quantile curves, for each location and day of the week. For the Wavetronix data, we applied the methodology to estimate the quantiles using the observed traffic speeds and volumes. For the INRIX data, we smoothed the observations by setting the weight to the quality score first, because some observations had a quality score of zero, indicating that the datum was not reliable. As such, we applied the method to the smoothed data.

To demonstrate the methodology, we used volume data from each Thursday in 2013 at sensor location I-35/80 EB to Merle Hay Road WB. The results of the estimated median and 10th quantile are shown in Figure 2.2. The median curve captures the overall trend structure of the speed and volume data for each location by day of the week. The results for the Wavetronix and INRIX data are similar.

A brief introduction to the function used to perform the multivariate quantile curve estimation is included in Appendix B.

2.4 Extreme Value Detection

In order to focus solely on how a winter storm affects traffic speeds, we initially examined the difference between the observed traffic speeds during a winter storm and the median curve for traffic speed for the specific location and day of the week, as described above. If traffic speeds are impacted by the winter storm, we expect that the response variable would be consistently negative.

For any given location and for a particular time of day and day of the week, we produced a distribution of differences between the observed traffic speeds (or volumes) and the median traffic speeds. If the resulting difference was zero, then the observed speed was exactly the same as the median speed, which indicated that the observed speed was indicative of normal traffic conditions. We expected that, in general, most of the differences would be around zero, with slightly different amounts of variation for each location.

In order to flag whether an observed speed was an extreme value (in the sense that it was lower than expected and very unlikely to be observed under normal traffic conditions), we compared it with the speed at that particular time of day from the 10% quantile curve. Then, we examined the results by four-hour periods as follows: 00:00 to 4:00, 4:01 to 8:00, etc. In this way, we could flag an entire four-hour period and visualize how the affected areas changed over time during a winter weather event. Because the INRIX dataset included more observations and locations than the Wavetronix dataset, we used the proportion of INRIX values that were lower than the corresponding 10% quantile curve for each four-hour period to detect abnormal traffic during a winter weather event.

An online app was developed using the Shiny package in R. The user interface is shown in Figure 2.3.

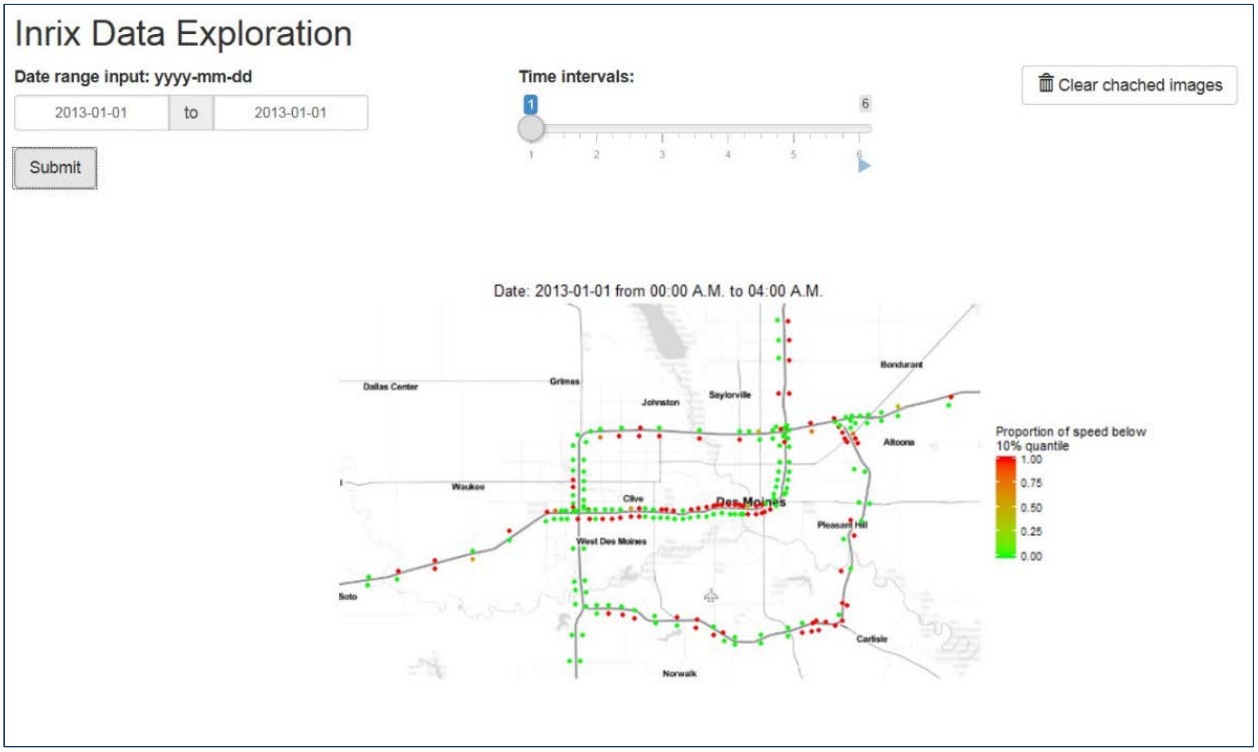


Figure 2.3. User interface of the extreme value detection app

Users can select the range of the dates they would like to examine. For example, if the range 2013-01-01 to 2013-01-31 is chosen, there should be 186 results because 31 days are selected and there are 6 results per day. Depending on the speed of the server, it may take several minutes to prepare the results, as shown in Figure 2.4.

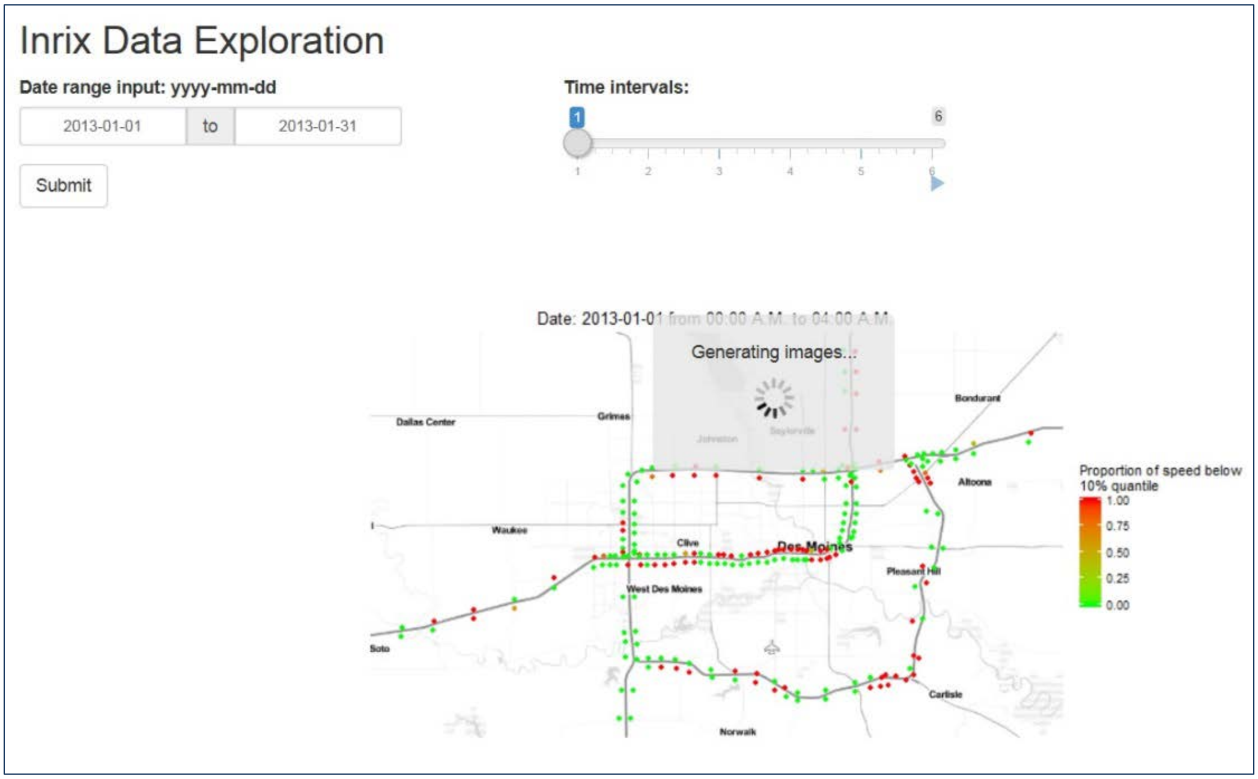


Figure 2.4. System preparing results for the user-selected dates, 2013-01-01 to 2013-01-31

When the results are ready, the "Generating images…" message disappears, and the user can click the play icon, i.e., the triangle under the "Time intervals" panel, to view the results. The procedure for starting this app is briefly introduced in Appendix C.

While there did not seem to be a strong spatial pattern evident in the normal traffic data, as described in Section 2.2, there was a spatial pattern in the decrease in traffic speeds induced by winter weather events. We demonstrate the results obtained for January 31, which were impacted by winter weather events, below. For illustration purposes, Figure 2.5 and Figure 2.6 show the detection results for the first two four-hour periods, i.e., from 00:00 a.m. to 08:00 a.m.

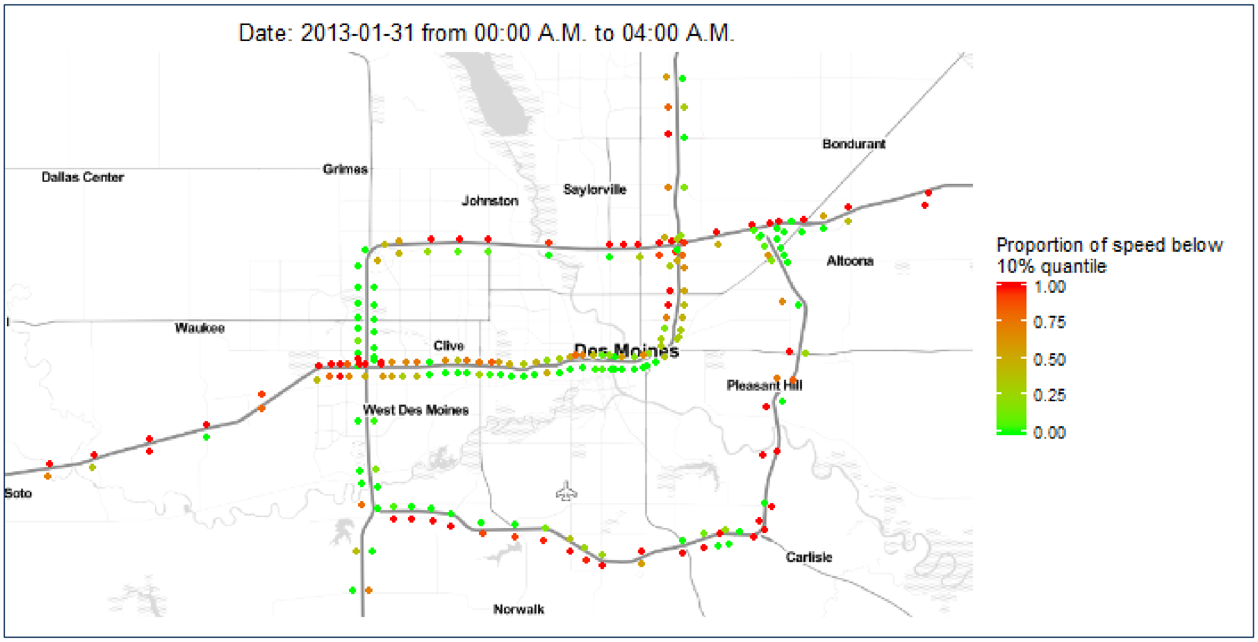


Figure 2.5. Locations that reported INRIX data from 00:00 a.m. to 04:00 a.m. on January 31, 2013

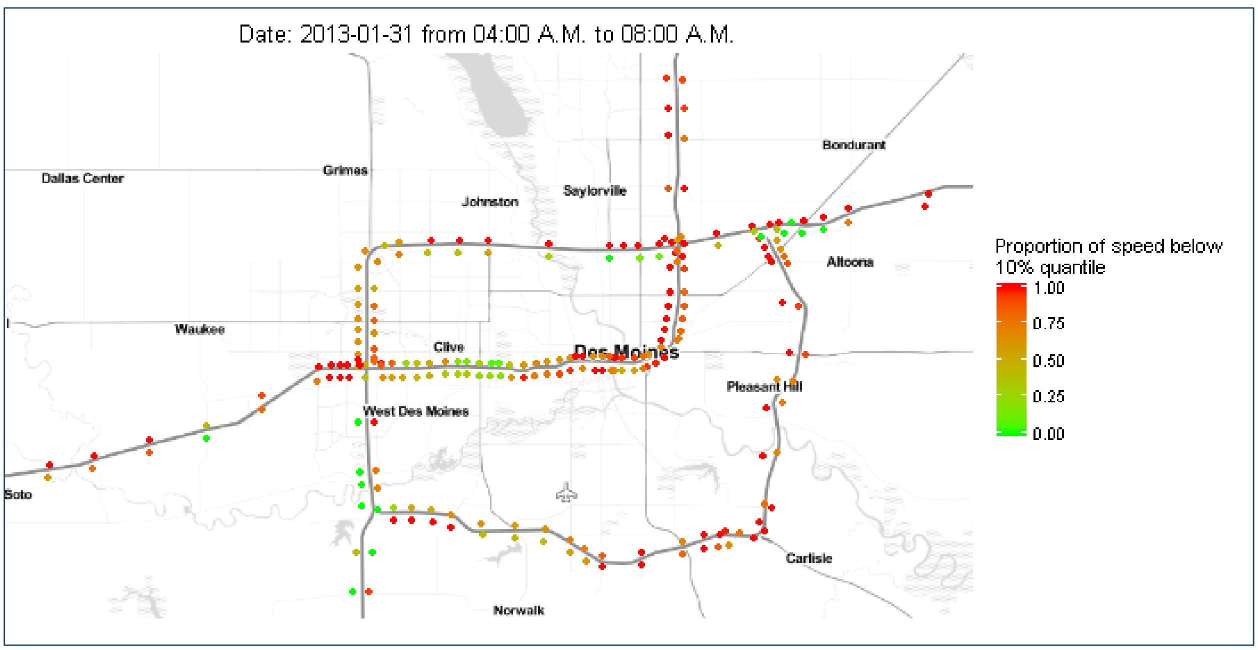


Figure 2.6. Locations that reported INRIX data from 04:00 a.m. to 08:00 a.m. on January 31, 2013

Each location is color-coded by the proportion of the observations that fall below the corresponding 10% quantile curve. Two dots near the same location indicate results from two directions.

From Figure 2.5, we see a moderate speed reduction on westbound I-35 and I-80, eastbound IA 5, northbound US 65, and eastbound I-235 from 00:00 a.m. to 04:00 a.m. on January 31. From Figure 2.6, we clearly see that traffic speed is largely affected by the winter weather event from 04:00 a.m. to 08:00 a.m., as is the case for the entire day of January 31.

2.5 Dynamic Model for Traffic Speed for Locations with Corresponding RWIS

Of the RWIS locations in the Des Moines area, two were close to locations provided in the Wavetronix dataset. These were Location 1: I-80 WB at US 65 SB LOOP and Location 2: IA 5 at MILE MARKER 103.55. Each location provided data for multiple directions and lanes. For other locations, there was no direct measurement available for lane condition that would indicate whether a decrease in traffic speeds was due to icy/wet/other conditions.

We selected Wavetronix speed data from the winter weather event on January 29 through January 31, which was characterized by initial periods of wet snow with a transition to blowing snow. We used the difference between the observed traffic speeds and the estimated median speeds as the response variable to remove deviations in traffic speeds that were due to the day of the week or the time of day. Previously, we had had success modeling deviations in traffic speed through a model that uses visibility, pavement temperature, lane condition (wet, ice, dry), and wind speed as covariates. However, in the previous analysis the data for pavement temperature and lane condition were collected at the same location as the data for traffic speed. In that case, we were confident that there would not be a mismatch, although there may have been a discrepancy between the lane conditions reported at the RWIS site and what was truly observed at the Wavetronix location for this project.

To develop the model relating weather conditions to traffic speeds, we first partitioned the data into four-hour periods, as mentioned above, with the last period potentially being longer than four hours. The general model structure applied to each partition was as follows:

$$y_t = \beta_0 + \beta_1 x_{t,wind} + \beta_2 x_{t,visib} + \beta_3 x_{t,|temp-32|} + w_t$$

$$w_t = \gamma w_{t-1} + v_t$$

where $w_o = 0$ and $v_t \sim N(0, \tau^2)$.

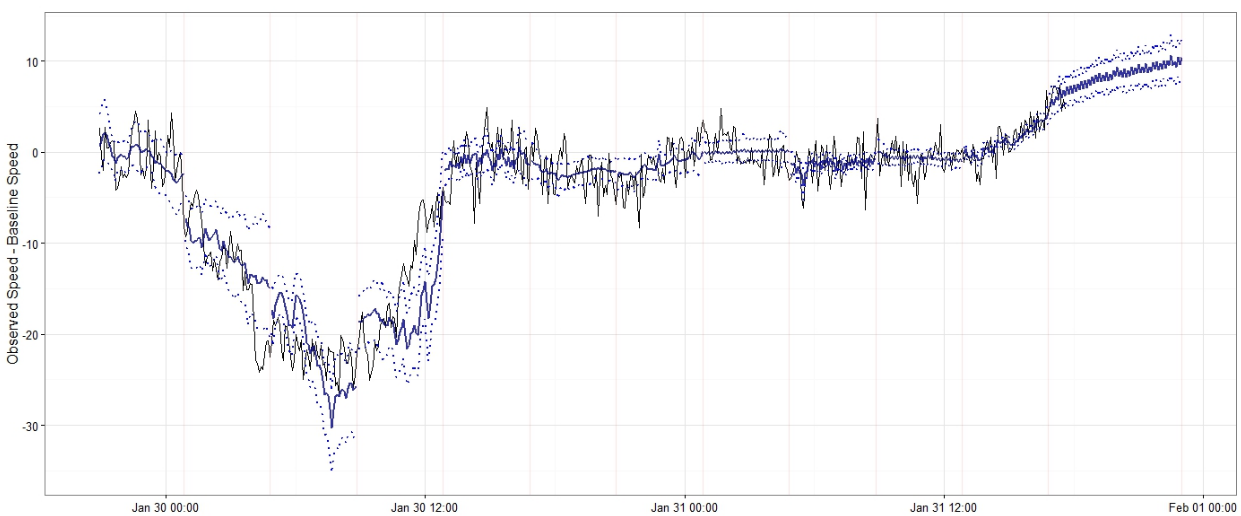
For the first four-hour period, we set diffuse normally distributed prior distributions for the βs, a diffuse inverse gamma distribution for τ, and a uniform prior, from -1 to 1, for γ. Given the posterior distributions for the parameters during the first four-hour period, we then updated the prior distributions for the βs for the second period and continued this process through the last time period. In this manner, we allowed the effects of wind speed and the other covariates to slowly evolve over time, if needed, because the effect of, for instance, wind speed can vary drastically under different weather conditions. We applied the model to the two locations with RWIS sites nearby, as discussed below.

Location 1: I-80 at US 65 SB LOOP

At Location 1, the winter weather event was found to cause a steady decrease in traffic speeds below the baseline traffic speed, reaching the minimum at around 7:00 a.m. to 8:00 a.m. Traffic speeds slowed to over 30 miles per hour less than the baseline traffic speed in both directions. After 8:00 a.m., a steady increase in traffic speeds was observed in both directions. The eastbound traffic speeds returned to around the baseline traffic speed, with a small dip in traffic speeds, around 5:00 p.m. to 6:00 p.m. However, the westbound traffic speeds experienced another dip around 4:00 p.m.

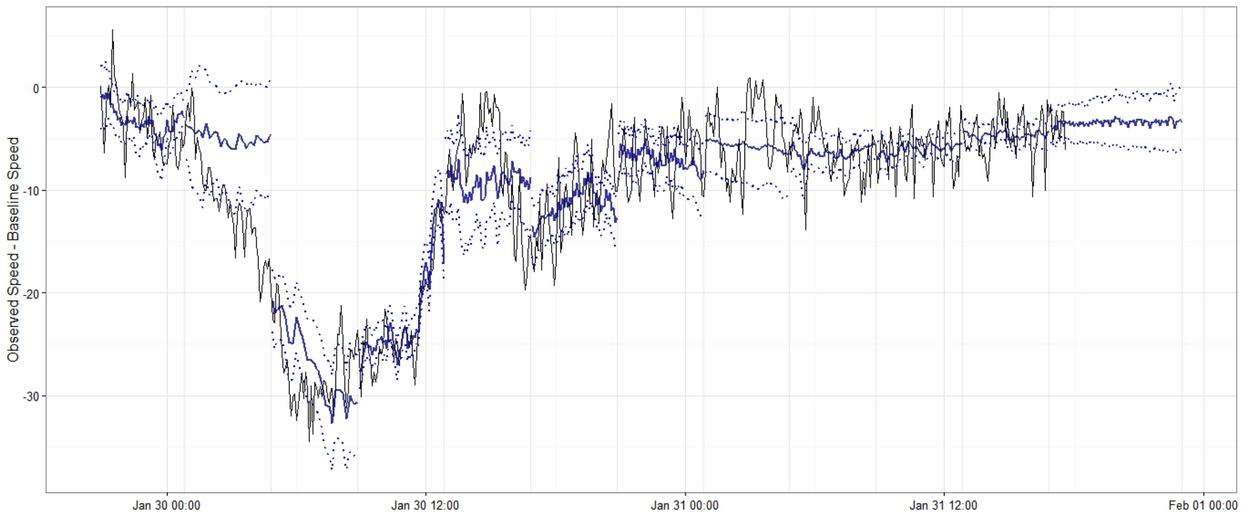
For Location 1, the values of visibility and wind speed were taken from the closest AWOS/automated surface observing system (ASOS) location, which in this case was Ankeny. Here again there may have been a disconnection between what happened at Location 1 versus what was observed in Ankeny. The closest lane condition reports showed periods of wet and icy conditions for the first 28 hours of the winter storm and only dry lane conditions for the last 20+ hours.

The model results for Location 1 are displayed in Figure 2.7 and Figure 2.8.



Black line = Observed speed reduction from the baseline
Blue line = Fitted results based on the Bayesian dynamic model
Blue dotted lines = Corresponding 90% credible region

Figure 2.7. Eastbound I-80 at US 65 SB LOOP



Black solid lines = Observed speed reduction from the baseline
Blue solid lines = Fitted results based on the Bayesian dynamic model
Blue dotted lines = Corresponding 90% credible region

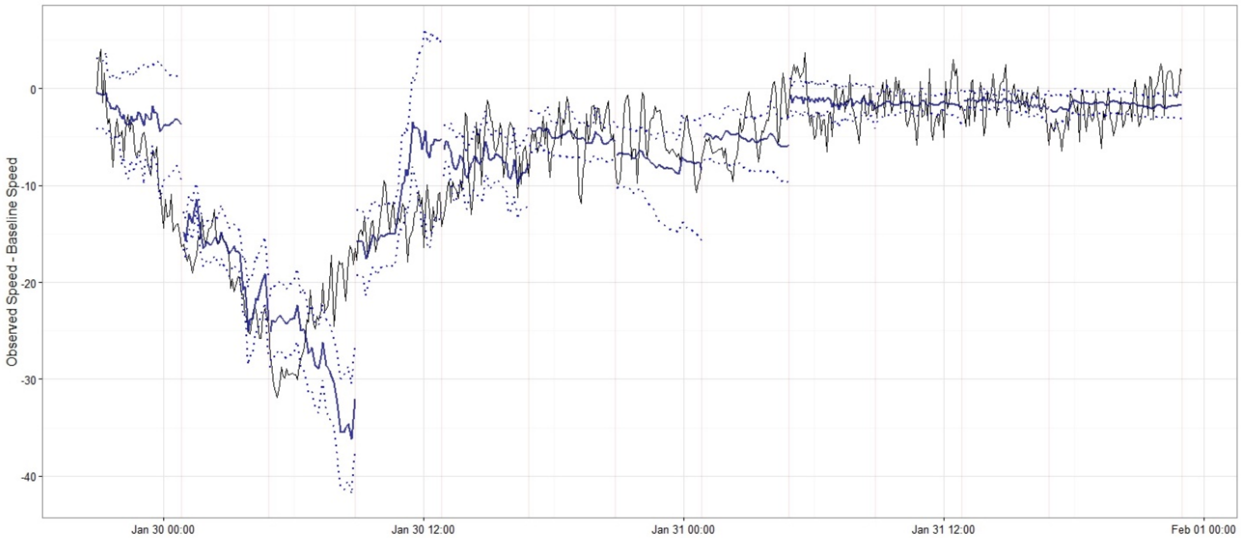
Figure 2.8. Westbound I-80 at US 65 SB LOOP

For the eastbound direction (Figure 2.7), we see a generally adequate fit for most of the four-hour periods. The fitted models for the third and fourth four-hour periods, which coincide with the highest decreases in traffic speed, do not seem to match exactly with what was occurring during this period. At around 1:00 p.m. on January 30, traffic speeds plateau and do not deviate too far from zero, nor do they experience large increases or decreases. While the model performs well during this period, we would have liked the model to capture trends in terms of large deviations from the baseline traffic patterns.

For the westbound direction (Figure 2.8), the model has similar issues. During the second four-hour period, the fitted model does not capture the decrease in traffic speeds, while in the third and fourth four-hour periods the models generally capture the trend much better than the models in the eastbound direction. However, the fitted model underestimates the increase in traffic speeds following the large decrease. We again see that the model adequately describes minor changes to traffic speeds. Here, however, we see that the traffic speeds are consistently negative and do not include zero, and therefore there is a deviation from the baseline traffic speed in the westbound direction throughout the majority of this weather event.

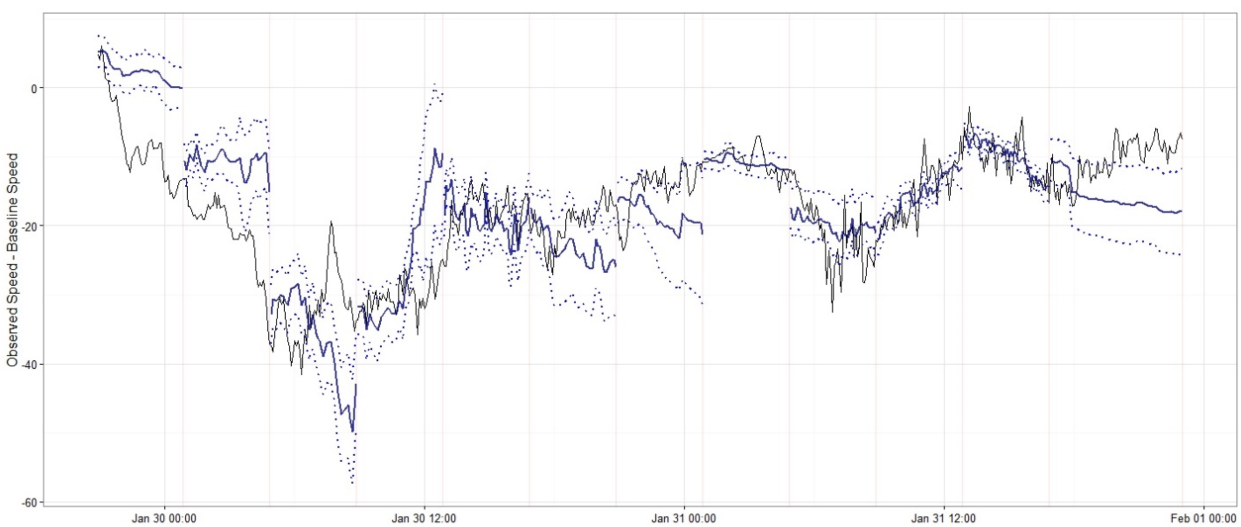
Location 2: IA 5 at MILE MARKER 103.55

Similarly to Location 1, at Location 2 the winter weather event was found to cause a steady decrease in traffic speeds below the baseline traffic speed, reaching the minimum at around 7:00 a.m. to 8:00 a.m., as shown in Figure 2.9 and Figure 2.10.



Black solid lines = Observed speed reduction from the baseline
Blue solid lines = Fitted results based on the Bayesian dynamic model
Blue dotted lines = Corresponding 90% credible region

Figure 2.9. Eastbound for IA 5 at MILE MARKER 103.55



Black solid lines = Observed speed reduction from the baseline
Blue solid lines = Fitted results based on the Bayesian dynamic model
Blue dotted lines = Corresponding 90% credible region

Figure 2.10. Westbound for IA 5 at MILE MARKER 103.55

Traffic speeds slowed to more than 30 miles per hour less than the baseline traffic speed in both directions, again similarly to Location 1. After 8:00 a.m., a steady increase in traffic speeds is evident in both directions, although it takes much longer for traffic speeds to increase to baseline levels than at Location 1.

For Location 2, the values of visibility and wind speed were taken from the Des Moines AWOS station. As mentioned above, there may have been a disconnection between the values of wind speed and visibility recorded at the Des Moines AWOS station and those observed at Location 2. The closest lane condition reports showed periods of mostly icy conditions for the first 28 hours of the winter storm and mostly dry lane conditions for the last 16+ hours. Wet lane conditions were only reported during the first four-hour period.

Model performance in terms of deviations in traffic speed from baseline values is subpar for some of the four-hour intervals in the eastbound direction (Figure 2.9). It is difficult to assess the adequacy of the model, however, because the weather variables observed at the corresponding stations may have been much different than what was actually occurring at the location at which traffic speed was recorded. The fitted model for the westbound direction (Figure 2.10) does not capture the initial decrease in traffic speeds. By the third four-hour period, we see that the fitted model overestimates the decrease, and the model then overestimates the increase in the subsequent period. The trend is captured much better in the second half of the storm event, with the decrease and increase in traffic speeds being adequately modeled.

Generally, in past analyses we found that the covariates included in this model are good indicators of traffic speed trends. When the combination of wind speed, visibility, pavement temperature, and lane condition cannot predict traffic speed well, we may be able to attribute decreases in speed to other factors, such as traffic incidents or maintenance. However, it is difficult to determine whether the model’s inadequacies are due to a mismatch between the lane conditions, wind speeds, and pavement temperatures recorded at nearby locations and those experienced at the actual locations where traffic speed data were collected.

An example of this mismatch can be seen in Figure 2.11, where the point estimates for the coefficients of the fitted model do not always match expectations.

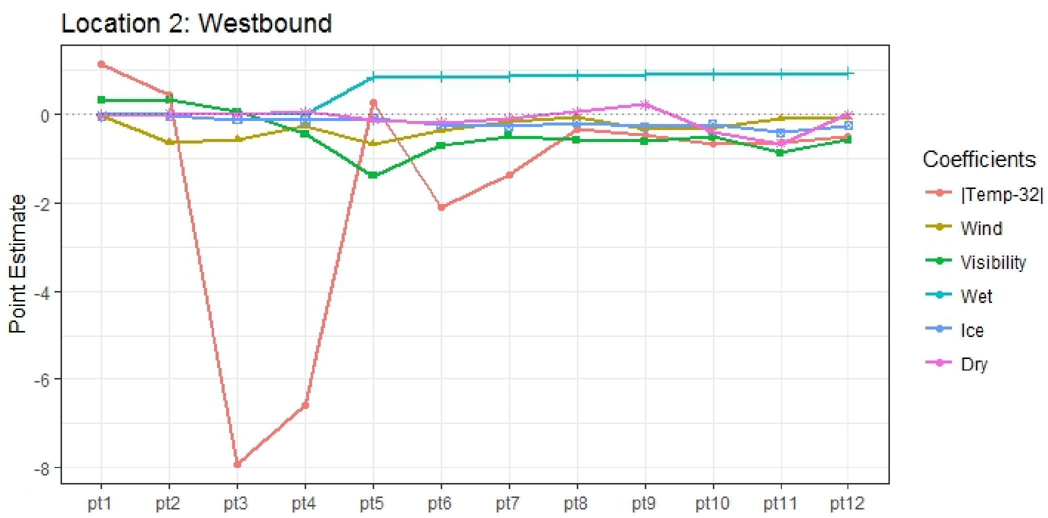


Figure 2.11. Point estimates for the coefficients of the fitted models for the westbound direction of Location 2

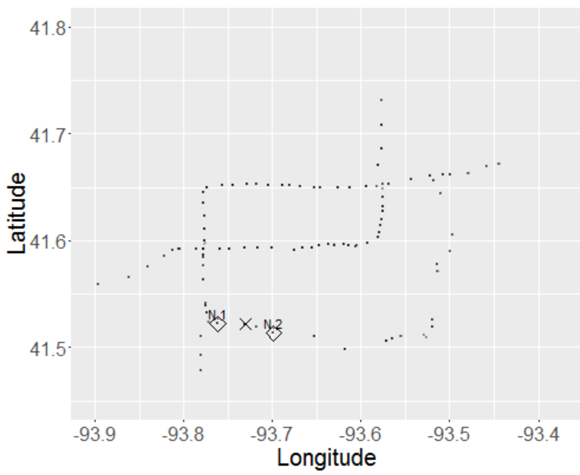
The point estimates for lane conditions, in particular, are counterintuitive in some cases. Take, for instance, the point estimate for wet conditions in Figure 2.11. Typically, we expect that when the lane is wet, traffic speeds decrease. However, the estimate for each fitted model in Figure 2.11 is consistently around zero or positive, which indicates that somehow wet lane conditions correspond to increases in traffic speed. This situation similarly occurs for the parameter estimates for icy road conditions. When the road is dry, traffic speeds should increase, or at least remain around zero. This occurs for the most part.

2.6 Curve Kriging

In spatial settings, Kriging methods (Cressie 1993) are often used for prediction; the methodologies are well developed for univariate observations. Given observations at different locations, Kriging methods, such as ordinary Kriging and universal Kriging, provide a weighted average of these observations such that the predictor for a new location is unbiased and the variance of the predicted value is minimized. Kriging for functional observations has been studied by Ramsay (2006) and many others (Goulard and Voltz 1993, Nerini et al. 2010, Giraldo et al. 2011, Franco-Villoria and Ignaccolo 2015, Menafoglio et al. 2013, Menafoglio and Petris 2016).

In this project, we used the curve Kriging method (Giraldo et al. 2011) to predict traffic speeds and volumes in the Des Moines metropolitan area. Because the Wavetronix sensor data were too sparse to provide information on the spatial dependence structure of the area, we derived the spatial structure from the INRIX data. More specifically, parameters for the trace-variogram, including the range and sill, were estimated based on the INRIX data to identify the spatial model. The same spatial model was then used in the curve Kriging method to predict the traffic speed and volume using the Wavetronix data. Historical data were used to derive the pointwise confidence band.

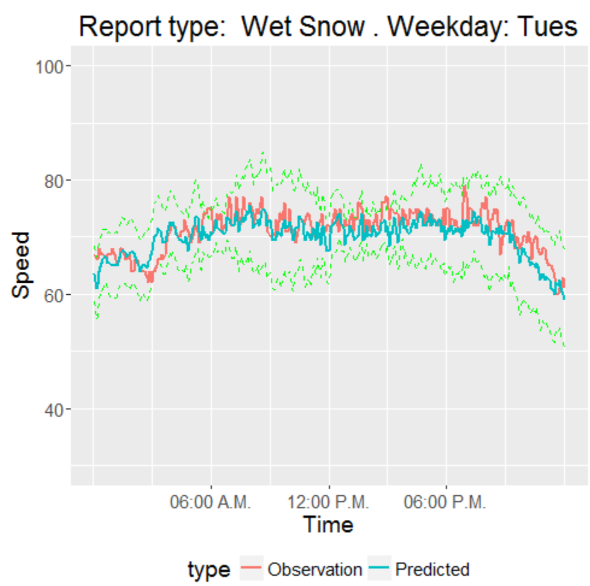
For any given location, we can use curve Kriging to predict the speed and volume of traffic using data from the two surrounding Wavetronix stations. As an illustration, we predicted the speed and volume of traffic at sensor IA 5 EB to SW CONNECTOR-EB on January 29, 2013 based on the observations at the sensor's two nearest neighbors. Figure 2.12 shows the geometric information of the three locations.

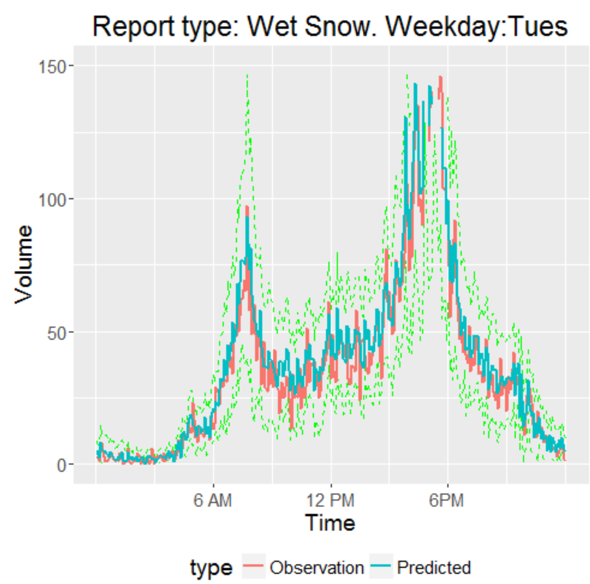


X = Location of the target sensor
N1 and N2 = Locations of two nearest neighbors among the available sensor locations

Figure 2.12. Locations of the target sensor (IA 5 EB to SW CONNECTOR-EB) and the two nearest neighbors

Figure 2.13 shows the prediction results for traffic speed and volume.





Orange/red line = Observation at the target location
Teal line = Predicted result
Green dotted lines = Corresponding 95% pointwise confidence bands

Figure 2.13. Prediction of speed (left) and volume (right) based on curve Kriging

As Figure 2.13 shows, the original observations fall into the pointwise 95% prediction confidence band. The predicted traffic speed and volume closely follow the actual speed and volume, which indicates that the methodology works well in this case. We have examined other cases, and the results are similar. A method for using the curve Kriging estimation function in R is given in Appendix B.

2.7 Conclusions

In this part of the report, we introduced several models based on two sources of traffic data, INRIX and Wavetronix, and limited weather information. Because the INRIX data included more locations than the Wavetronix data and the corresponding observations were self-consistent, we used the INRIX data to detect abnormal traffic areas.

We developed a method to estimate the multivariate quantiles for the INRIX observations, and the INRIX data were compared with the estimated quantiles to identify abnormal traffic patterns in both space and time.

An online interactive app was developed to visualize the results and help the Iowa DOT make informed decisions about winter weather maintenance.

A dynamic Bayesian model was implemented at two Wavetronix sensor locations where weather information was available, with the corresponding median curve as the baseline. The fitting results were satisfactory.

Furthermore, we also explored the spatial structure of the traffic data using the INRIX dataset and used curve Kriging to predict traffic speed and volume at any location. The prediction method was tested at the Wavetronix locations and was found to work well.

REFERENCES

Abdous, B. and R. Theodorescu. 1992. Note on the Spatial Quantile of a Random Vector. *Statistics & Probability Letters*, Vol. 13, No. 4, pp. 333–336.

Babu, J. G. and C. R. Rao. 1988. Joint Asymptotic Distribution of Marginal Quantiles and Quantile Functions in Samples from a Multivariate Population. *Journal of Multivariate Analysis*, Vol. 27, No. 1, pp. 15–23.

Chaudhuri, P. 1996. On a Geometric Notion of Quantiles for Multivariate Data. *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 862–72.

Cressie, N. 1993. Statistics for Spatial Data: Wiley Series in Probability and Statistics. *Wiley-Interscience*, Vol. 15, pp. 105–209.

Franco-Villoria, M. and R. Ignaccolo. 2015. Bootstrap Based Uncertainty Bands for Prediction in Functional Kriging. *ARXIV*. www.arxiv.org/abs/1505.06966.

Greenfield, T., M. Haubrich, M. Kaiser, Z. Zhu, D. Fortin, and J. Li. 2012. Winter Performance Measurement Using Traffic Speed Modeling. *Transportation Research Circular Number E-C162: International Conference on Winter Maintenance and Surface Transportation Weather*. April 30–May 3, Coralville, IA, pp. 187–197.

Kemperman, J. H. B. 1987. The Median of a Finite Measure on a Banach Space. International Conference on Statistical Data Analysis based on the L1-norm and related Methods, August 31–September 4, Neuchâtel, Switzerland, pp. 217–230.

Liu, Regina Y. 1990. On a Notion of Data Depth based on Random Simplices. *The Annals of Statistics*, Vol. 18, No. 1, pp. 405–414.

López-Pintado, S. and J. Romo. 2009. On the Concept of Depth for Functional Data. *Journal of the American Statistical Association*, Vol. 104, No. 486, pp. 718–734.

Giraldo, R., P. Delicado, and J. Mateu. 2011. Ordinary Kriging for Function-valued Spatial Data. *Environmental and Ecological Statistics*, Vol. 18, No. 3, pp. 411–426.

Goulard, M. and M. Voltz. 1993. Geostatistical Interpolation of Curves: A Case Study in Soil Science. *Geostatistics Tróia '92*, pp. 805–816.

Menafoglio, A. and G. Petris. 2016. Kriging for Hilbert-space Valued Random Fields: The Operatorial Point of View. *Journal of Multivariate Analysis*, Vol. 146, pp. 84–94.

Menafoglio, A., P. Secchi, and M. D. Rosa. 2013. A Universal Kriging Predictor for Spatially Dependent Functional Data of a Hilbert Space. *Electronic Journal of Statistics*, Vol. 7, pp. 2209–2240.

Nerini, D., P., Monestiez, and C. Manté. 2010. Cokriging for Spatial Functional Data. *Journal of Multivariate Analysis*, Vol. 101, No. 2, pp. 409–418.

Qiu, L. and W. Nixon. 2009. *Performance Measurement for Highway Winter Maintenance Operations*. IIHR-Hydroscience & Engineering, University of Iowa, Iowa City, IA.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

Ramsay, J. O. 2006. *Functional Data Analysis*. John Wiley & Sons, Inc., Hoboken, NJ.

APPENDIX A: BAYESIAN HIERARCHICAL MODEL

All models presented here are fit using the statistical software R (R Core Team 2016) and JAGS. The R and JAGS software and the following R packages should be installed and loaded before using the functions listed below.

Packages required: rjags, ggplot2, dplyr, lubridate, reshape2, gridExtra

Forecast Traffic Speed Deviations from Baseline Speeds - DOTforecast.R

Use: draws from prior distribution of β terms

Function: beta_forecast

Usage: beta_forecast(means, vars)

Arguments:

means: vector of estimated means for the prior distributions of the β terms in the model
vars: vector of estimated variances for the prior distributions of the β terms in the model

Use: forecast traffic speeds

Function: data_model

Usage: data_model(betpriors, tauprior, gamprior, data)

Arguments:

betpriors: matrix of draws from the prior distributions of the β terms obtained through beta_forecast
tauprior: vector of draws from prior distribution of the $\boldsymbol{\tau}$ parameter
gamprior: vector of draws from prior distribution of the $\boldsymbol{\gamma}$ parameter
data: dataset corresponding to winter weather event of interest

JAGS Code for Sequential Bayesian Model - JAGSModels_Sequential.R

Use: fitting model in JAGS

JAGS Function: mod5.seq

Usage: must be used using JAGS language

Arguments: example provided in Example_FitModel.R

Plotting the Results of the Bayesian Model - MCMCresults_plots.R

Use: plotting values of covariates

Function: cov.plots

Usage: cov.plots(event)

Arguments:
event: event is the data corresponding to the winter weather event of interest

Use: plotting results from fitted model

Function: mcmcresults

Usage: mcmcresults(mcmc, model, event)

Arguments:
mcmc: posterior draws from JAGS object
model: model option – default is option model5
event: event is the data corresponding to the winter weather event of interest

Dealing with Missing Covariate Values during a Winter Weather Event - missing.R

Use: imputing missing covariate values

Function: missing.val

Usage: missing.val(event)

Arguments:
event: event is the data corresponding to the winter weather event of interest

Applying a Threshold for Proportion of Values Corresponding to Lane Condition - propthres.R

Use: adjusting values for lane condition

Function: prop.thres

Usage: prop.thres (event, thresh)

Arguments:
event: event is the data corresponding to the winter weather event of interest
thresh: value of threshold – should correspond to a proportion (in Example_FitModel.R default is 13%)

Discounting Variance - variance_adj.R

Use: discounting variance terms

Function: var_fun

Usage: var_fun (data, prvar, tau, gam, thres)

Arguments:

data: data for one segment of winter weather event
prvar: variance of the prior distribution
tau: expected value from posterior distribution of τ
gam: expected value from posterior distribution of γ
thres: value of threshold – should correspond to a proportion (in Example_FitModel.R default is 13%)

Step-by-step Procedure to Fit the Model to One Winter Weather Event – Further Functions Defined: Example_FitModel.R

Instructions: It is important to first source the R functions and load the R packages required. The R code should then be run piece by piece. First, the data must be processed and subsetted. Then, the model should be fitted to the data, and finally the forecast estimates of the expected values under the observed covariate values can be produced.

APPENDIX B: SHORT MANUAL FOR THE FUNCTIONS DEVELOPED FOR MULTIVARIATE QUANTILE CURVE ESTIMATION AND CURVE KRIGING

The following packages should be installed and loaded before using the functions listed below.

Packages required: fda, ggplot2, geofd

The functions listed below are available in Auxiliary_function.R.

Multivariate Quantile Estimation Function

Function: median.curve.BMR

Usage: median.curve.BMR(data, alpha = 0.1, ind.plot = FALSE, ind.value = "Speed")

Arguments:

> data: the matrix of the traffic observations. Each column represents one traffic observation, that is, speed or volume. Each row represents one specific observation time. There could be missing values in the data matrix, but observations with missing values are not used in the quantile curve estimation at this point. Therefore, it is recommended that observations with missing values should be handled before passing to data.
>
> alpha: the quantile value. For example, alpha = 0.1 means that the 10% quantile is estimated. The default value for alpha is 0.1.
>
> ind.plot: logic value, indicating whether the plot with the estimated quantile curves is shown. The default value is FALSE.
>
> ind.value: auxiliary variable using to handle the outliers in the estimation procedure. If ind.value = "Speed", any estimated curve with a maximum value greater than 100 will be re-estimated. Additionally, this value appears in the y-axis of the plot if ind.plot = TRUE. ind.value = "Speed" corresponds to the speed observations. If the data matrix contains the volume observation, ind.value = "Volume".

Values:

> quantile.curve: The estimated quantile curve corresponding to the alpha quantile.
> quantile.curve.smoothed: The smoothed version of the quantile.curve using B-spline with 9 equally spaced inner knots.

Example:

data = matrix(rnorm(4000), nrow = 100)

median.curve.BMR(data, alpha = 0.1, ind.plot = TRUE, ind.value = "Speed")

Curve Kriging Estimation Functions

Two functions for performing curve Kriging estimation are listed below. The first is used to estimate the trace variogram, and the second is used in the curve Kriging estimation.

Function: trace.variog.inrix

Usage: trace.variog.inrix(trace.data, trace.geom, geom.tag)

Arguments:

trace_data: the matrix of the observations at the available locations. Each column represent one observation. Observations with missing values are not used at this stage.

trace.geom: the data frame containing the location information of the observations in trace_data.

geom.tag: the column names of trace.geom that correspond to the location information of the observations in trace_data. The default value for geom.tag is c("Longitude_Corrected","Latitude_Corrected").

Values:

best.model: the best fitted trace variogram model.

Example:

```
data = matrix(rnorm(5000), ncol = 50)
trace.geom = data.frame(lat= rnorm(50), lon = rnorm(50))
geom.tag = c("lat", "lon")
trace.variog.inrix(trace.data = data,trace. geom = trace.geom, geom.tag = geom.tag)
```

Function: curve.krig

Usage: curve.krig(geom.target, geom.two.selected, candidata.curve, best.model, geom.tag, data.1.his, data.2.his)

Arguments:

geom.target: the data frame that contains the location information of the target station.

geom.two.selected: the data frame containing the location information of the two nearest stations. Note that we only use two nearest station to predict the traffic values for the target station.

candidata.curve: the matrix that contains the observations for the two nearest locations in columns.

best.model: the fitted model by trace.variog.inrix using the INRIX data.

geom.tag: the column names of geom.two.selected that correspond to the location information of the observations in candidata.curve. The default value for geom.tag is c("Longitude_Corrected","Latitude_Corrected"). Notice that geom.target must have the same location names as geom.two.selected.

data.1.hist: the matrix of historical observations at the first location in geom.two.selected. Each column represents one observation. This data is used to compute the confidence band.

data.2.hist: the matrix of historical observations at the second location in geom.two.selected. The format is the same as data.1.hist, and the observations in the same column of data.1.hist and data.2.hist should be made at the same time.

Values:

cur.krig: the predicted traffic values for the target location by curve Kriging method.
sigma.band: the standard error of the predicted curve Kriging results.

Example:

```
data = matrix(rnorm(5000), ncol = 50)

trace.geom = data.frame(lat= rnorm(50), lon = rnorm(50))

geom.tag = c("lat", "lon")

best.model = trace.variog.inrix(trace.data = data,trace.geom = trace.geom, geom.tag = geom.tag)

geom.target = data.frame(lat = 0, lon =0)

dis.to.target = sqrt(trace.geom$lat^2 + trace.geom$lon^2 )

two.nearest.index = sort(dis.to.target,index.return=TRUE)$ix[1:2]

geom.two.selected = trace.geom[two.nearest.index,]

candidate.curve = data[,two.nearest.index]

data.1.his = matrix(rnorm(5000),ncol = 50)

data.2.his = matrix(rnorm(5000),ncol = 50)

predicted.result = curve.krig(geom.target = geom.target, geom.two.selected = geom.two.selected, candidata.curve= candidate.curve, best.model = best.model, geom.tag = c("lat","lon"), data.1.his = data.1.his,data.2.his = data.2.his)

plot(predicted.result$cur.krig,type = "l",ylim = c(-4,4))

lines(1:100,predicted.result$cur.krig + 1.96 * predicted.result$sigma.band,col = "red",lty=2)

lines(1:100,predicted.result$cur.krig - 1.96 * predicted.result$sigma.band,col = "red",lty=2)
```

APPENDIX C: PROCEDURE FOR STARTING THE ONLINE APP

The following packages should be installed and loaded before running the following code.

Packages required: shiny, dplyr, ggplot2, ggmap.

Before using the online app, the following preliminary files and folders should be prepared.

TMC_Identification.txt: This file contains the location index for INRIX data. The name of the column containing INRIX location index should be tmc.

A folder named as Reformation_files should be prepared, and the files in it are named as yyyy_tmc.csv. For example, 2013_118+04613.csv contains the observations made at INRIX location 118+04613 in 2013.

The data format of yyyy_tmc.csv: The first column is the tmc of this location, i.e., repeat tmc, although it has no effect on calculations. The second column is the observation time, and it should have the format of 2013-01-01 03:32:14. The third column contains the observed speed. The fourth to sixth columns contain the information for the average speed, reference_speed, and travel_time_minutes. The last two columns contain the confidence score and cvalue.

After preparing the data as shown above, run the code in combined.R. Notice that lines 4 to 6 should be modified if necessary.

By typing the following commands, this app can be generated:

```
setwd("../traffic/") #The working directory should be set to "traffic/"

shiny::runApp('/shiny/quantile/')
```