Agent-based Modeling for Recovery Planning after Hurricane Sandy

by

Elham Hajhashemi

Thesis submitted to the faculty of the Virginia Polytechnic and State University
in partial fulfillment of the requirements for the degree of

Master of Science

In

Civil Engineering

Pamela M. Murray-Tuite

Susan Hotle

Kevin Heaslip

Kathleen Hancock

August 2nd, 2018

Blacksburg, Virginia

Keywords: Agent-based modeling; hurricane Sandy; adaptation

Agent-based Modeling for Recovery Planning after Hurricane Sandy

Elham Hajhashemi

Abstract

Hurricane Sandy hit New York City on October 29, 2012 and greatly disrupted transportation systems, power systems, work, and schools. This research used survey data from 397 respondents in the NYC Metropolitan Area to develop an agent-based model for capturing commuter behavior and adaptation after the disruption. Six different recovery scenarios were tested to find which systems are more critical to recover first to promote a faster return to productivity. Important factors in the restoration timelines depends on the normal commuting pattern of people in that area. In the NYC Metropolitan Area, transit is one of the common modes of transportation; therefore, it was found that the subway/rail system recovery is the top factor in returning to productivity. When the subway/rail system recovers earlier (with the associated power), more people are able to travel to work and be productive. The second important factor is school and daycare closure (with the associated power and water systems). Parents cannot travel unless they can find a caregiver for their children, even if the transportation system is functional. Therefore, policy makers should consider daycare and school condition as one of the important factors in recovery planning. The next most effective scenario is power restoration. Telework is a good substitute for the physical movement of people to work. By teleworking, people are productive while they skip using the disrupted transportation system. To telework, people need power and communication systems. Therefore, accelerating power restoration and encouraging companies to let their employees' telework can promote a faster return to productivity. Finally, the restoration of major crossings like bridges and tunnels is effective in the recovery process.

Agent-based Modeling for Recovery Planning after Hurricane Sandy

Elham Hajhashemi

General Audience Abstract

Natural and man-made disasters, cause massive destruction of property annually and disrupt the normal economic productivity of an area. Although the occurrence of these disasters cannot be controlled, society can minimize the effects with post-disaster recovery strategies. Hurricane Sandy hit New York City on October 29, 2012 and greatly disrupted transportation systems, power systems, work, and schools. In this research, commuter behavior and adaptation after the hurricane were captured by using a survey data that asked questions from people living in NYC metropolitan area about their commuting behavior before and after Hurricane Sandy. An agent-based model was developed and six different recovery strategies were tested in order to find effective factors in returning people to normal productive life faster.

In the NYC Metropolitan Area, transit is one of the common modes of transportation; therefore, it was found that the subway/rail system recovery is the top factor in returning to productivity. The next important factor is school and daycare closure. Parents are responsible for their children, therefore; they may not travel to work when school and daycares are closed. The third important factor is power restoration. To telework, people need power and communication systems. By teleworking, people are productive while they skip using the disrupted transportation system. The final important factor is the restoration of major crossings like bridges and tunnels.

Dedication

To my family, Davoud Hajhashemi, Parvaneh Azarbayjani, and Arezou Hajhashemi, that without their love and support I would not be here. Thank you for always being there for me and encouraging me to pursue my academic ambitions.

Acknowledgement

I would like to thank my advisers Dr. Pamela Murray-Tuite and Dr. Susan Hotle for giving me the opportunity to work on this interesting topic and motivated me to enhance my skills and research capabilities. I am especially grateful to my advisers that made research enjoyable for me by their continuous support, kind suggestions and valuable comments.

I would also like to thank my committee members, Dr. Kevin Heaslip and Dr. Kathleen Hancock for their valuable suggestion and comments.

I really appreciate feedback given by the CRISP project team, Dr. Kris Wernstedt, Dr. Seth Guikema, and Dr. Edward Fox. . I would like to thank my friend, Navid Mirmohammadsadeghi for giving useful advice on the MATLAB codes.

Partial funding was provided by National Science Foundation grant CMMI – 1638207 CRISP Type 2 Coordinated, Behaviorally-Aware Recovery for Transportation and Power Disruptions. National Science Foundation grant CMMI – 1313674 RAPID: Commuter Adaptation to Transportation Disruption in Hurricane Sandy's Aftermath provided funding for the original survey that formed the basis for this thesis, for which the author is grateful. The content of this thesis does not necessarily reflect the views of NSF and the author is solely responsible for its contents.

Contents

List of Figures viii
List of Table ix
CHAPTER 1: INTRODUCTION 1
1.1 Research Objectives 2
1.2 Contribution 2
1.3 Outline 4
CHAPTER 2: LITERATURE REVIEW 5
2.1: Commuter Changes and Behavior during Disruption 5
2.2: Agent-Based Modeling 7
2.2.1: Agent-Based Modeling in Evacuation 8
2.2.2: Agent-Based Modeling in Demand for Transportation Systems 9
2.2.3: Agent-Based Modeling in Recovery Process 10
CHAPTER 3: DATA AND METHODOLOGY 12
3.1: Agents and their characteristics 12
3.1.1: Survey data 12
3.1.2: Census Data 18
3.2: Environment 19
3.2.1: Power Outage Data 20
3.2.2: School and Daycare Closure Data 22
3.2.3: Work Condition Data 23
3.2.4: Bridges and Tunnels Condition 25
3.2.5: Transit System 25
3.2.6: Post-Impact Policies 28
3.3: Agents' Behavior and Methods of Interaction 29
3.3.1: Logit Models 29
3.3.2: Decision Frameworks 32
3.3.3: Simulation Assumptions 45
3.3.4: Population Synthesis 46
CHAPTER 4: RESULTS 52
Chapter 4.1: Base Condition Results 54
Chapter 4.2: Comparing Scenarios 58

CHAPTER 5: CONCLUSION 64
References:.. 69
Appendix A: Method of Generating Population by PopulationSim 72

List of Figures

Figure 3. 1 Agent-based Model Components 12
Figure 3. 2 Home and Work Locations... 14
Figure 3. 3 Car Ownership Percentage 19
Figure 3. 4 Percentage of households without power in each day in New York.......................... 21
Figure 3.5 Percentage of households without power in New Jersey 22
Figure 3. 6 Work Closure percentage 24
Figure 3. 7 MTA Network 26
Figure 3. 8 LIRR Network 27
Figure 3. 9 MNRR Network 27
Figure 3. 10 New Jersey Rail road... 28
Figure 3. 11 Work Condition and Telework Flow Charts 33
Figure 3. 12 Issue with Childcare Flow Chart 34
Figure 3. 13 Rail and Subway Commuters Flowchart.. 35
Figure 3. 14 Mode Choice Model Flowchart... 38
Figure 3. 15 Car, Carpool or Taxi Commuter Decision Framework 40
Figure 3. 16 Bus Commuters Decision Framework.. 41
Figure 3. 17 Choose Departure Time Decision Framework 44
Figure 3. 18 Distribution of Depart Earlier (minutes) 44
Figure 3. 19 Distribution of Depart Later (minutes)... 45

Figure 4. 1 Transportation Mode of People in Normal Condition... 52
Figure 4. 2 Number of People Using Each Link in the Normal Condition-MTA Subway 52
Figure 4. 3 Number of People Using Each Link in the Normal Condition-MNRR 53
Figure 4. 4 Number of People Using Each Link in the Normal Condition-LIRR........................ 53
Figure 4. 5 Number of People Using Each Link in the Normal Condition-NJ Transit and Path rail
.... 54
Figure 4. 6 People Adaptation 55
Figure 4. 7 Commute Mode of People... 56
Figure 4. 8 Departure Time Distribution 57
Figure 4. 9 Percentage of People that Depart in each Time... 58
Figure 4. 10 Cumulative Lost Person-Work Days.. 58
Figure 4. 11 Number of People That Do Not Work on Each Day... 59
Figure 4. 12 Number of People That Telework on Each Day 60
Figure 4. 13 Number of People that Change Their Mode on each Day....................................... 61
Figure 4. 14 Number of People that Change Their Route on each Day 62
Figure 4. 15 Number of People that Change Their Departure Time on each Day 63

List of Table

Table 3. 1 Mode Usage for Survey Participants 16
Table 3. 2 Variables Definition for Income Model 17
Table 3. 3 Income Prediction Model 18
Table 3. 4 Population and household numbers in New York [30] 21
Table 3.5 Percentage of work closure 24
Table 3. 6 Bridges and Tunnels Timeline [2] 25
Table 3.7 Recovery Policies Timeline [2] 29
Table 3. 8 Telework Model 30
Table 3. 9 Mode Choice Model 31
Table 3. 10 Delay and Crowding News after Hurricane Sandy [[2],[44]] 46
Table 3. 11 Percentage of People Having Flexible Working Hours and Telework Option 48
Table 3. 12 Transportation Modes of People in Normal Condition 49
Table 3. 13 Transportation Modes of the Agents in the Normal Condition 50

CHAPTER 1: INTRODUCTION

Natural and man-made disasters, occurring at a growing rate [1], cause massive destruction of property annually and disrupt the normal economic productivity of an area. Although the occurrence of these disasters cannot be controlled, society can minimize the effects with post-disaster recovery strategies.

This study evaluates the events surrounding Hurricane Sandy and its effects on household-level economic productivity. Hurricane Sandy, a recent high-impact natural disaster that was termed "Superstorm Sandy" due to its intensity, affected 24 states in some form with severe damage predominately in New Jersey and New York that caused simultaneous mode disruption. The storm hit New York on October 29, 2012 and greatly disrupted the transportation systems (including flooding streets, tunnels, bridges and subway lines) and power systems needed for the region to be economically productive [2].

The main goal of disaster recovery and this study is to "restore households, business, and government activity to the 'normal' patterns that existed before the disaster struck" [3] as quickly as possible for the community. Returning to productivity, for the purposes of this study, means the "ability to work a full day" for a given job. Some jobs require one to be physically present while others allow employees to work remotely. Those that must be present need the transportation system to travel, while those working remotely need both the power and communication systems to be in working order; therefore, transportation system recovery and power recovery are critical factors in the context of post-disaster recovery.

Recovery is a dynamic process that depends on many intertwined factors, including the environment, behavior and previous experiences of individuals, and their chosen methods for adaptation after disaster. Therefore, effective recovery planning requires city officials to have a deep understanding of this dynamic nature [4]. After each disruption, people try to adapt themselves to a new situation by changing their behavior and using what is available of the disrupted system. There have been several studies about people's behavior after disruption that have used survey approaches and statistical models. However, statistical models alone, like logit models, do not have the ability to capture dynamics for different scenarios over time (i.e., how past behavior affected the environment, which affects future behavior) and check the effects of small changes on the overall process of recovery. Agent-based models are capable of simulating time-based situations that are complicated and dynamic; therefore, statistical models may be used in conjunction with agent-based models for identifying significant factors in the recovery process.

1.1 Research Objectives

In this research, an agent-based model is developed based on a telephone survey in the New York City Metropolitan Statistical Area in January 2013 that includes questions about pre- and post-Hurricane Sandy commuting patterns and basic sociodemographic characteristics.

This research presents an agent-based model for capturing people's behavior and adaptation after Hurricane Sandy and specifically addresses how different recovery scenarios affect the timeframe of when people return to a productive state. That is, what policies can officials and agencies implement that promote the return to productivity earlier following a disaster? Moreover, this agent-based model helps in understanding the commuter decision-making process relative to the environment and interactions with other agents by identifying significant factors that define people's behavior.

The objectives of this research are to:

- Develop an agent-based model for commuters' adaptation in Hurricane Sandy's aftermath for survey respondents and the larger population of New York and New Jersey commuters using the affected transportation systems.
- Test different timings for restoration activities and recovery policy scenarios in order to find effective factors for returning to productivity. That is, which systems are most critical to recover first to promote a faster return to productivity?
- Identify data and information needed for developing an improved agent-based model in future research

1.2 Contribution

Transportation systems and power systems are tools that help people meet their needs. Therefore, recovery strategies can be more effective by considering people's behavior and adaptation. There is existing literature about capturing people behavior but very limited studies have used these finding in order to improve recovery processes. People are highly adaptive and after disruption, they could change their commuting patterns in order to meet their needs. Available recovery strategies do not completely account for adaptation. However, after big disruptions like Hurricane Sandy, many different factors affect people's decisions. For instance, parents are responsible for their children and this responsibility may cause them to cancel their work trip even if the transportation system has completely recovered. Therefore, it may be better to put more effort to recover daycare/schools and transportation systems at the same pace. Moreover, telework may be a good substitute for travelling to work after disruption because people can skip traffic, delay and crowding this way. Also, subway/rail system recovery does not have the same importance in an area like New York where people are highly dependent on subway/rail and Houston where the rail/subway has fewer commuters; so it is important to consider people's preferences and the availability of transportation options. Prioritizing recovery of transportation systems can be more realistic and effective by considering people's adaptation and preferences. To do so, first there is a need in understanding how people react to different kinds of disruption. There have been studies about capturing people's behavior by using survey approaches, but most of these studies only conclude about how people's behavior changes after disruption and have not used these findings for capturing the effect of these behaviors on recovery processes.

In this study, a combination of statistical models and if then rules are used in an agent-based model framework to model the condition of transportation and power systems after the disruption and simulate people's behavior and adaptation in this situation. By using this agent-based model, the effect of different recovery strategies on the area's overall productivity is examined. Most of the previous research prioritizes recovery based on available resources and budget without considering people's adaptation. This research is addresses this gap.

Moreover, the modeling approach in this research overcomes the limitations of previous studies about commuters' behavior after disruption by building an agent-based model that considers route and departure time choices for each agent while capturing changes in daily travel patterns. Moreover, in the model created for this thesis, commuters learn from their previous travel experience and adjust their travel decisions based on this previous experience.

1.3 Outline

This thesis has five chapters. Chapter 2 presents the literature review, which has two parts. The first part is about commuter adaptation and behavior during a disruption and the second part is about agent-based modeling applications in different areas of transportation. Chapter 3 is about the data and methodology. In the first part of Chapter 3, different components of the agent-based model and data that are needed for model development are discussed. Then in the second part, agents' behavior and methods of interaction are explained. Chapter 4 presents results of the base condition model (i.e., what happened in reality) and different scenarios regarding system recovery, followed by a discussion to compare their outputs. Chapter 5 includes the conclusions and recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

This chapter presents a literature review in two parts. The first section presents research associated with the behavior of commuters during different types of disruptions like hurricane, earthquake, bridge collapse and workforce strikes in public transportation. The second section is about the advantages of agent-based models and their applications in transportation.

2.1: Commuter Changes and Behavior during Disruption

Existing literature includes several studies about the recovery process and people's adaptation in response to each disaster. For example, Kontou, Murray-Tuite, and Wernstedt [5] conducted a telephone survey in January 2013 of residents in the New York metropolitan area. The survey included questions about regular commuting patterns, post-hurricane commuting patterns, and disruptions that affected commuters and their socio-demographic characteristics. They developed five multi-variable binary logit models for changing mode, canceling a work trip, changing route and changing departure times (earlier or later) for home to work trips. Based on these models, having transit as the primary mode of transportation increased the likelihood of people canceling their work trips, changing modes and departing earlier compared to commuters normally using other modes. People who are able to telecommute in a normal (i.e., undisrupted) situation are more likely to cancel their work trips and less likely to depart earlier from home to work. Women tend to be less likely to change modes or depart later than men. Families with more children are more likely to cancel their work trips, and people who encounter daycare or school closures are more likely to change their routes. Besides these characteristics that predict the different adaptations for different commuters, the environment is also important. For instance, tunnel closures cause people be more likely to cancel their trips and delay and crowding increases the probability of changing routes and departing earlier.

Based on Giuliano and Golob's [6] findings about commuter adaptation after the Northridge earthquake, commuters are more likely to shift their routes and departure times rather than changing their modes. Even the probability of canceling the trip is higher than that for changing modes. In addition, the authors noticed that for people living in impacted areas, temporarily changing residential location is more probable than changing modes. Changes that commuters made during the freeway reconstruction after the Northridge earthquake were temporary and commuters returned to their normal routine when reconstruction finished.

After the I-35W Mississippi River bridge collapse, Zhu et al. [7] used survey data and traffic counts to study aggregated travel demand changes and commuter adaptations. Based on people's responses to the survey, changes in travel patterns from most common to least common are change departure time, change route, choose an alternative destination for their activity, cancel the trip, telecommute and change mode. The authors believed that changing travel mode is harder and people are less likely to change their modes because car ownership and service availability constrain it. There was not a significant difference in total demand since commuters mainly changed routes or departure times rather than canceling trips and these changes only modify the trip demand distribution not total demand.

Mokhtarian, Ye, and Yun [8] studied the effects of a major freeway reconstruction project on commuter behavior by using data from two internet-based surveys. The authors developed a binary logit model to identify factors associated with the increased use of transit during a disruption. In agreement with previous studies, results of this study also indicate that people are more likely to change their departure times and routes than other changes like modes. Only 8 percent of respondents changed their mode of transportation while 48 and 45 percent changed their departure times and routes, respectively. Women were more likely to use vacation days during a disruption in comparison to men. Also, they were more likely to change their departure time and carpool in comparison to men. The estimated binary logit model showed that persuading people who already use transit to increase their ridership is much easier than persuading non-transit users to start using transit as their mode of transportation.

Van Exel and Rietveld [9] reviewed 13 studies about workforce strikes in the public transportation systems. They found that mode choice is highly dependent on factors like car ownership and a person's work and home locations. Most of the people shift to the car if it is possible for them, but many commuters without alternative transportation modes cancel their trips.

Small changes in the transportation network cannot cause significant differences in behavior since commuters usually follow their routines and there is a need for significant disruption in order to disrupt habitual behavior [10]. Most of the disruptions to transportation networks only affect one small part of the system and situations like Hurricane Sandy that affect all modes of transportation simultaneously in an area like New York are rare.

Based on Levinson and Zhu's [10] review of 16 papers about behavioral responses to transportation network disruptions, there are some limitations in these studies:

1) Although there are many papers about exploring travel behavior after disruption and all of them have concluded that changing routes and departure times are the first two common adaptations for people, most of them did not provide a good description of the route and departure time choices.

2) Most of the surveys only report commuter adaptation like changing routes, departure times, or modes but rarely combined these changes; in reality, one person may make several different changes in their commuting pattern (e.g., depart earlier and change routes).

3) Moreover, studies in the current literature have shown that experience impacts travel decisions. However, surveys are not completely capable of capturing travel patterns over time. Most of the studies about travel during disruption did not include this experience and learning process in their modeling approach.

This research expands on existing literature by building an agent-based model that considers route and departure time choices for each agent while capturing changes in daily travel patterns. Moreover, in the model created for this thesis, commuters learn from their previous travel experience and adjust their travel decisions based on this previous experience.

2.2: Agent-Based Modeling

The recovery process is dynamic and complex as it is impacted by the behavior and adaptation choices of surrounding individuals. Therefore, agent-based modeling is one of the best ways to realistically model disaster recovery situations.

Agent-based modeling is an approach that simulates a group of autonomous decision-makers that interact with each other and the environment surrounding them based on a set of rules [11]. The advantages of agent-based modeling include:

1) By modeling each agent's behavior and interactions with other agents, it is possible to observe behavior that is based on the interaction between agents and the environment. Also, characteristics that lead to different behaviors can be captured by agent-based modeling [12].
2) Agent behavior can be specified by simple rules defined by if-then statements or statistical models like Multinomial logit models (MNL), neural networks or genetic algorithms [12].

3) The individual's behavior is affected by personal characteristics and environmental situations. In order to predict human adaptation after each disruption (e.g., evacuation or recovery modeling), human interaction with the environment is essential. The environmental component in an agent-based model, including the impact of behaviors of other commuters, allows for more accurate simulation of this complex system [13].

Agent-based modeling is becoming more popular in many fields like ecology, computer simulation, biology etc. [14]. Bernhardt [15] wrote an article about the use of agent-based modeling in different aspects of transportation like highway traffic [16], pedestrian movement [17] and demand modeling [18]. Based on Bernhardt's [15] conclusion, agent-based modeling is an appropriate approach for modeling transportation related problems. Agent-based modeling is specially an efficient option in any research area where human decision making is important and can cause significant differences in total system function [19].

2.2.1: Agent-Based Modeling in Evacuation

Agent-based modeling is becoming more popular in evacuation modeling.

Yin et al. [13] studied travel demand and decision-making behavior of people throughout an evacuation by using an agent-based modeling approach. In this model, each household makes six different evacuation decisions: whether to evacuate or stay, accommodations if they decide to evacuate, destination, mode, number of vehicles, and departure time. A post-Hurricane Wilma, a hypothetical hurricane in Miami, and post-Hurricane Ivan telephone surveys were used for developing model components. The agents in this model were the households with the behavior described by different econometric and statistical models. The models were incorporated into a case study in a Miami-Dade area for a hypothetical hurricane. In order to develop a disaggregate population, the population synthesizer in the TRANSIMS package was used [13].

Chen, Meaker, and Zhan [20] developed an agent-based model with the VISSIM microscopic simulation package to find the minimum evacuation time for the Florida Keys. In the evacuation process, each person needed to make decisions regarding departure time and routes. Selection of departure time and route depended on congestion and other factors. Similarly, congestion occurred as a result of people's departure time and route choice. For instance, if everyone who wanted to evacuate decided to depart at the same time and travel the same route, then that route would be congested. Agent-based simulation adds greater understanding of how this loop-of-causality can influence overall group behavior [20].

Chen and Zhan [21] developed an agent-based model for comparing staged and simultaneous evacuation strategies by using Paramics (microscopic simulation system) in different types of road networks and population including ring road, grid road and real road structure. The agent-based model captured the behavior of a group of agents that is hard to capture in aggregated models [21].

Lamel and Klupfel [22] developed an agent-based model for evacuation process in the city of Hamburg. In this studied they modeled 1500 artificial agents that wanted to evacuate from Hamburg by using MATSim toolkit. Two different scenarios were compared with each other. In the first scenario, people would evacuate immediately after receiving evacuation notice and in the second one, they depart in a time frame within two hours from evacuation notice therefore each person can have different departure time while in the first scenario all people depart at the same time. Based on their finding in second scenario the overall evacuation process is better and faster [22].

2.2.2: Agent-Based Modeling in Demand for Transportation Systems

Many other studies have dealt with growing demand for transportation system problems by using agent-based modeling technics:

Nam et al. [23] modeled each individual as an agent to study transportation demands, in Sydney, Australia. The TRANSIMS simulator was used for calculating the travel time and density of each road in that area. For collecting agent characteristics, the authors used travel diaries that reported sequences of daily trips, mode of transportation, and purpose of each trip and departure time for each responding individual. Each agent encountered several decision-making processes. Decisions were based on agent characteristics and environmental conditions. A MNL was used to model agents' decisions about relocation, and another MNL model was used for transportation mode choice. The simulation used a synthetic population from census data. This agent-based model showed how existing transportation infrastructure is used based on current population and how future transportation demand can be calculated based on agent characteristics, land use, and environmental change [23].

Rossetti et al. [24] evaluated road congestion problems through maximizing use of current transportation systems' capacity by changing user behavior patterns. The agent-based model approach used by the author, modeled route and departure time choices for drivers based on the traveler experience and learning process while accounting for the uncertainty inherent in people's behavior.

2.2.3: Agent-Based Modeling in Recovery Process

Some studies have used agent-based modeling in areas related to the recovery process and its impacts:

Nejat and Damnjanovic [4] used an agent-based model for home reconstruction after disasters that cause home damage. Homeowners were the agents whose decisions were whether to reconstruct their home now or wait, conditioned upon game theory and their neighbor's activity including reconstruction and relocation. For developing this model, NetLogo, Google Earth and GIS were used.

Grinberger and Felsenstein [25] developed an agent-based model using the Repast Symphony platform to test the effectiveness of policy choices in the restoration of urban equilibrium after a hypothetical earthquake in Jerusalem, Israel. Agents were individual citizens that decided their residential locations and activity participation. The environment included the buildings that are commercial or residential. On each day, agents made two decision. The first decision is about residential location; they can decide to move to a new location within the city or out of the city. Next, each agent may participate in up to three activities daily all of which are located in one of the buildings in the study area. Based on agents' decisions, land use can change from residential to commercial, commercial to unoccupied, residential to unoccupied, or unoccupied to residential, and all of these lead to a new urban equilibrium. Based on this research, it is not always easy to return to the pre-disaster situation and, sometimes a new equilibrium may arise after a disaster like an earthquake.

Srikukenthiran et al. [26] used the Nexus platform to simulate short-term disruption to a transit network and test different handling strategies. The modeling area was in Toronto with an artificial problem that causes some delay in the transit system, and passenger movement. Behavior and crowdedness was modeled to test different response strategies. Based on this research, in short-term disruptions (less than 30 minutes) without any intervention or with simple solutions like asking passengers to find alternative routes, it is possible to decrease crowding levels.

All of these studies show how agent-based models are capable of simulating real-life situations that are complicated and dynamic. Each person's decision and small changes in recovery measures and the environment can have significant impacts on the timeline of recovery and returning people to productivity. Alone, statistical models, like logit models, representing people's adaptation after a disruption do not have the ability to capture dynamics for different scenarios and check the effects of small changes on the overall process of recovery. However, these models may be used in conjunction with agent-based models.

Previous agent-based models have been developed for evacuation processes, longer-term demand changes, travel behavior, and housing recovery after disruption. A limited number of studies have tried to use agent-based modeling in the area of behavioral response to the major transportation system disruptions. Therefore, in this research, a combination of statistical models and an agent-based model are used to capture people's travel behavior and adaptation in response to Hurricane Sandy's impacts for the first nine working days after Hurricane Sandy. By using an agent-based model, it is possible to examine how people deal with the post-hurricane situation and capture their decision-making process relative to the environment and interactions with other agents. This model helps to understand how people's behavior can change based on the environment and how different people in the same environmental situation can act differently based on their family and personal characteristics. Finally, with this model, it is possible to examine the effect of the restoration timeline and its impact on the overall commuting pattern.

CHAPTER 3: DATA AND METHODOLOGY

Each agent-based model has three major components: 1. Agents and their characteristics 2. Environment 3. Agent behavior and methods of interaction [12]. The first step in the model development is defining each of these components and finding all the needed model inputs with regard to Hurricane Sandy. Figure 3.1 presents an overview of the agent-based model components in this study.

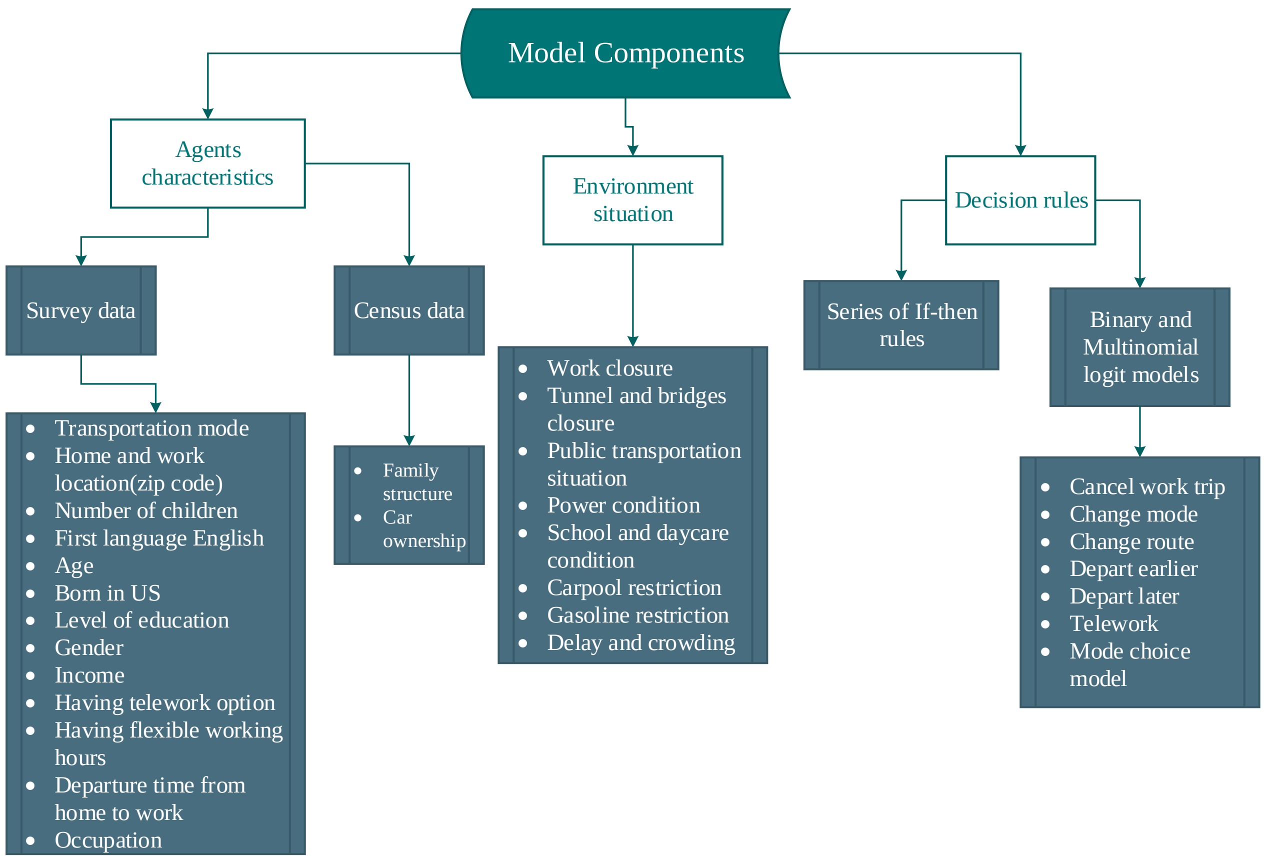


Figure 3. 1 Agent-based Model Components

3.1: Agents and their characteristics

In this model, agents are people living in the New York City metropolitan area that are employed and commute from home to work at least once per week.

3.1.1: Survey data

This study used previously collected data from a post-Hurricane Sandy telephone survey designed to explore how residents of the New York City metropolitan area changed their commuting behavior in response to disruptions in every mode of transportation. This survey was conducted in January 2013 with residents of 23 counties within New York City Metropolitan area as the survey region. This data was used in the present study to define the agent characteristics and behavioral responses when adapting to Hurricane Sandy's disruptions.

This survey included 31 questions about pre-hurricane normal commuting patterns, basic socio-demographic characteristics, post-hurricane commuting patterns and how their commuting changed after the hurricane until the time that the transportation system returned to the normal pre-hurricane situation. People changed their usual commuting patterns because of the disruptions and their adaptations to the disaster were measured in six ways: change route, change mode, change departure time (depart earlier or later from home to work), telework, and cancel the work trip. There are 397 records available from the survey data that are used for developing the agent-based model. More information about this survey data is available in [5]. Variables from the survey that are used in this study include home and work zip codes, transportation mode, age, income, gender, number of children, level of education, occupation, departure time from home to work, having the option of flexible working hours in normal situations, having the option of teleworking in normal situations, whether they were born in the US and whether their first language is English. Minor adjustments to the data are described in the following subsections.

3.1.1.1: Home and work location zip code

The home and work locations of agents from the survey are shown in Figure 3.2. The red triangles are home locations, and navy circles are work locations. As shown in the figure, home locations are more dispersed than work locations. There are many work locations in the Manhattan area.

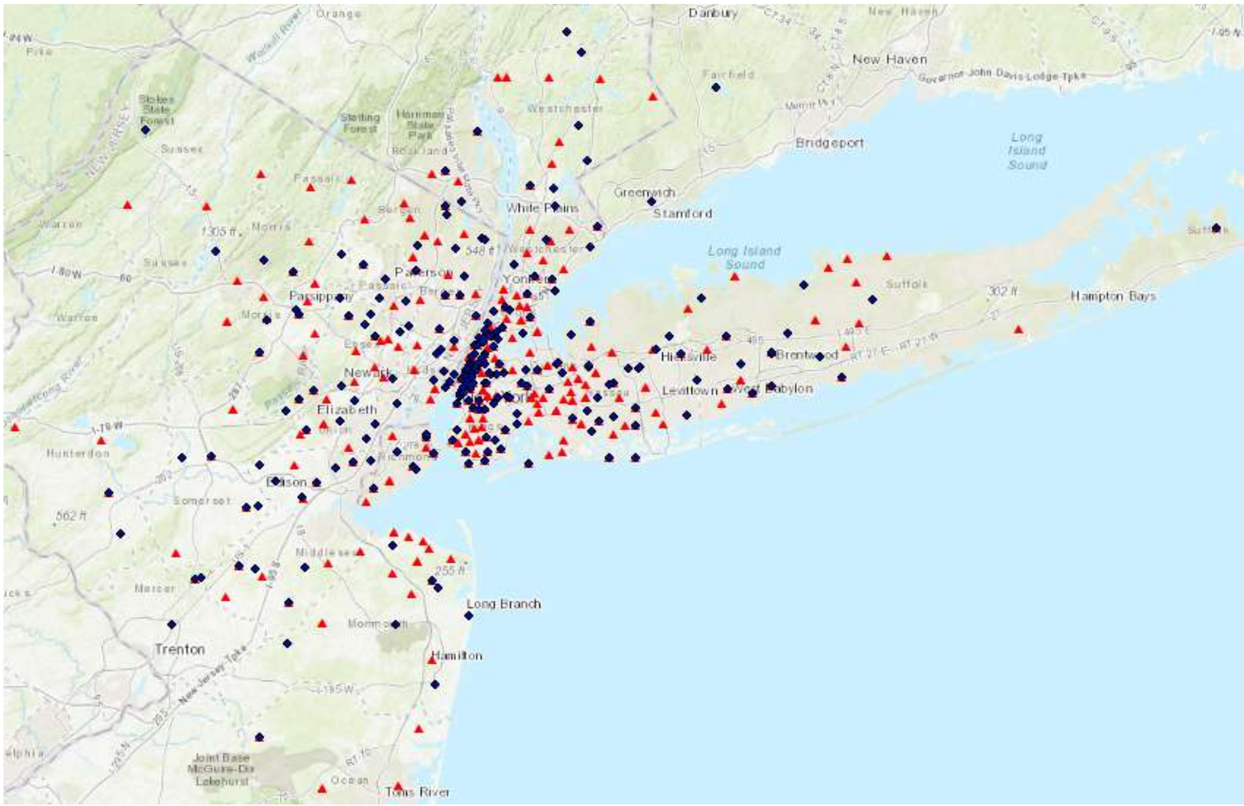


Figure 3. 2 Home and Work Locations

Home and work location zip codes included some missing responses. There are three types of missing zip codes:

- *People who did not answer at least one of the questions about their home and work zip codes.* If only one of their home or work locations is known, then based on their mode of transportation and the duration of their trip between home and work, a plausible zip code replaced their missing zip code. Arc GIS and Google Maps were used to find missing zip codes for the transit commuters based on the transit network layer and known home or work location. If their main mode of transportation is subway or rail, it is assumed that this person used the closest subway or rail line for moving from home to work. However, they can move in two different directions on that subway line to reach the unknown zip code location. If, by moving in any of these directions for the duration of their travel time from home to work, they pass any of the bridges and tunnels their way, their answer to the question “Was your commute affected by a tunnel closure or carpool restriction on bridges?” helped choose the correct direction. If bridges and tunnels are on their way, they move on the subway or rail line toward the bridge and tunnels, and, based on the travel time, the missing zip code is estimated. Otherwise, they move in the opposite direction of bridges and tunnels. Since most of the missing zip codes are for work location and number of work location increases while moving toward Manhattan area, for people that there is not a bridge or tunnel on their way, the direction toward Manhattan is assumed to be correct direction. For car commuters and other modes of transportation, Google Maps was used to find missing zip codes using a similar method.

- *People whose home and work locations are both missing or at least one of the zip codes and the main mode of transportation is missing.* These observations were omitted from our study dataset. Location is one of the important components for the agent-based model in this thesis and the home and work zip codes were needed. Therefore, 14 respondents with missing home and work zip codes were omitted from our dataset and the total number of observations was reduced to 383.
- *People who gave their zip codes, but one zip code is not located close to either the New Jersey and New York area.* For instance, some zip codes were for Stockholm and Boston. Since it is more likely that people made a mistake in stating their zip code in comparison to the mode of transportation that they use and their trip duration, the given zip code was assumed to be wrong and a new zip code was assigned to this person in a similar procedure as mentioned above for the missing zip codes.

Zip codes for home and work were converted to latitudes/longitudes of the zip code centroid, and then these latitudes/longitudes were converted to x y coordinates. The Euclidean distance from the home to work location is calculated based on x y coordinates for each person. Based on home and work location zip codes, the counties in which each person lives and works are found as well.

3.1.1.2: Transportation mode

In the survey, respondents were able to report their selected mode(s) of transportation. One person could choose more than one mode of transportation; therefore, the sum of the numbers in Table 3.1 is more than 383. For instance, if one person reported both car and MTA subway as his or her selected mode of transportation, this person is counted in both car and MTA subway groups in Table 3.1. The distribution of the different mode choices is presented in Table 3.1.

A variable was created indicating whether the respondent was a transit commuter. This variable represents people that use one or more of the transit systems including: New York City (MTA) Bus, New Jersey Transit bus, New York City subway (MTA), PATH rail, Long Island Railroad (LIRR), Metro-North Railroad (MNRR), New Jersey Transit rail or any other rail or bus system.

If a person reported more than one transportation mode, one of the below situations occurs:

- One of these modes is transit, and the other one is car, carpool, taxi, walk or bike. In this case, it is assumed that this person has used the non-transit mode for reaching the transit station and the main mode of transportation is assumed to be transit.
- Both of the modes are rail and subway. In this case, both of the modes are considered as a primary mode of transportation.
- One of the modes is bus and the other is rail or subway. In this case, the mode in which they have spent most of their time is assumed to be the main mode of transportation.

Table 3. 1 Mode Usage for Survey Participants

Mode of Transportation	**Number of observations**
Car	190
Carpool	18
MTA subway	100
MTA bus	53
New Jersey transit bus	6
New Jersey transit rail	9
MNRR	10
LIRR	12
Path rail	4
Another rail	1
Another bus	5
Taxi	7
Ferry	5
Bike	1
Walk	15
Other	7

3.1.1.3: Age

There were 35 missing values for age. The mean substitution method from [27] was used for dealing with missing data. Missing values were replaced with the mean of observed ages.

3.1.1.4: Income

Income included 106 missing values. A simple linear regression model was developed by using RStudio software to predict missing income values. Variables that were used in this model are shown in Table 3.2 and the model is shown in Table 3.3.

Table 3. 2 Variables Definition for Income Model

Variables	**Definition**
Level of education	College and above 1, 0 otherwise
Age	Continuous (years old)
Log Age	Logarithm of age
Occupation	If work in computers, engineering, science, management, business, and financial 1, 0 otherwise
Gender	Female 1, 0 otherwise
US_English	If born in US or first language English 1, 0 otherwise
County group I	If live in counties that their average income is more than $125,000 1, 0 otherwise
Travel cost	If cost of travel from home to work more than $20 1, 0 otherwise
County group II	If live in county that average income is less than $80,000 1, 0 otherwise

Table 3. 3 Income Prediction Model

Independent variables	β	Pr(>|t|)
Intercept	-89628.04	0.02
Level of education	44810.1	0
Age	6685.91	0
Log Age	-68.93	0
Occupation	24816.17	0.001
Gender	-18457.22	0.004
US_English	21591.59	0.049
County group I	22455.74	0.002
Travel cost	45006.2	0
County group II	-25633.31	0.002
Multiple R-squared	0.3941	----------
Adjusted R-squared	0.3733	----------
p-value	< 2.2e-16	----------

3.1.2: Census Data

Car ownership and family structure are other variables needed as agent characteristics in the modeling process, but the survey data did not include information about these two variables. Therefore, the census data was used for defining these variables.

3.1.2.1: Car ownership

Percentages of car ownership in New York and New Jersey were collected from census data and are presented in Figure 3.3. A random number was generated for each agent, and, based on their home county, if the random number was less than the percentage of car ownership for that county, that person was assumed to own a car, otherwise that person did not own a car.

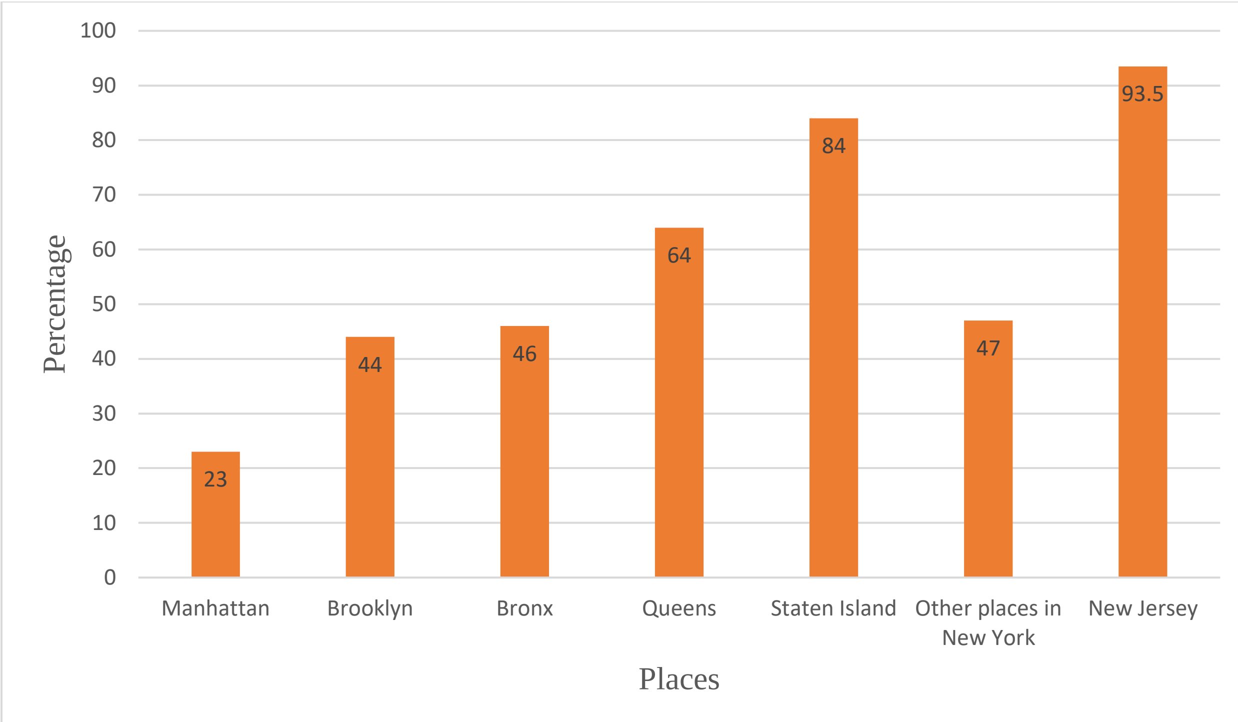


Figure 3. 3 Car Ownership Percentage

3.1.2.2: Family structure

The family structure is an important factor in predicting agents' behavior after Hurricane Sandy. Daycare and school closures can have more effect on single parents and dual-career households in comparison to the married families where only one of the parents works. Detailed data on the employment of other household members were not part of the original survey. Data about marital status and the number of working people in each household were obtained from the census data. In the US, 68 percent of families with children under age of 18, are married couples and among these married-couple families, 61.1 percent had both parents employed [28]. Based on this percentage, the number of families that are married couples with both parents working was calculated.

3.2: Environment

Hurricane Sandy hit New York City on October 29, 2012, and significantly disrupted the transportation and power systems. The modeled environment includes the condition of power, school and daycare, transit system, bridges, tunnels, workplace, and policies like carpool restrictions and gasoline restrictions.

3.2.1: Power Outage Data

Although teleworking relies on power and communications, this survey did not include questions about the power condition of each household. To obtain the power outage data, the residential zip code of each household was used to determine which power company provided service for this household. The number of customers without power in each service area was collected from official websites and reports from each company. Based on the total number of customers, the percentage of people without power in each service area was obtained. These percentages were used to determine which households were without power in the modeled population for each day. A random number was generated for each household, and, based on the power provider, this number was compared with the power outage percentage daily. If this random number was less than the percentage of people without power, that household was assumed to be without power. Otherwise, that household had electricity.

Survey respondents live in either New York or New Jersey. In New York, Con Edison provides electric distribution to all five boroughs except the Rockaways, which are served by the Long Island Power Authority (LIPA). Power outage data for the five boroughs, Westchester and Long Island were obtained from a report [29]. This report represents the total number of customers without power for each day in New York, and each customer was assumed to be a household. The total number of households in New York was needed in order to calculate the percentage of households without power on each day. Based on the census data, there are 2.63 persons per household on average in New York [30]. By using this number and the population of the five boroughs, Westchester and Long Island, the total number of households and percentage of households without power on each day were calculated so that these percentages can be used for finding the total number of households without power in our modeled population. Table 3.4 and Figure 3.4 represent the population and percentage of households without power in New York, respectively.

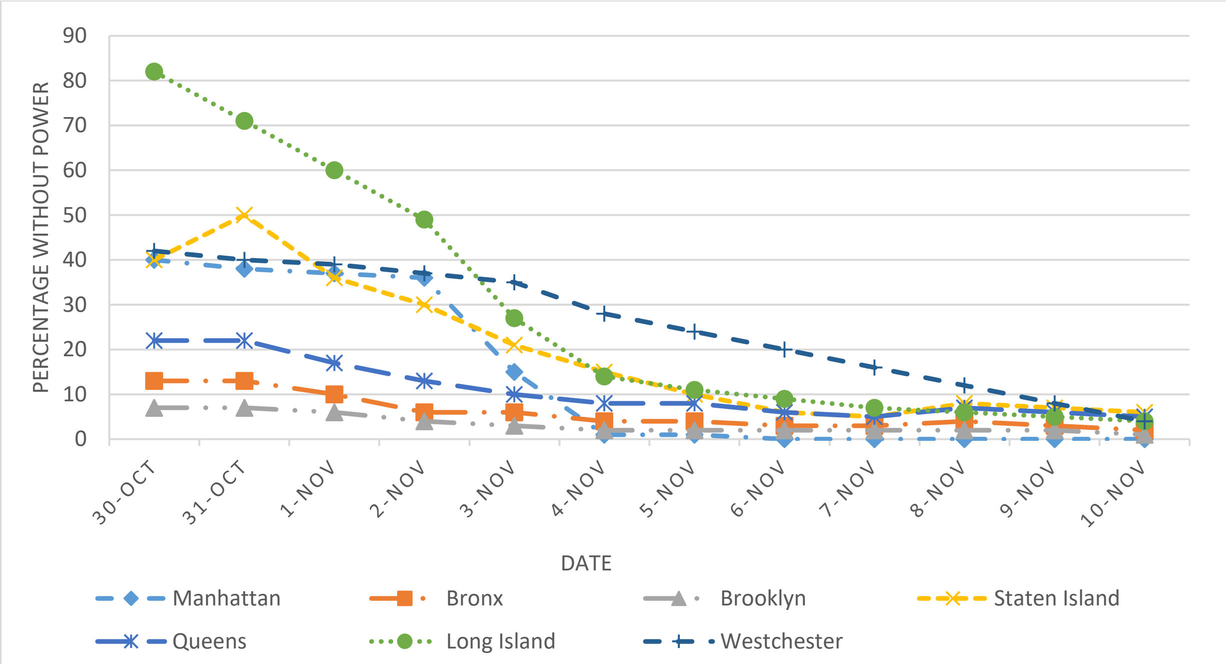


Figure 3. 4 Percentage of households without power in each day in New York

Table 3. 4 Population and household numbers in New York [30]

Borough	**Population**	**Number of Households**
Manhattan	1,643,734	625,000
Bronx	1,455,720	553,510
Brooklyn	2,629,150	999,680
Queens	2,333,054	887,100
Staten Island	476,015	181,000
Long Island	2,863,000	1,088,600
Westchester	976,369	371,243

There are four different power provider companies in New Jersey, each of which serves different counties. These four companies are PSE&G, JCP&L, Orange & Rockland and Atlantic City Electric. Power outage data for each of these companies were obtained from reports [31], [32],

[33], and [34] and the total number of customers were found on each of these companies' websites [35], [36], [37], and [38] . Each customer is assumed to be a household; therefore, the percentage of households without power is calculated. Figure 3.5 shows the percentage of households without power in each service area.

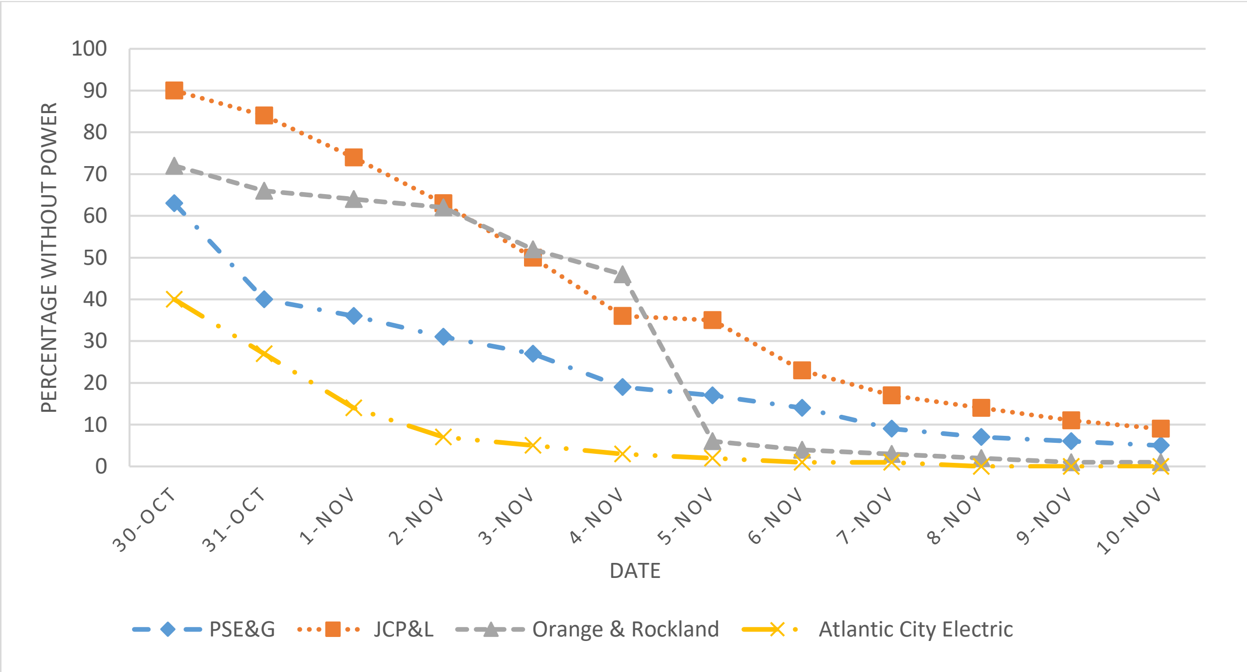


Figure 3.5 Percentage of households without power in New Jersey

3.2.2: School and Daycare Closure Data

Out of 397 respondents, 181 indicated that they had to cope with daycare and school closures in the aftermath of Hurricane Sandy. Therefore, the condition of daycare and schools could be used as an important factor in predicting the behavior of each agent after Hurricane Sandy.

No report was available about the daily operating conditions of schools after Hurricane Sandy, but limited information was available from websites [[39],[40]] about the percentage of school and daycare closure for some specific days. Based on the known percentage, hypothesized values were assigned to other days for the percentage of school closures. All public schools were closed for a full week in New York City. November 5th was the first day of school after Hurricane Sandy, and most of the schools reopened in their normal locations. However, 86 schools remained closed. By November 10, almost all of the schools were reopened either in their normal locations or an alternative location [39]. By November 5, in New Jersey, 18 percent of schools were still closed.

Almost all schools reopened by November 13 [40]. For each household with children under the age of 15, a random number was generated and this number was compared with the percentage of daycare and school closures on that day; if the random number was less than the school closure percentage, it was assumed that school was closed for the children in the household. Otherwise, school was considered to be open.

3.2.3: Work Condition Data

For agent-based modeling, the daily condition of the workplace for each agent is needed because if the workplace is closed, there is no need for the agent to travel to work and they can either telework or not work at all. There is not a direct question about the daily work conditions in the survey. In addition, there is not any information about the percentage of closed and open offices online. Therefore, a combination of three questions in the survey was used to figure out the percentage of closed work locations every day after Hurricane Sandy. The first question is "Hurricane Sandy hit the New York City metropolitan area on Monday, October 29, 2012. Did the Hurricane affect your work schedule?" The second question is "during the days that you did not work your normal schedule and at your normal location was: Your work closed or your normal work hours changed?' If respondents answered yes to both of these questions, they answered the third question that is "On what day did you return to your normal work schedule and location after October 29th?" Although these three questions did not specifically ask about work closure, they can help lead us to the percentage of work closures. Hurricane Sandy affected the work schedule of 244 people, and they did not work their normal schedules because of work closures for some days. Table 3.5 shows how many of these people return to normal work conditions on each day after Hurricane Sandy. Finally, based on 397 total respondents, the percentage of work closure on each day was defined. Figure 3.6 presents these percentages.

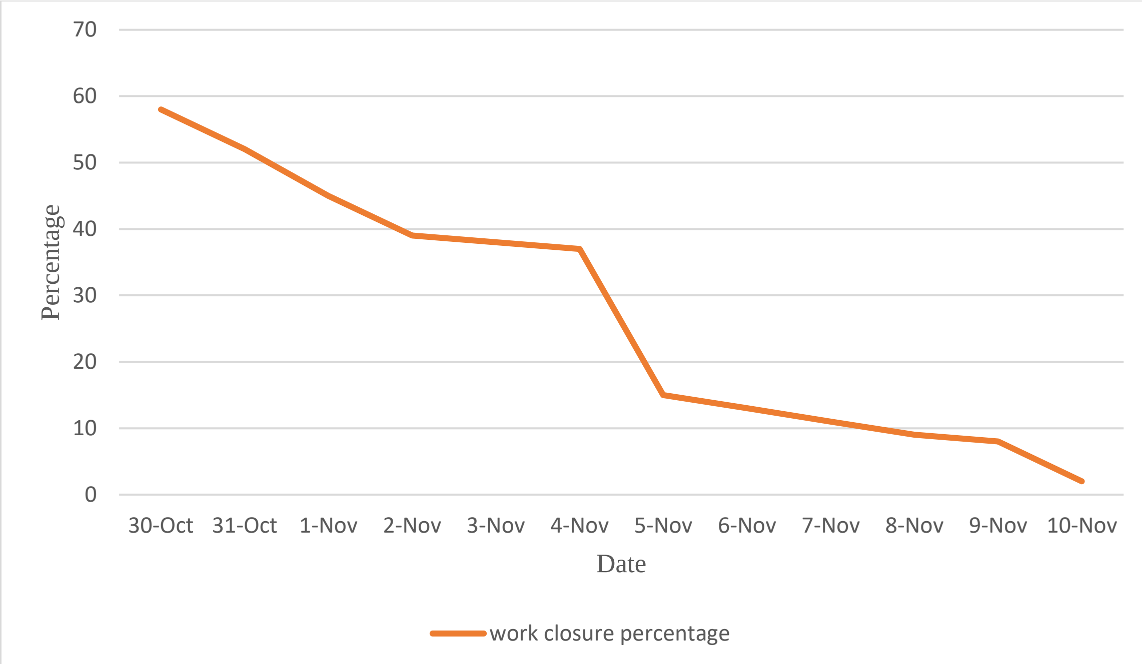


Figure 3. 6 Work Closure percentage

Table 3.5 Percentage of work closure

Day	# returning to normal work conditions	# work closed	% work closed
Oct 30	13	231	58
Oct 31	24	207	52
Nov 1	27	180	45
Nov 2	22	158	39
Nov 3	6	152	38
Nov 4	3	149	37
Nov 5	88	61	15
Nov 6	8	53	13
Nov 7	9	44	11
Nov 8	8	36	9
Nov 9	1	35	8
Nov 10	26	9	2

3.2.4: Bridges and Tunnels Condition

Manhattan is connected to New Jersey via the Lincoln Tunnel, Holland Tunnel, and George Washington Bridge from one side and it is connected to Queens and Brooklyn with the Queensborough Bridge, Queens Midtown Tunnel, Williamsburg Bridge, Manhattan Bridge, Brooklyn Bridge, Hugh L. Carey Tunnel, and Robert Kennedy Bridge from the other side. After Hurricane Sandy, many of these bridges and tunnels were either closed or were under policies like a carpool restriction for some days. The timeline of the tunnel and bridge closures were collected from a transportation report during and after Hurricane Sandy [2]. The Queensborough Bridge, Williamsburg Bridge, Manhattan Bridge, and Brooklyn Bridge were closed on October 29 and were reopened on October 30. The George Washington Bridge, Robert Kennedy Bridge, and Lincoln Tunnel were open all the time during the disruption. The timeline for the Holland Tunnel, Queens Midtown Tunnel, and, Hugh L. Carey Tunnel is shown in Table 3.6.

Table 3. 6 Bridges and Tunnels Timeline [2]

Bridges and tunnels	**Closed**	**Reopen only for bus**	**Reopen to all traffic**
Holland Tunnel	10/29/2012	11/2/2012	11/7/2012
Queens Midtown Tunnel	10/29/2012	11/6/2012	11/9/2012
Hugh L. Carey Tunnel	10/29/2012	11/12/2012	11/13/2012

3.2.5: Transit System

The transit system includes the New York City (MTA) Bus, New Jersey Transit bus, New York City subway (MTA), PATH rail, Long Island Railroad (LIRR), Metro-North Railroad (MNRR) and New Jersey Transit rail. Shape files for transit lines and stations were available from websites [41], [42], and [43] were used in ArcMap. Figures 3.7 to 3.10 show an overview of each provider's rail network and stations in ArcMap. Bus services were suspended for two days completely and they recovered on October 31. Subway and rail systems were completely suspended for two days, and some services started to reopen on October 31 and the process of subway and rail system recovery took a while especially for the New Jersey systems.

In order to model the recovery of rail and subway lines, more details about each subway line condition and timeline of the reopening were needed. Therefore, the timeline of the subway and rail service restoration after Hurricane Sandy was gathered from data available on the SubwayNuts website [44]. In addition, information about alternative transportation modes that people were able to use instead of their disrupted mode of transportation was collected from supplemental news and government websites [2, 24]. There were alternatives like temporary bus shuttles for some disrupted subway lines, such as the Manhattan to Brooklyn subway service that was completely disrupted until the 3rd of November.

Figure 3. 7 MTA Network

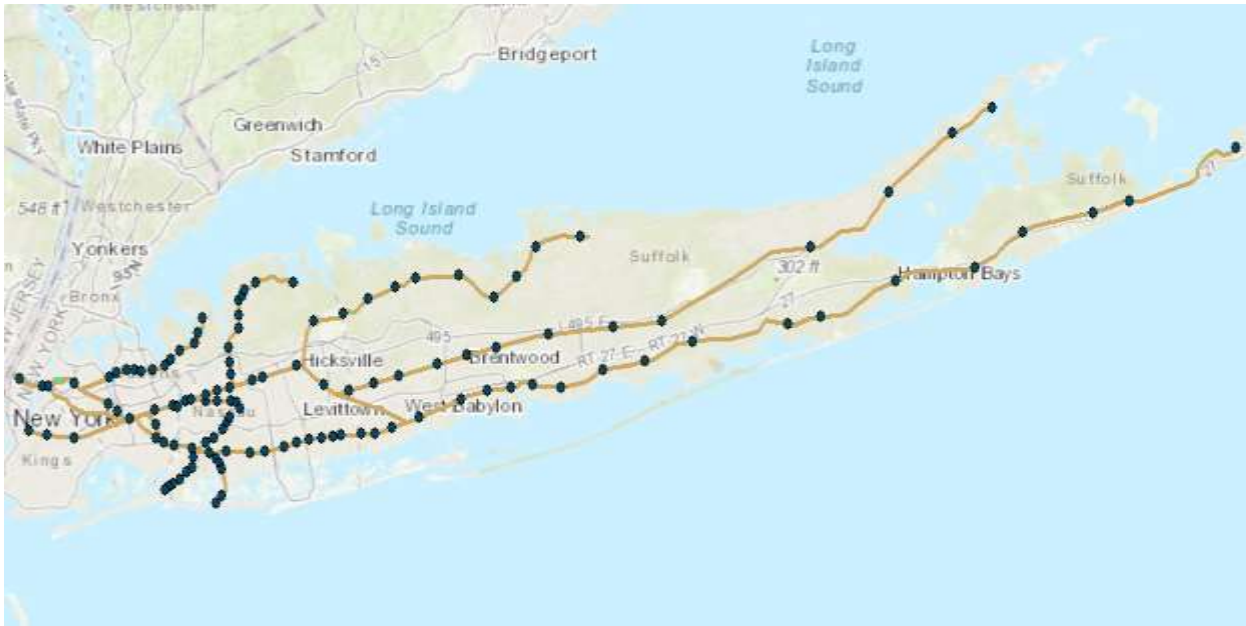


Figure 3. 8 LIRR Network

Figure 3. 9 MNRR Network

Figure 3. 10 New Jersey Rail road

3.2.6: Post-Impact Policies

After Hurricane Sandy, millions of people tried to commute to and within the city with a disrupted transportation system. Since many of the bridges and tunnels were closed and many subway and rail lines were disrupted, traffic and gridlock were observed at major crossings (open bridges and tunnels) [2]. Moreover, power and supply outages caused gasoline shortages across the New Jersey and New York metro area, and there were severe traffic backups at open gas stations [2].

Carpool restrictions and gasoline purchase restrictions were policies that were implemented in order to solve traffic and gas shortage problems [2]. The timeline of these policies is presented in Table 3.7. Carpool restrictions applied to the Queensborough Bridge, Williamsburg Bridge, Manhattan Bridge, Brooklyn Bridge and Lincoln Tunnel. Gasoline restrictions applied to New Jersey, New York City, Suffolk County, and Nassau County in New York.

Table 3.7 Recovery Policies Timeline [2]

Recovery Policies	**Start Date**	**End Date**
Carpool restriction	10/31/2012	11/3/2012
Gasoline purchase restriction in NJ	11/3/2012	11/13/2012
Gasoline purchase restriction in NY	11/9/2012	11/23/2012

3.3: Agents' Behavior and Methods of Interaction

Series of if-then rules and statistical models are used for defining agents' behavior and methods of interaction. This agent-based model simulates nine working days starting from the day after the hurricane (October 30) until November 9. Agents could choose from six different adaptations when their usual commuting pattern was disrupted. These adaptations are change route, depart earlier from home to work, depart later from home to work, change mode, cancel work trip and telework.

3.3.1: Logit Models

For predicting the probability of each of these changes, six different Binary logit models were used. Kontou, Murray-Tuite, and Wernstedt [5] developed five of these multivariable binary logit models for commuting changes (changing mode, canceling work trips, changing routes and changing departure times for home to work trips). Results of these models can be find in [5].

In addition to these five models, another Binary logit model was developed by using RStudio to predict the probability of teleworking. Potentially significant variables for teleworking were found by correlation matrices. Variables with a correlation of 0.25 or more with the change were considered as the primary variables for developing the model and independent variables that were highly correlated with each other were not used in the model. The likelihood ratio test was used to identify the preferred model. This model is presented in Table 3.8. The final model is significant as the adjusted r square value is 0.3367, and all the variables are significant at the 95% confidence level.

Table 3. 8 Telework Model

Independent variables	**β**	**Pr(>|z|)**
Intercept	-3.5935	0
Transit commuter (binary)	0.8312	0
Have option of telecommuting (binary)	1.6990	0
Have option of flexible working hour (binary)	1.3871	0
Level of education(binary)	0.8573	0.019
Management, business, and financial occupation (binary)	0.7077	0.041
Observations	331	-----------
Adjusted R-square	0.3367	-----------
Log likelihood restricted	200.46	-----------
Log likelihood unrestricted	141.66	-----------

Based on the model that predicts the probability of teleworking, being a transit commuter, having the option of teleworking and flexible working hours during normal situations, having a college degree or above and working in a management, business and financial occupation increases the probability of teleworking during the disruption.

Being a transit commuter under normal conditions increases the probability of teleworking during the disruption. The predicted odds for teleworking after disruption for transit commuters is 2.29 ($e^{0.8312} = 2.29$) times the odds for those who were not. As expected, commuters who have the option of flexible working hours and teleworking in normal situations had higher probabilities of teleworking during the disruption. Many companies allowed their workers to telework after Hurricane Sandy [2], but having this option in a regular situation can indicate that this job type has telework as an option. The predicted odds for those who had the option of teleworking and flexible working hour before disruption were 5.46 ($e^{1.699} = 5.46$) and 4 ($e^{1.387} = 4$) of those who did not, respectively.

Higher education levels and management occupations indicate mostly office-related jobs and, not surprisingly, people in these positions have a higher chance of teleworking.

The change mode model only calculates the probability of change mode and does not show the mode that people switch to from their normal mode of transportation if they end up changing their mode. Therefore, an MNL is developed by using Easy Logit Modeler software to predict the probability of choosing each mode. Mode options are drive alone, carpool, bus, rail (include subway), taxi, and walk. Table 3.9 presents the mode choice model.

Table 3. 9 Mode Choice Model

Alternative Specific Parameters		**Estimated value**	**t-statistics**
Constant	Carpool	-2.5076	-1.77
Constant	Bus	0.9263	1.209
Constant	Rail	2.1692	3.672
Constant	Taxi	-8.0799	-0.912
Constant	Walk	2.7774	1.863
Age (continuous)	Carpool	0.0114	0.498
Age (continuous)	Bus	-0.0004	-0.031
Age (continuous)	Rail	-0.0361	-3.441
Age (continuous)	Taxi	-0.0962	-1.251
Age (continuous)	Walk	-0.0089	-0.3
Income (continuous)	Carpool	-0.0005	-0.103
Income (continuous)	Bus	-0.0177	-4.646
Income (continuous)	Rail	0.0031	1.352
Income (continuous)	Taxi	0.0721	1.403
Income (continuous)	Walk	-0.0105	-1.478
Distance from home to work (continuous)	Carpool	-0.0002	-0.167
Distance from home to work (continuous)	Bus	-0.0006	-0.417
Distance from home to work (continuous)	Rail	-0.0006	-0.412
Distance from home to work (continuous)	Taxi	-0.5085	-1.6
Distance from home to work (continuous)	Walk	-0.6023	-2.37

Born in US (binary)	Carpool	-0.525	-0.648
Born in US (binary)	Bus	-0.8032	-1.691
Born in US (binary)	Rail	-1.4421	-4.208
Born in US (binary)	Taxi	-2.5165	-1.597
Born in US (binary)	Walk	-2.7752	-3.417
Log likelihood at zero	------	-628.819	
Log likelihood at constants	------	-454.4771	
Log likelihood at convergence	--------	-405.5537	
R-squared w.r.t. zero	--------	0.3551	
R-squared w.r.t. constants	--------	0.1076	
Adjusted R-squared w.r.t. zero	--------	0.3153	
Adjusted R-squared w.r.t. zero	--------	0.0629	

3.3.2: Decision Frameworks

Decision flow-charts detail the agent behavior estimation for the post-Hurricane Sandy period. Figure 3.11 shows the work condition and telework sub-model flowcharts. In this ABM, at the start of each day, people check their work location's condition to see whether it is closed or open. If it is closed, some people may telework anyway, and all others are considered unproductive that day. Agents need power if they want to telework. If power is available, the probability of teleworking is calculated based on telework model. A random number is generated and compared with the telework probability. If the generated random number is less than the telework probability, that person teleworks; if not, that person does not telework.

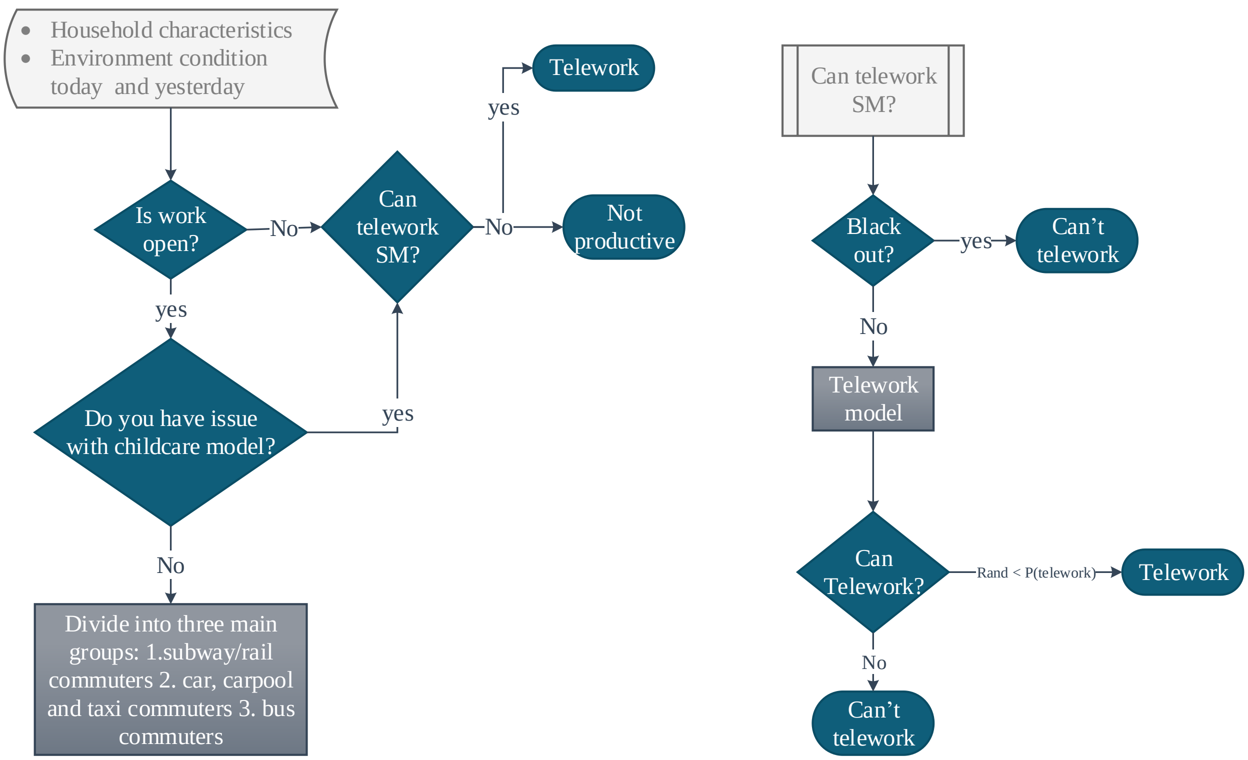


Figure 3. 11 Work Condition and Telework Flow Charts

If their work is open, the agents check the condition of their main mode of transportation. However, people with children, even if their work is open, cannot go to work when daycares and schools are closed unless they can make other care arrangements. Figure 3.12 presents the decision framework for families with children under the age of 15. If daycares and schools are open, having children does not cause any problems for the parents' work trips. However, in the situation that daycares and schools are closed, the family structure plays an important role in defining an agent's behavior. For married-couple families where only one of them works, school and daycare closure do not cause any problems because one of the parents is always at home and can care for the children. If schools and daycares are closed, dual-income families need to find an alternative caregiver or one of them needs to stay home and take care of children while the other one goes to work. Single-parent families need to find another caregiver if they want to go to work. Therefore, for dual-income families and single-parent families, if daycares and schools are closed, first, the probability of canceling work is calculated based on the cancel work model. Then a random number is generated. If the random number is less than cancel work probability, they cancel work. If not, the square of the cancel work probability is compared by a random number. In dual-income families, if the random number is smaller than the square of the cancel work probability, they are considered to have found another caregiver. Otherwise, the spouse is caring for the children. While in single-parent families, if the random number is smaller than the square of the cancel work probability, they are considered to have found another caregiver. Otherwise, they cancel their work trip.

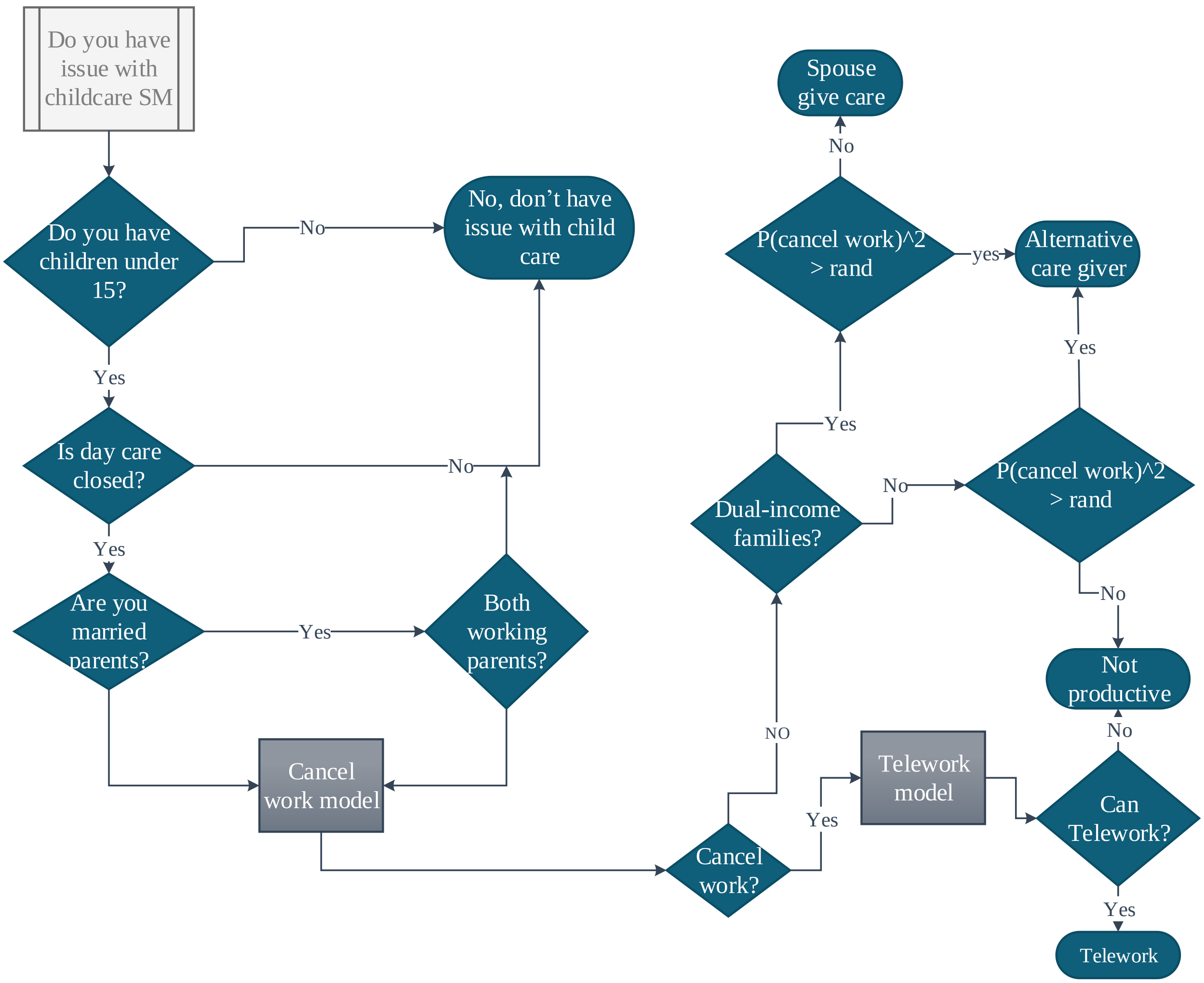


Figure 3. 12 Issue with Childcare Flow Chart

When their work is open and there is not an issue with childcare, people are grouped based on their main mode of transportation into three main groups: 1. rail and subway commuters, 2. car, carpool, and taxi commuters and 3. Bus commuters. People first check the condition of their normal transportation mode. Figure 3.13 presents the flowchart for subway/rail commuters. Each subway/rail system consists of different lines, and each of these lines can have different recovery durations. Therefore, we need to know the subway/rail line that each agent uses while traveling from home to work daily and vice versa.

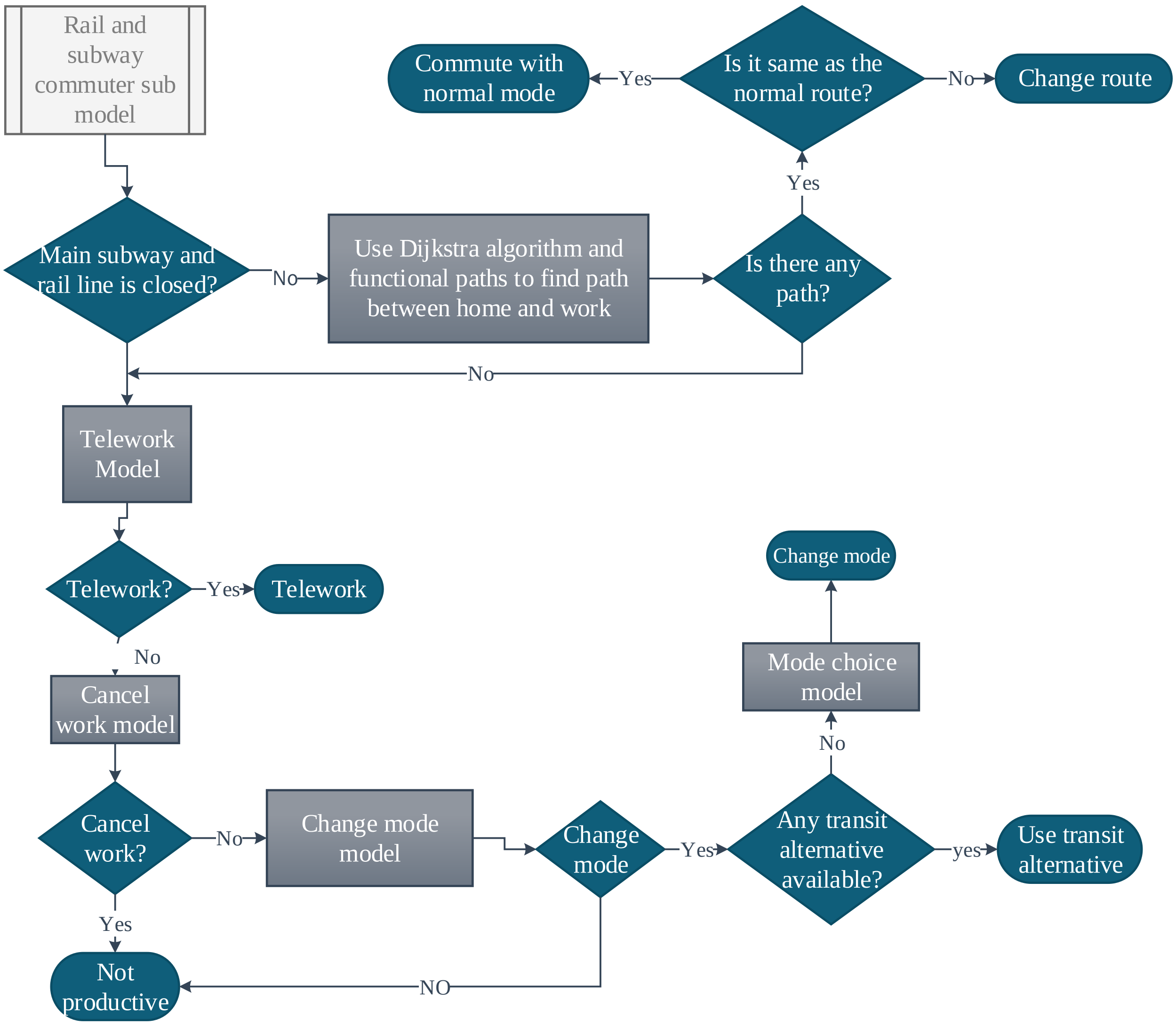


Figure 3. 13 Rail and Subway Commuters Flowchart

From the available shapefiles of subway/rail stations [41], the latitude and longitude of each subway/rail stop are converted to x y coordinates. The distance between each home and work location and subway/rail stops is calculated, the two closest stations to home are chosen as the probable origins, and the two closest stations to the work location are chosen as the probable destinations for each person.

In order to find the subway lines that people use to move from their origins to destinations, Dijkstra's algorithm (shortest path) is used. To use this algorithm, the origin, destination, edge connections (links) and their costs are needed as input files where the cost of each line is the length of each link in the actual network. The distance costs and chosen paths are outputs of this algorithm.

Shapefiles for subway/rail lines and stations available from websites [41], [42], and [43] are used in ArcMap. To calculate the distance between stations (links lengths) the closest facility tool was used. Some subway/rail networks like MTA and NJ Transit include stations that are in walking distance of each other. During their daily trip from home to work, many people need to change their subway/rail lines and to do so they may walk from one station to another. Between close stations, there is no link available in the shapefiles of these subway/rail lines. Therefore, a walkable path is added between the stations with distance less than 0.3 miles (station complex) in ArcMap.

Moreover, some people need to use more than one subway and rail system to move from home to work. For instance, people who live in Brooklyn, NY with jobs in New Jersey need to use both MTA subway and NJ Transit to reach their workplace. In these cases, people walk between stations that are in different subways and rail lines, so a walkable path is added between the stations with distance less than 0.3 miles in different subway and rail systems.

The shortest path algorithm is run for all four combinations of origins and destinations for each person in the normal situation, and the paths that they used before Hurricane Sandy for moving from home to work is found. The path that is chosen as the normal commuting path for each person for moving from home to work is the path that is shorter, with start and stop stations that are closer to work and home location and with fewer line changes than other paths. This is because people usually do not like to change lines and prefer to use a subway/rail line that directly takes them to their destination.

The walkable distance between stations is considered as the length of a straight line that connects these stations. Therefore, sometimes the shortest path algorithm chooses paths with too many walkable paths as the best path because they are shorter based on their length. To solve this problem a penalty is applied for choosing the walkable path, and their cost in the system is considered to be the length of straight line between two stations plus 1000 meters as the penalty so that shortest path algorithm only chooses these paths when the commuter really needs to change lines.

After Hurricane Sandy, many of these subway lines lost their functionality, and all of them were entirely closed from October 28th (one day before Sandy struck) and remained closed for several days. At the start of each day, the subway/rail lines that are closed in that day are omitted from the edge connection (link file). The edge connection is updated daily based on the subway/rail lines available for that day. Then the shortest path algorithm is used again and the cost and path matrices for each day are developed. If the cost for the path is infinity, this indicates that one or more of closed subway lines are in the agent's trip, so they cannot use the regular transportation mode for traveling from home to work. If the path cost is not infinity, it is compared to the path that agent used before Hurricane Sandy. If they are the same, the agent can travel with their regular route , but if they are different, they have adapted to a new situation by changing routes.

For people that cannot use their regular mode of transportation, only three options remain. They can change their mode of transportation, cancel their work trip, or telework. Of the 397 respondents, 169 canceled their work trips, 100 changed modes, and 94 people teleworked at least once after Hurricane Sandy. Based on these numbers, the order of preference appears to be cancel their work trip, change mode and telework. Changing mode is constrained by the availability of another mode. Teleworking depends on several different factors like the availability of power and communication systems and is highly dependent on the occupation, so it is not an option for everyone.

The abovementioned numbers from the survey data and results of previous literature prioritize agent preferences in these situations. Based on [6], [7] and [8], people's preference is to cancel their work trip, telework and change mode. Changing mode is the least preferred choice, and people mainly use this option when they have no other choice. Although the number of people that telework is fewer than the ones that cancel their work trip, this option is considered before canceling work in this agent-based decision framework because people can telework even if they cancel their work.

Therefore, for the people who cannot use their normal mode of transportation, first the probability of teleworking is calculated based on the telework model. Then a random number is generated; if the random number is less than the probability of teleworking, this person will telework; otherwise the probability of canceling work is calculated based on the cancel work model. If the random number is less than the probability of canceling work, this person is not productive; otherwise, the probability of changing mode is calculated based on the change mode model. If the random number is less than the change mode probability, then this person will change their mode. To find the selected mode, the mode choice model is used. Figure 3.14 shows the mode choice model flowchart.

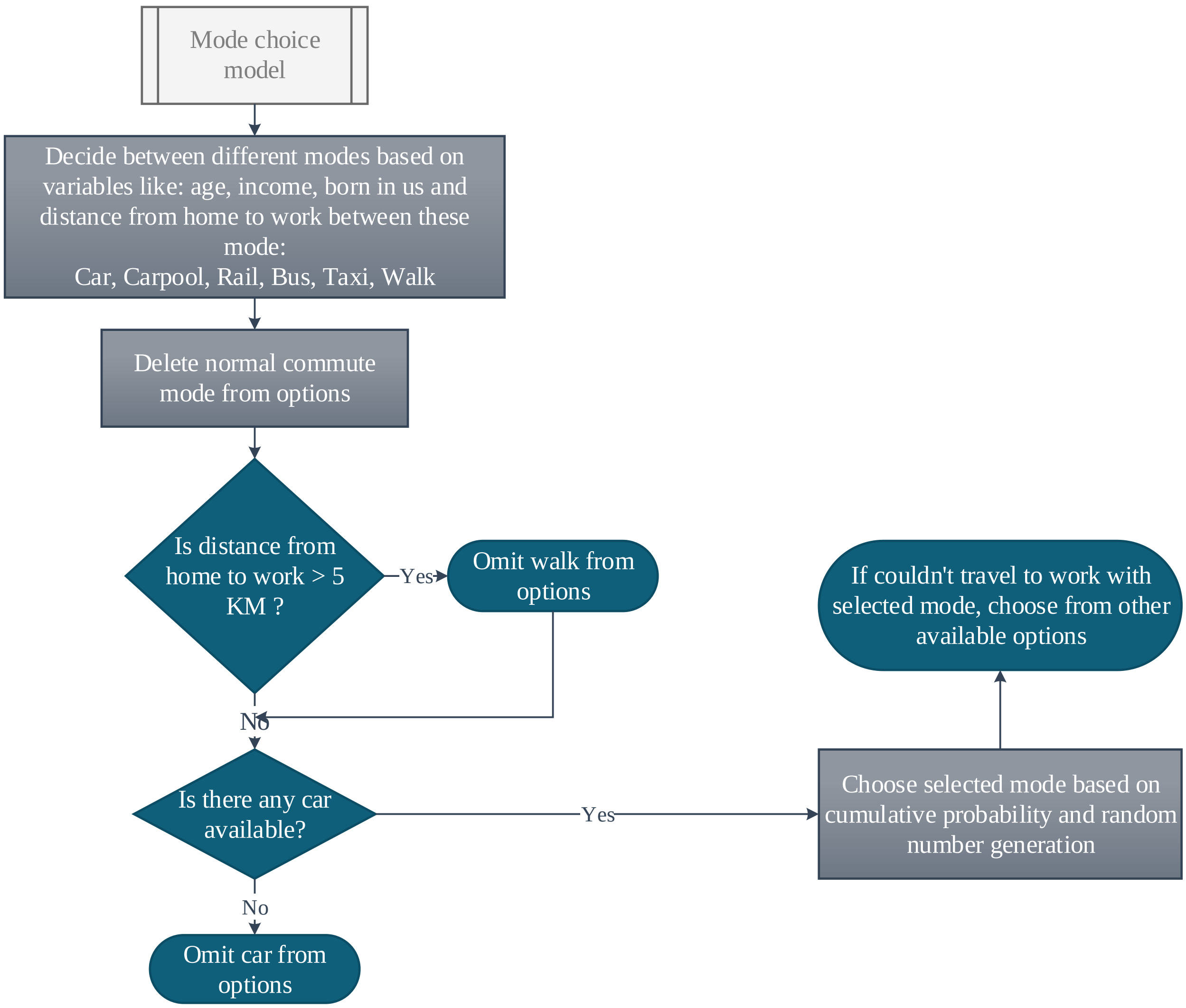


Figure 3. 14 Mode Choice Model Flowchart

Before using the mode choice model, a transit alternative option is considered. During the disruption, transportation agencies usually add some alternative transportation modes as a backup for closed subway/rail lines. For instance, after Hurricane Sandy, there were not any MTA transit lines between Brooklyn and Manhattan for several days due to flooding and power issues so, between Brooklyn and Manhattan, 330 buses ran to replace the missing subway service [2]. Moreover, some bus services ran in New Jersey to replace missing rail service in the NJ Transit system. In these cases with a backup system for disrupted rail and subway lines, a transit alternative is considered as the first option for people who have decided to change modes. If it is possible for them to use the transit alternative based on their home and work locations, they use this mode; otherwise the mode choice model is used to figure out which mode they choose.

Based on this model, the probability of choosing each of the six modes (drive alone, carpool, bus, rail, taxi and walk) is calculated for each person, but before that, options that are not available for each person are omitted. First, their normal commute mode is omitted because people want to change mode so the normal commuting mode should not be an option. Next, if the distance from home to work is more than 5 kilometers, the walk mode option is omitted too. Finally, if the person does not own a car, the drive alone option is omitted. Then, the probability of choosing the remaining modes is calculated and, based on random number generation and cumulative probabilities, the selected mode is identified.

Since the transportation system is disrupted, there is no guarantee that people can use their selected mode of transportation, so the condition of that mode should be checked. For instance, if they choose to travel by rail instead of their main mode of transportation, the shortest path algorithm is used to check the availability of a path for this person. If the cost of the selected path is not infinity, they can use rail to move from home to work; if not, rail is omitted from options. The probability of choosing the remaining modes is calculated and, again, based on the cumulative probabilities and the random number generation another mode is selected until the agent is able to move from home to work by one of these modes.

The decision framework for people whose main mode of transportation is car, taxi, or carpool is a little bit different. Figure 3.15 shows the decision framework for car, carpool or taxi commuters. Many major tunnels and bridges that connect Manhattan with New Jersey from one side and Queens, and Brooklyn from other side were closed after Hurricane Sandy. Several days after Hurricane Sandy struck, the Hugh L. Carey tunnel, Queens Midtown tunnel and Holland tunnel were reopened for buses only, and it took more time to reopen for all traffic [2]. Therefore, if someone had one of these tunnels or bridges on his or her way to work, they needed to change their route or mode to be able to reach their destination.

Therefore, all car and bus commuters who wanted to move from New Jersey to New York or wanted to move within New York but one end of their trip is in Manhattan would pass one of these bridges and tunnels on their way. The sum of distances between home and each of these tunnels and bridges and work and each of these tunnels and bridges is calculated, and the closest bridge to the home and work locations is chosen as the first priority for each agent and all other bridges, and tunnels are listed as alternatives based on their distance.

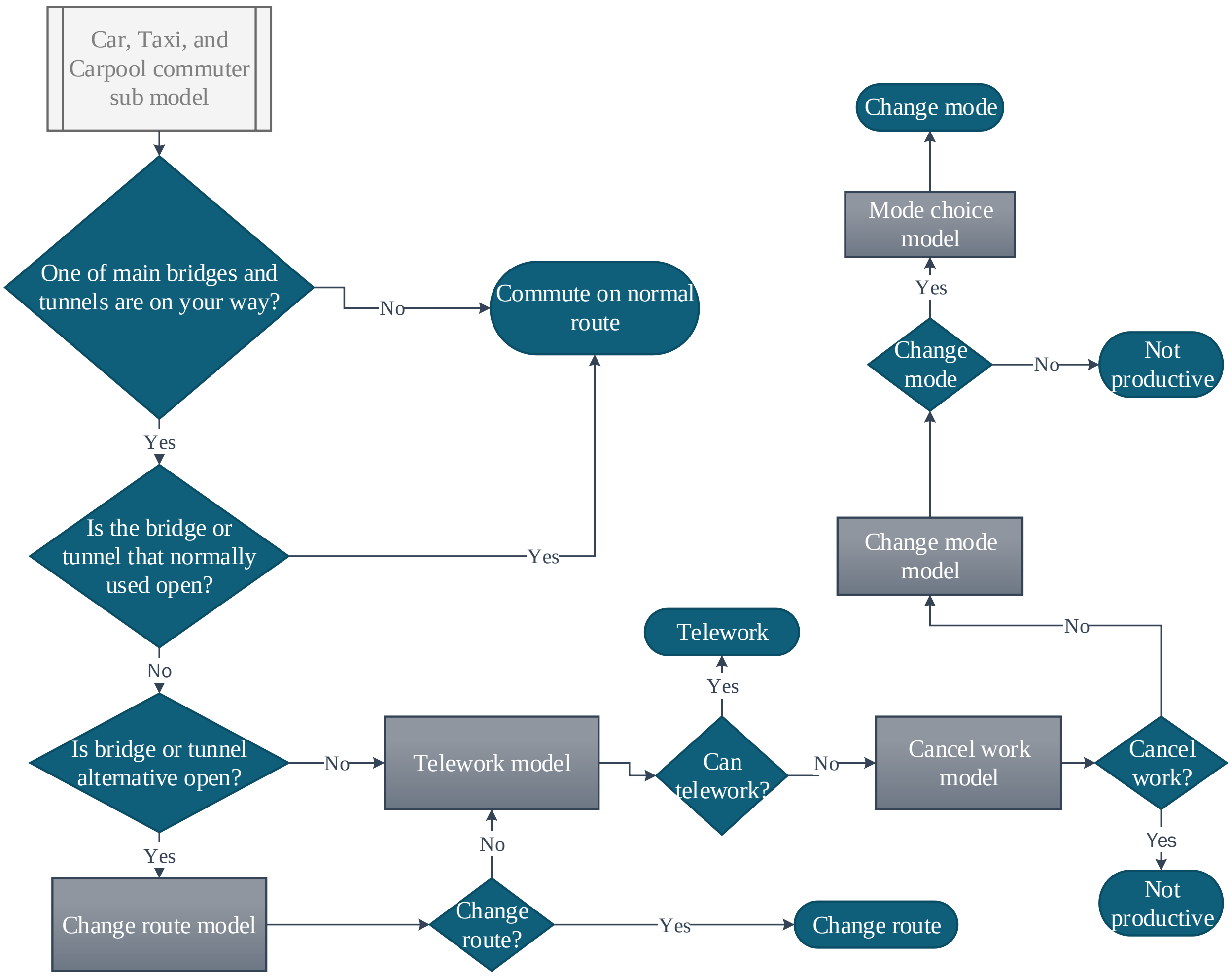


Figure 3. 15 Car, Carpool or Taxi Commuter Decision Framework

At the beginning of each day, the condition of the bridge and tunnel that is used by each agent in a normal situation for commuting to work is checked and, if it is closed, the probability of changing route is calculated based on the change route model and a random number is generated. If the random number is less than the probability of changing route, they will change route and move from home to work with the next closest open bridge or tunnel. If not, they first consider teleworking, next canceling the work trip and finally changing mode, similar to rail and subway group.

If an agent's main mode of transportation is bus, all of the steps are similar to the car, carpool and taxi group with small differences. At the beginning of each day, first they check the condition of the bus system as to whether it is functional or not. All bus services were closed for the first day of the simulation (October 30), and all of them restored their service by October 31. If the bus service is restored, the distance between home and all of the bus stops is calculated, and the closest stop to the home location is considered as the bus stop that this agent starts his/her trip from. If the bus service is disrupted, they will consider teleworking, canceling the work trip and changing mode, in that order. Figure 3.16 represents the decision framework for bus commuters.

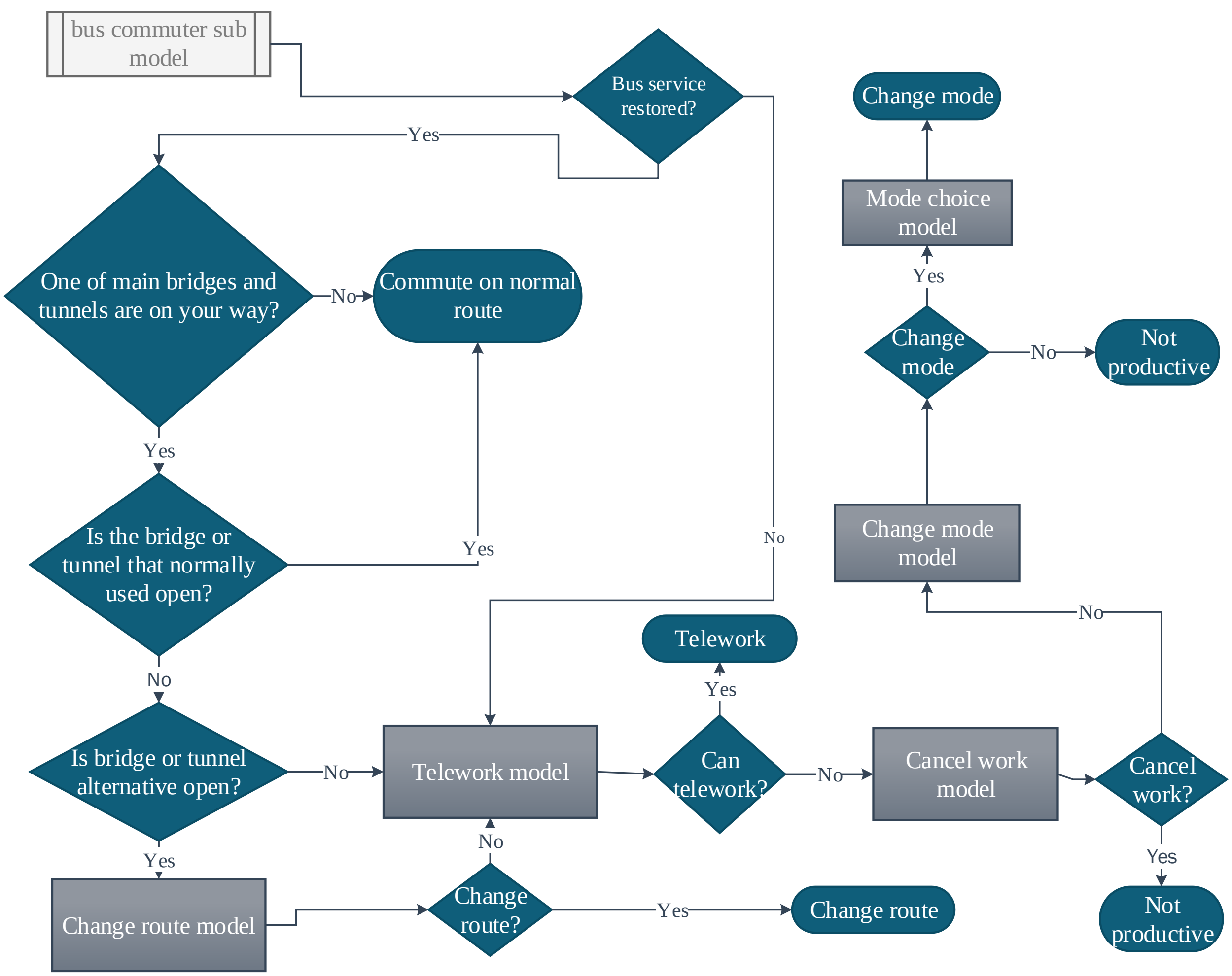


Figure 3. 16 Bus Commuters Decision Framework

Some of the variables that are used in the logit models depend on the environment and agent characteristics, which need to be updated daily. These variables are tunnel closure, carpool restrictions, gasoline restrictions, and delay and crowding.

A tunnel closure is a variable that is defined for car, carpool, taxi and bus commuters. This variable is binary and takes a value of one if the tunnel closure has affected the commuting pattern. A carpool restriction is a variable that is defined for people whose main mode of transportation is drive alone, and gasoline restriction is a variable defined for car, carpool, and taxi commuters. These two variables are also binary and take a value of one if a carpool restriction or gasoline restriction have affected commuting patterns.

Tunnel closure and carpool restriction for each person is defined based on the main bridge and tunnel that they use to commute from home to work in a normal situation. At the start of the day, if the bridge/tunnel that a person normally used for commuting is closed or there is a carpool restriction, it is assumed that the commuter encounters a tunnel closure and carpool restriction that day, so the value of these variables is one for that day for this person. This value can change the next day if that bridge or tunnel reopens or if the carpool restriction is lifted.

For each person that has a car, a random number is generated and if that number is less than 0.5, it is assumed that plate number of their car is even. Otherwise, the plate number is assumed to be odd. Typical gasoline consumption of a car is around 24 miles per gallon that is 10 kilometer per liter [45], and fuel capacity of cars is around 45 liters (around 12 gallons) [46]. So each time the car is fueled, it can be used for around 400 kilometer (248 miles). Twice the distance from home to work is considered as distance that people drive daily. For the first day of simulation, a random number from 0 to 40 is considered as available fuel (in liters) in the car of each person, and, based on the distance from home to work, the next day that this person needs fuel is calculated. If on that day, the gasoline restriction policy is in effect and the plate number and day number are one even and the other odd, it is assumed that they encounter gasoline restrictions on that day and value of gasoline restriction variable is one. This value is also updated daily like tunnel closure and carpool restriction.

At the end of each day, people who are able to travel to their work learn from their experience and this experience can affect their decisions tomorrow. One important aspect of this experience is delay and crowding.

Based on the survey data, delay and crowding caused 59 people to cancel their work trips, 89 people to leave earlier, 18 people to leave later, 43 people to change mode and 79 people to change route at least once after Hurricane Sandy. So to avoid delay and crowding, people prefer to change routes and leave earlier more than all of the other changes. Delay and crowding is a binary variable that is defined for all of the agents, and it takes the value of one if people encounter delay and crowding; otherwise it is zero.

For subway and rail commuters, the number of people that use each link of the public transit before disruption is compared with the number of people that use each link after disruption. If one link is used more than in the normal situation, this link is considered crowded. For car and bus commuters, the total number of users for each tunnel and bridge is compared with the number of daily commuters after Hurricane Sandy and if one tunnel/bridge is used more than in normal situation this route is considered as crowded. Moreover, for bus commuters, the total number of people that use each bus station in the normal situation is compared with the total number of people that use each station after disruption. If one station is used more than in the normal situation, this station is considered crowded. All the people that have one of those crowded subway lines, bridges and tunnels, or bus stations in their commute will consider delay and crowding for the next day. Therefore, the delay and crowding variable value changes to one and people may prefer to depart earlier the next day in order to avoid delay and crowding or change routes if it is possible for them.

In this model, two different elements of time are needed. One is day, and the other is the different time frames within a day. On each day, people need to choose a departure time. Departure times are grouped in 11 different time frames starting from 4: 30 AM ending at 10 AM and each of them are half an hour.

Based on significant variables for the change of departure time, people usually depart earlier in order to avoid delay and crowding. All of the significant variables for change departure time are used to decide whether this agent departed earlier or later and the amount of this change in departure time is defined by using the distribution of people’s answers in the survey to the question about by how much they left earlier or later. Figures 3.17, 3.18, and 3.19 show the departure time decision framework and distribution of depart earlier and later, respectively.

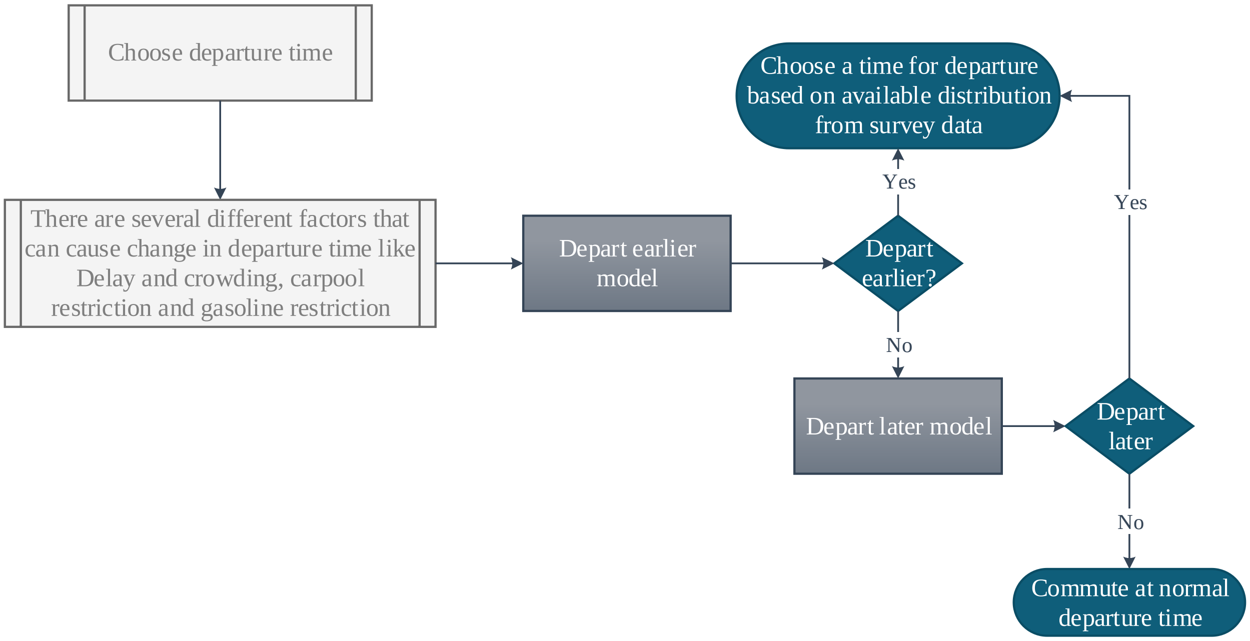


Figure 3. 17 Choose Departure Time Decision Framework

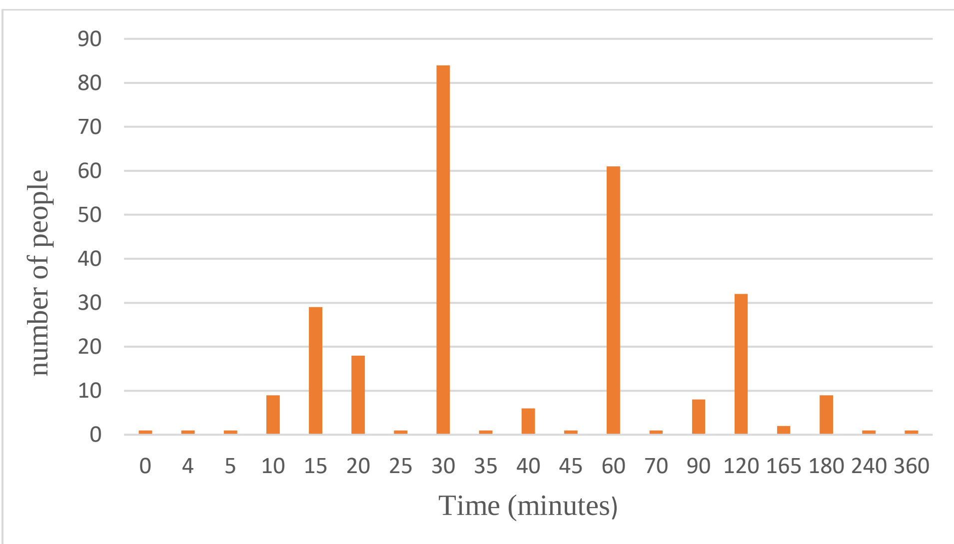


Figure 3. 18 Distribution of Depart Earlier (minutes)

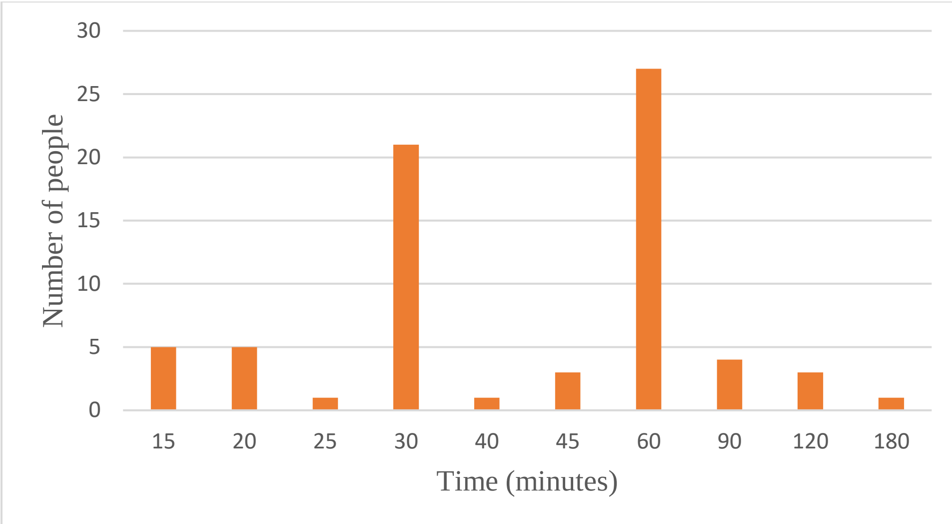


Figure 3. 19 Distribution of Depart Later (minutes)

3.3.3: Simulation Assumptions

This model is developed using MATLAB. After running the preliminary model, some assumptions were made to make the model more representative and prevent behavior that is not realistic. These assumptions are mentioned below:

- When people decide to change their departure time, the new departure time is chosen based on the distribution of change departure time in survey data. If this distribution were used every time, large jumps could occur in departure time. In order to address this problem, the change departure time distribution is only used for the first time that they decide to change their departure time and, for next time, they change it only one time step (half an hour) at a time.
- For the first day of work after the disruption, people consider news about the delay and crowding on the roads for choosing their departure time. Information about the delay and crowding on roads and subway lines were available in [[2], [44]]. Table 3.10 presents data available about the delay and crowding after Hurricane Sandy in reports.
- After the first time of returning to work, people cannot cancel their work trip anymore. The exception to this rule is for dual-income families with children that may need to take turns canceling their work trips until the time that schools and daycares reopen.

Table 3. 10 Delay and Crowding News after Hurricane Sandy [[2],[44]]

Date	Delay and crowding
10/31/2012	1.Traffic gridlock in Manhattan, Queens, Brooklyn and all open tunnels and bridges 2. Long wait time for bus commuter
11/1/2012	1. Traffic in all bridges and tunnels 2. crowdedness in MTA subway 3. Long wait for bus shuttle that connected Brooklyn and Manhattan
11/2/2012	1. Crowding along MTA subway functional lines
11/3/2012	1. Delay and crowding in LIRR and NJ rail

- Since the adaptation that people choose depends on a probability and a comparison of that probability to a generated random number, there are some cases that although people cannot travel to work with their normal commuting pattern, they do not choose any of the adaptations. In these cases, people choose what they did the previous day.

3.3.4: Population Synthesis

The preliminary model includes 383 commuters, which are the survey respondents. This number is not enough to represent a large area like the New York metropolitan area. Therefore, the PopulationSim package was used to generate a synthetic population for the modeling region.

Inputs of this package are the disaggregate population sample that is obtained from the Census Public Use Microdata Sample (PUMS) [47] and geographic levels and their relationship. The PUMS includes answers to the American Community Survey (ACS). However, the home location of respondents in the PUMS data is shown at the Public Use Microdata Areas (PUMA) level. PUMAs are geographic units within each state including more than 100,000 people. Geographic level should include PUMAs and other geographic levels that are needed. For instance, because in this project, the zip code of each person’s home is needed, geographic levels are PUMAs and zip codes.

An input file that includes the list of zip codes, their populations and the relationship of zip code and PUMA level is used as an input for this population synthesizer to get the zip code of each person's home location in the output. Some of the zip codes' boundaries did not fit perfectly within the PUMA boundary. This means some zip codes are within more than one PUMA area. Therefore, populations within a zip code were divided in proportion to the area of each zip code within the different PUMA areas. This relationship file was developed with GIS by using the TIGER/Line Shapefiles [48] and the ZIP Code Tabulation Area (ZCTA) Relationship Files [49] from United States Census Bureau.

The outputs of this synthesizer included the person and household level data, as well as almost all of the agents' characteristics that are needed for the agent-based model. These variables included home zip codes, transportation mode, age, income, gender, number of children, level of education, occupation, departure time from home to work, whether they were born in the US, whether their first language is English, car ownership, family structure, and work location.

The output for the work location is in units even bigger than PUMA. Some PUMAs are aggregate and work location is reported in that aggregated level. Since the zip code of each person's work location is needed, information about the number of people working in each zip code were obtained from ZIP Code Business Statistics from United States Census Bureau [50]. Moreover, the relationships between the zip code, PUMA, and aggregated PUMA level was obtained from United States Census Bureau and IPUMS USA websites [51]. The percentage of people working in each zip code was calculated based on the number of people working in each zip code and relationship files. Based on the random number generator and cumulative probability, a zip code was assigned to each person. For instance, one of the work locations output is the aggregated PUMA 3000, where this area includes the 3001, 3002, and 3003 PUMAs. Based on the relationship files, 25 zip codes are within these PUMAs. By summing the number of people that work in each of these zip codes, the number of people that work in the 3000 area is found. The number of people that work in each zip code is divided by the total number of people working in the 3000 area and the percentage of employees in each zip code is found. For each person whose work location is in the 3000 area, a random number is generated and this random number is compared with the cumulative probabilities of working in each zip code. In this way, a zip code is assigned to each person's working location.

If z_{ij} is zip code i in aggregated PUMA j, pz_{ij} is the number of people working in zip code i obtained from ZIP Code Business Statistics from United States Census Bureau [50] and N_j is total number of zip codes within aggregated PUMA j. The total number of people that work in aggregate PUMA j is $\sum_{i=1}^{N_j} pz_{ij}$. Therefore the percentage of people that work in each zip code i in aggregated PUMA j is $pz_{ij} / \sum_{i=1}^{N_j} pz_{ij}$. For each person whose work location is in aggregated PUMA j a random number is generated and this random number is compared with the cumulative percentage of zip codes within that aggregated PUMA j. Cumulative percentages are [0, $pz_{1j} / \sum_{i=1}^{N_j} pz_{ij}$), [$pz_{1j} / \sum_{i=1}^{N_j} pz_{ij}$, $pz_{1j} / \sum_{i=1}^{N_j} pz_{ij} + pz_{2j} / \sum_{i=1}^{N_j} pz_{ij}$),….[$pz_{1j} / \sum_{i=1}^{N_j} pz_{ij} + pz_{2j} / \sum_{i=1}^{N_j} pz_{ij} + \ldots + pz_{i-1j} / \sum_{i=1}^{N_j} pz_{ij}$,1].

Two more characteristics are needed in the agent-based model, having the option of flexible working hours in normal situations and having the option of teleworking in normal situations. ACS does not ask about these two characteristics. In order to find out whether each person has these two options or not, data from our original survey was used. The percentage of people having the option of flexible working hours and teleworking in a normal situation were calculated by occupation. Table 3.11 shows these percentages. A random number is generated for each person and based on their occupation this random number is compared with the percentages of having flexible working hours and teleworking in that occupation. If the random number is less than those percentages, it is assumed that this person has the option of flexible working hours and teleworking, otherwise, the person does not have these options.

Table 3. 11 Percentage of People Having Flexible Working Hours and Telework Option

Occupation	**Total number**	**#having telework option**	**#having flexible working hour**	**%having telework option**	**%having flexible working hour**
Computers, engineering, and science	29	12	22	41	76

Construction and extraction	9	1	4	11	44
Education, legal, community service, arts, and media	93	17	41	18	44
Farming, fishing, and forestry	2	0	1	0	50
Healthcare-related	65	9	29	14	45
Installation, maintenance, and repair	15	0	4	0	27
Management, business, and financial	70	31	53	44	76
Military	3	0	2	0	67
Office and administrative support	18	4	9	22	50
Production	7	4	4	57	57
Sales-related	30	10	19	33	63
Transportation and material moving	11	0	2	0	18
Another occupation	38	5	18	13	47

ACS includes information about all groups of people in all age ranges, but in this model, only people that are employed and travel from their home to work are modeled. Therefore, all unemployed people, and people that work from home are omitted and this reduced the total number of people from 19 million to around 7.6 million. Table 3.12 shows the transportation modes of these 7.6 million people in normal conditions.

Table 3. 12 Transportation Modes of People in Normal Condition

Mode of Transportation	Number of observations
Car	4,649,517
Carpool	257,418
Taxi	54,086
Bus	701,585

Subway/Rail	1,983,300

Many of these people that are car, carpool, taxi and bus commuters, are not in the affected area, meaning that they do not have any disrupted bridges and tunnels on their way therefore they would commute normally even after the transportation disruption (assuming that work is open and schools are in session). Therefore, people whose commuting patterns are not affected by Hurricane Sandy are omitted as well. In the end, after the population synthesis, the total number of agents in the simulation was 2,456,835. Table 3.13 shows the transportation modes of the agents in the normal condition.

Table 3. 13 Transportation Modes of the Agents in the Normal Condition

Mode of Transportation	Number of observations
Car	256,477
Carpool	22,014
Taxi	3,059
Bus	191,985
Subway/Rail	1,983,300

The MATLAB code was first run for the normal (undisrupted) condition to find the paths that people use normally to travel from home to work and the number of people that use each bridge, tunnel and subway/rail link in normal conditions. Next, the code was run for the base, disrupted situation. In the base recovery situation, the environment uses the recovery events that really happened after Hurricane Sandy. Results of base recovery model are shown in chapter 4.

In order to examine the effects of different recovery processes on population productivity for policy purposes, six different scenarios were defined. In each scenario, only one factor was changed and all other factors remained the same as the base model, as outlined below:

1. In the first scenario, the effect of electricity recovery is examined on the overall process of returning to productivity. That is, what if the electrical system recovered one day earlier compared to what happened in reality? For instance, the power outage percentages in Westchester were Day 1=42%, Day 2=40%, Day 3=39%, Day 4=37%, Day 5=24%, Day

6=20%, Day 7=16%, Day 8=12%, Day 9=8%. In this scenario these percentages would change to Day 1=40%, Day 2=39%, Day 3=37%, Day4=24%, Day5=20%, Day6=16%, Day 7=12%, Day 8= 8% and for Day 9 the percentage of power outage in day 10 in the real situation would be used (i.e. 4%). All other systems would recover at the rates seen in reality, which is the same as the base recovery model.

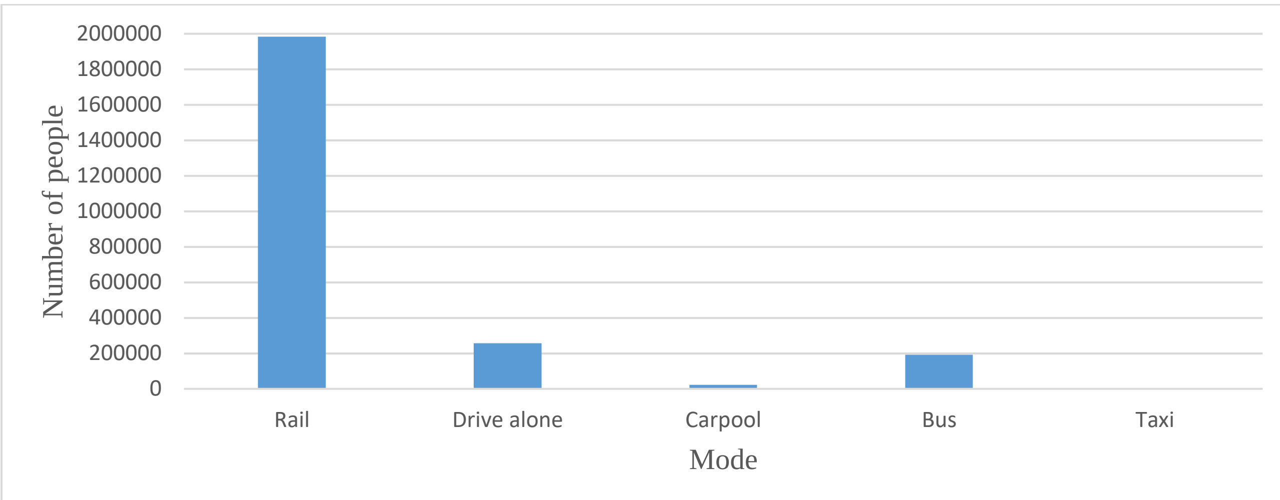
2. In the second scenario, the effect of daycare and school closures on the recovery process is studied. In this case, schools and daycares would recover one day faster than the real situation and all other systems would recover at the rates seen in base model.
3. In the third scenario, tunnels would reopen one day faster and carpool restrictions would start and end one day earlier while the condition of all other systems would remain the same as the base condition.
4. In the fourth scenario, all subway/rail links would reopen one day earlier.
5. In the fifth scenario, the New Jersey area rail/subway would recover as fast as the New York area rail/subway. The total number of links in the subway/rail system in the New York Transit network (including MTA subway, LIRR, and MNRR) is 1481 and the total number of links in the New Jersey Transit network (including NJ Transit and Path rail) is 528. The number of links is 2.82 times more in New York. In this scenario, these subway/rail systems should recover at the same rate, therefore if on day 1, 20 percent of the New York area rail/subway system is recovered, 20 percent of the New Jersey area rail/subway should be recovered too. The, ratio of functional links in these transportation systems is always 2.82. For instance if 282 links are recovered in the New York rail/subway on the third day, 100 links should recover in the New Jersey rail/subway on that day and prioritizing links for recovery are based on the real condition.
6. In the sixth scenario, New York area rail/subway would recover as slowly as the New Jersey area rail/subway. For instance if 100 links are recovered in New Jersey rail/subway on third day, 282 links should recover in New York rail/subway on that day and prioritizing links for recovery are based on the real condition.

The code was run for each of these scenarios and results are presented and compared in chapter 4.

CHAPTER 4: RESULTS

Following the steps outlined above, the outputs of the normal (undisrupted) condition, base condition and six different scenarios were obtained. Figure 4.1 shows the transportation modes of people in the normal condition.

Figure 4. 1 Transportation Mode of People in Normal Condition



Figures 4.2 to 4.5 show the total number of people that used each link of the subway/rail system before the disruption. These numbers were used for defining delay and crowding after disruption.

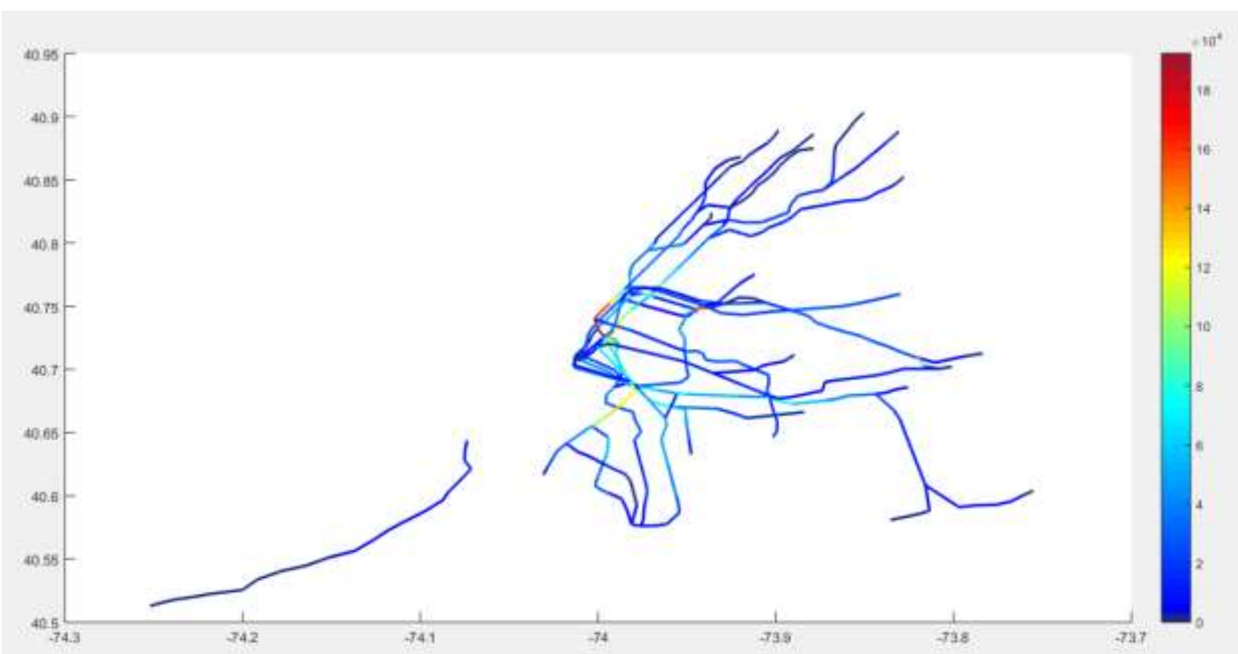


Figure 4. 2 Number of People Using Each Link in the Normal Condition-MTA Subway

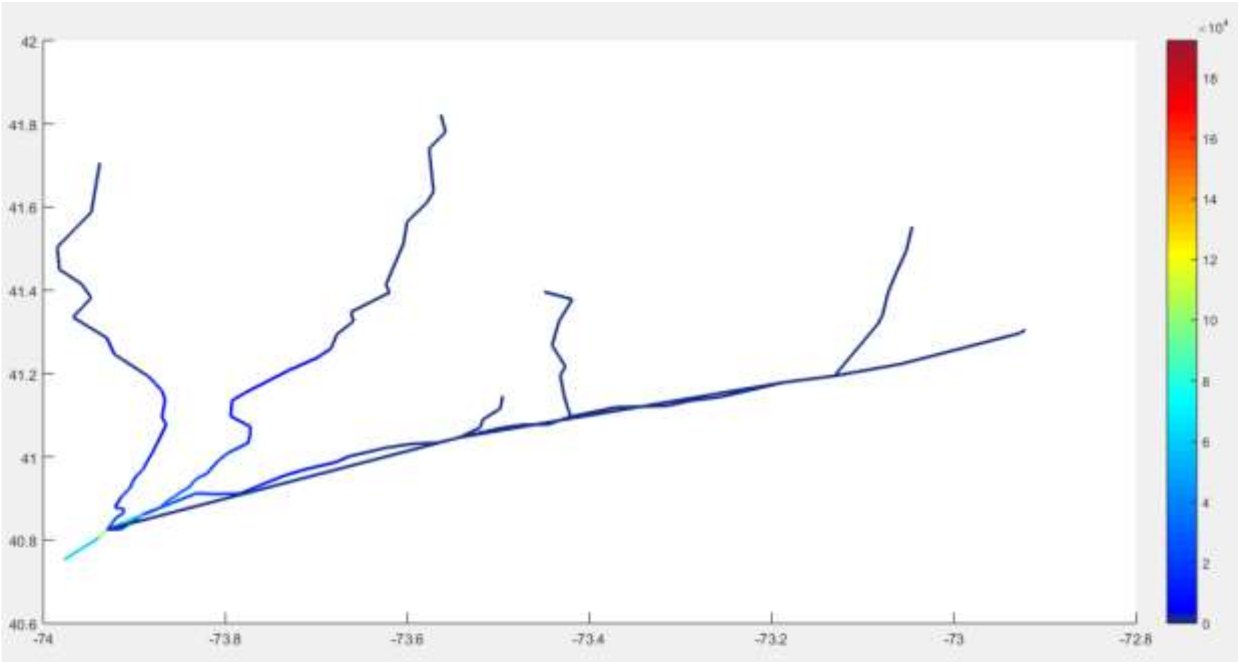


Figure 4. 3 Number of People Using Each Link in the Normal Condition-MNRR

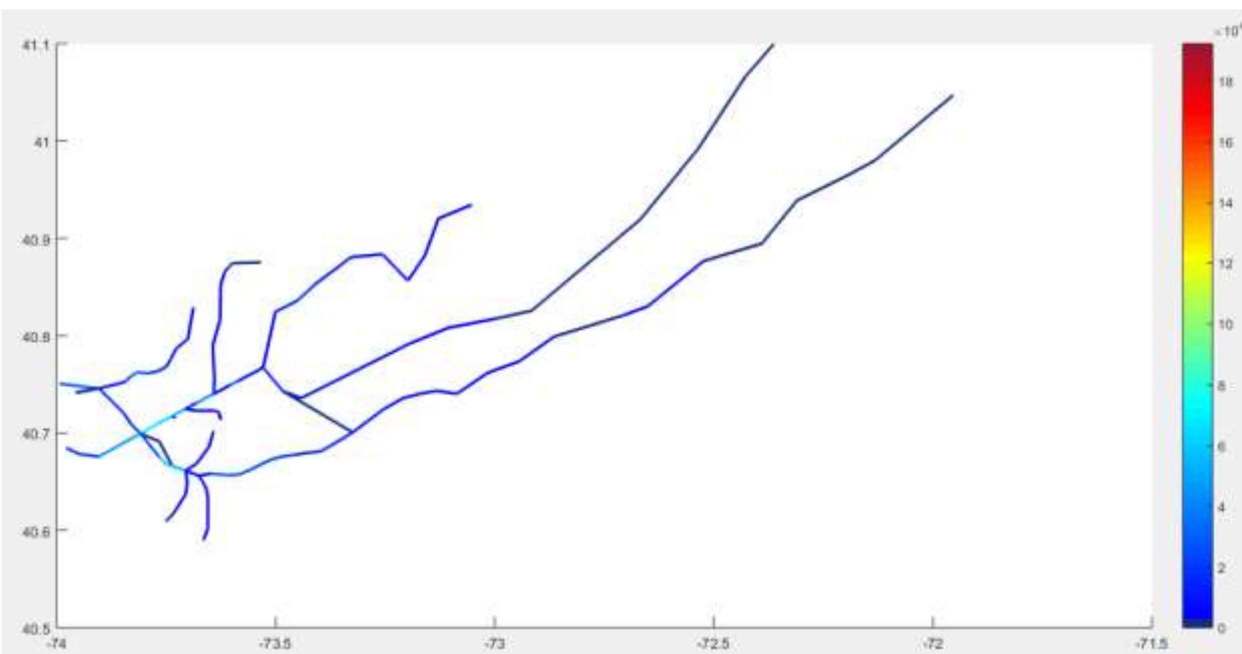


Figure 4. 4 Number of People Using Each Link in the Normal Condition-LIRR

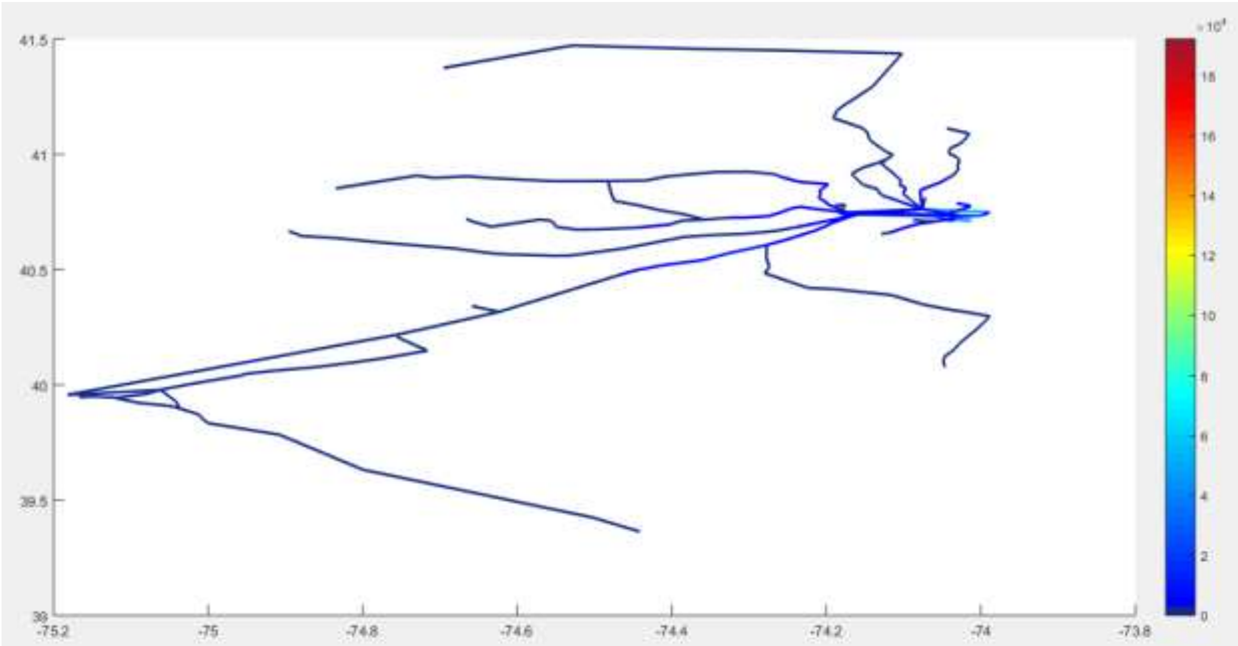


Figure 4. 5 Number of People Using Each Link in the Normal Condition-NJ Transit and Path rail

Chapter 4.1: Base Condition Results

In this section, results of the base condition simulation are presented. Figure 4.6 shows people's adaptation on different days after disruption in the base condition. Although changing departure time and changing route are the first two preferred options [[6], [7], [8]], on day one, bus and rail/subway systems were completely disrupted and on day two, the rail/subway system was completely disrupted; therefore, bus, and rail/subway commuters did not have the option of changing route. To be productive, these people had to change mode, or telework otherwise they would cancel work and in agreement with previous literature [[6], [7], [8]], on the first two days most common to least common options were cancel work trip, telework, and change mode.

From the third day, the subway/rail system started to recover, the number of people that change departure time and change route increases while the number of people that change mode decreases. Change mode is the least preferred option from third day to the last day of simulation (9^{th} day) as expected.

As the work closure percentage decreased, the transportation system became more crowded; therefore, more people changed departure time to deal with delay and crowding. On the 5^{th} day of the simulation, a big proportion of rail/subway links recovered, moreover, this day was the first full day of school and work. Therefore, there was a big jump in the number of people that changed their departure times on this day; they wanted to reach work on time and avoid the delay and crowding. In addition, the number of people that teleworked and canceled their work decreased on this day because most of the schools were reopened and families with children did not need to telework or cancel their work anymore. Also, many people traveled to work by subway/rail system instead of teleworking and canceling the work trip because many of subway/rail links were recovered on this day.

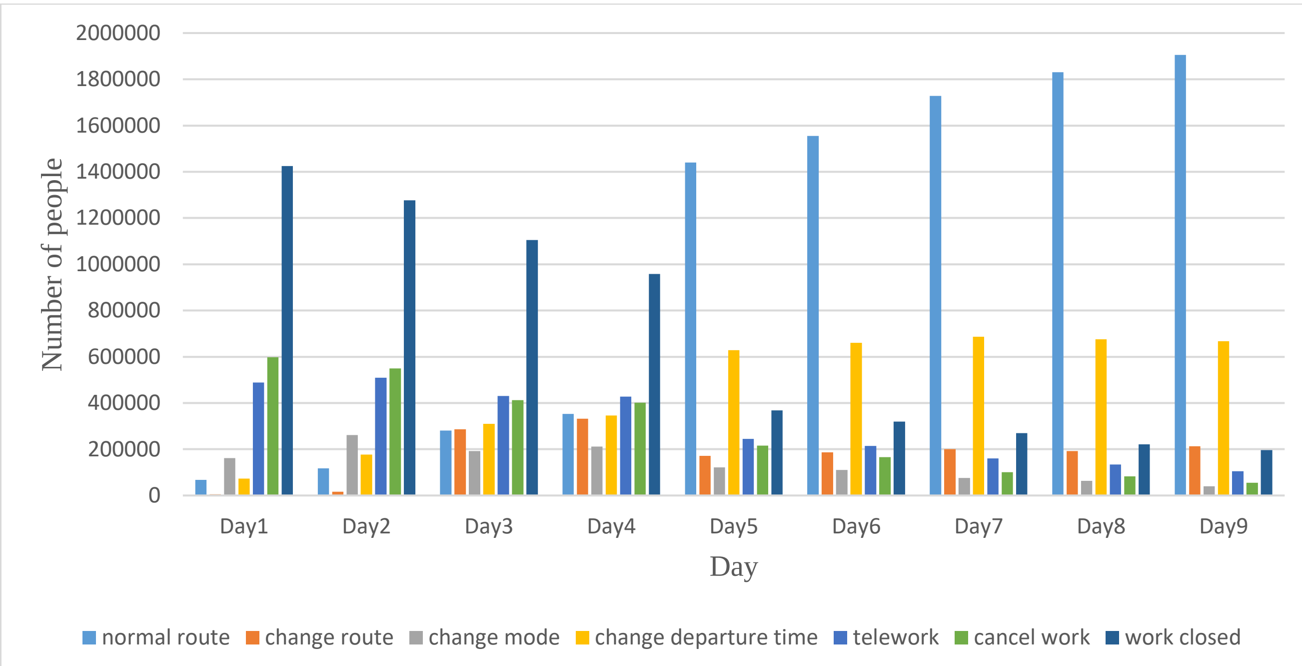


Figure 4. 6 People Adaptation

People whose work was opened and traveled to work had to choose one of the available transportation modes. Figure 4.7 presents the number of people using each commute mode. On the first day, the available options were to drive alone, carpool, taxi, and walk where the first two most common options were to drive alone and carpool. However, many people canceled work and teleworked because the transportation system was disrupted and it was not possible for everyone to change mode. On the second day, when the bus service recovered, the first two most common options changed to drive alone and bus. On day five, when a large portion of subway/rail system recovered, the number of people that used the rail/subway system increased substantially and the number of people that teleworked, and canceled work decreased noticeably. Toward the end of the simulation, the number of users of each mode gets closer to the normal condition, where this can show that people want to go back to their normal routine as soon as possible.

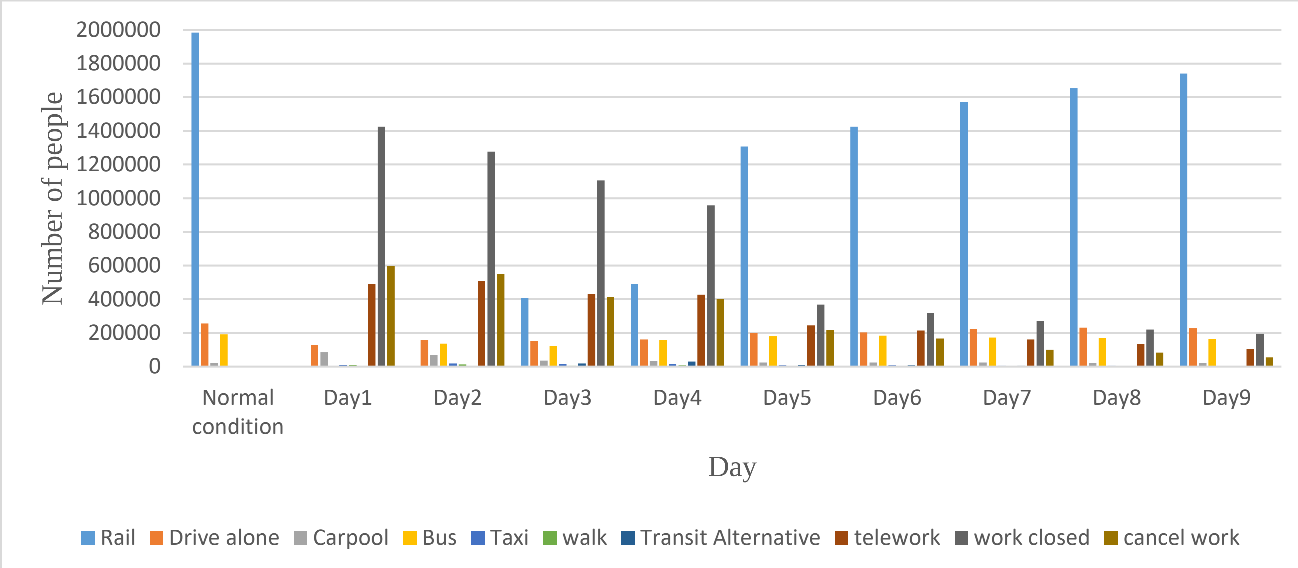


Figure 4. 7 Commute Mode of People

People chose a departure time daily based on their previous day's experience, environment condition and their usual departure time before disruption. Figure 4.8 shows the departure time distribution of people in normal condition and nine days in the base recovery condition. After disruption, departure times were more spread and there was not a distinct peak as most of the people adjusted their departure time in order to deal with delay and crowding, gasoline restrictions and other problems that were in the disrupted network.

In the base recovery condition, the total number of people that move from home to work on each day were different since the number of people that canceled their work trip and work closures were different on each day. Therefore, Figure 4.9 shows the percentage of people that departed in each timeframe in the normal condition and nine days in base recovery condition in order to compare them better.

The number of people that departed in earlier times (before 7 am) was more and the number of people that departed later than 9 am was less in base recovery condition in comparison to the normal condition because people wanted to skip delay and crowding. The number of people that moved from home to work in the peak hours were still more than other times even in the base recovery condition.

However, even on day 9, departure times were more spread than the normal condition and the number of people who traveled earlier than the peak hour was more than in the normal condition since the transportation system is still disrupted. The number of people who were moving to work by the transportation system was still less than the normal condition because some jobs were still closed and many people teleworked and canceled their work. In addition, if people decided to change departure time for the first time, the amount that they changed their departure time is based on distribution of people that answered the question about how much they changed their departure time. However, after the first time, if they decided not to change their departure time, their departure time would move toward their original departure time one step (half an hour) at a time. Therefore, it takes time for all people to go back to their normal departure time.

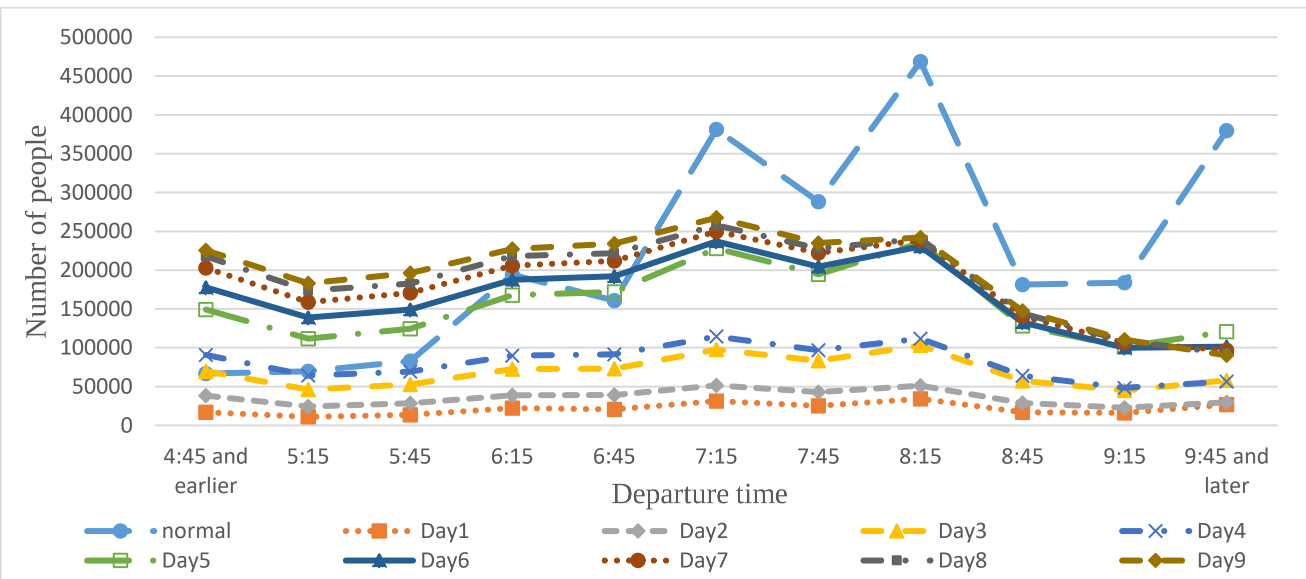


Figure 4. 8 Departure Time Distribution

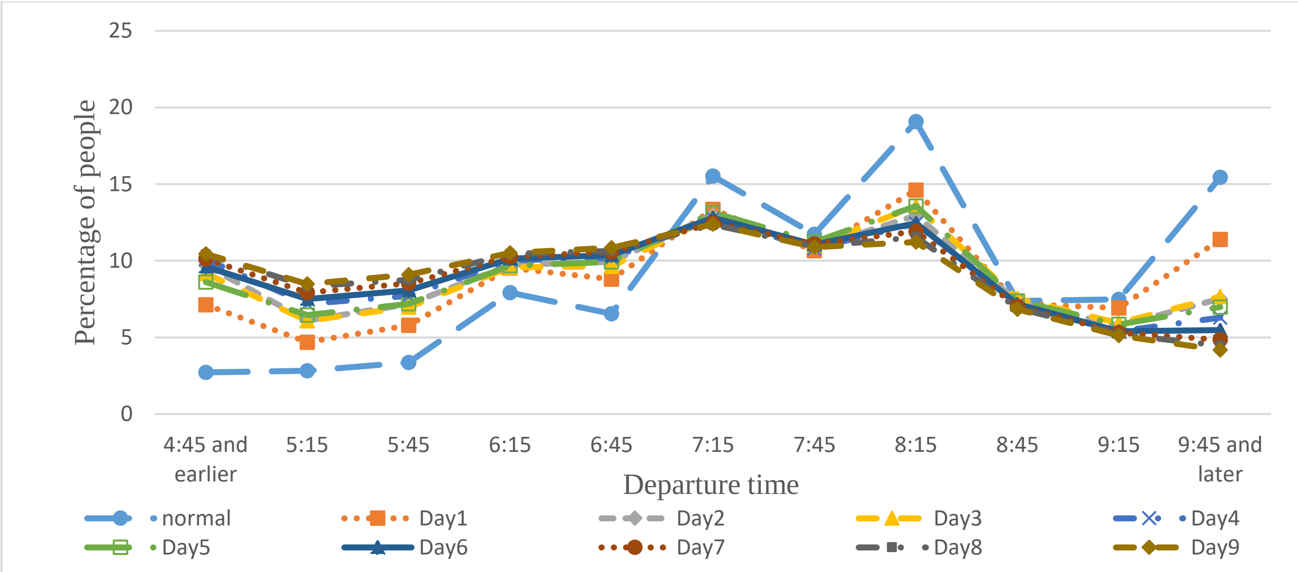


Figure 4. 9 Percentage of People that Depart in each Time

Chapter 4.2: Comparing Scenarios

Results of the base condition and six different scenarios are compared in this section. Figure 4.10 shows cumulative lost person-work days in the different scenarios. If a single person does not work on day 1 and day 2, he/she is counted twice in the graph.

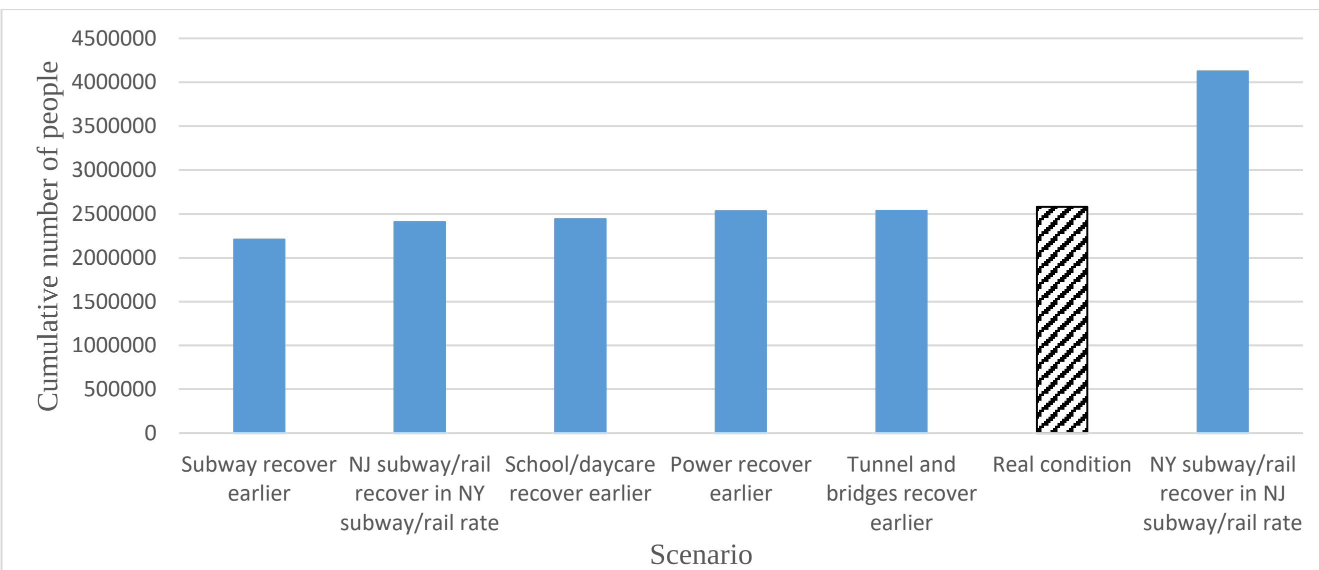


Figure 4. 10 Cumulative Lost Person-Work Days

Based on Figure 4.10, the most effective policy to the least effective one in regards to productivity are:

1. Subway/rail links recovered one day earlier
2. New Jersey area rail/subway recovered as fast as New York area rail/subway
3. School/daycares reopened one day earlier
4. Power recovered one day earlier
5. Tunnel and bridges recovered one day earlier
6. Base condition
7. New York area rail/subway recovered as fast as New Jersey area rail/subway

Based on census data, 31 percent of people are transit commuters in the New York Metropolitan Area. After Hurricane Sandy, the subway and rail system were disrupted for two days completely and NJ Transit recovered slowly. Not all of the transit commuters owned a car or were able to travel to work with the other modes of transportation; therefore, many of them had to cancel their work trip. As a result, when the rail/subway system recovered faster and NJ rail/subway recovered as fast as the NY rail/subway system, productivity increased. Figure 4.11 shows the number of people that do not work on each day after disruption. When subway/rail recovered one day earlier, the number of people that canceled their work decreased noticeably on the second day because in all other scenarios there was not any subway/rail system on the second day while in this scenario some subway/rail links recovered on the second day.

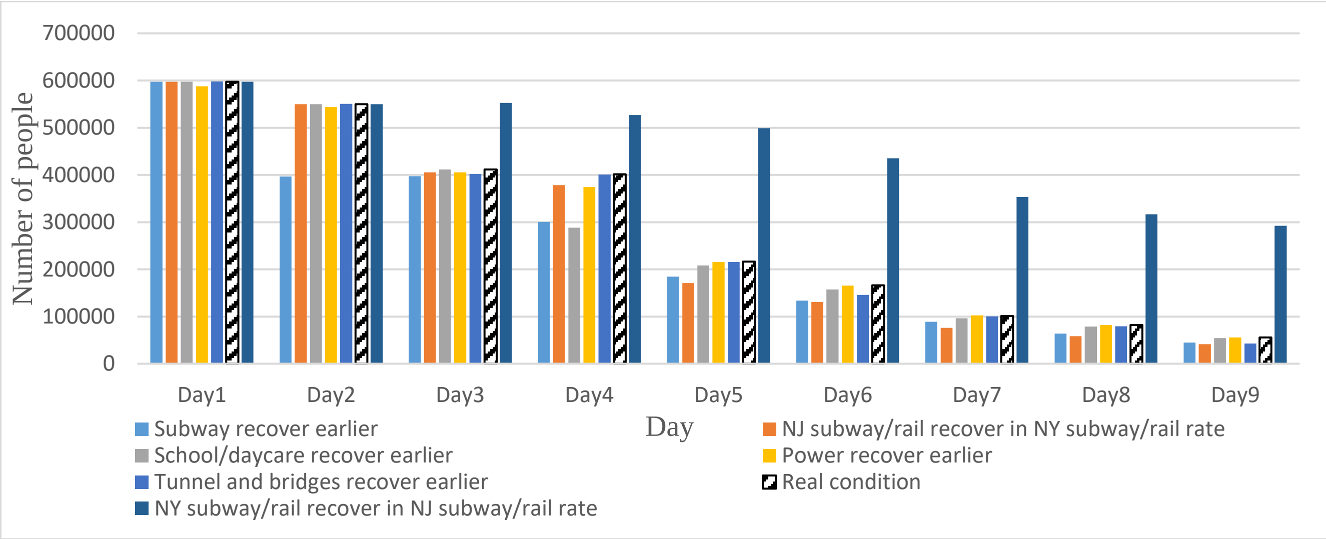


Figure 4. 11 Number of People That Do Not Work on Each Day

The third most effective strategy was recovering school/daycare one day faster. Although, schools and daycares are not part of transportation system, they directly affect the behavior of people who are transportation system users. When schools and daycares are closed, parents cannot travel to their work unless they can find another caregiver. The first day that most of the schools reopened after hurricane was day 5 of the simulation, where the number of people who canceled their work trip decreased. However, in the scenario that schools and daycares recovered one day faster, the number of people who canceled their work trip decreased noticeably from day 4. These numbers show how effective, schools and daycares conditions are in the recovery process.

The next effective scenario was recovering electricity one day earlier. Based on the survey data, 94 people who usually traveled to work teleworked after Hurricane Sandy, from which only 54 had the option of telework in normal conditions; therefore, after the disruption, many companies let their employees telework. If power is available, teleworking can be a good substitute for commuting to work because teleworkers can skip traffic, delay and crowding. Moreover, they do not need to shift to other modes. In addition, teleworking is a good option for families that have children while schools and daycares are closed. Figure 4.12 presents the total number of people that teleworked in different scenarios.

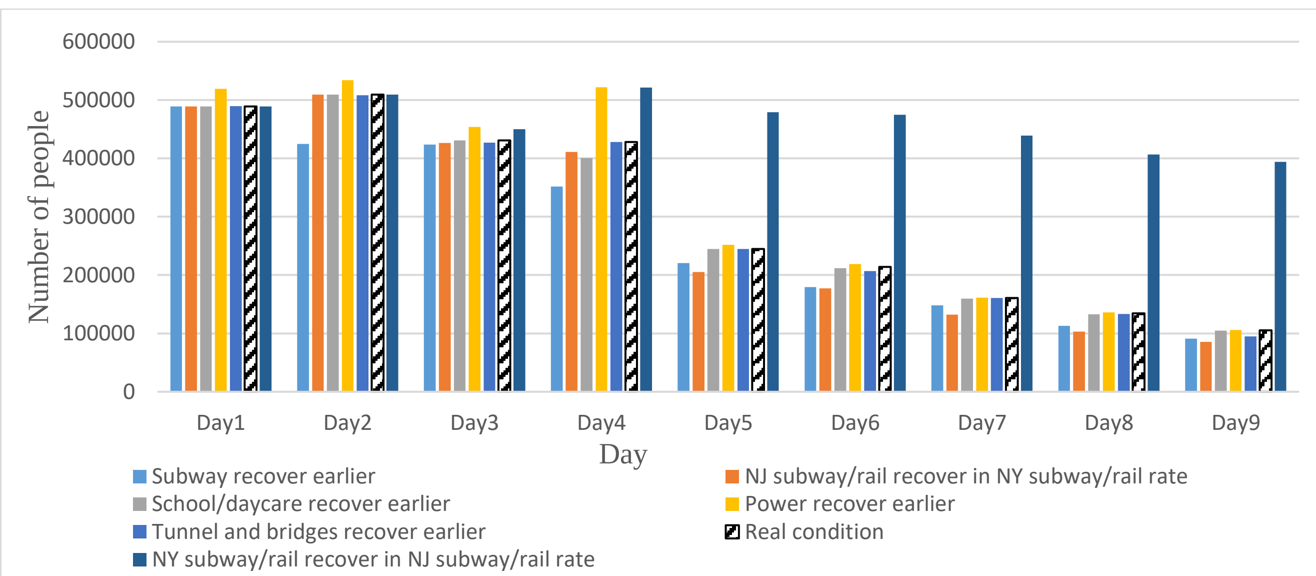


Figure 4. 12 Number of People That Telework on Each Day

Until day 5, the number of people who telework in the scenario where power recovered one day earlier is more than all of the other scenarios. Because teleworking depends on the power condition, when more people have power, the number of people that can telework increases. From day 5, the number of people that teleworked decreased in all scenarios except for the scenario where the NY subway/rail recovered at the NJ subway/rail system's pace (last scenario). Since in all of the scenarios the transportation system condition improved by day 5, people travel to work by using transportation systems except the last scenario where even on day 5 many of the subway/links were not functional. Therefore, people had to telework, cancel work, change mode or try to find a way by changing route in the available functional links and paths. Figure 4.13 and 4.14 show the number of people that change modes and change routes after the disruption in different scenarios. As it is clear in the figures, starting on day 5, the number of people that change route and change mode decreases in all of the scenarios except for the last scenario where the number of people who change their route and mode are still considerable.

The number of people that changed mode are less when the subway recovered earlier, NJ subway/rail recovering at the NY subway/rail rate, and power recovered sooner scenarios. In the first two scenarios, people were able to travel to work with their normal mode more because the subway/rail system recovered faster so they did not need to change mode. In the third one, people were able to telework more, so they did not need to change mode that much.

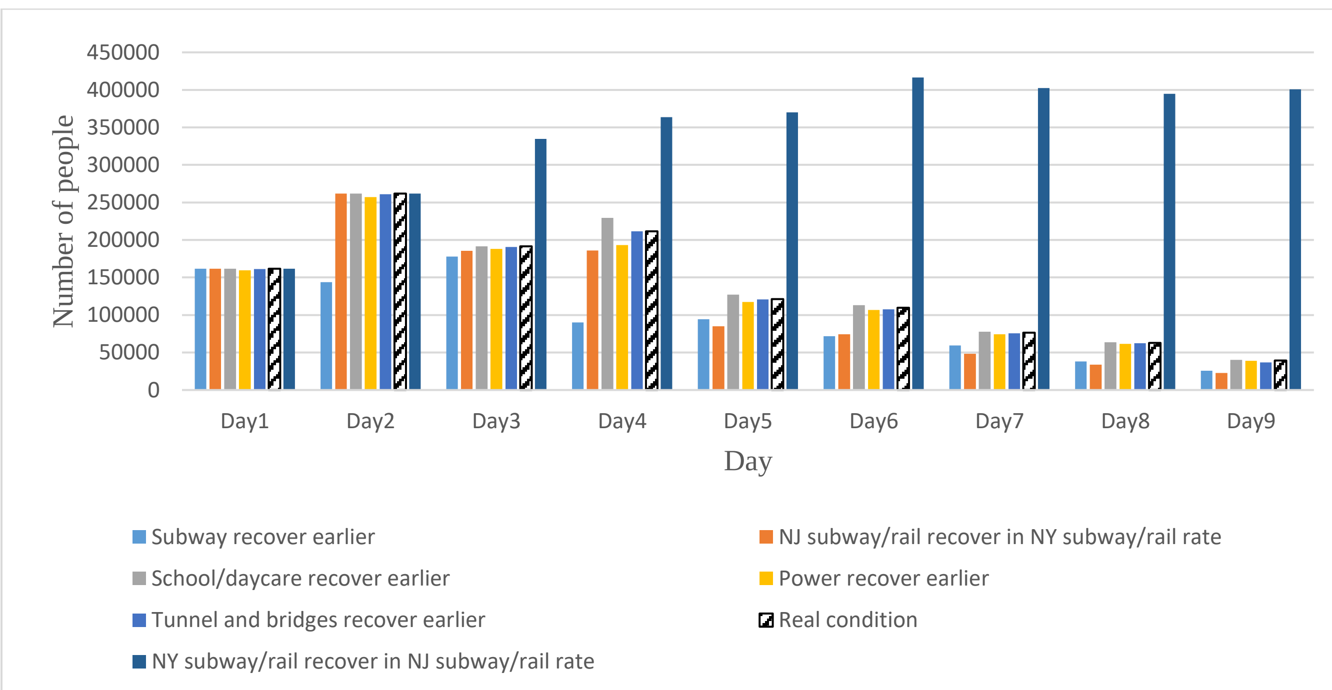


Figure 4. 13 Number of People that Change Their Mode on each Day

The fifth most effective scenario was the one where tunnels recovered one day earlier and carpool restrictions started and ended one day earlier. The number of people that return to a productive lifestyle in this scenario was not as much as in the first three scenarios. The number of people driving through one of the bridges and tunnels was less than the number of rail/subway commuters. In addition, out of three bridges and tunnels that connect New York to New Jersey, only one of them was closed for several days and all other were opened. Moreover, out of the seven bridges and tunnels that connect Brooklyn and Queens to New York, only two of them were closed for some days. Therefore, people always had a chance to reach to their work by changing route.

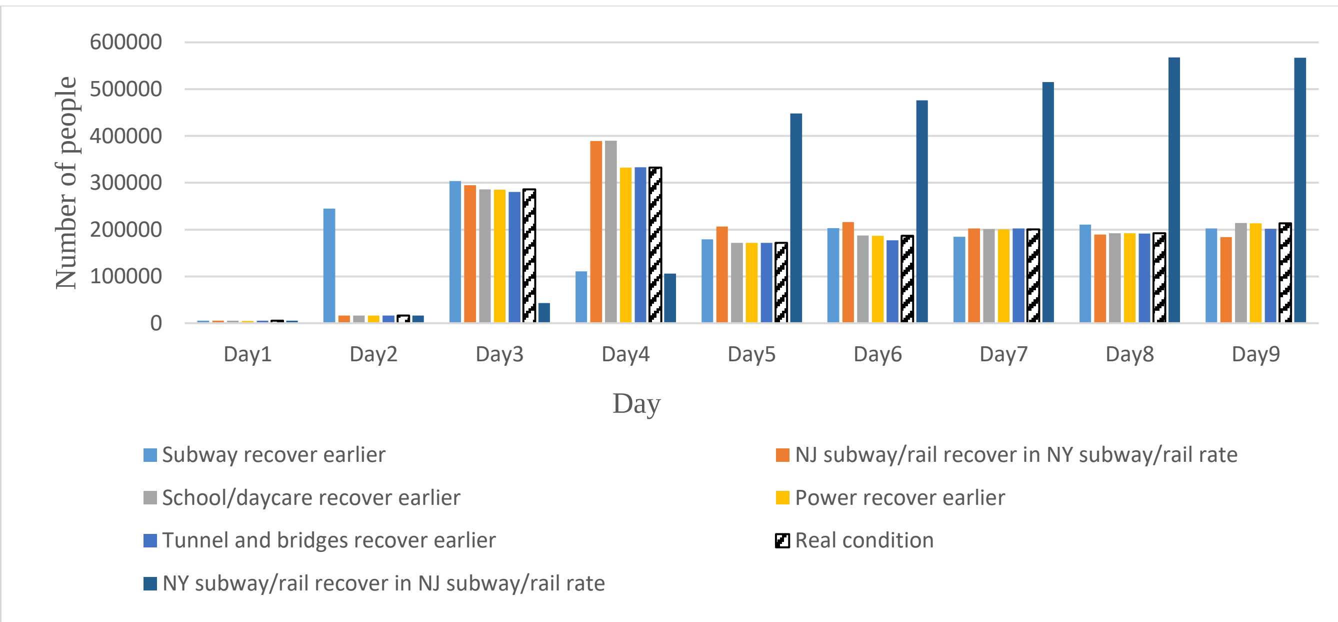


Figure 4. 14 Number of People that Change Their Route on each Day

The worst scenario that was much worse than even the base condition was the last scenario. In this case, there were no rail/subway lines available for many people to move from home to work and not all of the subway/rail commuters had a car or access to other modes of transportation. Moreover, telework is not an option for all kinds of occupations. Therefore, in this scenario, the number of people that canceled their work is more than all other scenarios.

Figure 4.15 shows the number of people who changed their departure time in different scenarios. People changed their departure time in most of the cases because of delay and crowding. Therefore, in all of the scenarios, the number of people that change their departure time is close to each other except for last scenario, where the number of people who changed their departure time is less than other cases. Since in the last scenario more people canceled their work and teleworked, fewer people are using the transportation systems, therefore there was not much delay and crowding so fewer people changed their departure times.

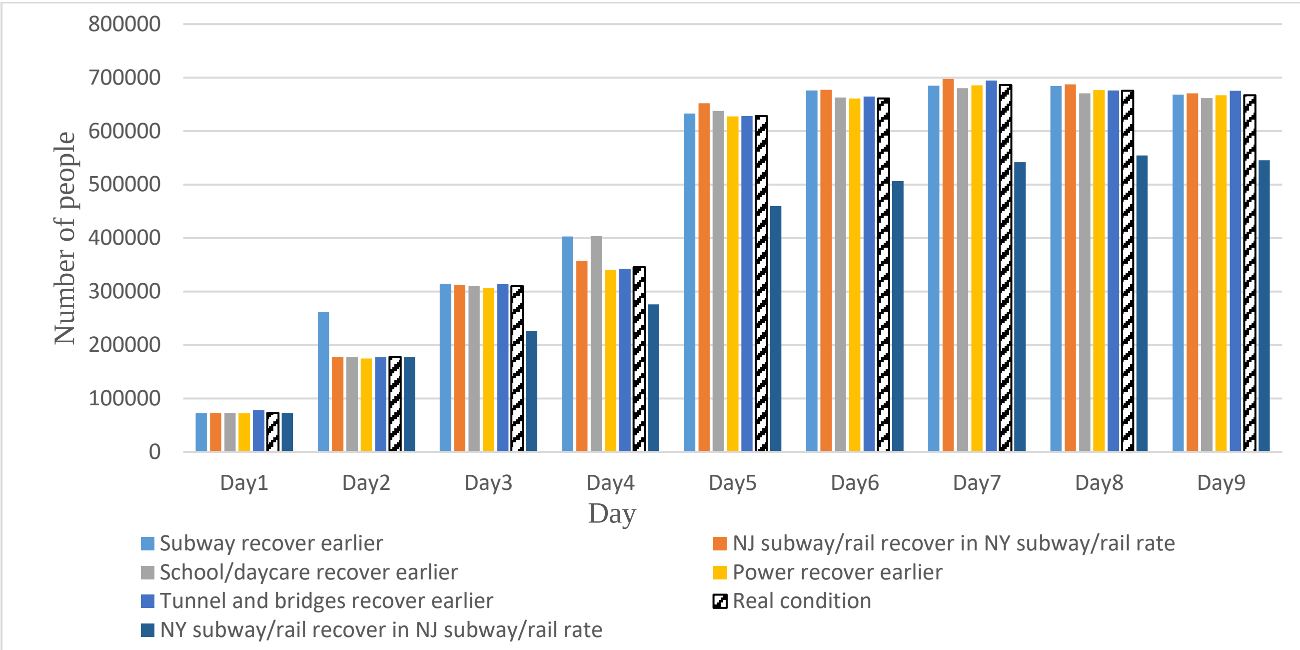


Figure 4. 15 Number of People that Change Their Departure Time on each Day

CHAPTER 5: CONCLUSION

In this research, data from a survey of 397 respondents in NYC Metropolitan Area was used to develop an agent-based model that captures commuter behavior and adaptation and simulates their behavior for nine working days after Hurricane Sandy. In this agent-based model, a series of if-then rules and logit models were used for defining agents' behavior and methods of interaction. A description of route, mode and departure time choice was presented in this thesis for each agent and they learned from their experience and changed their behavior based on their experience (crowdedness). In this model, each agent was able to adapt to the new situation by choosing one or more of the modifications including: change route, change departure time, change mode, telework and cancel work trip.

Six different recovery scenarios were tested by using this model to find critical factors that promote a faster return to productivity. Cumulative lost person-work days were calculated for all six scenarios and base condition. Change in productivity was calculated based on the percentage change in cumulative lost person-work days in each recovery scenario in comparison to the base recovery condition. Many of the people in NYC Metropolitan Area have rail/subway as their main mode of transportation and not all of them are able to change mode or telework. Therefore, rail/subway system disruption can stop many people from moving to work. After Hurricane Sandy, the subway/rail system was completely disrupted for two days and then took a while to recover completely. However, recovery process of NJ Transit and Path rail took longer. As a result, three out of six recovery scenarios were about subway/rail system. In the first one, subway/rail system recovered one day earlier. In the second one, NJ rail/subway system recovered at the same rate as the NY rail/subway system while in the last one NY rail/subway system recovered at the same rate as the NJ rail/subway system. The first two recovery scenarios were the most effective in promoting productivity by 14.5 and 6.7 percent, respectively while the last one was the worst scenario that decreased productivity by 59 percent.

It is important to consider users' preferences and needs while planning for recovery. Humans are adaptive and many factors can change their behaviors and reactions. The recovery process for working parents is not the same as families without children. Parents are responsible for their children and this responsibility may cause them to cancel their work trip even if the transportation system has completely recovered. The other recovery scenario examined the effect of school and daycare closure by comparing productivity in base condition and a condition that daycare and schools recovered one day earlier than based condition. This scenario promoted productivity by 5.4 percent and was the third best policy.

After disruption, telework can be a good substitute for physical movement of people. By telework, people can be productive while they skip delay, crowding and struggling with a disrupted transportation network. Power and communication systems are needed for telework. In the next scenario, the effect of power and telework on recovery was examined by comparing productivity in the base recovery condition and a condition where the power system recovered one day earlier than the base condition. This scenario promoted productivity by 1.9 percent. Therefore, by recovering electricity faster and encouraging employers to let their employees' telework, people can return to productivity faster.

Closure of tunnels and carpool restrictions affected people's commuting pattern after Hurricane Sandy but their effect was not as much as subway/rail disruption because all the time there were some open bridges that people could move to their work by changing route. However, people that did not want to change their route, could change their mode, telework or cancel their work trip. In another scenario, all the tunnels were opened one day earlier and carpool restrictions started and ended one day earlier. In this scenario, productivity increased by 1.8 percent compared to the base condition.

As agent-based models represent situations more realistically, the outputs of the simulation become more reliable. Survey data and data about environment condition form the basis of an agent-based model. Therefore, one important factor in improving agent-based models is the way a survey is designed. Surveys include some questions that people do not like to answer or do not know the answer. For instance, questions that had many missing answers in this survey were about income, home zip code and work zip code. Missing and wrong answers in work zip code were more than home zip code. This can show that people do not like to give information about their income and work location or they do not recall their work zip code correctly. Therefore, it is important to have enough survey respondents so if some respondents are omitted due to missing data, enough observations still remain. Moreover, it is useful to ask another question about location besides zip code, like city or county where people work and live since people may give answers that are more accurate in this way. Also, knowing the city or county can help to predict missing zip codes more easily.

Asking about family structure can be useful in predicting behavior after disruption. The structure can help inform assumptions about resources and impediments for returning to productivity. In particular, the modeling could be improved by knowing how many people live in the household; how many of them work; if they have any children, are they married parents or a single parent; and if they are married parents, are they dual-income or not.

There are some questions that could not be answered by different sources or even the American Community Survey. A survey can be a good way for finding answers to these questions. One of these questions is about presence of any other caregiver for children in the household. It is important to know, what parents do in situations when schools and daycare are closed and they need to go to work. Another question that is not included in the American Community Survey is options that employees have, like teleworking or flexible working hours. Questions about telework and flexible working hours were included in the survey used in this study and they were very helpful in predicting the probability of telework after disruption.

Some information that is needed for developing the agent-based model is about the environment condition. However, there are not enough sources readily available for finding this information including school/day care condition, work condition and power condition after disruption. Perhaps, the best way for capturing information about these environment components is asking the survey respondent when school/day cares reopened for children in the household, when work reopened for workers in the family, and when power restored in their home. On the other hand, there can be questions that ask when children in the household went back to school/daycare, and when workers in the family returned to work. If the date that the children went back to school/daycare was later than the date that they reopened or the day that workers in family returned to work was later than the day that work reopened, there can be a question that asks for the reasons for the differences. Although people may not like to answer open-ended questions in the survey, asking some open-ended questions may help in capturing unpredictable behaviors.

Some other open-ended questions can be asked in the form of travel diaries. Asking for travel diaries can help in capturing behavior and decision-making processes. It is useful to ask people to explain why they made each change. For instance, they departed earlier because of delay and crowding or they depart earlier because they needed to drop their children somewhere on their way to work.

There are people that use more than one mode of transportation for traveling from home to work. To make the agent-based model more realistic, it is important to know how people shift from one mode to other or if they use bus, rail or subway how they reach the station. After the disruption, some people change mode. It is important to ask which transportation modes other than their normal mode is accessible for them; for instance, do they have a car or not. In addition, if they change their mode, they would use which mode of transportation instead of their normal mode.

Finally, it is important to gather data needed for modeling as soon as possible. Because in the case of surveys, people may forget answers to some of the questions and in regards to environment situation, some information may not be available anymore after some time.

More work remains to be done with this model for capturing effective factors on recovery and presenting a comprehensive recovery model. For future research, this agent-based model can be extended by modeling the environment, particularly power and transportation systems in more detail. All power system and transportation system components that are affected by disruption can be modeled. In this agent-based model, only power condition is modeled (whether people have power in their home or not) but in future research components of power system can be modeled as well. Then, different power recovery timeline can be examined while accounting for human adaptation. Moreover, in some cases power and transportation system problems can affect each other. For instance, disruption of some subway lines depends on both power outage and flooding in tunnels. Also, power outage in traffic lights can cause delay and crowding problems. By modeling power and transportation components both, a more comprehensive plan for recovery that accounts for human behavior can be developed.

In addition, some modeling components related to family structure can include more details for families with children. In this research, for all the people, same change departure time model is used also if people want to change their departure time for all of them distribution of people answer to the question that by how much you will change your departure time from survey data is used. However, in families with children, different variables may affect their decision about change departure time and amount that they change their departure time in comparison to families without children since they are responsible for their children. Therefore, it may be more accurate to develop two different models for people with and without children in case of changing departure time.

More detail related to daycare and schools can be added into future model. First, ability of people working in school and daycare to get to work should be considered. Because schools are, actually reopen when school/daycare building has a good condition and all people working there are able to be present to their job. A daycare and school with power, water, and gas and in a good condition but without a teacher cannot be considered as open since no one is present to take care of students. Second, a study about school bus provider can be conducted for future research in order to figure out ability of children for reaching to school and daycare after disruption through disrupted transportation system. Moreover, using travel diaries can help capture more detailed information from people's behavior and develop an agent-based model that simulate situations after the disruption in more detail.

References:

1. Wamsler, C., *Managing urban risk: perceptions of housing and planning as a tool for reducing disaster risk.* 2004.
2. Kaufman, S., et al., *Transportation During and After Hurricane Sandy.* 2012.
3. Lindell, M.K., *RECOVERY AND RECONSTRUCTION AFTER DISASTER.*
4. Nejat, A. and I. Damnjanovic, *Agent-Based Modeling of Behavioral Housing Recovery Following Disasters.* Computer-Aided Civil and Infrastructure Engineering, 2012. **27**(10): p. 748-763.
5. Kontou, E., P. Murray-Tuite, and K. Wernstedt, *Commuter Adaptation in Response to Hurricane Sandy's Damage.* 2016.
6. GIULIANO, G. and J. GOLOB, *Impacts of the Northridge Earthquake on Transit and Highway Use.* 1998.
7. Zhu, S., et al., *The traffic and behavioral effects of the I-35W Mississippi River bridge collapse.* Transportation Research Part A: Policy and Practice, 2010. **44**(10): p. 771-784.
8. L. Mokhtarian, P., L. Ye, and M. Yun, *The Effects of Gender on Commuter Behavior Changes in the Context of a Major Freeway Reconstruction.* 2010.
9. Van Exel, N.J.A. and P. Rietveld, *Public transport strikes and traveller behaviour.* 2001.
10. Levinson, D.M. and S. Zhu, *Disruptions to Transportation Networks: A Review*. Transportation Research, Economics and Policy. 2012.
11. Bonabeau, E., *Agent-based modeling: Methods and techniques for simulating human systems.* 2002.
12. Macal, C.M. and M.J. North, *Tutorial on agent-based modelling and simulation.* Journal of Simulation, 2010. **4**(3): p. 151-162.
13. Yin, W., et al., *An agent-based modeling system for travel demand simulation for hurricane evacuation.* Transportation Research Part C: Emerging Technologies, 2014. **42**: p. 44-59.
14. Evans, M.R. and D.s. Elston, *A Primer for Agent-Based Simulation and Modeling in Transportation Applications*, in *The Exploratory Advanced Research Program*, U.S.D.o. Transportation and F.H. Administration, Editors.
15. Bernhardt, K.L.S., *Agent-Based Modeling in Transportation* TRANSPORTATION RESEARCH CIRCULAR E-C113, 2007.
16. Dia, H., *An Agent-Based Approach to Modeling Driver Route Choice Behavior Under the Influence of Real-Time Information.* Transportation Research Part C, 2002.
17. Kukla, R., et al., *Development of an Autonomous Agent Model of Pedestrian Flow.* Journal of the Transportation Research Board, 2001.
18. Zhang, L. and D. Levinson., *Agent-Based Approach to Travel Demand Modeling.* Journal of the Transportation Research Board, 2004.
19. Hager, K., J. Rauh, and W. Rid, *Agent-based Modeling of Traffic Behavior in Growing Metropolitan Areas.* Transportation Research Procedia, 2015. **10**: p. 306-315.
20. Chen, X., J.W. Meaker, and F.B. Zhan, *Agent-Based Modeling and Analysis of Hurricane Evacuation Procedures for the Florida Keys.* Natural Hazards, 2006. **38**(3): p. 321-338.
21. Chen, X. and F.B. Zhan, *Agent-based modelling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies.* Journal of the Operational Research Society, 2006. **59**(1): p. 25-33.
22. Lammel, G. and H. Klupfel, *Slower is faster: the influence of departure time distribution on the overall evacuation performance*. 2012.

23. Nam , H., et al., *An Agent Based Model for the Simulation of Transport Demand and Land Use.* SMART Infrastructure Facility, University of Wollongong, NSW, Australia.
24. Rossetti, R.J.F., et al., *An Agent-Based Framework for the Assessment of Drivers' Decision-Making.* IEEE, 2000.
25. Grinberger, A.Y. and D. Felsenstein, *Bouncing Back or Bouncing Forward? Simulating Urban Resilience and Policy in the Aftermath of an Earthquake.*
26. Srikukenthiran, S. and A. Shalaby, *Enabling large-scale transit microsimulation for disruption response support using the Nexus platform.* Public Transport, 2017. **9**(1-2): p. 411-435.
27. Enders, C.K., *Applied Missing Data Analysis*. 2010.
28. *The Majority of Children Live With Two Parents, Census Bureau Reports*. United State Census Bureau 2016; Available from: https://www.census.gov/newsroom/press-releases/2016/cb16-192.html.
29. Bloomberg , M.R., *A Stronger, More Resiliant New York.* 2013.
30. *Current and Projected Populations*. Available from: http://www1.nyc.gov/site/planning/data-maps/nyc-population/current-future-populations.page.
31. *Effects of Hurricane Sandy in New Jersey*. 2012 2017; Available from: https://en.wikipedia.org/wiki/Effects_of_Hurricane_Sandy_in_New_Jersey.
32. Caroom, E. *Electricity not likely to come back soon for N.J.* 2012; Available from: http://www.nj.com/news/index.ssf/2012/10/electricity_not_likely_to_come.html#incart_m-rpt-2.
33. Lee, E. *In the wake of Sandy, 775K in New Jersey still in the dark even as recovery continues*. 2012; Available from: http://www.nj.com/news/index.ssf/2012/11/more_than_775k_customers_in_nj.html.
34. *Nearly 400K in N.J. still without power as new storm closes in*. 2012; Available from: http://www.nj.com/news/index.ssf/2012/11/nearly_400k_in_nj_hope_coming.html.
35. *PSEG*. Available from: https://www.pseg.com/.
36. *Jersey Central Power & Light.*
37. *Company History & Statistical Information*

Available from: https://www.oru.com/en/about-us/media-center/corporate-facts.

38. *Company Information*. Available from: https://www.atlanticcityelectric.com/AboutUs/Pages/CompanyInformation.aspx.
39. Chakrabarti, R. and M. Livingston. *The Impact of Superstorm Sandy on New York City School Closures and Attendance*. 2012; Available from: http://libertystreeteconomics.newyorkfed.org/2012/12/the-impact-of-superstorm-sandy-on-new-york-city-school-closures-and-attendance.html.
40. Rundquist, J. *2 weeks after Hurricane Sandy, most schools are finally back in session*. 2012; Available from: http://www.nj.com/news/index.ssf/2012/11/2_weeks_after_hurricane_sandy.html.
41. *NYC Mass Transit Spatial Layers*. 2017; Available from: https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Mass+Transit+Spatial+Layers.
42. Romalewski, S. *MTA subway data in GIS format*. 2010; Available from: https://spatialityblog.com/2010/07/08/mta-gis-data-update/.

43. *New Jersey Geographic Open Data*. 2017; Available from: http://njogis-newjersey.opendata.arcgis.com/.
44. *Hurricane Sandy Effects on Subway & Rail Service*. 2012; Available from: https://subwaynut.com/sandy/.
45. Editors, C.c. *Best and Worst Gas Mileage* 2018; Available from: https://www.cars.com/articles/best-and-worst-gas-mileage-2018-1420698621218/.
46. *Fuel tank*. Available from: https://en.wikipedia.org/wiki/Fuel_tank.
47. *United States Census Bureau*. American Community Survey; Available from: https://www.census.gov/programs-surveys/acs/data/pums.html.
48. *United States Census Bureau*. TIGER/Line Shapefiles; Available from: https://www.census.gov/cgi-bin/geo/shapefiles/index.php.
49. *United States Census Bureau* ZIP Code Tabulation Area (ZCTA) Relationship Files; Available from: https://www.census.gov/geo/maps-data/data/zcta_rel_download.html.
50. *United States Census Bureau* ZIP Code Business Statistics; Available from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=BP_2012_00CZ1&prodType=table.
51. *IPUMS USA*. Available from: https://usa.ipums.org/usa-action/variables/PWPUMA00#codes_section.
52. *PopulationSim*. Available from: https://rsginc.github.io/populationsim/.
53. *2010 Census Public Use Microdata Area (PUMA) Reference Maps - New York*. United State Census Burea; Available from: https://www.census.gov/geo/maps-data/maps/2010puma/st36_ny.html.
54. *2010 Census Public Use Microdata Area (PUMA) Reference Maps - New Jersey*. United States Census Bureau; Available from: https://www.census.gov/geo/maps-data/maps/2010puma/st34_nj.html.
55. *PUMS Technical Documentation*. Available from: https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html.

Appendix A: Method of Generating Population by PopulationSim

The PopulationSim package was used for generating the population. Useful information about this package, installation and instruction of how to use it can be found in [52]. It is useful to first read instructions provided in [52] and then this step by step method of generating a population with this package that is outlined below:

1. This package runs in the python environment; therefore, Anaconda Python2.7 should be installed. Ortools library (a library that is needed for running PopulationSim) is only work on 64 bit so 64 bit Anaconda python 2.7 should be installed.

2. Check whether you need to add python and Anaconda to environment variables. Open the command prompt and write python and click enter. If you get this message “python is not recognized as internal or external command”, you need to add them. There are two ways for doing this. In both of the ways, first we need to know where Python and Anaconda are. Open the Anaconda prompt and write “where python” and click enter. Again, write “where anaconda” and click enter.

```
(base) C:\Users\Elham>where python
C:\Users\Elham\Anaconda2\python.exe

(base) C:\Users\Elham>where anaconda
C:\Users\Elham\Anaconda2\Scripts\anaconda.exe
```

When copying these addresses you only need to copy C:\Users\Elham\Anaconda2 and C:\Users\Elham\Anaconda2\Scripts parts.

- In the first approach, you can add these variables manually by following this direction: control panel---system and security---system---advanced system settings---environment variable. If you already have a variable with name of Path, click on that and click edit otherwise, click on new and write Path in variable name part. In value part, copy the address of where python is installed from the Anaconda prompt; copy the address of where Anaconda is installed and click ok.
- In second approach, open command prompt and use the setx command to add them to environment variable.

```
Microsoft Windows [Version 10.0.17134.165]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Elham>SETX PATH "%PATH%;C:\Users\Elham\Anaconda2;C:\Users\Elham\Anaconda2\Scripts"

SUCCESS: Specified value was saved.

C:\Users\Elham>
```

Close the command prompt and open it again. Write "python" and click enter to check whether they have added to path correctly. If they have added correctly, this message would be shown.

```
C:\Users\Elham>python
Python 2.7.14 |Anaconda, Inc.| (default, Nov  8 2017, 13:40:45) [MSC v.1500 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

3. Close and open the command prompt again to create an Anaconda environment that includes all the libraries that are needed for using the PopulationSim package. Write conda create –n name python=2.7 and press enter. Instead of "name" you can write your desired name. After this environment is created, type activate name and press enter. Then in this environment you need to install all required libraries that are listed in the getting started section of [52].

4. Download example_calm from [52]. This example include files that need to be modified for each different population generation. In the data folder in example_calm there are 6 different .csv files that need to modified. For making seed_households and seed_persons files, PUMS data is used. In this website [47], population records and housing unit records can be found for the desirable year and state in .csv format. New York population records, New York housing unit records, New Jersey population records, and New Jersey housing unit records were used for this research. Population records are used for making the seed_persons file and housing unit records are used for making the seed_households file. The definition for PUMA values indicating which number is for which county in that state is available in [53] for New York and in [54] for New Jersey. PUMAs that are not located in the modeling area can be omitted from seed_households and seed_persons files.

PopulationSim should run for each region separately, once for New Jersey and once for New York. After downloading population and housing unit records, the PUMS dictionary [55] can be used to read each variable definition and choose the ones that are needed for modeling purposes. Any variable that is not needed can be omitted. The way that one variable is defined can be modified.

For instance, if being female is shown by one and being male is shown by two but desired way is to have female as one and male as two, this can be modified, or occupations can be grouped together. New variables can be added, for instance variable ESR shows employment status records that has six groups.

```
ESR         1
     Employment status recode
            b .N/A (less than 16 years old)
            1 .Civilian employed, at work
            2 .Civilian employed, with a job but not at work
            3 .Unemployed
            4 .Armed forces, at work
            5 .Armed forces, with a job but not at work
            6 .Not in labor force
```

If people that work is one of the desired variables, this variable can be created, with a value of 1 if ESR is 1 or 4, and zero otherwise.

SERIALNO is a variable that shows the id of each household, people having the same SERIALNO are in the same household. A variable name hhnum should be created based on SERIALNO. In the seed_households file, this variable starts from one to the total number of households. In the seed_persons file, all the people that are in the same household get the same value for hhnum. Another variable that should be created in the seed_persons file is wgtp. This variable is the same as the WGTP variable in the seed_households file. All the people in the same household get the same value for the wgtp variable that is equal to the WGTP value in the seed_households file for that household.

5. For each geographic level, a control file is needed, except for the PUMA level. In the example_calm, there are four geographic levels, TAZ, Tract, PUMA, and Meta region. In this research, three different geographic level were used, zip code, PUMA, and Meta region. The TAZ file includes the smallest geographic level; therefore in this research, the TAZ file includes zip code information. Therefore, two-control files are needed control_total_taz and scaled_control_total Meta. Any other naming format can be used. For instance, it is possible to use a naming format like geography1, geography2 instead of TAZ and PUMA but then the names of control files should be updated too.

In control_total_taz there are six different columns. The first column is TAZ which is a list of geographic divisions that is used for the project; in this case, it is a list of zip codes. POPBASE shows the population of each TAZ, and HHBASE shows the number of housing units (number of households) in each TAZ. STATEFP, PUMA, and REGION show state, PUMA and region that each TAZ is in it. The region value is one since only one region is modeled at a time. Values in TAZ should be unique, so if one zip code is in more than one PUMA for each part of this zip code that is within a different PUMA a different name should be used. For instance zip code 7601 is within PUMAs 301 and 307; since it is not possible to have two 7601 under the TAZ column, this zip code was renamed to 107601 and 207601, corresponding to the parts of the zip code in PUMAs 301 and 307, respectively. Any other naming format can be used.

There can be controls on some variables like the number of households with different sizes in each TAZ. For each control, a column should be added to the control_total_taz file that includes the population of that group in each TAZ. For instance, to control different household sizes, four different columns named HHSIZE1 to HHSIZE4 should be created in the control_total_taz file that includes the number of households with population of 1 to 4 in each of the TAZs. Having any of these controls is optional except for one control that is mandatory. The mandatory control is the number of households in the smallest geographic level and the HHBASE variable is in the control_total_taz file for this reason.

Scaled_control_total_meta includes list of PUMAs and their populations and if any other optional control is needed in the Meta level. The Geo_cross_walk file includes all geographic levels and their relationships. For instance, each TAZ is within which PUMA and Region. This relationship file was developed with GIS by using the TIGER/Line Shapefiles [48] and the ZIP Code Tabulation Area (ZCTA) Relationship Files [49] from the United States Census Bureau.

6. In the configs folder, there is a controls.csv file. In this file, variables that the researcher wants to be controlled are defined. As mentioned above, one variable that is necessary to control is the number of households. This variable is called hhnum. The controls.csv file includes six columns. In the target column, the name of controlled variables is written; geography shows the geographic level of control. For hhnum, geography should be the smallest geographic level, that is TAZ. Seed_table column shows control variable (for instance hhnum) is in which file, seed_persons or seed_households. In the control_field column, the name of the variable that shows the number of

this variable in control files should be written. For hhnum, the number of households in each zip code is in the TAZ control file with the name of HHBASE. Therefore, HHBASE is written under the control_field column. Finally, in the expression column, the range and definition of variable is written. For hhnum, housing units that their weight is more than zero and less than infinity should be used for control because housing units with weight of zero are vacant. Therefore, this expression (households.WGTP > 0) & (households.WGTP < np.inf) is written under the expression column. There can be control for other variables if needed. There are some examples in example_calm and their explanation is available in [52].

7. If any name other than the ones in example_calm were used for files name or variables, the settings.yaml file under the configs option needs to be updated. To open this file, first open anaconda navigator. Then launch spyder. Click in the file---open and then open the settings.yalm file. Part of the naming and variables that are different can be updated in spyder environment. One part of the code includes geographies. This should include the name of all available geographic levels in model. In example_calm geographies are [REGION, PUMA, TRACT, TAZ] but in this model TRACT is not a geographic level so this part of code is updated to [REGION, PUMA, TAZ]. In addition, in the input_table_list part of the code, tract_control data is omitted. In the output table part of the code, summary_TRACT is deleted. In the output_synthetic_population part of the code, all the variables that are needed in output files should be listed. Moreover, in the run_list part, sub_balancing.geography=TRACT is deleted. All the files related to TRACT have been deleted because TRACT is not a geographic level in this model.

8. After updating all files and codes based on desired geography level, controls and variables, open the command prompt and navigate to the folder that includes all the model files. For instance if you want to run example_calm, you need to navigate to the example_calm folder. To do this, first the address of "example calm" is found. If the address is C:\Users\Elham\Desktop\files\example_calm after opening the command prompt, we are in C:\Users\Elham. The cd command is used to navigate to the desired folder. Then, the anaconda environment that was created earlier needs to be activated. Next, write "python run_populationsim.py" and press enter. This command will run the code and all the outputs will save in .csv format in the output folder. Then input files of MATLAB code should become ready from outputs of populationSim in a way that is needed for the code.