Grant/Award #: HDTRA1-16-0043
PI Name: William G. Kennedy
Organization/Institution: George Mason University
Project Title: A Framework for Modeling Society Following a Nuclear WMD Event

What are the major goals of the project?

Original statement of work (SOW): "The objective of this research is to develop a framework for modeling societal responses to a nuclear WMD event, for use in support of a humanitarian response. This framework will facilitate projections of how an affected society might react to a WMD event both in the short term (e.g. evacuation) and longer term (e.g. shelters, food distribution, migration), helping decision makers plan and structure their relief efforts."

<u>Revised objective:</u>
"The objective of this research is to characterize the reaction of the population of a megacity and surrounding region to a nuclear WMD event."

<u>Why revised:</u>
We revised the objective in response to discussions with the sponsor clarifying the sponsor's interest in the short-term "reaction," not long-term "response" and the general inadvisability of studying a nuclear WMD event in a foreign country. Given this is basic research, we dropped the discussion of potential uses of the resulting understanding of the population's reaction.

What was accomplished under these goals?

Summary:
In this first year, we explored understanding a large population's reaction to a nuclear WMD event in four major activities: (1) reviewing existing social theories of disasters, (2) collecting data and modeling the infrastructure of a mega-city and surrounding region, (3) generating synthetic population generation, and (4) developing an agent-based model of all the individuals in the region. We reviewed of social science theories and data on individual/group behavior during disasters. This work led to the publication of a case study and preparation of two review papers. Concerning the environment of a mega-city and surrounding area, we collected spatial, demographic, and workforce data for the selected mega-city of New York from several sources and devised methods and algorithms to make the data useful. The road data was preprocessed to achieve one connected network forming the transportation layer of the models. Using demographic data, we synthesized individuals, their households, and their social networks. Other datasets were utilized so that children attend nearby schools or daycare constrained with actual capacities and people are employed in workplaces located nearby matching workforce data. Finally, we began modeling individuals' movement in three counties, two rural counties and one in the heart of New York City.

Activities, Objectives, and Results:

In order to achieve our overall research objective, we needed databases of the social, spatial, and behavioral characteristics of our study area relevant to the context of a nuclear WMD event. These allowed us to create realistic human and physical terrain layers as the foundation of our computational model. More specifically in this first year, we explored the scientific frontiers in understanding the reaction of a mega-city's population to a nuclear WMD event in four major research areas: (1) **Literature Review**: reviewed existing social science theories and available data on reactions to a variety of disasters, (2) **Mega-city's Infrastructure**: collected data and modeling of a mega-city's environment and surrounding region, (3) **Mega-city Population**: collected data and synthesis of the population of the region, and (4) **Agent-based Modeling**: started development of agent-based models of the individuals in the region. We selected New York City (NYC) and surrounding area as the prototype area of interest with the intention of characterizing the reactions of any mega-city's population of approximately 25 million people to such an event. Our accomplishments and new findings are discussed in these four thrust areas.

1. Literature Review

We first reviewed social science theories and data, both historical and modern, on individual and group behavior during disasters. The definition of disaster we are using to ensure we have captured the theories and data relevant to our research is that of the United Nations Office for Disaster Risk Reduction (UNISDR). They define a disaster as a sudden extreme event that disrupts the functioning of a community or society, causing human, material, economic or environmental losses that exceed the ability of the affected community or society to cope. The type of disaster we are studying is a no-warning disaster, meaning we do not need to deal with preparations prior to the event, but will represent the preparations "as is."

In the social sciences, most theories are represented as verbal models or descriptions of how people, crowds, or societies behave. We have collected such theories, organized them, and will make them operational in computer models, somewhat comparable to mathematical formulations describing the physical sciences. The research thus far has led to the development of two manuscripts not yet published. The first manuscript reviews the status of relevant disaster research and organizes the literature and theories around a sequence of phases of reactions leading to responses to the event. The second manuscript reviews the social impact and the usefulness of social networks in supporting individuals' reactions and ability to cope with the disaster.

In the first manuscript, based on evolving understandings of the complex, interconnected characteristics of disasters and social adaptations over time, we organized disaster theories in the interdisciplinary context of complex adaptive systems (CAS). Within this context, disasters can be understood to consist of three sets of interacting systems: the socio-ecological system, the system of collective social behavior, and the individual actor's cognitive system. Ongoing dynamics within each set build to disruptive events that cause failures in the overall system. In this paper, we chronologically explore theories and frameworks used in disaster studies from the perspective of the three interacting systems, demonstrate how they have developed to explain disasters in modern times, and briefly discuss how Computational Social Science methodologies can support theory-building in disaster studies. By identifying current understandings and research questions co-located in CAS and by illustrating how computational methods based on qualitative work can provide deeper understandings, we introduce new lines of research targeting the linkages and interdependencies between human, social, natural and technological systems for the field of disaster studies.

In the second manuscript, we address the social impacts of a disaster. The social impacts of disasters are multifaceted -- psychosocial, sociodemographic, socioeconomic, and sociopolitical. Their economic impacts have been trending, and projections studies suggest that the risk will continue increasing. Yet, the current empirical base of disaster research on human behavior can be inconclusive and contradictory, partly due to the complexity and un-observability of disasters. Also, earlier research findings regarding social responses to disasters become outdated as the social, cultural, and political norms in societies change over time. We argue in this manuscript that the digital revolution, the open data trend, and the advancements in data science provide new opportunities for social scientific disaster research. To this end, we introduce the term *computational social science of disasters (CSSoD)* and discuss and showcase the opportunities and challenges in this new avenue of disaster research. We first briefly review the fields relating to CSSoD; i.e. traditional social scientific studies of disasters, computational social science, and crisis informatics. We then identify gaps in the literature which CSSoD could help fill, and suggest ways to advance our understanding of social and behavioral aspects of disasters. In doing so, our goal is to close the gap between data science and the sociology of disasters.

2. Mega-city Infrastructure

Concerning the environment of a mega-city and surrounding area, during this year we collected spatial, demographic, workforce, and education data for the selected mega-city of New York from various sources and devised methods and algorithms to clean up the data and make it useful for our purposes. We tested and refined our algorithms by focusing on two counties approximately 100km up the Hudson river from NYC (namely, the counties of Ulster and Sullivan) and a small area in Manhattan itself. The road data is preprocessed to achieve one connected network of roads to form the transportation layer of the forthcoming models. Using demographic data, we also created individuals and their households. Education data is utilized so that children attend nearby schools and daycare constrained with actual capacities of those institutions and people are employed in workplaces placed nearby matching available workforce and commuting data.

Road Networks
Since we are building a spatially explicit model at a detailed level, we needed to build our transportation layer using real world data. To do so, we gathered road network data from US Census Bureau's Geography Division for all roads in six states: New York, New Jersey, Connecticut, Pennsylvania, Rhode Island, and Massachusetts. We realized that the geographic data files provided are not in very good shape regarding topological connections. Specifically, roads are not touching each other, either falling short or long when we expect the endpoint of one road to be connected to another road. More interesting is the fact that two roads intersecting each other spatially (or having a common internal point) does not mean that they are topologically connected. We needed to get the roads topologically connected to be able to simulate the agents moving along them successfully. To clean the road data files and to create a network topology, we used the available GRASS (Geographic Resources Analysis Support System) C++ code libraries (also available in QGIS software). The GRASS code scales well because it processes data files incrementally instead of reading all in once into the memory. We first cleaned the data files in four steps:

- snapping lines to points with a specified threshold for closeness (v.clean.snap),
- breaking lines at each intersection (v.clean.break),
- removing duplicate geometry features (v.clean.rmdupl),
- removing small angles between lines at nodes (v.clean.rmsa).

Using this process, we were able to create a giant connected component of the road network (v.net.components) and saved it as a single data file. In
Figure 1 below is the giant connected road component for two counties. These methods will be applied across the entire study area in the next year.

Figure 1. Giant Connected Component Road Network for Ulster and Sullivan Counties (source: NWMD Project Team)

3. Mega-city Population

In parallel, we created a synthetic population at 1:1 scale, one software agent for each person in the area. We will later experiment and investigate the theories of social reactions to a nuclear WMD event in our agent-based model. In creating our population, we made use of the following data sources:

- Road network data (2010 TIGER/Line® Shapefiles: Roads) (discussed above)
- Demographic profile data (TIGER/Line®, 2010 Census, Census Tracts)
- Workforce data (Longitudinal Employer-Household Dynamics, Origin-Destination Employment Statistics)
- Education data (US Environmental Protection Agency, Office of Environmental Information and Oak Ridge National Lab, Education, US, 2015, ORNL Freedom, Shared Enterprise Geodata and Services (SEGS) dataset)

Demographic Profile

The demographic profile dataset we retrieved from the U.S. Census Bureau contains information about sex and age of the people living in a census tract, as well as the household's type. Individuals are identified either as male or female, and their ages are given in 5-year bands (the first band is under 5 years and the last band is 85 years and over). These individuals live in one of ten different household types; i.e., husband-and-wife families, male/female/nonfamily householders, households with a child less than 18, single householders over 65, or households in group quarters. Group quarters can be institutional, such as correctional facilities for adults, juvenile facilities, or nursing facilities/skilled-nursing facilities, or non-institutional, such as college/university student housing, or military quarters. We first generated synthetic individuals with a sex and age, and then populated households (strictly matching the number of each household type) with these individual agents.

Workforce

The U.S. Census Bureau also provides data on where people work. The Longitudinal Employer-Household Dynamics (LEHD) program within the Center for Economic Studies provides yearly Origin-Destination Employment Statistics (LODES) data at census-block level. (A census block is the smallest geographic unit the U.S. Census Bureau uses.) In particular, this dataset provides total number of people commuting between every pair of census blocks. We aggregated this information to the tract level. Then, using this information, we identified the number of employed individuals in a tract, and assigned them to workplaces in their own tracts or in other tracts (strictly following the origin-destination statistics).

Education

The daytime locations of children are accounted for using school and daycare data. We assigned children under 18 to a nearby school or daycare using the educational institution data retrieved from the US Environmental Protection Agency (EPA) Office of Environmental Information (OEI) (Education, US, 2015, ORNL Freedom, SEGS). The dataset contains geographic coordinates of educational institutions, enrollments, grade levels, and start and end grades of each institution.

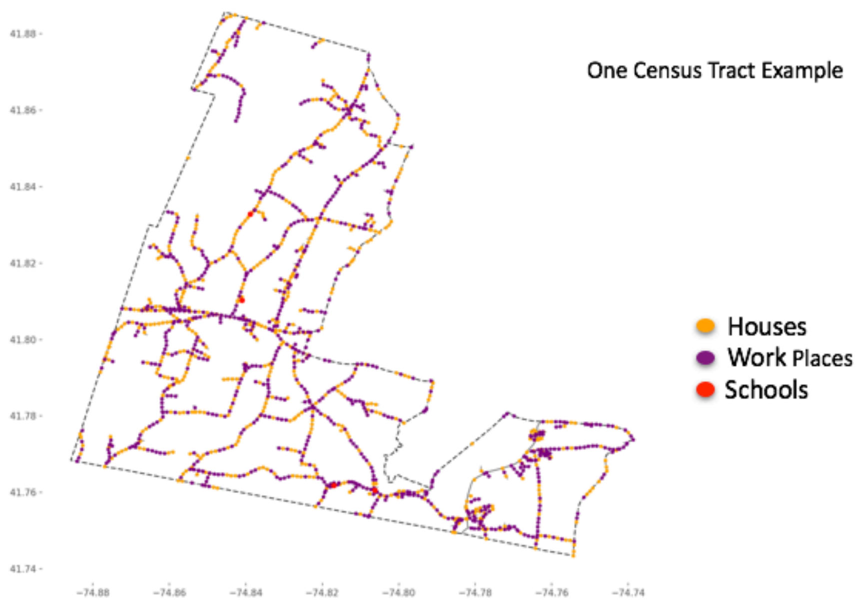

Figure 2. Example Placement of Houses, Workplaces, and Schools in One Track
(source: NWMD Project Team)

Social Networks

Disaster literature regarding human behavior emphasizes the importance of social networks for information dissemination and decision-making. One of the first reactions of individuals exposed to a disaster is to try to communicate with their family members and close friends. To model these relationships, we synthesized networks from household, work, and school location data. Individuals were connected by network ties based on groups living in the same household, working in the same workplace, or attending the same school. When a group size was less than seven, a complete network was generated for that group (i.e., a clique). If the number of house members, employees, or students in a group was seven or more, we generated a Newman–Watts–Strogatz small-world graph with each individual having on average four nearby links and a 10% chance of a longer link. An example network in Figure 3 shows the social ties within a census tract by the type of bond (house, work or school). For visualization purposes network links that cross census track boundaries are excluded.

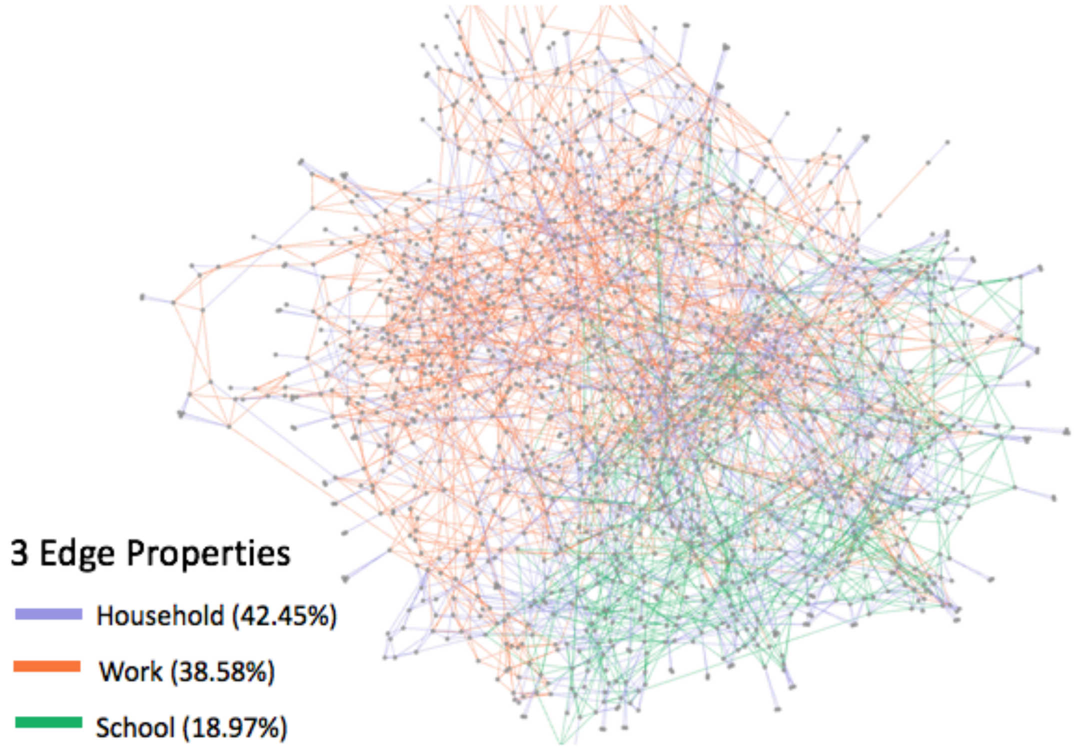

Figure 3. Social Network of a Census Tract (source: nWMD Project Team)

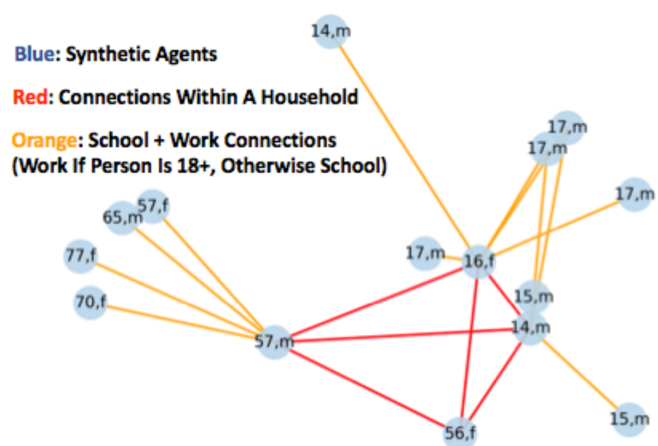

Figure 4. Synthesized Social Network Example (source: NWMD Project Team)

Validation
To validate our population synthesis process, we compared the synthesized population and the census demographic profile with metrics not used in the synthesize process. These are (1) average family size, (2) households with minors (under 18), and (3) households with seniors (above 65). Plots below show the errors in our two sample counties (Figures 5-6). Each point represents a census tract. Both axes are in percentage errors. This will allow us to improve the heuristics used in the synthesis and indicate how well our synthesized population matches the actual population.

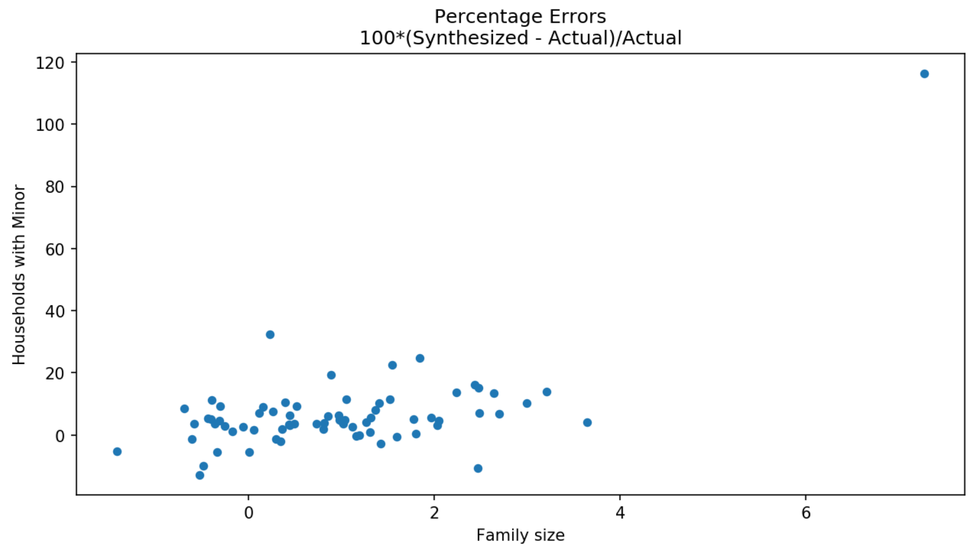

Figure 5. Percentage Error of Households with Minor (source: NWMD Project Team)

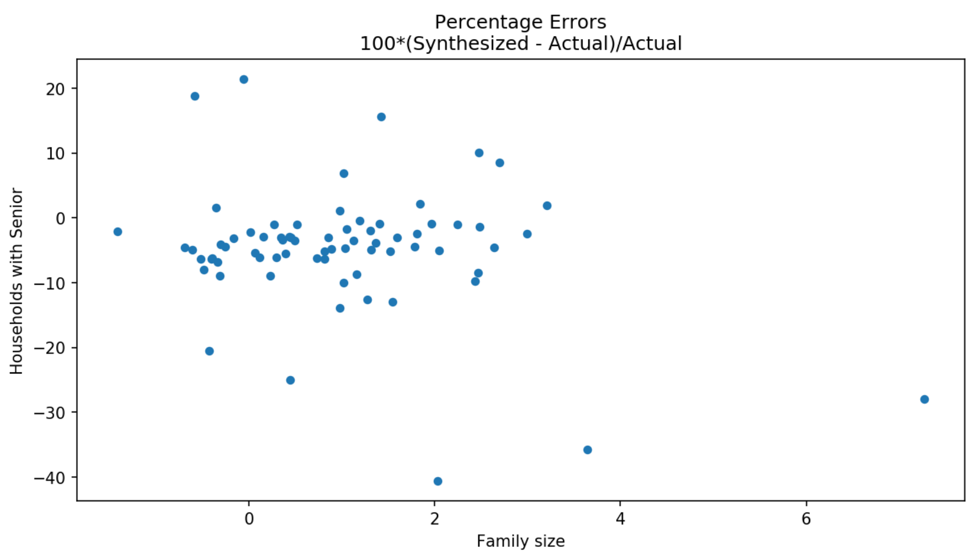

Figure 6. Percentage Error of Households with Senior (source: NWMD Project Team)

4. Agent-based Modeling

Finally, we started building models of agents in the mega-city area. We began modeling individuals in three counties: two rural counties (Sullivan and Ulster Counties, approximately 100km north of NY city) and part of one in the heart of the city (Manhattan County). These are being used as initial testbeds for processes we will scale up to the entire region. Two separate models over different areas have been developed during this period. The first will be the basis of the model of the whole region and the second will be of a small area around the nuclear WMD. The models use the infrastructure data and synthesized population data previously discussed. The modeling included assembling households from the synthesized population based on family makeup heuristics that replicate the census track's statistics. We also created and placed homes, workplaces, and schools in each census track on the road network using similar guidelines. Analysis of differences between our generated environment and the census track's data guided minimizing such differences. We then built the first models using the assembled and processed data with agents moving on the road networks as shown in Figures 5 through 9.

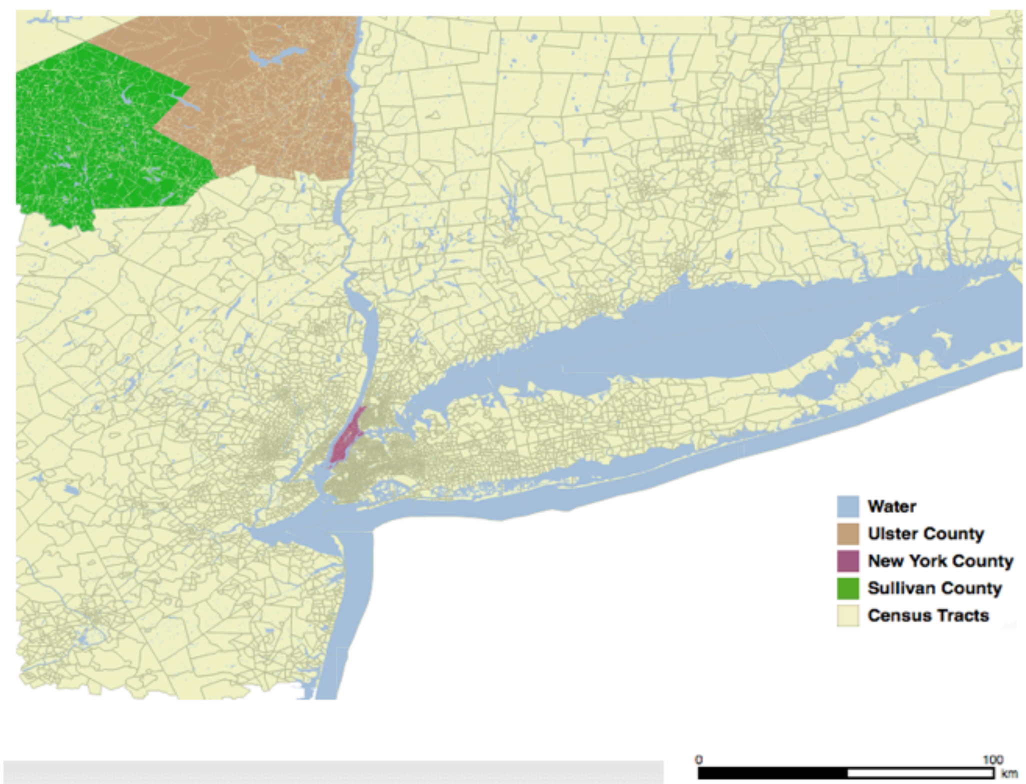

Figure 5. Mega-City Region and Counties Modeled in the First Year (source: NWMD Project Team)

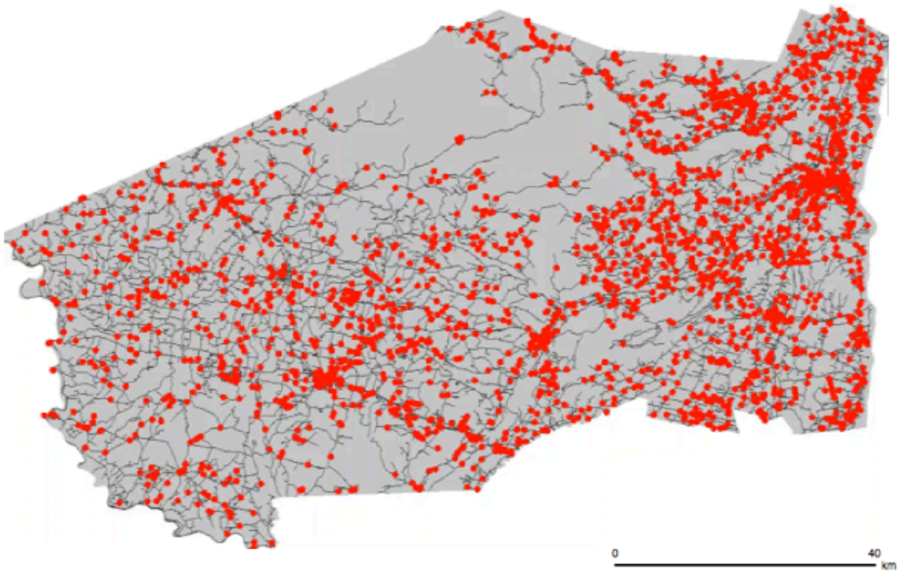

Figure 6. Agents in Two Counties Moving on their Road Network (source: NWMD Project Team)

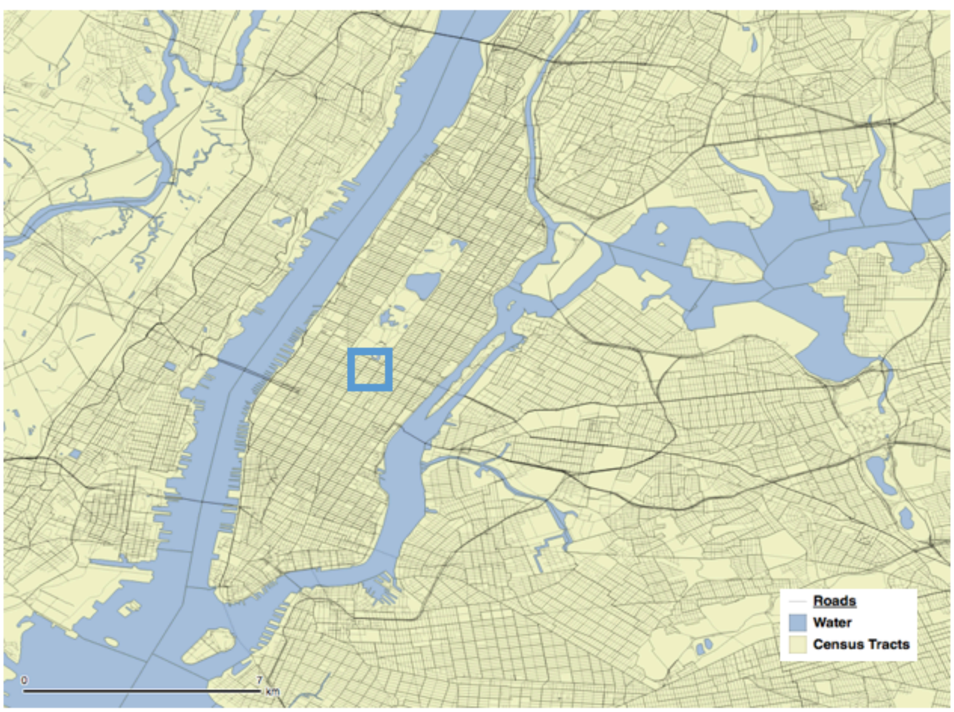

Figure 7. Small Scale Region within Manhattan (source: NWMD Project Team)

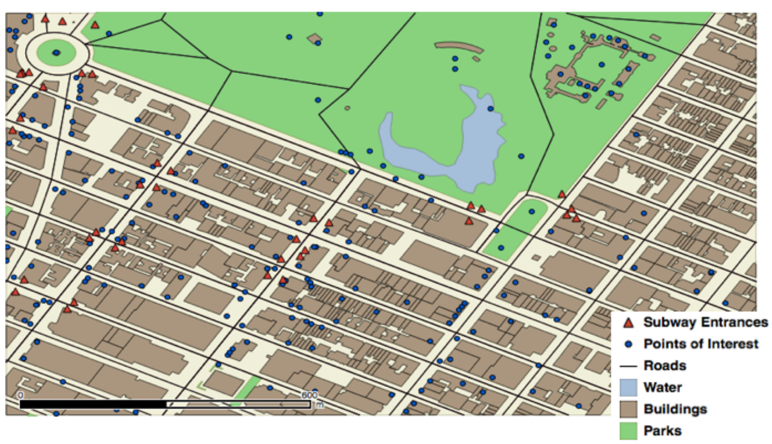

Figure 8. Map of Small Scale Region of New York City (source: NWMD Project Team)

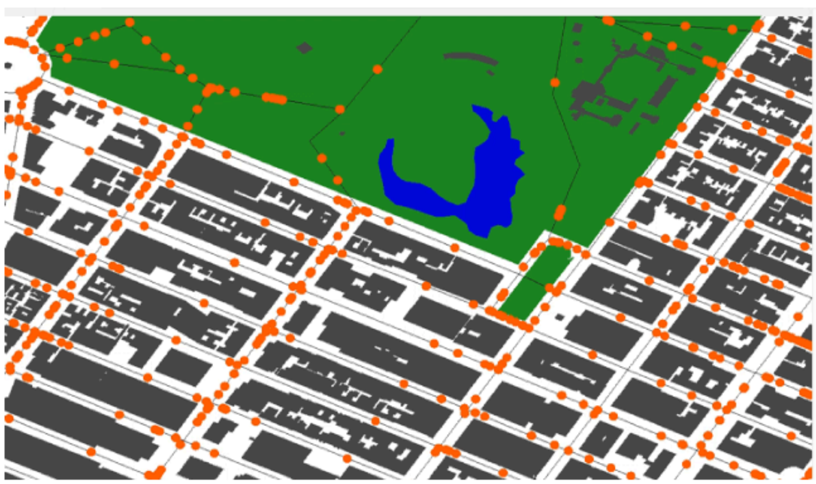

Figure 9. Model and Agents moving in Small Scale Region of New York City (source: NWMD Project Team)

What opportunities for training and professional development has the project provided?

Training and professional development included one-on-one mentoring on conducting research on this grant and both graduate student's dissertations, review of manuscripts for publication, participation in conferences, and participation on dissertation committees.

Annetta Burger (graduate student):

- was advanced to doctoral candidacy on 6 January 2017 based on a dissertation proposal titled "From Networks to Recovery: Effects of social networks on community recovery in the face of flooding disasters."

- participated in NIST's Global Cities Team Challenge Public Safety Super Cluster Workshop in Washington, DC, in March 2017.

Talha Öz, (graduate student):

- was advanced to doctoral candidacy on 9 January 2017 based on a dissertation proposal titled "Collective Stress in the Digital Age."

- presented an analysis of public sentiment and blame in response to the Flint, Michigan, water crisis at an international workshop on disaster management in Indianapolis in October 2016.

How have the results been disseminated to communities of interest?

In this first year, the PI and Co-PI made presentations to two Department of Defense organizations and a variety of academic communities in both local seminars and international conferences and workshops:

William G. Kennedy (PI) briefed the cognitive modeling considerations of the grant to the Post-Graduate Summer School of ACT-R (a cognitive modeling architecture workshop) in Lancaster, PA, August 2016.

William G. Kennedy (PI) presented the status and plans for the research project to the Computational Social Science Program's Friday Seminar at George Mason University in September 2016.

William G., Kennedy (PI) presented the status and plans for the research project to FY17 Basic Research Office (BRO) Forum at the Air Force Office of Scientific Research (AFOSR) spaces in Arlington in October 2016.

William G. Kennedy (PI) presented the status and plans for the research project to the Krasnow Institute for Advanced Study at their Monday Seminar at George Mason University in March 2017.

William G. Kennedy (PI) was interviewed by the press on the research project and the interview published by the Atlantic magazine in March 2017. The story was picked up by other news outlets in the same month.

Andrew Crooks (Co-PI) presented the status and plans for the research project to the Association of American Geographers at their annual meeting in Boston in April 2017.

William G. Kennedy (PI) and Andrew Crooks (Co-PI) briefed the Army Science Board in Crystal City, VA in April 2017.

What do you plan to do during the next reporting period to accomplish the goals?

The objectives for the second year of this grant are to:

1. **Scale a model up to entire region and the other down to pedestrians.**
    - Build road network for whole area.
    - Synthesize population of whole area including social networks.
    - Place homes, schools, and workplaces in whole area.
    - Submit conference paper, journal article, tutorial, and code on population synthesis.

2. **Instantiate normal and event reaction behaviors in both models.**
    - Craft behavioral rules for normal patterns of life and reactions to no-warning disaster.
    - Build cognitive architecture to execute behavioral rules for both models.
    - Submit paper on resulting ABM cognitive architecture of disaster behavior.

3. **Integrate nuclear WMD effects and both models.**
    - Develop zones of weapon effects affecting human behaviors.
    - Implement weapon effects in both models.

4. **Make first run of both models for the first minutes after the event.**