Civil Engineering and Environmental Systems

ISSN: 1028-6608 (Print) 1029-0249 (Online) Journal homepage: https://www.tandfonline.com/loi/gcee20

Using opportunities in big data analytics to more accurately predict societal consequences of natural disasters

Jessica Boakye, Paolo Gardoni & Colleen Murphy

To cite this article: Jessica Boakye, Paolo Gardoni & Colleen Murphy (2019) Using opportunities in big data analytics to more accurately predict societal consequences of natural disasters, Civil Engineering and Environmental Systems, 36:1, 100-114, DOI: 10.1080/10286608.2019.1615480

To link to this article: https://doi.org/10.1080/10286608.2019.1615480

CIVIL ENGINEERING AND ENVIRONMENTAL SYSTEMS
2019, VOL. 36, NO. 1, 100–114
https://doi.org/10.1080/10286608.2019.1615480

Using opportunities in big data analytics to more accurately predict societal consequences of natural disasters

Jessica Boakye[a], Paolo Gardoni[a] and Colleen Murphy[b]

[a]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA; [b]College of Law, University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT
The availability of data sources has greatly increased due to advances in technology and data sharing. With these new data sources and significantly larger volume of data, engineers have been presented with a unique opportunity to create more realistic and informative models that can be used in real world applications. This paper presents a probabilistic framework for using big data to assess and predict the well-being of individuals before and in the aftermath of a hazard. Data are used to inform a Capability Approach (CA) where capabilities are defined as important dimensions of well-being reflecting what individuals have a genuine opportunity to do or become. The paper also addresses three of the grand challenges presented by big data: privacy, source validity, and accuracy. As an example, the probabilistic framework is used to study the ability of households in a coastal community to be sheltered in the aftermath of a hypothetical earthquake.

ARTICLE HISTORY
Received 24 April 2019
Accepted 27 April 2019

KEYWORDS
Hazard management; big data analytics; spatial capability approach

1. Introduction

Communities can experience devastating impacts from natural hazards. In addition, often these impacts are not evenly distributed over the hazard area due to spatial variability in the intensity measures of the hazard, as well as spatial differences in the vulnerabilities of the physical systems and of individuals from different social groups. Therefore, predictive models should be able to capture not only the average impact on a region, but also the spatial variability in the impact (Gardoni and Murphy 2018). To model the spatial variability in the intensity measures, we can use different models including sophisticated 3-dimensional physics-based hazard models that capture the variability in seismic intensity measures (Guidotti, Tian, and Gardoni 2019). The spatial variability in the physical systems comes from the variability in the vulnerabilities of the components of the physical systems, which could be also affected by differential aging and deterioration (Jia and Gardoni 2018a, 2018b; Jia, Tabandeh, and Gardoni 2017; Kumar and Gardoni 2011, 2013, 2014; Kumar, Gardoni, and Sanchez-Silva 2009, 2015). We can use fragility and repair-rate curves to model the time-dependent vulnerability of each component of the physical systems (Gardoni, Der Kiureghian, and Mosalam 2002; Gardoni, Mosalam, and

CONTACT Jessica Boakye boakye3@illinois.edu

Der Kiureghian 2003; Iannacone and Gardoni 2018). The accuracy of these fragilities and repair rate curves depend on detailed, real time inventories of infrastructure which can be difficult to obtain. Extensive work has been done to create detailed inventories (Cleveland, Elnashai, and Pineda 2007), however, these inventories are often static and have to be manually updated. For the social environment, it is difficult to capture the spatial variability due to issues with data availability. To accurately model individuals' vulnerabilities, information on socio-economic characteristics of the individuals is needed. However, such information is traditionally not available at the household level but at a higher level of aggregation.

Current models are often reliant on assumptions or simplifications that are necessary because of limitations in available data. Over the past few years, the availability of data sources has greatly increased due to advances in technology and data sharing (Wang and Ye 2018). With these new data sources and significantly larger volume of data, engineers have been presented with a unique opportunity to create more realistic and informative models that can be used in real world applications. This is exemplified by the growth of data science companies and research institutes who collect large data and study relevant applications. For example, the European research institute, SoBigData, is used by companies, policy makers, and researchers alike to develop novel research related to big data analytics (Grossi et al. 2018). Moreover, the United States government started Data.gov in 2009 which is a website containing datasets related to multiple applications including healthcare, education, and transportation (Kim, Trimi, and Chung 2014).

This paper presents a probabilistic framework for using big data and big data analytics to assess and predict the well-being of individuals in the aftermath of natural disasters. The term 'big data' usually refers to a large volume of data that is often hard to store, difficult to visualise, and is highly variable in format and type. Although storage is not often a significant problem in civil engineering applications, difficulties in visualisation and variability present significant challenges. Big data analytics is the process of examining large and varied data sets. Big data analytics is especially salient in disaster mitigation, and risk and resilience analysis where insufficient or missing data has traditionally forced researchers to develop simple models often applied to overly simplistic examples. In addition, decision makers need to predict or assess on an ongoing basis the well-being of affected individuals in the aftermath of a natural disaster to decide where to allocate resources for mitigation or recovery. The presented framework proposes to use data to inform a Capability Approach (CA) where capabilities are defined as important dimensions of well-being reflecting what individuals have a genuine opportunity to do or become. The CA uses indicators to quantify the capabilities.

The paper starts with a review of the CA for risk assessment (including the development of probabilistic predictive models) and risk evaluation. Then, the paper provides a literature review of big data usage in the context of hazard management and discusses the usage of big data analytics within a CA along with the challenges of privacy, accuracy, and validity. Finally, the paper illustrates the framework considering the opportunity of households to have a shelter in the aftermath of a hypothetical earthquake.

2. Capability approach for risk assessment

Risk analysis (or assessment) involves estimating the consequences and associated probabilities of a hazardous scenario. The societal consequences can be conceptualised and quantified using a CA. First developed by Amartya Sen (1989, 1992, 1993, 1999a, 1999b) and Martha Nussbaum (2000a, 2000b, 2001) in the context of development economics and policy, the CA offers a conception of some of the constitutive components of individual well-being. According to the CA, the well-being of individuals should be measured and evaluated based on the opportunities people have to live valuable lives (Robeyns 2006). This differs from a utilitarian approach where emphasis is put on individual preferences. There are two key definitions in a CA: *functionings* and *capabilities*. Functionings refer to what an individual does or becomes in his or her life that is of value, such as being educated. In the CA, functionings are dimensions of well-being. Capabilities refer to the functionings that are feasible for an individual to choose to achieve given what he/she has and what he/she can do with that given, for example, the built infrastructure, legal norms, and economic institutions (Sen 1993). Whether education is feasible for an individual can depend on whether an individual has income to pay for books or fees, whether a school is located nearby, and whether social norms equally encourage the education of girls and boys.

Murphy and Gardoni (2006, 2007, 2008, 2010, 2011, 2012a, 2012b) and Gardoni and Murphy (2008, 2009, 2010, 2013, 2014) proposed a CA to quantify the societal impact of hazards on individual well-being. This impact is defined as the effect the hazard has on the functionings that individuals have an opportunity to achieve (i.e. capabilities.) For example, the impact of an earthquake can be measured in terms of the changes in opportunities individuals have, such as the opportunity to be sheltered or the opportunity to be educated (Boakye et al. 2019). Although the capability approach is traditionally limited to a discussion on individuals, the importance of group dynamics on the well-being of individuals is recognised. Sen (1999b) acknowledges that groups may be instrumentally important for enhancing individual capabilities and Stewart (2005) argues that the capability approach should be extended to explicitly account for group dynamics by considering group capabilities as well as individual ones. Within our framework, this can easily be done by adding an additional group capability or a capability which can measure an individuals' relationship to a prominent community group. The steps needed to implement a CA for risk assessment are: (1) selection of capabilities, (2) selection of indicators, (3) development of probabilistic predictive models, and (4) development of an aggregate measure. This section discusses these steps in detail.

Gardoni and Murphy (2009, 2010) proposed three criteria for selecting capabilities in the context of hazard risk analysis. First, capabilities should be potentially affected by the hazard. This affected capability is identified by either a theoretical justification or empirical evidence. Second, capabilities should be parsimonious to limit issues with data collection (or storage). Third, capabilities should be orthogonal meaning that each capability should provide information that cannot be ascertained from another capability. This avoids giving disproportionate weight to one capability over another because of their orthogonality. Capabilities can be assumed to be incommensurable and to be all necessary to the well-being of an individual (Gardoni and Murphy 2009, 2010; Sen 1993).

Since capabilities cannot be directly measured, indicators are chosen as proxies (Raworth and Stewart 2003). Indicators can be real-valued or categorical. Real-valued indicators take numerical values whereas categorical indicators take qualitative values. Indicators need to be estimated before and after a natural hazard. These indicators can serve as quantification metrics for the different capabilities that can be measured through the disaster impact and recovery (Boakye, Murphy, and Gardoni 2018). Table 1 gives some examples of capabilities and possible corresponding indicators. In practice, indicators have to be chosen based on data availability in the region of interest as detailed in Boakye, Murphy, and Gardoni 2018.

Predictive models are then needed for the indicators to estimate the values of the indicators as a function of influencing factors (regressors.) These regressors can describe, for example, the state of the functionality of structures and infrastructure and/or various socio-economic factors identified from qualitative studies on social vulnerability. Tabandeh et al. (2019) proposed time-varying probabilistic predictive models for several indicators.

To define the overall state of well-being, the indicators need to be combined to create an aggregate measure of achievement. Because the capabilities are incommensurable, the well-being of each individual can be seen as a series system where each capability (and the corresponding indicator) is a component of the system (Tabandeh, Gardoni, and Murphy 2017). Following a system reliability approach (Gardoni 2017), the system fails if any component fails to reach a desired level as shown in Figure 1. Green, orange, and red represent acceptable, tolerable, and unacceptable levels respectively. The small arrows yellow and red arrows in the figure denote acceptability and tolerability limits. The definition of the desired levels (or limits) is discussed further in the next section.

3. Capability approach for risk evaluation

After the risk has been assessed, we can use a CA to evaluate if the risk is acceptable or not. The risk can be assessed for both a single indicator and the risk to well-being (assessed using system reliability displayed as displayed in Figure 1). Murphy and Gardoni (2008) defined an acceptable threshold for a capability based on the minimum level of capabilities a community should allow in principle (Nussbaum 2000b). In the case of a special condition like a natural hazard, a lower level of attainment could be allowed as long as it is temporary, reversible, and does not fall below an ever lower tolerability threshold (Murphy and Gardoni 2008). Figure 2 (adapted from Gardoni and Murphy 2018) shows the acceptability and tolerability limits for a given indicator. This process is repeated for each indicator and then system reliability can be used to assess the overall risk to well-being as discussed in Section 2.

Table 1. Example capabilities and indicators.

Capability (Opportunity to …)	Indicator	Indicator type
Maintaining health	Access to a hospital	Categorical
Being sheltered	Access to a permanent residence	Categorical
Being mobile	Travel time	Real-valued

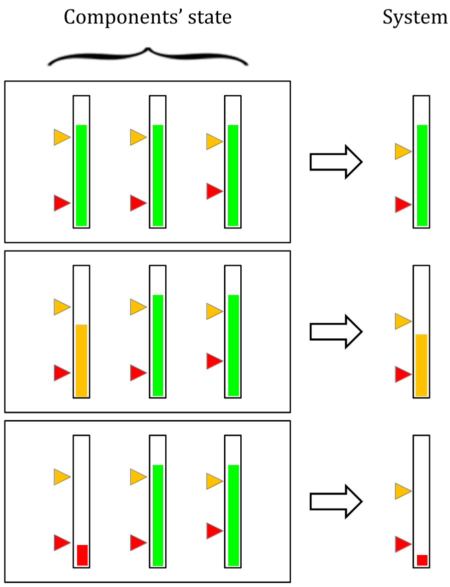

Figure 1. System reliability approach.

To mathematically model the recovery of well-being over the occurrence of a disruptive event, Tabandeh et al. (2019) integrated the probabilistic models for the indicators into a Dynamic Bayesian Network (DBN). The statistical inference problem is to estimate the probability that the state of well-being at any time during the recovery is either acceptable, tolerable, or intolerable. The graphical structure of the DBN visualises the role of each regressor and indicator in the overall state of well-being.

Resilience and sustainability goals can be defined to consider recovery time, environmental justice, and social justice (Gardoni and Murphy 2018). In terms of capabilities, resilience and sustainability goals are concerned with how quickly the capabilities return to their pre-hazard conditions (or better) and the inequalities in the distributions of capabilities across space and time (Boakye, Murphy, and Gardoni 2018).

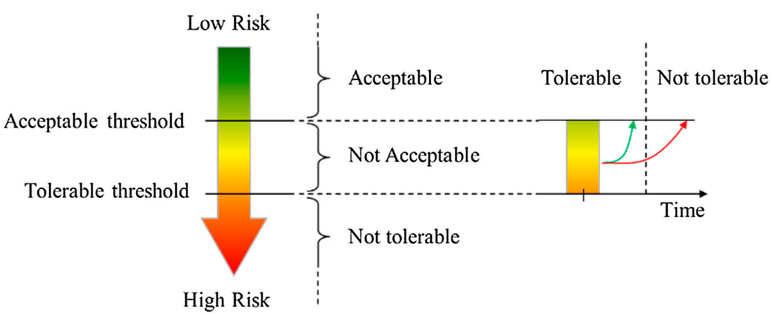

Figure 2. Acceptability and tolerability limits.

4. Big data in hazard management

Big data sources can be grouped into two large categories: free and for purchase. In general, the data can be in many forms (numerical, text, etc.) and require algorithms for post processing. Examples of free sources that can be used in big data analytics are government sponsored data (like the United States Census) and social media data (such as Twitter or Facebook.) Although many free big data sources exist, they are often incomplete or in an aggregate form. Using incomplete data as an input into models can result in biased outputs and should be avoided whenever possible. Aggregate data, although informative, is often at an aggregation level that is too large to be useful. For example, a community with over 6,000 people may only have 6 aggregation levels for socio-economic data such as median income (explained further in Section 7.) For these reasons, researchers are often dependent on for purchase data. An example of for purchase data are call detail records (CDRs), which require an agreement with a telephone company (Arslan et al. 2017).

Since social media is one of the largest sources of big data, researchers have recognised the importance of social media within disaster management. Many governments have active Twitter pages that provide real time information for the public during a disaster period. Cenni et al. (2017) have created a multi-user tool to analyze Twitter data for early warning systems, sentiment analysis, and connectivity analysis. Wang and Ye (2018) have identified four dimensions in social media data: space, time, context, and network. Spatial information is useful to study the spatial distribution of risk. Gupta et al. (2013) use geo-referenced tweets to visualise tweets about Hurricane Sandy on a world map. Time information is especially salient in disaster mitigation. Since all social media posts come with a time stamp, many studies have focused on identifying time patterns of posts by governments and emergency organisations (Sakaki, Okazaki, and Matsuo 2010). Content information can be used to characterise public sentiment or response to disaster. Qu et al. (2011) use content information from a Chinese social media platform to see if people are asking for situational updates, expressing opinions, or asking for help/support. Finally, network information can be used to identify behaviours of various agents in disaster situations. Researchers use a social media analysis to detect network patterns (Starbird and Palen 2010) and identify the main information sharers (Kogan, Palen, and Anderson 2015).

Once big data sources have been identified, methods from machine learning can be used to recognise patterns and/or combine heterogeneous data sources to enhance existing models. Machine learning is a branch of artificial intelligence that focuses on algorithms for prediction and classification. These algorithms are either supervised (when the response is known) or unsupervised (when the response is unknown.) Least squares and nearest neighbour analyses are commonly used supervised methods while cluster analyses are widely used for unsupervised methods (Trevor, Robert, and Friedman 2009). Both supervised and unsupervised methods have been used to inform engineering models. For example, Asencio–Cortés et al. 2018 compared different learning algorithms to see which one performed best for the prediction of earthquake magnitude in California using a 1 GB catalog of ground motions.

5. Big data analytics within a capability approach

In the previous sections, we detailed a CA and how it can be used for risk analysis. In this section, we discuss how big data analytics can be used in a capability approach. In general, data analytics can be used to inform indicators or be used to define high resolution regressors for indicators.

Data analytics can be used to inform indicators. Often the data that can be found from analytics (Twitter, cell phone, etc.) do not account for the entire population. However, the additional information can be integrated into engineering models to generate more accurate predictions. For example, consider mobility. The mobility of a community is a complex but important variable for well-being (Boakye et al. 2019). Understanding individual mobility is complicated and in the past researchers have used statistical models (e.g. random walk and diffusion) to approximate human mobility. With growth in technology, cell phone records have emerged as a leading tool to measure human mobility. Detailed records can track the movements of individuals and can be used to check the accuracy of statistical models. Gonzalez, Hidalgo, and Barabasi (2008) studied the trajectory of 100,000 anonymized mobile phone users whose position is tracked for a six-month period. Their findings were in contrast to statistical models.

Following disasters, the mobility of people becomes even more salient and issues of mass displacement are of interest. Martin and Singh (2018) used over 700 million publically available media articles, in-person interviews, and Twitter data to analyze the patterns and reasons for forced migration and mass displacement following disasters. With this vast amount of data, they were able to create early warning and simulation tools for decision makers. The goal of the early warning tool is to inform decision makers that mass displacement or forced migration may happen soon while the simulation tool is a prediction model that decision makers can use for mitigation purposes.

Data analytics can be used to create high resolution regressors for indicators. As noted previously, Tabandeh et al. (2019) developed models that can couple social vulnerability factors with high resolution, detailed engineering models. However, these models require accurate information on the built and social environments (Boakye, Murphy, and Gardoni 2018; Sharma, Tabandeh, and Gardoni 2018). To accurately measure the damage to the built environment, we depend on complex engineering models which require detailed information about the structures. For example, Gardoni, Der Kiureghian, and Mosalam (2002), Gardoni, Mosalam, and Der Kiureghian (2003) developed physics-based models for infrastructure damage. The inputs to these models require information on the material and geometry of each component that can be difficult to obtain from publically available data. Further complications occur because the data needed to complete the analysis usually comes from multiple sources. One possible solution is to combine data mining techniques (such as convolutional neural networks) with image processing to create real time inventories of the built environment. Of course, the addition of data mining could add additional uncertainties or error that need to be accounted for. This is further discussed in Section 6. In all applications, it is imperative to clearly define which engineering models are used and to propagate the uncertainties throughout the models.

In addition to requiring real-time inventories, accurate modelling of indicators requires high resolution information on socio-economic regressors. Knowing age, race, gender, and other vulnerability factors at each household could allow for household-level predictions of consequences such as loss of shelter or power. These household-level predictions of consequences could help decision makers in the planning and mitigation of hazards. Unfortunately, publically available data are historically found in aggregated forms making it difficult to obtain household-level data. Data analytics can be used to create high resolution prediction models socio-economic regressors using data from a multitude of resources including but not limited to social media and CDRs. These prediction models for socio-economic regressors can then be used as input in models for indicators which can account for spatial variability within the socio-economic regressors.

6. Grand challenges: privacy, source validity, and accuracy

Although big data analytics presents opportunities for more realistic models, there are also three grand challenge areas that big data presents: privacy, source validity, and accuracy. The goal of the prediction models is to provide decision makers with information they can use to protect and serve the public. In order to do that, we argue that methods must protect the public's privacy and be as accurate as possible. Although challenging, this section discusses our recommendations in these areas.

Many argue that privacy is a right or something that should be preserved, however, the accumulation of personal data has an incremental adverse effect on privacy. As popularity in data sharing has increased, more personal information has been revealed (Tene and Polonetsky 2012). Although sensitive data are traditionally aggregated to try to preserve privacy, there is a trade-off that should be considered. As noted in Section 5, household-level socio-economic regressors are needed to account for spatial variability. As a result, the creators of these models should try to balance privacy concerns with the need for accurate models especially if sensitive data is used to relevant models. To accurately model and prevent concerns with social justice (Gardoni and Murphy 2018), it is important to have household-level predictions of socio-economic regressors. Moreover, it is important to note that many machine learning techniques would produce *predictions* of socio-economic information. Therefore, it is possible to create high resolution regressors without directly dealing with or releasing privileged or private information. It is important that this distinction is clearly communicated with the public and the users of the models so that it is clear what is predicted.

The growth in data sharing has led to numerous data sources that contain false information. The information provided by these sources may be especially salient for researchers and decision makers to include in their model formulations. If no effort is made to examine source validity, prediction models may produce results that have large errors. This is especially dangerous in the field of risk management when the goal is to provide prediction models that can be used to mitigate societal consequences. Therefore, data sources must be screened for validity before they are included in any prediction models. Moreover, the creators of the models should be transparent and report which data sources are being used to allow for scrutiny and/or consensus from colleagues.

As shown in Section 4, social media data has been identified as a useful source in many studies on hazard management. However, the inclusion of social media data can create errors since users may post invalid things or false statements threatening the accuracy of the prediction models. Unlike traditional data sources that can come from governmental organisations, social media websites provide little to no screening on false information making it difficult to capture high-quality content (Agichtein et al. 2008). Additionally, the trend of using social media data only should cause concern. Using social media data alone would introduce inherent bias against people who do not have access or use social media. These groups (e.g. elderly or young children) are especially vulnerable to disproportionate disaster impacts so it is important to ensure that models using social media data take this into account.

Further, the combination of sources that is used in many data mining techniques often leads to errors related to noise or disagreement that need to be measured and accounted for the in the final prediction model. These errors may ultimately affect the model accuracy. As researchers, one of our major priorities must be to ensure the accuracy of our predictive models. If the error from the data source is known explicitly we can mathematically treat it like measurement error as done in Gardoni, Der Kiureghian, and Mosalam (2002). Here the addition of measurement error increases the model variance (and the overall model uncertainty.) If the additional uncertainty coming from the data source is too large (and increases the model uncertainty too much), it may be necessary to remove that data source to create a useful model.

7. Example

In this example, we study the effect of a hypothetical earthquake on the capability of being sheltered. The study area is the coastal community of Seaside, Oregon with a population that fluctuates between 6,000 and 14,000 residents. Based on counts from the 2010 Decennial Census, 6,440 inhabitants are assigned to various buildings throughout the

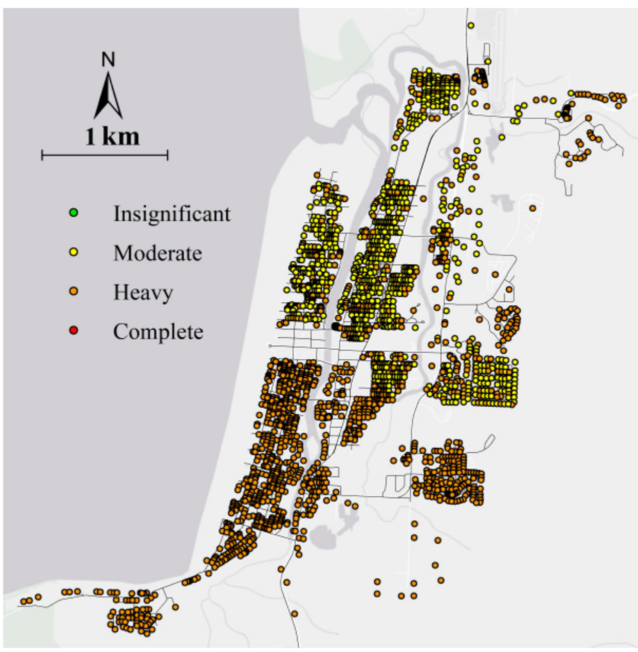

Figure 3. Mean damage to buildings.

city (Guidotti, Gardoni, and Rosenheim 2019). The earthquake has magnitude $M_w = 7.0$ and an epicentre located 25 km southwest of the city. Ground Motion Prediction Equations (Boore and Atkinson 2008) are used to generate maps of the ground motion intensity measures over the relevant study region.

To estimate the structural damage to the buildings of Seaside, fragility curves (as defined in Gardoni, Der Kiureghian, and Mosalam 2002) are used for a given earthquake intensity measure at the site to find the conditional probability that each building is in one of four different damage states (FEMA 2015). The mean damage is then calculated following Bai, Hueste, and Gardoni (2009). Figure 3 shows the mean damage for each residential building in Seaside. The definitions of insignificant, moderate, heavy and complete used in Figure 3 follow Bai, Hueste, and Gardoni (2009).

After the earthquake occurs, people may choose to dislocate from their permanent residence due to structural damage. For the same level of damage, the likelihood of dislocation is affected by other socio-economic factors (Gladwin 1997; Lin 2009). People dislocation is estimated using a logistic model with regressors based on the expected structural damage and race (Guidotti, Gardoni, and Rosenheim 2019). Structural damage and people dislocation are then used to define the indicator for the capability of being sheltered (having access to a permanent residence.) Lower values of the indicator indicate that the residents have a higher shelter need because they may not have the ability to stay at his/her permanent residence due to high structural damage. An even lower value of the indicator is given to the residents who have the highest shelter need because of high structural damage and an inability to dislocate to a temporary residence (estimated by the dislocation model.) The locations of the lowest indicator values can be used to

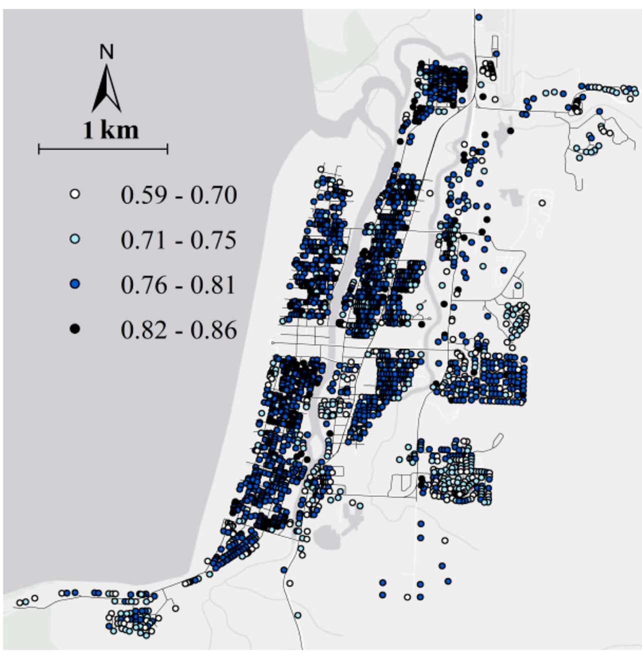

Figure 4. Shelter indicator.

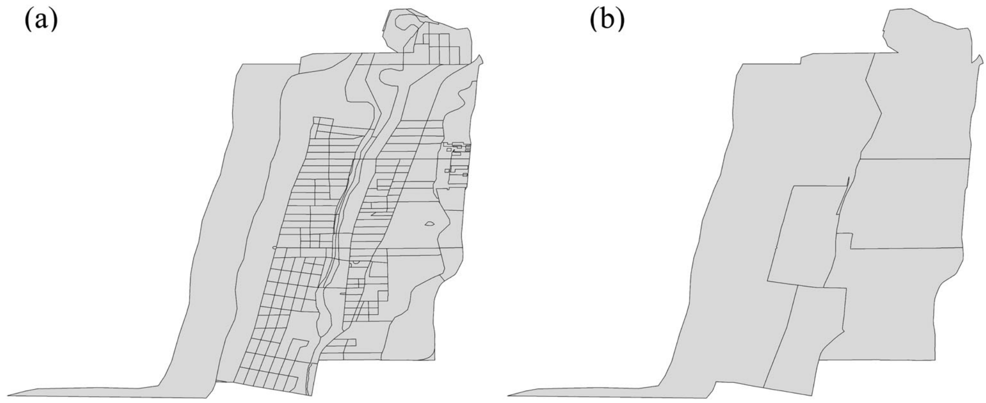

Figure 5. Granularity of (a) race and (b) income for Seaside, Oregon.

optimise resources for temporary shelters. Figure 4 shows the values of the indicator for shelter.

Since we can predict structural damage at the household level, we would like the socio-economic regressors (i.e. race) to also be at that level of granularity. However, data related to race are available only at the census block level. Census blocks are the smallest statistically defined geographical areas defined by the Census Bureau for tabulation in the 100-percent data (data is collected at every household as opposed to a sample of households in the study area.) The data available at the census block level are count data such as age and race. The census block is smallest aggregation provided by the Census Bureau and the size varies by population density. For data related to other socio-economic factors such as income, the aggregation available is at the census block group level. Census block groups are statistically defined geographical areas that contain 600–3,000 people (1500 is optimal.) Figure 5(a and b) show the levels of aggregations for race and income respectively for Seaside. These figures show that the census provides 217 and 6 different values for race and income respectively for the city of Seaside. Given that our model for building damage includes estimates for each household in Seaside, we would also want household-level predictions of the socio-economic conditions to use in a model for people dislocation.

Data analytics can assist in improving the granularity of the socio-economic regressors. Data mining and machine learning techniques can be used to create a model that predicts income (for example) at the household level. The inputs of this model would come from multiple data sources and be at different granularities. To check the accuracy of the prediction, one can aggregate the prediction to the census block group level and compare it to the available median household income estimate provided by the census and/or check it against some available household level incomes (e.g. income of public employees.) This household-level prediction would allow us to better understand and predict the spatial variability in the probability of dislocating due to salient socio-economic conditions.

8. Conclusions

In this paper, a CA is used to quantify the impact of a natural disaster incorporating the spatial variability in the intensity measures of the hazard, in the vulnerabilities of the physical systems, and the socio-economic conditions of individuals. There are models available to accurately capture the spatial variability in the intensity measures and the physical systems. Capturing the spatial variability in the socio-economic conditions is challenging due to limitations in data availability. Data analytics are proposed as a tool to estimate the social regressors at the desired granularity and capture the spatial variability in socio-economic conditions. We also discuss opportunities, challenges, and recommendations for incorporating large data methods into disaster studies. To illustrate the described framework, the effect of a hypothetical earthquake on the need for shelter is examined. This example is for a single consequence and an earthquake of given magnitude and location. The process could be repeated for multiple earthquakes of different magnitudes and/or locations. To fully examine the risk of an earthquake (which can have varying degrees of magnitude and location), there are two possible approaches. The first is to do a fully coupled analysis where we use the probability distribution functions of both the location and magnitude. Then, the total probability rule can be used to find the mean impact by integrating over all of the possible values. Second, we could examine the worst case scenario and use that for mitigation and policy decisions.

Acknowledgements

This work was supported in part by the Graduate College at the University of Illinois at Urbana Champaign, the Civil and Environmental Engineering Department at the University of Illinois at Urbana Champaign and the SURGE program at the University of Illinois at Urbana Champaign. Opinions and findings presented are those of the writers and do not necessarily reflect the views of the sponsor.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by the Graduate College at the University of Illinois at Urbana Champaign, the Civil and Environmental Engineering Department at the University of Illinois at Urbana Champaign and the SURGE program at the University of Illinois at Urbana Champaign.

References

Agichtein, E., C. Castillo, D. Donato, A. Gionis, and G. Mishne. 2008, February. “Finding High-quality Content in Social Media.” Proceedings of the 2008 international conference on web search and data mining, ACM, 183–194.

Arslan, M., A. M. Roxin, C. Cruz, and D. Ginhac. 2017, December. “A Review on Applications of Big Data for Disaster Management.” Signal-image technology & Internet-based systems (SITIS), 2017 13th International Conference on, IEEE, 370–375.

Asencio–Cortés, G., A. Morales–Esteban, X. Shang, and F. Martínez–Álvarez. 2018. "Earthquake Prediction in California Using Regression Algorithms and Cloud-based Big Data Infrastructure." *Computers & Geosciences* 115: 198–210.

Bai, J.-W., M. Hueste, and P. Gardoni. 2009. "Probabilistic Assessment of Structural Damage due to Earthquakes for Buildings in Mid-America." *Journal of Structural Engineering* 135 (10): 1155–1163.

Boakye, J., R. Guidotti, C. Murphy, and P. Gardoni. 2019. "Quantifying the Role of the Transportation Network on the Societal Impact of Natural Hazards." In preparation

Boakye, J., C. Murphy, and P. Gardoni. 2018. "Resilience and Sustainability Goals for Communities and Quantification Metrics." In *Handbook on Sustainable and Resilient Infrastructure*, edited by P. Gardoni, 50–69. New York: Routledge.

Boore, D. M., and G. M. Atkinson. 2008. "Ground-motion Prediction Equations for the Average Horizontal Component of PGA, PGV, and 5%-Damped PSA at Spectral Periods between 0.01 and 10.0 s." *Earthquake Spectra* 24 (1): 99–138.

Cenni, D., P. Nesi, G. Pantaleo, and I. Zaza. 2017, August. "Twitter Vigilance: A Multi-user Platform for Cross-domain Twitter Data Analytics, NLP and Sentiment Analysis." 2017 *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 1–8.

Cleveland, L. J., A. S. Elnashai, and O. Pineda. 2007. New Madrid Seismic Zone Catastrophic Earthquake Response Planning. *MAE Center Report 07-03*.

FEMA (Federal Emergency Management Agency). 2015. HAZUS 2.1 Technical and User's Manuals. https://www.fema.gov/media-library/assets/documents/24609.

Gardoni, P., ed. 2017. *Risk and Reliability Analysis: Theory and Applications*. Cham: Springer.

Gardoni, P., A. Der Kiureghian, and K. M. Mosalam. 2002. "Probabilistic Capacity Models and Fragility Estimates for Reinforced Concrete Columns based on Experimental Observations." *Journal of Engineering Mechanics* 128 (10): 1024–1038.

Gardoni, P., K. M. Mosalam, and A. Der Kiureghian. 2003. "Probabilistic Seismic Demand Models and Fragility Estimates for RC Bridges." *Journal of Earthquake Engineering* 7 (Special Issue 1): 79–106.

Gardoni, P., and C. Murphy. 2008. "Recovery from Natural and Man-Made Disasters as Capabilities Restoration and Enhancement." *International Journal of Sustainable Development and Planning* 3 (4): 1–17.

Gardoni, P., and C. Murphy. 2009. "Capabilities-based Approach to Measuring the Societal Impacts of Natural and Man-made Hazards in Risk Analysis." *ASCE Natural Hazards Review* 10 (2): 29–37.

Gardoni, P., and C. Murphy. 2010. "Gauging the Societal Impacts of Natural Disasters using a Capabilities-based Approach." *Disasters* 34 (3): 619–636.

Gardoni, P., and C. Murphy. 2013. "A Capability Approach for Seismic Risk Analysis and Management." In *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*, edited by Solomon Tesfamariam and Katsu Goda, 255–267. Cambridge: Woodhead Publishing Ltd.

Gardoni, P., and C. Murphy. 2014. "A Scale of Risk." *Risk Analysis* 34 (7): 1208–1227.

Gardoni, P., and C. Murphy. 2018. "Society-based Design: Promoting Societal Well-being by Designing Sustainable and Resilient Infrastructure." *Sustainable and Resilient Infrastructure*. doi:10.1080/23789689.2018.1448667.

Gladwin, Hugh. 1997. "Warning and Evacuation: A Night for Hard Houses." In *Hurricane Andrew: Ethnicity, Gender, and the Sociology of Disasters*, edited by Walter Gillis Peacock, Betty Hearn Morrow, and Hugh Gladwin, 52–74. London: Routledge.

Gonzalez, M. C., C. A. Hidalgo, and A. L. Barabasi. 2008. "Understanding Individual Human Mobility Patterns." *nature* 453 (7196): 779–782.

Grossi, V., B. Rapisarda, F. Giannotti, and D. Pedreschi. 2018. "Data Science at SoBigData: The European Research Infrastructure for Social Mining and big Data Analytics." *International Journal of Data Science and Analytics* 6 (3): 205–216.

Guidotti, R., P. Gardoni, and N. Rosenheim. 2019. "Integration of physical infrastructure and social systems in communities' reliability and resilience analysis." *Reliability Engineering and System Safety* 185: 476–492. doi:10.1016/j.ress.2019.01.008.

Guidotti, R., S. Tian, and P. Gardoni. 2019. "Simulation of Seismic Wave Propagation in the Memphis Metropolitan Statistical Area (MMSA). In Preparation

Gupta, A., H. Lamba, P. Kumaraguru, and A. Joshi. 2013, May. "Faking Sandy: Characterizing and Identifying Fake Images on Twitter During Hurricane Sandy." Proceedings of the 22nd international conference on World Wide Web, ACM, 729–736.

Iannacone, L., and P. Gardoni. 2018. "Physics-based Repair Rates for Pipelines Subject to Seismic Excitations." Proceedings of 16th European conference on Earthquake Engineering, Thessaloniki, Greece.

Jia, G., and P. Gardoni. 2018a. "State-dependent Stochastic Models: A General Stochastic Framework for Modeling Deteriorating Engineering Systems Considering Multiple Deterioration Processes and their Interactions." *Structural Safety* 72: 99–110.

Jia, G., and P. Gardoni. 2018b. "Simulation-based Approach for Estimation of Stochastic Performances of Deteriorating Engineering Systems." *Probabilistic Engineering Mechanics* 52: 28–39.

Jia, G., A. Tabandeh, and P. Gardoni. 2017. "Life-cycle Analysis of Engineering Systems: Modeling Deterioration, Instantaneous Reliability, and Resilience." In *Risk and Reliability Analysis: Theory and Applications*, edited by P. Gardoni, 465–494. Cham: Springer.

Kim, G. H., S. Trimi, and J. H. Chung. 2014. "Big-data Applications in the Government Sector." *Communications of the ACM* 57 (3): 78–85.

Kogan, M., L. Palen, and K. M. Anderson. 2015. "Think Local, Retweet Global: Retweeting by the Geographically-vulnerable during Hurricane Sandy." Proceedings of the 18th ACM conference on computer supported cooperative work & social computing. ACM, 981–993.

Kumar, R., D. B. Cline, and P. Gardoni. 2015. "A Stochastic Framework to Model Deterioration in Engineering Systems." *Structural Safety* 53: 36–43.

Kumar, R., and P. Gardoni. 2011. "Modeling Structural Degradation of RC Bridge Columns Subjected to Earthquakes and their Fragility Estimates." *Journal of Structural Engineering* 138 (1): 42–51.

Kumar, R., and P. Gardoni. 2013. "Stochastic Modeling of Deterioration in Buildings and Civil Infrastructure." In *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*, edited by S. Tesfamariam and K. Goda, 410–434. Cambridge: Woodhead Publishing Ltd.

Kumar, R., and P. Gardoni. 2014. "Renewal Theory-based Life-cycle Analysis of Deteriorating Engineering Systems." *Structural Safety* 50: 94–102.

Kumar, R., P. Gardoni, and M. Sanchez-Silva. 2009. "Effect of Cumulative Seismic Damage and Corrosion on the Life-cycle Cost of Reinforced Concrete Bridges." *Earthquake Engineering & Structural Dynamics* 38 (7): 887–905.

Lin, Y. S. 2009. "Development of Algorithms to Estimate Post-Disaster Population Dislocation—a Research-Based Approach." PhD diss., Texas A&M University.

Martin, S., and L. Singh. 2018. "Data Analytics and Displacement: Using Big Data to Forecast Mass Movements of People." In *Digital Lifeline? ICTs for Refugees and Displaced Persons*, edited by Carleen F. Maitland, 185–206. Cambridge, MA: MIT Press.

Murphy, C., and P. Gardoni. 2006. "The Role of Society in Engineering Risk Analysis: A Capabilities-based Approach." *Risk Analysis* 26 (4): 1073–1083.

Murphy, C., and P. Gardoni. 2007. "Determining Public Policy and Resource Allocation Priorities for Mitigating Natural Hazards: a Capabilities-based Approach." *Science and Engineering Ethics* 13 (4): 489–504.

Murphy, C., and P. Gardoni. 2008. "The Acceptability and the Tolerability of Societal Risks: A Capabilities-based Approach." *Science and Engineering Ethics* 14 (1): 77–92.

Murphy, C., and P. Gardoni. 2010. "Assessing Capability Instead of Achieved Functionings in Risk Analysis." *Journal of Risk Research* 13 (2): 137–147.

Murphy, C., and P. Gardoni. 2011. "Evaluating the Source of the Risks Associated with Natural Events." *Res Publica (liverpool, England)* 17 (2): 125–140.

Murphy, C., and P. Gardoni. 2012a. "Design, Risk and Capabilities." In *The Capability Approach, Technology and Design*. Vol. 5., edited by Ilse Oosterlaken and Jeroen van den Hoven, 173–188. Dordrecht: Springer.

Murphy, C., and P. Gardoni. 2012b. "The Capability Approach in Risk Analysis." In *Handbook of risk theory : epistemology, decision theory, ethics, and social implications of risk*, edited by Sabine Roeser, 979–997. Dordrecht: Springer.

Nussbaum, M. 2000a. "Aristotle, Politics, and Human Capabilites: A Response to Antony, Arneson, Charlesworth, and Mulgan." *Ethics* 111 (1): 102–140.

Nussbaum, M. 2000b. *Woman and Human Development: The Capabilites Approach*. Cambridge: Cambridge University Press.

Nussbaum, M. 2001. "Adaptive Preferences and Women's Options." *Economics and Philosophy* 17: 67–88.

Qu, Y., C. Huang, P. Zhang, and J. Zhang. 2011. "Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake." Proceedings of the ACM 2011 conference on computer supported cooperative work, ACM, 25–34.

Raworth, K., and D. Stewart. 2003. "Critiques of the Human Development Index: A Review." In *Readings in Human Development*, edited by S. Fukuda-Parr, and A. K. Shiva Kumar, 140–152. Oxford, UK: Oxford University Press.

Robeyns, I. 2006. "The Capability Approach in Practice." *The Journal of Political Philosphophy* 14 (3): 351–376.

Sakaki, T., M. Okazaki, and Y. Matsuo. 2010. "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors." Proceedings of the 19th international conference on World Wide Web, ACM, 851–860.

Sen, A. 1989. "Development as Capabilities Expansion." *Journal of Development Planning* 19: 41–58.

Sen, A. 1992. *Inequality Reexxamined*. Cambridge: Harvard University Press.

Sen, A. 1993. "Capability and Well-being." In *The Quality of Life*, edited by M. Nussbaum, and A. Sen, 30–53. Oxford: Clarendon.

Sen, A. 1999a. *Commodoties and Capabilities*. Oxford: Oxford University Press.

Sen, A. 1999b. *Development as Freedom*. New York: Anchor Books.

Sharma, N., A. Tabandeh, and P. Gardoni. 2018. "Regional Resilience Analysis: A Multi-scale Approach to Model and Optimize the Recovery of Interdependent Infrastructure." In *Handbook of Sustainable and Resilient Infrastructure*, edited by P. Gardoni, 529–552. New York: Routledge.

Starbird, K., and L. Palen. 2010. *Pass it on?: Retweeting in Mass Emergency*, 1–10. Seattle: International Community on Information Systems for Crisis Response and Management.

Stewart, F. 2005. "Groups and Capabilities." *Journal of Human Development* 6 (2): 185–204.

Tabandeh, A., P. Gardoni, and C. Murphy. 2017. "Reliability-based Capability Approach: A System Reliability Formulation for the Capability Approach." *Risk Analysis* 38 (2): 410–424.

Tabandeh, A., P. Gardoni, C. Murphy, and N. Myers. 2019. "Societal Risk and Resilience Analysis: Dynamic Bayesian Network Formulation of a Capability Approach." *ASCE Journal of Risk and Uncertainty Analysis* 5 (1): 04018046.

Tene, O., and J. Polonetsky. 2012. "Big Data for All: Privacy and User Control in the Age of Analytics." *Northwestern Journal of Technology and Intellectual Property* 11: xxvii.

Trevor, H., T. Robert, and Friedman J. H. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*.

Wang, Z., and X. Ye. 2018. "Social Media Analytics for Natural Disaster Management." *International Journal of Geographical Information Science* 32 (1): 49–72.