University Transportation Research Center - Region 2

Final Report

Bayesian Multilevel Models for Ridership Demand using Rainfall

Performing Organization: City University of New York (CUNY)

October 2017

Sponsor:
University Transportation Research Center - Region 2

University Transportation Research Center - Region 2

The Region 2 University Transportation Research Center (UTRC) is one of ten original University Transportation Centers established in 1987 by the U.S. Congress. These Centers were established with the recognition that transportation plays a key role in the nation's economy and the quality of life of its citizens. University faculty members provide a critical link in resolving our national and regional transportation problems while training the professionals who address our transportation systems and their customers on a daily basis.

The UTRC was established in order to support research, education and the transfer of technology in the field of transportation. The theme of the Center is "Planning and Managing Regional Transportation Systems in a Changing World." Presently, under the direction of Dr. Camille Kamga, the UTRC represents USDOT Region II, including New York, New Jersey, Puerto Rico and the U.S. Virgin Islands. Functioning as a consortium of twelve major Universities throughout the region, UTRC is located at the CUNY Institute for Transportation Systems at The City College of New York, the lead institution of the consortium. The Center, through its consortium, an Agency-Industry Council and its Director and Staff, supports research, education, and technology transfer under its theme. UTRC's three main goals are:

Research

The research program objectives are (1) to develop a theme based transportation research program that is responsive to the needs of regional transportation organizations and stakeholders, and (2) to conduct that program in cooperation with the partners. The program includes both studies that are identified with research partners of projects targeted to the theme, and targeted, short-term projects. The program develops competitive proposals, which are evaluated to insure the mostresponsive UTRC team conducts the work. The research program is responsive to the UTRC theme: "Planning and Managing Regional Transportation Systems in a Changing World." The complex transportation system of transit and infrastructure, and the rapidly changing environment impacts the nation's largest city and metropolitan area. The New York/New Jersey Metropolitan has over 19 million people, 600,000 businesses and 9 million workers. The Region's intermodal and multimodal systems must serve all customers and stakeholders within the region and globally.Under the current grant, the new research projects and the ongoing research projects concentrate the program efforts on the categories of Transportation Systems Performance and Information Infrastructure to provide needed services to the New Jersey Department of Transportation, New York City Department of Transportation, New York Metropolitan Transportation Council , New York State Department of Transportation, and the New York State Energy and Research Development Authorityand others, all while enhancing the center's theme.

Education and Workforce Development

The modern professional must combine the technical skills of engineering and planning with knowledge of economics, environmental science, management, finance, and law as well as negotiation skills, psychology and sociology. And, she/he must be computer literate, wired to the web, and knowledgeable about advances in information technology. UTRC's education and training efforts provide a multidisciplinary program of course work and experiential learning to train students and provide advanced training or retraining of practitioners to plan and manage regional transportation systems. UTRC must meet the need to educate the undergraduate and graduate student with a foundation of transportation fundamentals that allows for solving complex problems in a world much more dynamic than even a decade ago. Simultaneously, the demand for continuing education is growing – either because of professional license requirements or because the workplace demands it – and provides the opportunity to combine State of Practice education with tailored ways of delivering content.

Technology Transfer

UTRC's Technology Transfer Program goes beyond what might be considered "traditional" technology transfer activities. Its main objectives are (1) to increase the awareness and level of information concerning transportation issues facing Region 2; (2) to improve the knowledge base and approach to problem solving of the region's transportation workforce, from those operating the systems to those at the most senior level of managing the system; and by doing so, to improve the overall professional capability of the transportation workforce; (3) to stimulate discussion and debate concerning the integration of new technologies into our culture, our work and our transportation systems; (4) to provide the more traditional but extremely important job of disseminating research and project reports, studies, analysis and use of tools to the education, research and practicing community both nationally and internationally; and (5) to provide unbiased information and testimony to decision-makers concerning regional transportation issues consistent with the UTRC theme.

Project No(s):
UTRC/RF Grant No: 49198-25-27

Project Date: October 2017

Project Title: An Agent-Based Disaster Response Inference Model for Assessment of Transportation Risk under Extreme Events

Project's Website:
http://www.utrc2.org/research/projects/agent-based-disaster-response

Principal Investigator(s):
Naresh Devineni, Ph.D.
Assistant Professor
Department of Civil Engineering
The City College of New York
New York, NY 10031
Email: ndevineni@ccny.cuny.edu

Mahdieh Allahviranloo, Ph.D.
Assistant Professor
Department of Civil Engineering
The City College of New York
New York, NY 10031
Email: mallahviranloo@ccny.cuny.edu

Performing Organization(s):
The City University of New York (CUNY)

Sponsor(s):
University Transportation Research Center (UTRC)

To request a hard copy of our final reports, please send us an email at utrc@utrc2.org

Mailing Address:

University Transportation Reserch Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: 212-650-8051
Fax: 212-650-8374
Web: www.utrc2.org

Board of Directors

The UTRC Board of Directors consists of one or two members from each Consortium school (each school receives two votes regardless of the number of representatives on the board). The Center Director is an ex-officio member of the Board and The Center management team serves as staff to the Board.

City University of New York
Dr. Robert E. Paaswell - Director Emeritus of NY
Dr. Hongmian Gong - Geography/Hunter College

Clarkson University
Dr. Kerop D. Janoyan - Civil Engineering

Columbia University
Dr. Raimondo Betti - Civil Engineering
Dr. Elliott Sclar - Urban and Regional Planning

Cornell University
Dr. Huaizhu (Oliver) Gao - Civil Engineering
Dr. Richard Geddess - Cornell Program in Infrastructure Policy

Hofstra University
Dr. Jean-Paul Rodrigue - Global Studies and Geography

Manhattan College
Dr. Anirban De - Civil & Environmental Engineering
Dr. Matthew Volovski - Civil & Environmental Engineering

New Jersey Institute of Technology
Dr. Steven I-Jy Chien - Civil Engineering
Dr. Joyoung Lee - Civil & Environmental Engineering

New York Institute of Technology
Dr. Nada Marie Anid - Engineering & Computing Sciences
Dr. Marta Panero - Engineering & Computing Sciences

New York University
Dr. Mitchell L. Moss - Urban Policy and Planning
Dr. Rae Zimmerman - Planning and Public Administration

(NYU Tandon School of Engineering)
Dr. John C. Falcocchio - Civil Engineering
Dr. Kaan Ozbay - Civil Engineering
Dr. Elena Prassas - Civil Engineering

Rensselaer Polytechnic Institute
Dr. José Holguín-Veras - Civil Engineering
Dr. William "Al" Wallace - Systems Engineering

Rochester Institute of Technology
Dr. James Winebrake - Science, Technology and Society/Public Policy
Dr. J. Scott Hawker - Software Engineering

Rowan University
Dr. Yusuf Mehta - Civil Engineering
Dr. Beena Sukumaran - Civil Engineering

State University of New York
Michael M. Fancher - Nanoscience
Dr. Catherine T. Lawson - City & Regional Planning
Dr. Adel W. Sadek - Transportation Systems Engineering
Dr. Shmuel Yahalom - Economics

Stevens Institute of Technology
Dr. Sophia Hassiotis - Civil Engineering
Dr. Thomas H. Wakeman III - Civil Engineering

Syracuse University
Dr. Baris Salman - Civil Engineering
Dr. O. Sam Salem - Construction Engineering and Management

The College of New Jersey
Dr. Thomas M. Brennan Jr - Civil Engineering

University of Puerto Rico - Mayagüez
Dr. Ismael Pagán-Trinidad - Civil Engineering
Dr. Didier M. Valdés-Díaz - Civil Engineering

UTRC Consortium Universities

The following universities/colleges are members of the UTRC consortium under MAP-21 ACT.

City University of New York (CUNY)
Clarkson University (Clarkson)
Columbia University (Columbia)
Cornell University (Cornell)
Hofstra University (Hofstra)
Manhattan College (MC)
New Jersey Institute of Technology (NJIT)
New York Institute of Technology (NYIT)
New York University (NYU)
Rensselaer Polytechnic Institute (RPI)
Rochester Institute of Technology (RIT)
Rowan University (Rowan)
State University of New York (SUNY)
Stevens Institute of Technology (Stevens)
Syracuse University (SU)
The College of New Jersey (TCNJ)
University of Puerto Rico - Mayagüez (UPRM)

UTRC Key Staff

Dr. Camille Kamga: *Director, Associate Professor of Civil Engineering*

Dr. Robert E. Paaswell: *Director Emeritus of UTRC and Distin*guished Professor of Civil Engineering, The City College of New York

Dr. Ellen Thorson: *Senior Research Fellow*

Penny Eickemeyer: *Associate Director for Research, UTRC*

Dr. Alison Conway: *Associate Director for Education/Associate Professor of Civil Enginering*

Nadia Aslam: *Assistant Director for Technology Transfer*

Nathalie Martinez: *Research Associate/Budget Analyst*

Andriy Blagay: *Graphic Intern*

Tierra Fisher*: Office Manager*

Dr. Sandeep Mudigonda*, Research Associate*

Dr. Rodrigue Tchamna*, Research Associate*

Dr. Dan Wan*, Research Assistant*

Bahman Moghimi*: Research Assistant;*
Ph.D. Student, Transportation Program

Sabiheh Fagigh*: Research Assistant;*
Ph.D. Student, Transportation Program

Patricio Vicuna*: Research Assistant*
Ph.D. Candidate, Transportation Program

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Social Media for Transportation Disaster, Preparation, Response, and Recovery		5. Report Date October 2017 6. Performing Organization Code
7. Author(s) Naresh Devineni, Ph.D. Mahdieh Allahviranloo, Ph.D.		8. Performing Organization Report No.
9. Performing Organization Name and Address The City College of New York Grove School of Engineering Steinman Hall 160 Convent Avenue New York, NY 10031		10. Work Unit No. 11. Contract or Grant No. 49198-25-27f
12. Sponsoring Agency Name and Address University Transportation Research Center The City College of New York 137th Street and Convent Ave, New York, NY10031		13. Type of Report and Period Covered Final, June 1, 2015- October 2017 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The Northeast United States, particularly New York State has experienced an increase in extreme 24-hour precipitation during the past 50 years (Horton et al., 2011). Recent events such as Hurricane Irene and Superstorm Sandy have revealed vulnerability to intense precipitation within the transportation sector. Stronger knowledge of extreme events and the resultant simultaneous regional network vulnerabilities can support emergency management division in creating more effective response systems. There is a necessity to understand the underlying reasons that generate the spatial-temporal demand. There is also a necessity to understand and forecast, based on climate, individual level behavior and their nodal functions during a simultaneous extreme rainfall event.

This project combined cutting edge hydroclimatology science, space-time statistical modeling expertise and state of the art transportation sector's modeling frameworks and decision support tools to address the following questions:

(a) How best can spatial distribution of extreme rainfall intensity be estimated using high-resolution radar rainfall data?
(b) How individuals, as intelligent agents with different demographics, react to different climate events and shift their ridership behavior?
(c) How best can the spatial-temporal demand for ridership in New York City be modeled using climate and demographic/landuse characteristics?

To answer these questions, the researchers conducted two investigations: 1) characterize the spatial variability of extreme rainfall events over Greater New York Area using radar rainfall data and 2) develop Bayesian multilevel models to estimate the impact of hourly and daily rainfall on subway ridership in Manhattan.
The research is cast in a practical context using Greater New York Area as an example. The researchers anticipate that these results and our current on-going projects have the potential for revolutionizing how real time transportation planning is conducted in the future, through dynamic risk assessment and management procedure.

17. Key Words Bayesian Models, Spatial-temporal demand variability, extreme weather patterns		18. Distribution Statement	
19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 26	22. Price

Form DOT F 1700.7 (8-69)

Disclaimer

Executive Summary

The Northeast United States, particularly New York State has experienced an increase in extreme 24-hour precipitation during the past 50 years (Horton et al., 2011). Recent events such as Hurricane Irene and Superstorm Sandy have revealed vulnerability to intense precipitation within the transportation sector. Stronger knowledge of extreme events and the resultant simultaneous regional network vulnerabilities can support emergency management division in creating more effective response systems. Current studies mostly focus on simulating traffic flow on the network or evaluating different dispatching and vehicle routing scenarios in response to disaster; it is not prognostic with underlying climate information (Koetse and Rietveld, 2009; Kyte et al., 2001; Maze et al., 2006). There is a necessity to understand the underlying reasons that generate the spatial-temporal demand. There is also a necessity to understand and forecast, based on climate, individual level behavior and their nodal functions during a simultaneous extreme rainfall event. This project combined cutting edge hydroclimatology science, space-time statistical modeling expertise and state of the art transportation sector's modeling frameworks and decision support tools to address the following questions:

(a) How best can spatial distribution of extreme rainfall intensity be estimated using high-resolution radar rainfall data?

(b) How individuals, as intelligent agents with different demographics, react to different climate events and shift their ridership behavior?

(c) How best can the spatial-temporal demand for ridership in New York City be modeled using climate and demographic/landuse characteristics?

To answer these questions, we conducted two investigations: 1) characterize the spatial variability of extreme rainfall events over Greater New York Area using radar rainfall data and 2) develop Bayesian multilevel models to estimate the impact of hourly and daily rainfall on subway ridership in Manhattan.

This project is an outgrowth of an ongoing initiative of the PIs for collaborative research between hydroclimatology and transportation fields for targeted scientific advancement. The work initiated a bidirectional conversation between these two fields along with consulting engineers to exchange ideas and develop scientific methods for transportation risk and resiliency assessment

under extreme events. The research is cast in a practical context using Greater New York Area as an example. We anticipate that these results and our current on-going projects have the potential for revolutionizing how real time transportation planning is conducted in the future, through dynamic risk assessment and management procedure.

Table of Contents

.

Introduction

Large cities around the world rely on public transportation infrastructure to maintain a good level of service and to increase mobility and economic productivity. In year 2015, more than 10.7 billion transit trips were reported in the United States (Matthew Dickens, 2016). According to the American Public Transportation Association (APTA) every $1 invested in public transportation generates approximately $4 in economic returns through increased employment rate, business sales, and enhanced property values. The full return on investment for transportation systems can only be achieved when cities optimize and plan the maintenance and growth of their public transportation systems. A better realization of the demand level for transportation along with the identification of other influential parameters on its performance are required to improve the operation. Transportation system performance depends on the geometry of the network, as well as external factors like accidents, operational tear and wear, disruptions on dependent systems (e.g. the power grid for subways), weather, etc. The objective of the work presented here is to progress further the understanding of impacts of rainfall events on the subway ridership level in Manhattan, New York.

During the past 50 years, New York State has experienced an increase in extreme daily precipitation (Horton et al., 2011) as well as unusual weather patterns such as Hurricane Irene and Superstorm Sandy. In New York City, with over 4 Million daily commute trips (Moss and Qing, 2012), the evidence in the aftermath of extreme weather patterns have revealed vulnerabilities to intense precipitation within the transportation network, yielding enormous economic losses for the City (Brian Tumulty, 2012). A better analysis to quantify the effects of various weather conditions on the transportation network would result in well-informed and efficient policy-making decisions. Our study focuses on the influence of rainfall conditions on subway ridership. It is worth noting that, despite its importance, the dependence between weather conditions and transit ridership has seldom been investigated, especially at a finer resolution using spatially distributed rainfall data.

A fully functional model of weather patterns over transit ridership requires an understanding of detailed spatio-temporal demand variability to predict individual behavior during extreme weather events. The current study is a step in this direction. One of the main goal of this project is to find out how subway ridership in Manhattan is dynamically changing during rainfall events based on the timing of service and spatial distribution of subway stations. The analysis is conducted on both hourly and daily ridership levels. In the hourly analysis, the effect of rainfall on the subway ridership level with 1-hour time lag is evaluated, targeting the people's decision to travel in the following hour after the occurrence of rainfall events. The dataset used include hourly subway ridership data from January 2010 to December 2011 for 116 stations in Manhattan, representing about 1.7 billion trips as originally recorded by the Metropolitan Transportation Authority-New York City Transit (MTA-NYCT). Additionally, we use high-resolution radar rainfall data, provided by the National Center for Environmental Prediction (NCEP) in the same period. The radar rainfall data enables us to establish a very detailed association between the system-wide subway ridership and rain condition on the daily and hourly basis.

A Bayesian multi-level regression model is employed to predict the subway ridership demand. The Bayesian model allows structuring of information within and across subway stations and provides inference on the posterior probability distribution of the ridership while considering the uncertainty of the regression parameters. In multi-level model, the parameter uncertainties propagate through appropriate conditional distribution (Gelman et al., 2004). The current multi-level model is structured into two levels: In the first level, the model relates the ridership at every time unit (which can be day or hour), as the response variable, to the rainfall and time, as the explanatory variables. Land-use and average daily ridership in stations are added to the model as explanatory variables of the response coefficients at the second level. The multi-level structure of our model provides a framework where the land-use of the subway station (spatial characteristics) and the temporal variables of the model can be treated on the same footing. Therefore, the model reduces the uncertainty in both parameter estimations and estimated ridership values.

The study can be beneficial to transit and planning agencies in multiple ways. By predicting change in the ridership level at each station based on the unique characteristics of stations. Furthermore, by allocating the service efficiently such that a relevant performance metric (e.g. total trip delay) is minimized while the excess demand can be met and the service can be adjusted accordingly both in hourly and daily scales. It also merges high-resolution rainfall data and subway ridership in a single analytical framework. Using high-resolution radar rainfall data, the model accounts for the spatial variability of rainfall in the study region, a factor that was not considered in transportation before.

Data Description

Study Area and Ridership Data

More than 1.5 billion annual trips make New York City's subway system the busiest rapid transit rail systems in the United States and the seventh largest worldwide.[1] The subway system operates 24 hours a day, 365 days a year and serves 469 stations in the five boroughs of Manhattan, Queens, Brooklyn, the Bronx, and Staten Island. Figure 1 indicates the subway lines within New York City Boroughs. Subway is the primary mode of travel for about half of the daily commuters (almost 1.6 million daily trips) to Manhattan (Moss and Qing, 2012). Transit system in New York City operates based on Automated Fare Collection (AFC) system, and fares can be paid using Metrocards. Use of AFC system enables the collection of time consistent ridership data. However, New York City transit system operates based on flat rates for trips, and the collected data is only limited to the passengers entering the stations and the information regarding the location or the time of exit from the stations are not recorded[1]. Therefore, our analysis is based on the ridership demand at origin stations. Overall, 1.7 billion ridership data were recorded over the period of the analysis (January 2010 to December 2011)[2]. In this study, we use hourly ridership data of Manhattan, including 116 stations, collected by the MTA-NYCT from January 2010 to December 2011. Peak hours were identified based on the maximum hourly ridership demand for all the stations, where during weekdays highest ridership occurred in [8:00-10:00] and [18:00-21:00] time periods. However, a different pattern was observed during the weekends. Demand during midnights was much higher than weekday nights.

[1] http://www.mta.info/

[2] It is worth noting that during emergency weather conditions (e.g. Hurricane Sandy) the transit system in NYC was completely shut down forcing us to exclude those dates from the analysis.

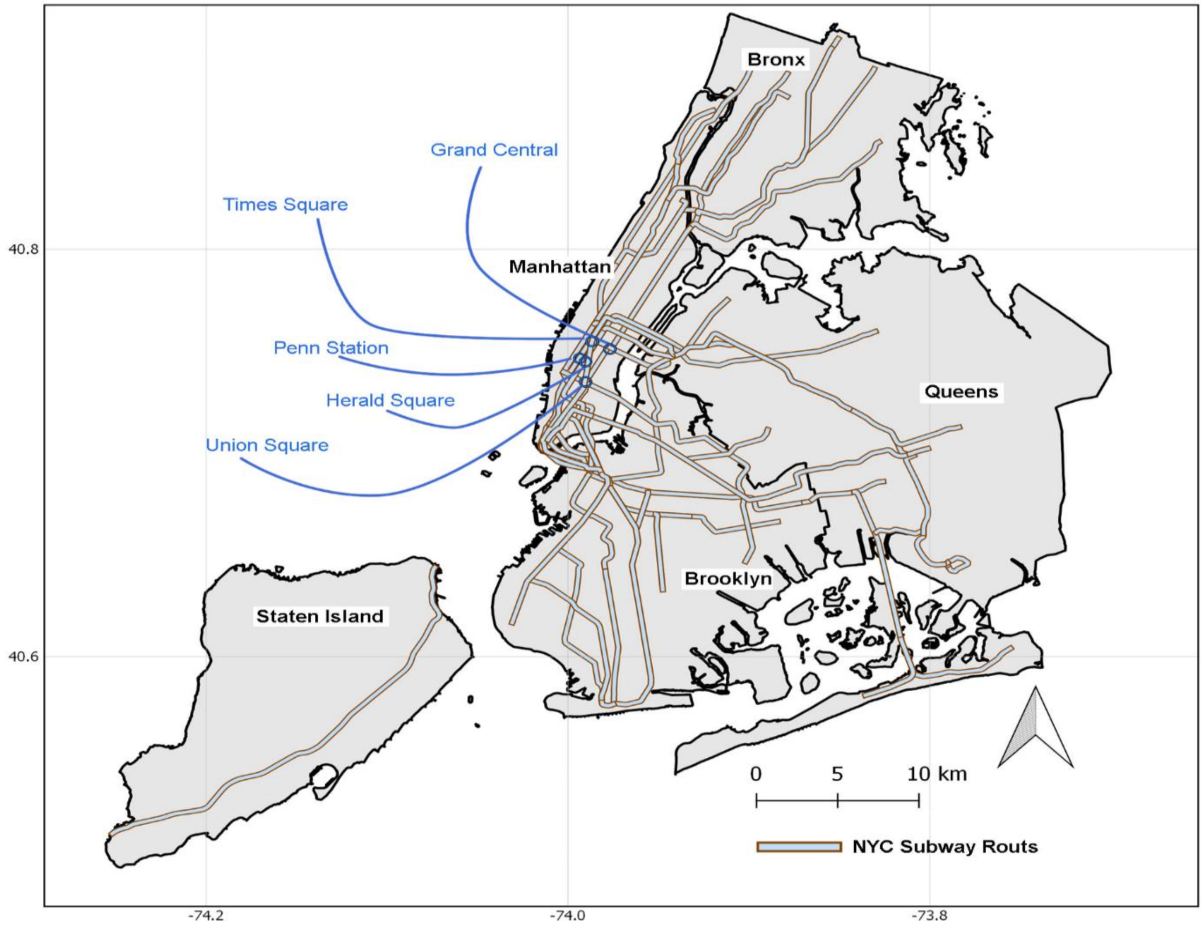


Figure 1 The subway lines in New York City and location of top 5 crowded stations in Manhattan during 2010-2011.

Radar Rainfall Data

In this study, we employed the NCEP Stage IV radar product (archived in GRIdded Binary, GRIB format) to generate rainfall fields over the study area (Leonard, 2002). The Stage IV radar data is a nationwide gridded data with a spatial resolution of 4km×4km, and a temporal resolution of 1 hour with a UTC timestamp and available from Earth Observing Laboratory[3]. The gridded rainfall data covers the whole study area with ~1.5 grid for width and ~5 grids for the length of Manhattan. While in the previous studies usually a single weather station data were considered as a proxy for the entire study area (Guo et al., 2003; Kalkstein et al., 2009; Singhal et al., 2014), such low resolution may result in reduced accuracy due to spatial variability of weather conditions in the area. Recently, Hamidi et al. (2017), presented that there is significant spatial variability of

[3] http://data.eol.ucar.edu

rainfall within New York City. They recognized different spatial patterns of the extreme rainfall for hourly vs. daily events. Significant finding from this investigation include 1) the summer extreme events have higher average rainfall intensities than the winter extreme events for the 1-hour duration, 2) The areal extent of the high intensity 1-hour extreme rainfall events during winter is larger than the areal extent of the high intensity 1-hour extreme events during summer, 3) the summer extreme events and the winter extreme events have similar intensities for the 24-hour duration. Given these statistically significant differences, we employed radar rainfall data in this study to understand the impact on subway ridership. During our study period, a total of 2328 hours of rain events (spread among 379 days of rain events) was detected.

Land-use Data

Land-use data, collected from NYC Department of City Planning[4], was utilized to identify land-use characteristics of each subway station. Buffer areas with the radius of 0.8 km (~0.5 miles) around each subway station are used to label the land-use type corresponding to each station. Then k-means clustering method (k=2) was employed (MacQueen, 1967) to divide subway stations into two groups of commercial and residential. Almost 70% of the subway stations (79 stations out of 116) were grouped into the commercial class, mostly located in the center and south of Manhattan, and 37 stations were grouped as residential class. Average hourly *(daily)* ridership values at commercial and residential zones during 2010-2011 were recorded as 1023 and 531 *(24554 and 12755)*, respectively. The stations grouped as commercial zones have significantly higher ridership demand value (93%) and larger variance compared to the residential zones.

[4] http://www1.nyc.gov

Methodology

We set up a Bayesian multi-level regression model to provide a simple strategy for combining the spatio-temporal relationship between ridership and rainfall, and to investigate the change in ridership demands with respect to the rainfall. The model relates the subway ridership at each station to the corresponding time indices, rainfall estimates and stations' characteristics (land-use and average daily ridership) in a multi-level framework, which can be tracked in Figure 2. Level-1 informs the ridership at each station by the explanatory variables. Level-2 allows the stations with the same land-use attributes and capacity (average daily ridership at each station) get common response characteristics by assigning regional predictor and common hyper distributions.

A Poisson distribution is assumed for ridership data that can handle over dispersed count outcome variables (Cameron and Trivedi, 1990):

$$y_{t,i} \sim Poisson(\lambda_{t,i}) \tag{1}$$

where i is subway station index, $i \in \{1,..,116\}$, $y_{t,i}$ is the subway ridership data at time *t* and station i, and $\lambda_{t,i}$ is the rate parameter, the mean of the ridership count data at time t and station *i*. The assumption of Poisson distribution for ridership data is validated using the nonparametric Kolmogorov–Smirnov (KS) test. Followings are the descriptions of the multi-level models for hourly and daily subway ridership:

Hourly Ridership

The Multi-level regression model at the first level describes how the rate parameter ($\lambda_{t,i}$) changes as a function of seven explanatory variables, equation 2.

$$\begin{aligned} Log(\lambda_{t,i}) = {} & \alpha_i + \beta_{1,i} \times [Ip_t] + \beta_{2,i} \times [Iw_t] + \beta_{3,i} \times [Ip_t \times Iw_t] + \beta_{4,i} \times [Rain_{t-1,i}] + \beta_{5,i} \times [Rain_{t-1,i} \times Ip_t] + \\ & \beta_{6,i} \times [Rain_{t-1,i} \times Iw_t] + \beta_{7,i} \times [Rain_{t-1,i} \times Ip_t \times Iw_t] \end{aligned} \tag{2}$$

where *t* is time in hour $t \in \{1,..,17500\}$, Ip_t is the peak hour binary indicator (takes the value of 1 if it is peak hour and zero otherwise), Iw_t is the binary weekday indicator (takes the value of 1 for weekdays, and zero otherwise), and $Rain_{t-1,i}$ is the amount of rainfall measured in mm/hour during the previous hour $(t-1)$ at the station *i*. α_i is an intercept of the regression model at each station, and can be interpreted as the average ridership at station i. $\beta_{k,i}, (k = 1:7)$ is a vector of regression coefficients interpreting the sensitivity of mean ridership to the predictors. Since the regression slopes may differ substantially for peak hours and off peak hours, and for the weekends and weekdays, we included the interaction terms, i.e., the product of the variables. This allows the slope to vary across these groups. α_i and $\beta_{k,i}$ are described by sets of hyper-parameters ($\mu_{\alpha_i}, \sigma_\alpha$ and $\mu_{\beta_{k,i}}, \sigma_{\beta_k}$, respectively) assuming a normal distribution:

$$\alpha_i \mid \mu_{\alpha_i} \sigma_\alpha \sim N(\mu_{\alpha_i}, \sigma_\alpha) \tag{3}$$

$$\beta_{k,i} \mid \mu_{\beta_{k,i}} \sigma_{\beta_k} \sim N(\mu_{\beta_{k,i}}, \sigma_{\beta_k}), k = 1:7 \tag{4}$$

μ_{α_i} represented the regression interception vector with the size of $i=116$, and $\mu_{\beta_{k,i}}$ indicated the regression slope vector with the size of $k \times i$ (7×116, which are the total numbers of explanatory variables and subway stations, respectively). In the multi-level framework, μ_{α_i} and $\mu_{\beta_{k,i}}$ are informed by station's land-use ($LanduseInd$) and average daily ridership ($Capacity$):

$$\mu_{\alpha_i} = a_\alpha + b_\alpha LanduseInd_i + c_\alpha Capacity_i \tag{5}$$

$$\mu_{\beta_{k,i}} = a_{\beta_k} + b_{\beta_k} LanduseInd_i + c_{\beta_k} Capacity_i \tag{6}$$

Noting that the ridership of the stations with the same characteristics are highly correlated (from section 2), the multi-level model allows for pooling information across the stations at the same zone (see figure 2-c) to reduce the associated uncertainty. This level is important in terms of choosing a proper pooling option between the two types of model; one where each station is estimated independently (no pooling), and one where all the stations are estimated together in one regression (full pooling). While the no pooling option may result in over-estimation of the effects, and the full pooling can ignore the variations among stations, the multi-level pooling we used here helps retain the regional characteristics and allows parameterizing across stations that may have a disparate range or scale of values (Devineni et al., 2013; Gelman and Rubin, 1992). We assumed a uniform prior distribution for the variance terms. The regression model parameters in Equations

(5)-(6), i.e., $a_\alpha, b_\alpha, c_\alpha$ and $a_{\beta_k}, b_{\beta_k}, c_{\beta_k}$ (level-2 in Figure 3) are presumed to be drawn from a common uninformative priors (Gelman et al., 2004).

Daily Ridership

The model structure presented for analysis of daily ridership demand is similar to the hourly analysis except that for daily analysis, the number of predictors decreased to 3 variables (*k=3*) as there is no peak/off-peak hourly indicators. Therefore, Equation (2) is changed to the following equation:

$$Log\left(\lambda_{t,i}\right) = \alpha_i + \beta_{1,i} \times [Iw_t] + \beta_{2,i} \times \left[Rain_{t,i}\right] + \beta_{3,i} \times \left[Rain_{t,i} \times Iw_t\right] \quad (7)$$

Where t is time in day $t \in \{1,..,730\}$, Iw_t is the binary weekday indicator (takes the value of 1 for weekday, and zero otherwise), $Rain_{t,i}$ is the rainfall (mm/day) on the same day of t and station i, α_i is an intercept of the regression model at each station, and $\beta_{k,i}, (k = 1:3)$ is the vector of regression coefficients. Equations (3) – (6) are applicable in the daily model as well. The only change is the size of coefficients changed from 7 for hourly model to 3 for daily model.

For both the models, the posterior distribution p (theta| data) of the complete parameter vector is derived by combining the distributions with the likelihood functions. Parameter estimation is implemented using Markov Chain Monte Carlo (MCMC) method for simulating the posterior probability distribution of the parameters conditional on the current choice of parameters and the data. We used JAGS (Just Another Gibbs Sampler (Plummer, 2003)) within R software for this purpose. The model runs four parallel MCMC chains for 5000 iterations retained after an initial burn-in of 1000 cycles for each chain to discard the initial state. Through the Gibbs sampler, the parameters are estimated by sequentially samples one parameter from the conditional distribution of that parameter relative to the others and provides an effective sampling-based numerical solution (Gilks et al., 1996). We verified the convergence of the posterior distribution based on the shrink factor suggested by Gelman and Rubin, 1992. The shrink factor compares the variance in the sampled parameters within the chains and across the chains to describe the improvement in the estimates for an increasing number of iterations. Gelman and Rubin, 1992 suggest running the chains until the estimated shrink factors are less than 1.1 for all the parameters.

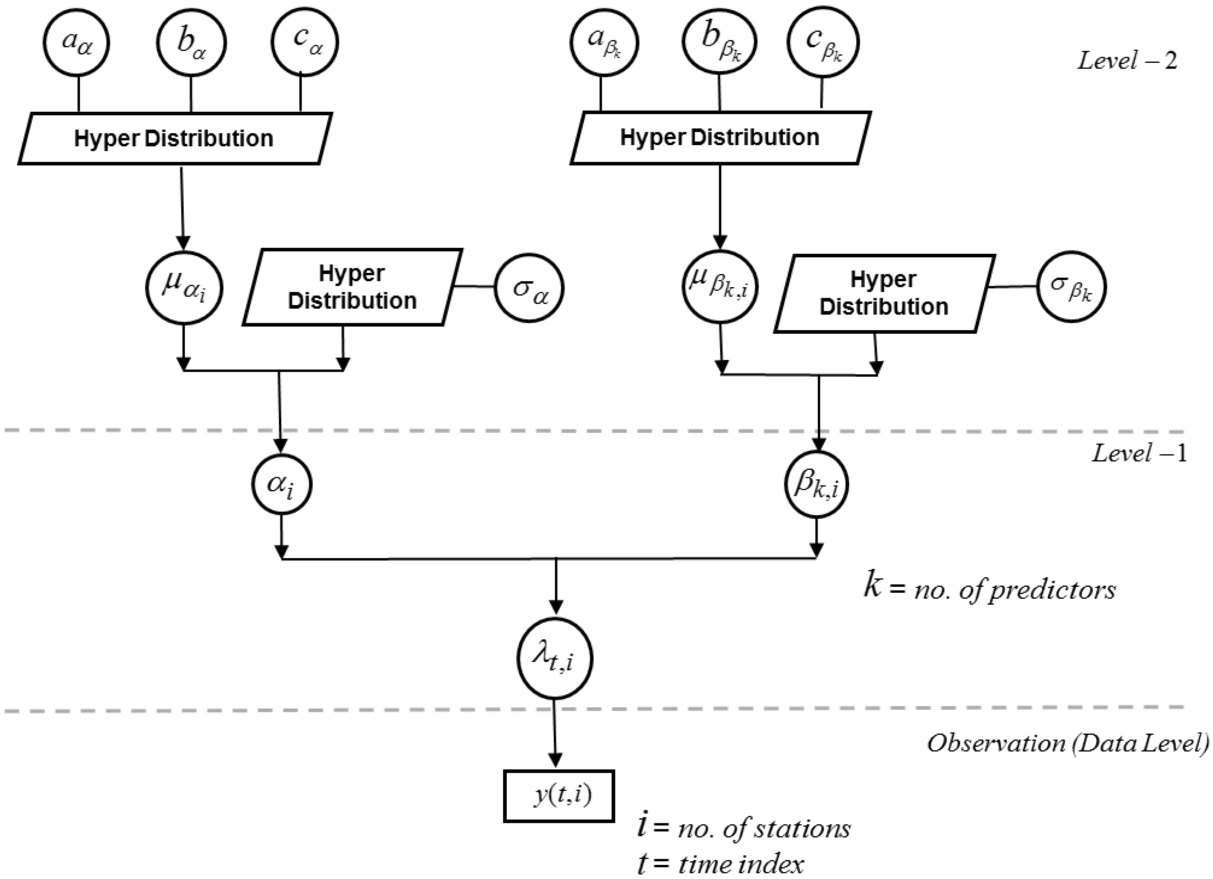


Figure 2 . Schematic algorithm of the Bayesian Multi-Level Regression Model.

Results and Discussion

The Bayesian multi-level regression model was run for hourly and daily ridership demand. The sensitivity of ridership to each of the explanatory variables across subway stations, i.e. $\beta_{k,i}$, are presented in this section for each station i, we investigate whether the response coefficient is significantly different from zero or not. We compute the p-value from the posterior distribution of the $\beta_{k,i}$ and evaluate its statistical significance at 10% level of significance (probability of incorrectly rejecting the null hypothesis that the response is 0). The results are presented in three parts as follow, noting that subscripts of H and D represent hourly and daily results, respectively:

Results of Hourly Analysis

Figures 3 presents the spatial distribution of hourly ridership sensitivity to the different explanatory variables. Each map shows the change of subway ridership in percentage with a unit change of the corresponding variable. For instance, β_{1H} indicates 0 - 200% change in subway ridership across all the stations in Manhattan when the time changes from off-peak to peak hours while the day type is not changing. Similarly, β_{7H} (the coefficient for interaction) indicates change in ridership from 0 - 10% during weekday peak hours, for one unit change of rainfall (1 mm/hour). The results demonstrate that for 90% confidence interval, the ridership in most of the stations is sensitive to changes in the values of explanatory variables, the significant parameters across the station are illustrated by open black circle in Figure 3. According to this figure, the sensitivity of hourly ridership to the peak hour, β_{1H}, and weekday, β_{2H}, indicators are extremely high for the stations located at the central and southern parts of Manhattan, where these regions are dominated by commercial land-use. The ridership demand during peak hours increases from weekends to weekdays (β_{3H}) especially in the northern part of the city, which mostly covers residential neighborhoods, while stations in commercial zones are insensitive to this change. For instance for this time period, station 181 St (A) located in far north part of the city has the highest sensitivity among all stations (β_{3H} ~ 176%). β_{4H} and β_{5H}, representing the sensitivity of hourly ridership to the rainfall and to the rainfall during peak hours respectively, indicate negative values. This reveals that hourly ridership slightly decreases during rainfall and peak hours (~ 0 - 9%). However, most of the stations are insensitive to the rainfall during weekday peak hours except the slight rise in the hourly ridership demand for few stations in the Southern part of the City (β_{7H}). As an example, Wall Street (2, 3) station located in Central Business District (in the southern part of the city) showed very high rate of sensitivity among all stations (β_{7H} ~ 12%). Hourly ridership sensitivity to the rainfall during weekdays, β_{6H}, slightly changes through the subway stations (~ -3 - 6%) and did not show any spatial pattern.

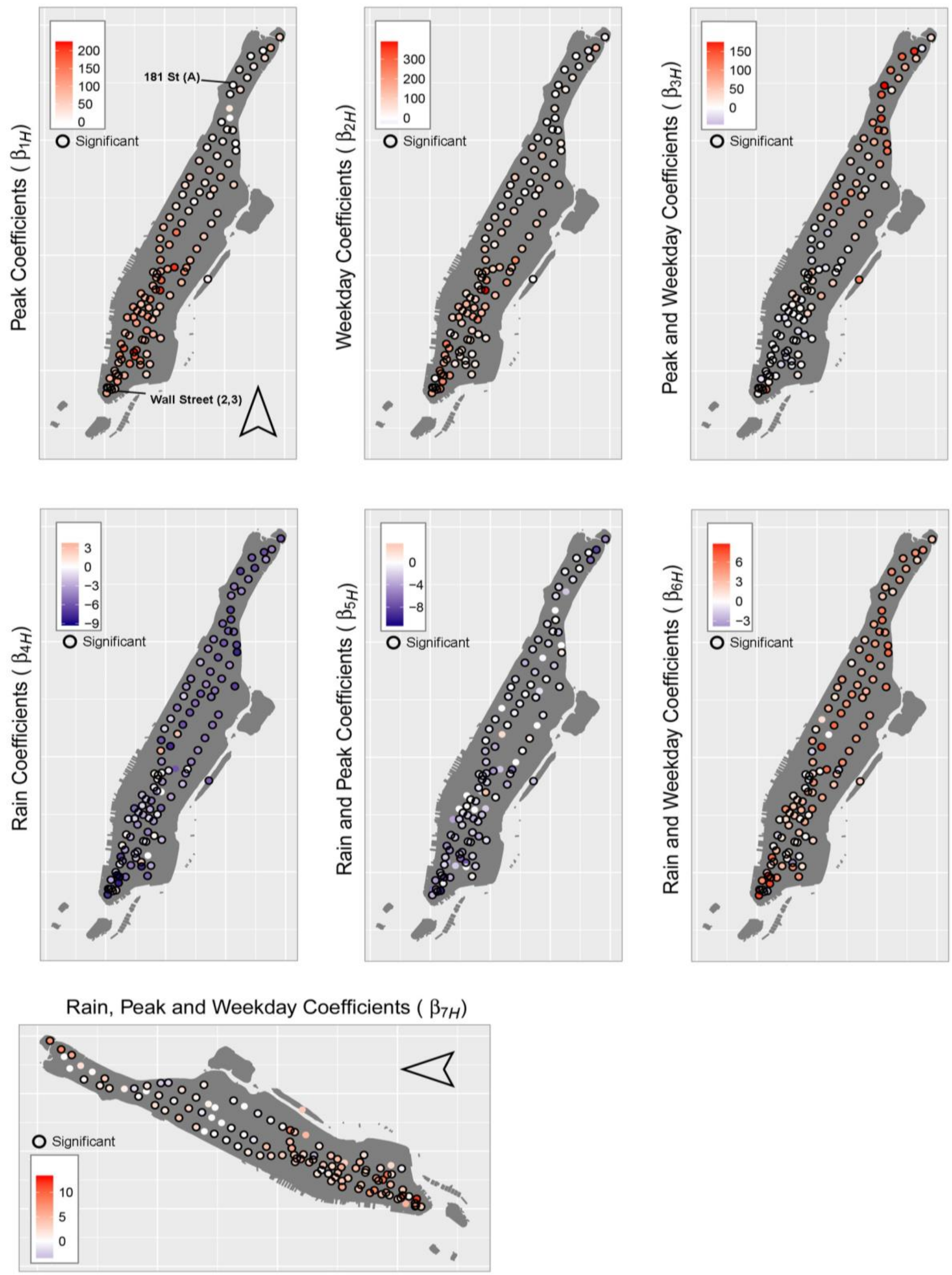


Figure 3 . Hourly results for the multi-level Bayesian regression model, showing the change of subway ridership demand in % with a unit change of the corresponding variable.

The daily analysis results are presented in Figure 4. Each map shows the change of subway ridership in percentage with a unit change of the corresponding variable. All the daily ridership sensitivities were highly significant (90% confidence interval) as presented with open black circle in the maps. Results indicate that daily ridership during weekdays are 10 - 300% higher than weekends (β_{1D}) especially for stations with commercial land-use attributes which is consistent with the hourly analysis results for the weekdays. We also find that a unit change of rainfall has negative impact on daily ridership up to 0.8%, mostly in the central and southern part of Manhattan (β_{2D}). For instance, City Hall (R,W) station located in the southern part of the city has the high sensitivity to the rainfall (β_{2D} ~ -0.8%). In the same area, daily ridership slightly increases with the rainfall during weekdays (β_{3D} changes up to 0.5%). Times Square 42-St station is a good example for this case showing 0.2 % increase of daily ridership with rainfall during weekdays (located in central part of Manhattan).

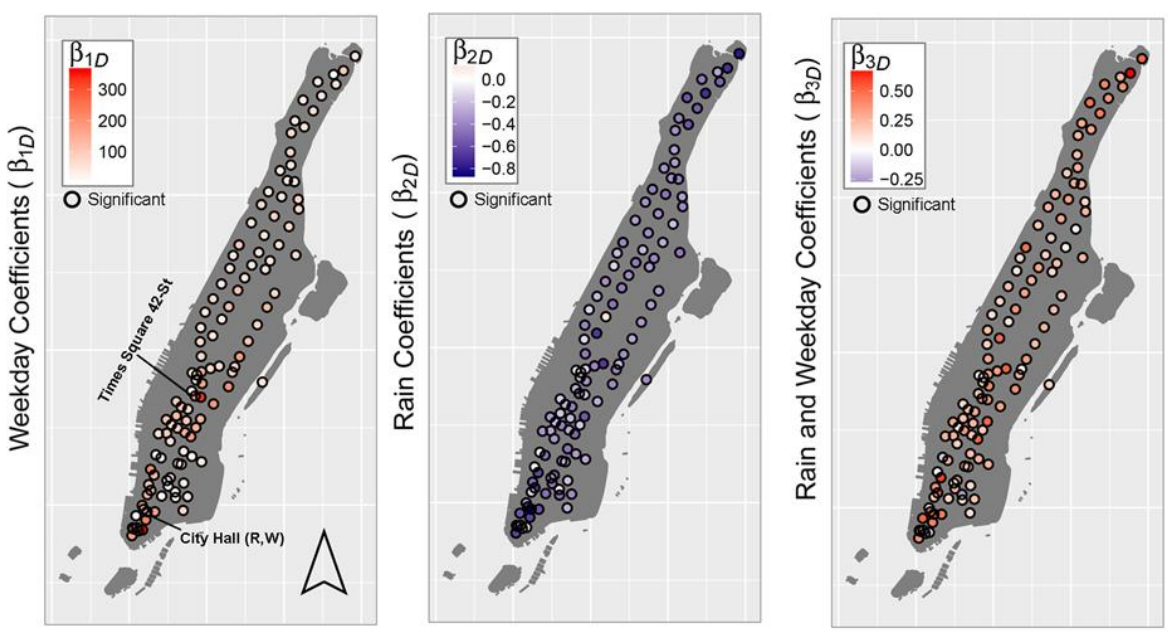


Figure 4 . Daily results for the Bayesian multi-level regression model, showing elasticity of subway ridership with respect to explanatory variables

Results of Analysis with Respect to Land-Use

Stations with different land-use attributes showed different level of sensitivity to each of the explanatory variables. For instance, in daily analysis, stations located in commercial zones have experienced higher change in ridership (about 30 percent more increase on average) during weekdays than stations located in residential zones (β_{1D}). This shows the high escalation in number of work-based trips during weekdays in commercial zones. Rainfall has shown negative influence on ridership in both hourly and daily analysis ($\beta_{4H} < 0, \beta_{2D} < 0$). These results are consistent with previous studies that found a negative impact of the rainfall on ridership for most of the stations in Manhattan. For both models, results present that stations in residential zones are more sensitive to rainfall than stations in commercial zones. It indicates that trips in residential zones (which are mostly home-based trips) are more flexible than trips in commercial zones (work-based trips).

To provide a comprehensive understanding of the daily and hourly demand of ridership in Manhattan, we used a non-parametric kernel density to estimate the density distribution of ridership sensitivity to each variable with respect to different land-use zones (Figure 5). The residential and commercial zones are presented in orange and green colors, respectively, and the vertical line in each plot shows the mode of the distribution (Y axis shows the frequency). Figure 5 indicates that, the sensitivity in commercial zones has higher frequency than residential for β_{3H} , implying the group of stations with commercial land-use attributes shows similar behavior. The residential stations are also showing similar behavior during rainfall (high frequency of β_{4H}). For β_{7H} we can see that stations in the commercial zones have higher sensitivity to the rainfall during weekday peak hours than residential stations (the absolute value of the mode for the β_{7H} density plot is larger for the stations at commercial zones). We can see the same behavior for β_{1H} and β_{2H} . In daily analysis results, Figure 5 demonstrates that stations at residential zones have higher frequency than stations at commercial zones for β_{1D} and β_{2D} . This shows the stations at the residential zones have similar behavior due to change from weekend to weekdays and change from no-rain to rain condition, respectively.

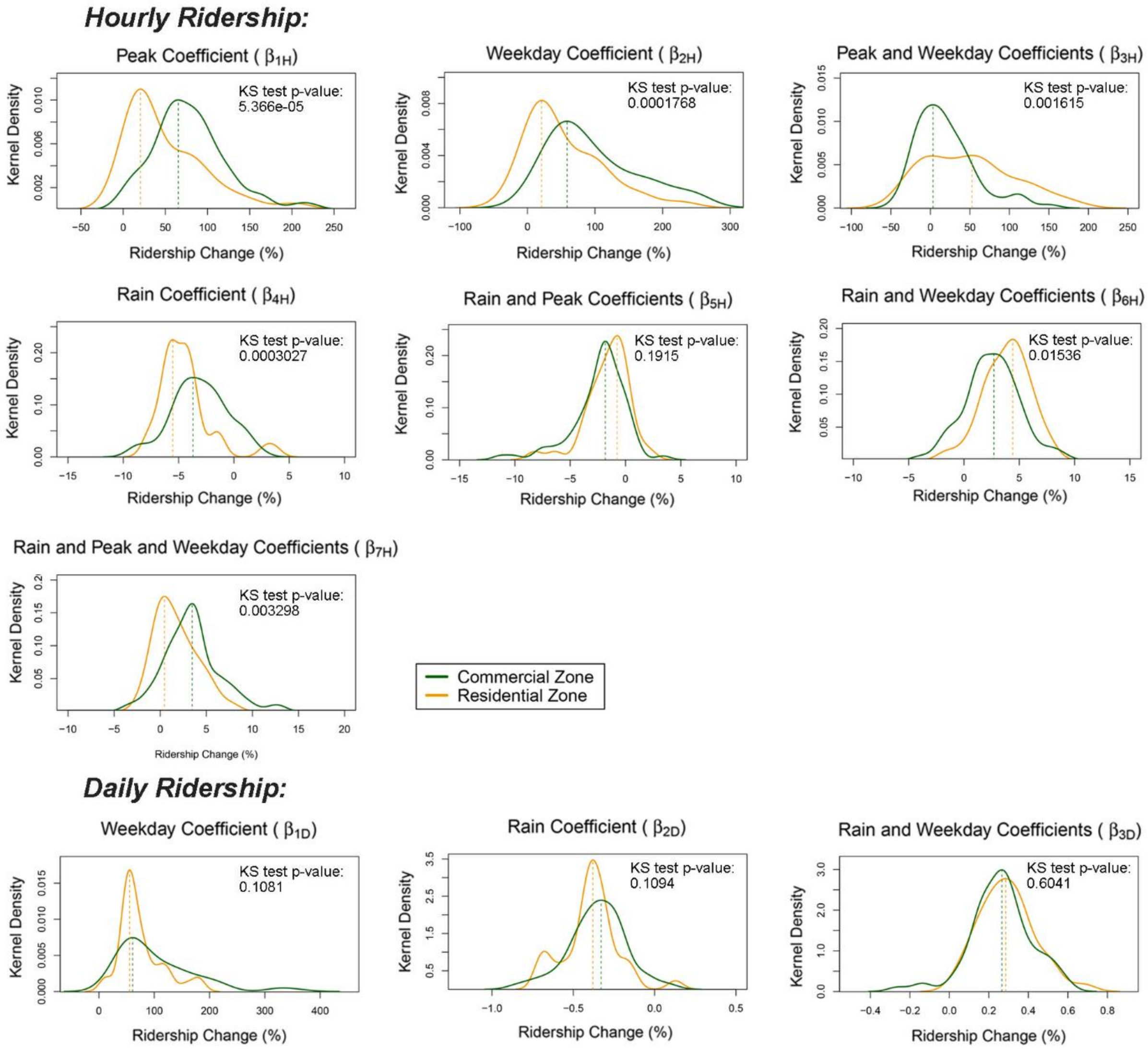


Figure 5 . Density plots of Sensitivity of ridership to the explanatory variables for hourly and daily demand of residential and commercial zones. Dashed lines indicate the highest frequency. The residential and commercial zones are presented in orange and green colors, respectively.

To assess groups of stations with similar behavior under each condition (for both hourly and daily models), we mapped the stations with high frequencies from the kernel density distributions shown in Figure 6. We considered different range of frequencies within the range of 5, 10, 15, 20, and 25 percent from the mode of the distributions. Stations density distribution maps for β_{7H} and β_{3D} are presented in Figure 6a, 6b respectively, display residential and commercial zones separately (L: low frequency, H: high frequency). The residential and commercial zones are presented in orange and green colors, respectively. In each map, stations with same color represent the same behavior. Figure 6a, indicates two different behavior for stations at commercial zones. First group are the stations located in the central and southern parts of the city with higher frequency and the second group of stations are located in the northern part of the Manhattan with lower frequency. Most of the residential stations in the northern part of Manhattan behave similarly with respect to β_{7H} (coefficient of weekday, peak hours and rainfall for hourly model, Figure 6a. Figure 6b, represents two sets of maps for the stations behavior in residential and commercial parts of the city with respect to β_{3D} (coefficient of rainfall and weekday for daily model). For stations at commercial zones in Figure 6b, we can see similar pattern for stations at the central part of the city while for the residential stations results do not represent a clear pattern within the city. Overall, Figure 6 determines that stations at the commercial parts of the city have more clear patterns compare to the stations with residential attributes.

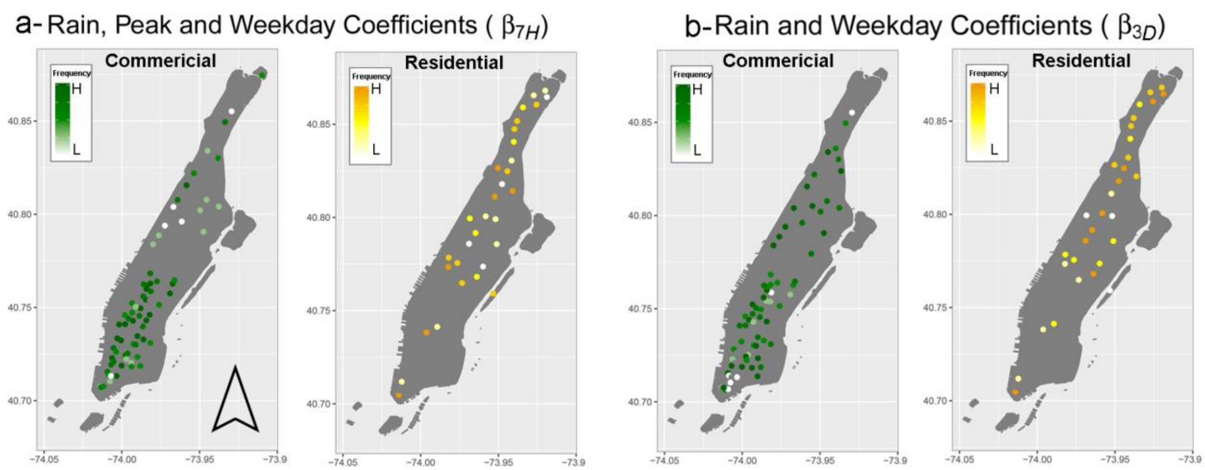


Figure 6 . Illustrating of the frequency of sensitivity for two selected variables (β_{7H} and β_{3D}); green: commercial, orange: residential; L: low frequency; H: high frequency

Conclusions

In this study the influence of rainfall conditions on subway ridership demand in Manhattan (across all 116 subway stations) was investigated. The analysis was conducted using the hourly subway ridership data and high-resolution radar rainfall data during 2010-2011. We developed a Bayesian multi-level regression model to assess the co-variability of subway ridership with rainfall across 116 stations. An important contribution of this paper is utilizing fine resolution radar rainfall data that enables us to capture spatial distribution of the rainfall within the study area and merge this information with land-use characteristics of each subway station under a unified analytical framework. Also, the multi-level structure of the model has advantage of allowing the subway stations with the same characteristics inform the ridership sensitivity in a similar way, and decrease the uncertainty in estimating the parameters. In order to capture people's decision to travel after rainfall events in the hourly analysis, the effect of rainfall on the subway ridership with one hour time lag is evaluated. Our study findings on the effects of rainfall on the hourly and daily subway ridership demands can be summarized as follows:

- The results obtained from Bayesian multi-level regression model refers to the existence of a high correlation between land-use attributes of the subway stations and its ridership sensitivity to the rainfall events.
- Both daily and hourly ridership demands decrease with rainfall (statistically significant) with the higher rate at the residential zones. This result is consistent with those obtained by previous studies in evaluating impact of various rainfall conditions on transit ridership and the finding that rainfall generally has a negative impact on transit ridership. Residential zones have more non-work-base trip rates compare to the commercial zones which results in higher rate of change in ridership for stations in this zones.
- For the daily model, the weekday ridership is higher than weekend ridership, with the higher positive rate of change during rainfall in the residential zones than the commercial zones (statistically insignificant, see Figure 6, β_{3D}). The fact that mandatory and work-based trips are dominant over the weekdays compared to weekends, leading to the conclusion that the demand for subway ridership in commercial zones are robust to fluctuations in rainfall conditions.

- Hourly ridership during weekday peak hours increases with rainfall slightly (statistically significant, see Figure 6, β_{7H}) in both commercial and residential zones, with higher rate in the commercial zones. An interpretation might be that there is a substantial change in percentage shares of other transportation modes under rainfall condition (e.g., biking and walking decrease during rain events causing an increase in subway ridership). Hourly ridership during weekend peak hours decreases with rainfall for all stations in both commercial and residential zones and it could be due to the high discretionary trip rates in weekends which have more flexibility.
- Daily demand in ridership does not change significantly due to rainfall, as we can see higher changes in hourly ridership. These results demonstrate the fact that hourly analysis is more comprehensive and reveals more information about ridership behavior of the people in different parts of the city.

The result supports the execution of a high-resolution analysis where land-use attributes, and hourly rainfall and ridership data should be taken into account. Considering magnitude of the transit ridership in Manhattan, almost 1.7 billion trips during 2010-2011, the complexity of the city, and the fact that the city was hit by several extreme weather conditions in recent years, analysis of the rainfall impacts on ridership and understanding the exposure of the transportation networks to spatially varying rainfall and simultaneous events is essential. Better knowledge on how rainfall events impact transportation system can support emergency management divisions on creating more effect response systems. This study provides proper decision tools for transit and planning agencies to optimally allocate limited time and resources during rainfall events. By quantifying different impacts of weather, transport and urban policies can be adjusted to accommodate the changes accordingly. Although the study presented here is focused solely on rainfall events, this method can be extended and applied to assess the impacts of other climate events as heavy snows or extreme heats on the travel behavior and daily choices of individuals. One can also extend the work to developing prognostic information (forecasts) based on climate to predict the individual level behavior and the nodal functions during simultaneous rainfall events. Clearly, those assessments are subject to the availability of longitudinal data and is part of the ongoing efforts of the authors.

People Involved and Publications

Naresh Devineni, Ph.D.

Assistant Professor

Department of Civil Engineering, NOAA-Cooperative Remote Sensing Science and Technology Center (NOAA-CREST), Center for Water Resources and Environmental Research (City Water Center), The City University of New York (City College) of New York – CUNY

Mahdieh Allahviranloo, Ph.D.

Assistant Professor

Department of Civil Engineering, The City University of New York (City College) of New York – CUNY

Shirin Najafabadi

Research Assistant

Department of Civil Engineering, The City University of New York (City College) of New York – CUNY

Ali Hamidi

Research Assistant

Department of Civil Engineering, NOAA-Cooperative Remote Sensing Science and Technology Center (NOAA-CREST), Center for Water Resources and Environmental Research (City Water Center), The City University of New York (City College) of New York – CUNY

Dissemination of the research results:

- Hamidi, A., N. Devineni, J. Booth, A. Hosten, R. Ferraro, and R. Khanbilvardi. (2017). Classifying Urban Rainfall Extremes Using Weather Radar Data: An Application to the Greater New York Area. Journal of Hydrometeorology, 18, 611–623, doi: 10.1175/JHM-D-16-0193.1.

- Shirin NajafAbadi (2016) "Quantifying the impacts of rainfall on subway ridership in Manhattan " − UTRC Transportation Technology Symposium: Innovative Mobility Solutions, New York Institute of Technology, 2016. − INFORMS Annual Meetings, Nashville.

- Najafabadi, Sh., Hamidi, A., Devineni, N., Allahviranloo, M. , Does Demand for Subway Ridership in Manhattan Depend on the Rainfall Events?, Journal of Transportation Research Part A, under review.

Acknowledgments

This research was supported, in part, by grants from the University Transportation Research Center Region 2 (UTRC) [grant numbers 49198-25 27]; and seed grant provided by CUNY Advanced Science Research Center (ASRC); their support is gratefully acknowledged.

Current and Future Projects from Our Team

PIS of this project are currently working on " Generating Multidimensional Disruption Index for Mobility during Extreme Climate Events" supported by CUNY Advanced Science Research Center (ASRC).

Based on the results and method developed through the current project,s PIs are planning to apply to National Science Foundation at 'Innovations at the Nexus of Food, Energy and Water Systems (INFEWS)' program.

References

Brian Tumulty, 2012. Cuomo: Sandy cost N.Y. $32B in damage and loss. USA Today.

Cameron, A.C., Trivedi, P.K., 1990. Regression-based tests for overdispersion in the Poisson model. J. Econom. 46, 347–364. doi:10.1016/0304-4076(90)90014-K

Devineni, N., Lall, U., Pederson, N., Cook, E., 2013. A tree-ring-based reconstruction of delaware river basin streamflow using hierarchical Bayesian regression. J. Clim. 26, 4357–4374. doi:10.1175/JCLI-D-11-00675.1

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, Chapman Texts in Statistical Science Series. doi:10.1007/s13398-014-0173-7.2

Gelman, A., Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. Stat. Sci. 7, 457–511. doi:10.1214/ss/1177011136

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Technometrics. doi:10.2307/1271145

Guo, Z., Wilson, N.H.M., Rahbee, A., 2003. Impact of Weather on Transit Ridership in Chicago, Illinois. Transp. Res. Rec. 2034, 3–10. doi:10.3141/2034-01

Hamidi, A., Devineni, N., Booth, F., Hosten, A., Ferraro, R., Khanbilvardi, R., 2017. Classifying Intensity and Area of Extreme Rainfall in Greater New York Area Using Weather Radar Data. J. Hydrometeorol. 18, 611–623.

Horton, R.M., Gornitz, V., Bader, D.A., Ruane, A.C., Goldberg, R., Rosenzweig, C., 2011. Climate hazard assessment for stakeholder adaptation planning in New York City. J. Appl. Meteorol. Climatol. 50, 2247–2266. doi:10.1175/2011JAMC2521.1

Kalkstein, A.J., Kuby, M., Gerrity, D., Clancy, J.J., 2009. An analysis of air mass effects on rail ridership in three US cities. J. Transp. Geogr. 17, 198–207. doi:10.1016/j.jtrangeo.2008.07.003

Koetse, M.J., Rietveld, P., 2009. The impact of climate change and weather on transport: An overview of empirical findings. Transp. Res. Part D Transp. Environ. 14, 205–221. doi:10.1016/j.trd.2008.12.004

Kyte, M., Khatib, Z., Shannon, P., Kitchener, F., 2001. Effect of Weather on Free-Flow Speed. Transp. Res. Rec. 1776, 60–68. doi:10.3141/1776-08

Leonard, C., 2002. Introduction to grib 1999, 1–3. doi:10.1126/scisignal.2001965

MacQueen, J.B., 1967. Kmeans Some Methods for classification and Analysis of Multivariate Observations. 5th Berkeley Symp. Math. Stat. Probab. 1967 1, 281–297. doi:citeulike-article-id:6083430

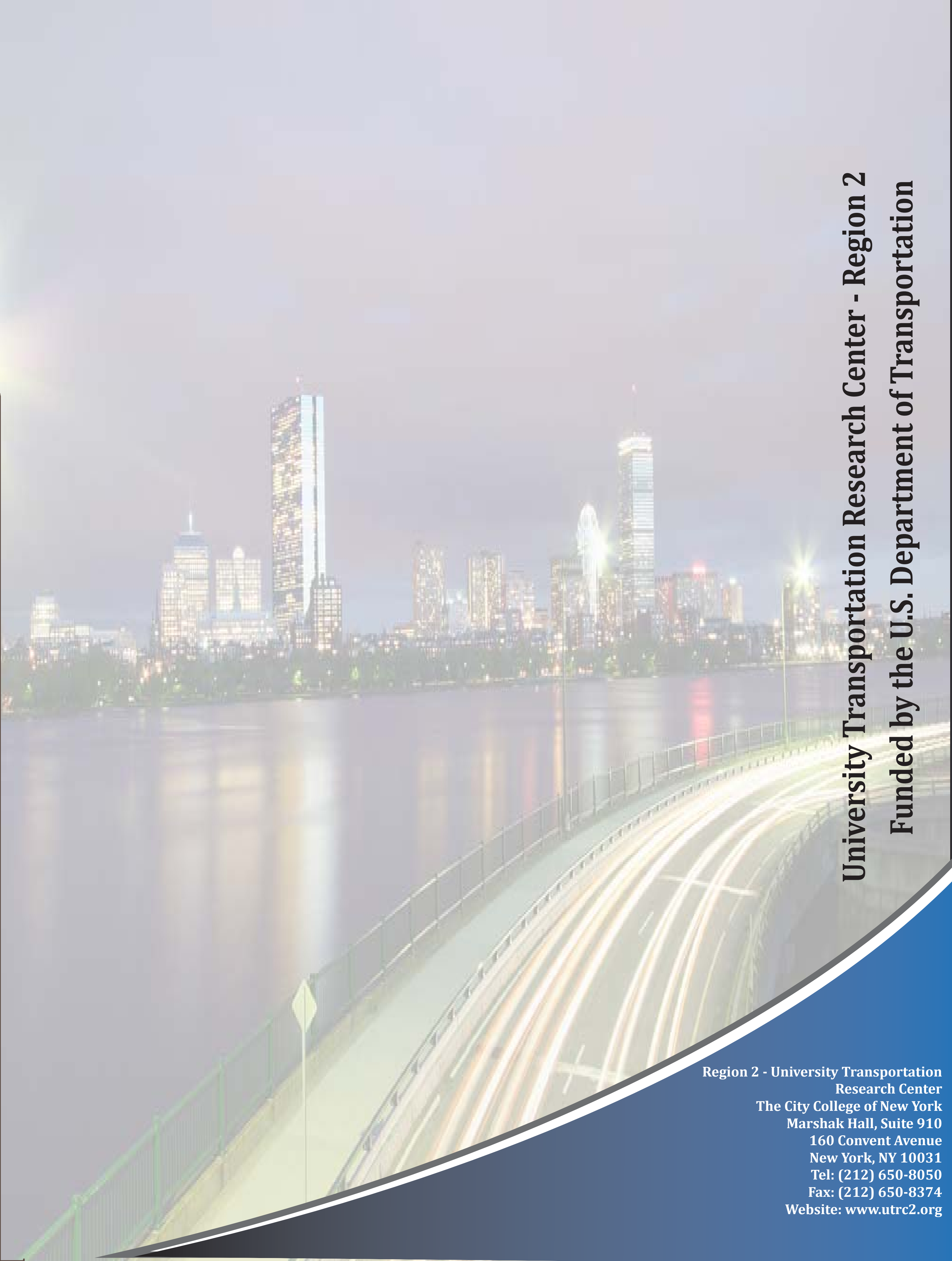
Matthew Dickens, 2016. Public transportation ridership report, Fourth Quarter 2015. Washington, DC.

Maze, T., Agarwai, M., Burchett, G., 2006. Whether Weather Matters to Traffic Demand, Traffic Safety, and Traffic Operations and Flow. Transp. Res. Rec. 1948, 170–176. doi:10.3141/1948-19

Moss, M., Qing, C., 2012. The Dynamic Population of Manhattan. New York University, New York.

Plummer, M., 2003. JAGS : A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling JAGS : Just Another Gibbs Sampler. Architecture 0.

Singhal, A., Kamga, C., Yazici, A., 2014. Impact of weather on urban transit ridership. Transp. Res. Part A Policy Pract. 69, 379–391. doi:10.1016/j.tra.2014.09.008



University Transportation Research Center - Region 2
Funded by the U.S. Department of Transportation

Region 2 - University Transportation Research Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: (212) 650-8050
Fax: (212) 650-8374
Website: www.utrc2.org