Live Simulations

Blue Sky Ideas Track

Samarth Swarup[†] and Henning S. Mortveit[†‡]

[†]Biocomplexity Institute and Initiative
[‡]Department of Engineering Systems and Environment
University of Virginia, Charlottesville, VA
{swarup,henning.mortveit}@virginia.edu

ABSTRACT

The next exciting step for large-scaled, data-driven, agent-based simulations is to make them *live*. In this article we describe what is meant by a live simulation, how this concept goes beyond the state of the art, why this would be transformative in multiple domains, and a path for achieving this vision. We discuss the major challenges to building live simulations covering aspects such as (i) data integration and unification, (ii) time scales and spatial resolutions, and (iii) simulation model scalability covering both computational tractability and sparse data, all with an eye on progress on various fronts that can be integrated towards realizing this vision.

KEYWORDS

agent-based simulation; scalability; live data integration;

1 INTRODUCTION

By the term "live simulation", we mean a large-scale agent-based simulation that is coupled with real-time data feeds. Supporting this has a number of implications, challenges, and applications, but before we describe those in detail, let us first discuss why this short statement constitutes a blue sky idea. We consider a blue sky idea to be one, (a) that is well beyond the state of the art, (b) that, if achieved, would be transformative, and (c) that, for which the path to achievement is possible, though very challenging. To make the case for live simulations, we consider three application domains:

- Disaster response
- Real-time epidemiology
- Computational social science

In each case, a live simulation would involve live data feeds as well as a simulation, i.e., a model that allows forward projection from current conditions. The three domains chosen are arranged in increasing order of complexity for creating live simulations. In subsequent sections, we discuss the challenges in more detail, as well as ongoing research threads in multiple domains that can be brought together to realize the vision of live simulations.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 2020, Auckland, New Zealand

https://doi.org/doi

1.1 Disaster Response

The state of the art: Traditionally, unsurprisingly, it has been hard to get data about population behaviors and mobility during disasters. Thus, the field of disaster response has been largely driven by surveys, both prospective and retrospective [24]. In recent years, there have been attempts at data mining of mobility traces [8, 84] and social media feeds [105] to enable some real-time tracking of people's movements and areas of need. Simulations of disasters have generally focused on hypothetical scenarios, even in cases of very detailed data-driven simulations [4]. This has limited the applicability of agent-based simulations to proofs-of-concept or simple policy evaluations [20, 102].

Why live simulations would be transformative: Live simulations would be transformative in our approach to relief and rescue efforts. We envision a scenario in which an agent-based simulation could be rapidly "spun-up" in a matter of minutes when a disaster occurs, would integrate data from multiple live feeds passively or actively, and would be used to continually generate short-term possible worlds based on the latest data updates. Such a platform could be used by first-responders, by incident managers, and by policy-makers for making decisions about where to allocate resources and effort, when to order evacuations and how to stage them, and how to minimize harm.

They would also be transformative as a research platform for developing and evaluating theories of behavior and its interaction with physical processes and physical infrastructure.

Path to achievement: The kinds of data to be gathered in this scenario include data about physical conditions, e.g., flooding levels in a hurricane, data about the condition of various infrastructures, e.g., bridge and building safety, and data about locations of people. These data can be gathered through various kinds of physical sensors, such as satellite images, drones, cellphones, street cameras, which are already in use, and sensors installed directly on infrastructure. The models needed for the simulation are mainly behavioral models that influence human mobility, such as when people choose to evacuate. These behaviors are well-studied in the transportation literature [59]. The timescales of such simulations would be on the order of hours and days, which highlights the need for high-performance computing, but also limits the scope of variability (as compared to scenarios which last for weeks or months).

However, there is an enormous amount of work to be done on the methodological front and the engineering front. New methods are needed for data fusion to combine data from multiple feeds into a consistent state estimate, data assimilation to integrate the state estimate into an agent-based model, risk assessment and active allocation of resources for further sensing, real-time discovery of implementable interventions, course of action analysis, and explanation of results. Engineering challenges include building robust platforms that allow interfacing with a variety of hardware, software, and data formats. Sensors can be noisy, unreliable, and can break in harsh conditions. Usability and usefulness also need to be emphasized so that the simulation can be run easily and results are presented in an understandable and actionable way.

1.2 Real-time Epidemiology

The state of the art: Mathematical and computational modeling in epidemiology has a long history [26]. In the Big Data age, digital epidemiology [78] has come to the fore, with many kinds of disease surveillance systems, including sentinel systems [16], participatory surveillance [13, 65], and social media-based systems [47]. Large-scale, agent-based simulation also has a long history in this field [27] and is well-known in the MAS community [90]. However, as in the case of disaster response, the use of agent-based simulations has been largely restricted to hypothetical scenarios, with the goal of policy evaluation or technical/methodological development. Even in cases of real epidemics where simulations were widely used, such as the H1N1 outbreak of 2009 [5] or the Ebola crisis of 2014 [73], setting up the simulations and generating initial conditions in the right format, etc., were done manually.

Why live simulations would be transformative: We envision a scenario where highly spatiotemporally-resolved disease surveillance would feed into a continually-running simulation platform that would allow projection of outbreaks, risk assessment, and generation of action recommendations, such as rankings of interventions. The surveillance can include an active or participatory component, in which case the platform itself can trigger sensing in areas where more information is judged necessary. This would work on a global scale, allowing rapid and efficient mobilization of resources. Simulations are necessary to understand which cases are likely to turn into large outbreaks, thus allowing appropriate allocation of scarce resources. They are also necessary to understand which kinds of interventions will work best for which cases.

Done correctly, this kind of live simulation platform has the potential to greatly reduce the risk of pandemics and epidemics, and thereby greatly alleviate the burden of infectious diseases in the world.

Path to achievement: Despite the large amount of research into disease surveillance systems and into large-scale agent-based simulations of epidemics, the path to a live simulation platform as outlined above is considerably more complex than the case of disasters, for several reasons.

One fundamental problem is that disease states cannot be accessed directly in the way that, e.g., mobility traces can. Diagnosis of a disease requires lab tests which may be expensive, time-consuming to administer and report to a data platform. Even symptoms cannot be accessed directly. Thus methods are needed to infer the true burden of the disease from sparse data, but the lack of ground truth makes this very difficult. However, at the very least, we can imagine that current global disease surveillance platforms can be extended to provide high-resolution spatiotemporal data. New methods are also needed to do active sensing using such platforms, where the sensing is driven by the model.

On the modeling and simulation front, we need to be able to build meaningful simulations for regions of the world where the data collection and availability is very poor. For example, when the Ebola crisis hit Guinea, Sierra Leone, and Liberia in 2014, Liberia was the only country which had a recent population census, from 2008. Guinea and Sierra Leone did not have any up to date official population data. If we make the effort to develop a global representation of the population and their activities in advance, we would be much better prepared for the next major outbreak. There are various efforts in this direction, with multiple data sources about populations [29, 68, 104] and activities [17, 28, 98] becoming available. Much work needs to be done, however, to create an automatically-updating data resource that integrates data from these sources and others to create a synthesized data set appropriate for a simulation model.

Since the scale of resources available to respond to pandemics does not match the scale of the problem [76], new methods are also needed for discovering interventions under very tight budget constraints. Methodological development is also needed for creating simulations that take into account human behavior [30], which is starting to be addressed [82]. However, live data about behaviors is very hard to obtain, and the problem of modeling behavior also includes socio-cultural modeling in a broader sense [56].

The timescale of pandemics generally are on the order of weeks to months, and there is less of a need for interfacing with physical hardware for sensing, so the engineering challenges are different from the ones for disaster simulations. However, there is still the same need for robustness, interfacing, usability and usefulness, and simulation analytics.

1.3 Computational Social Science

The state of the art: Agent-based modeling in social science has a long history [79] and is a very broad domain, which includes a diverse range of phenomena such as online discourse, crime, collective action, opinion dynamics, and more. However, we discuss it here as one domain because it is generally recognized as such [22] and because agent-based modeling is recognized as a broadly applicable technology in this domain [55]. This also helps highlight the breadth of application of live simulations as well as show the difficulty of applying simulation-based technology to sociological phenomena. It has also been theorized that many of these phenomena can be understood as facets of an underlying process of social interaction [37], which could be investigated with precisely the kind of platform we are proposing.

Recent years have seen increasingly successful application of computational methods. In particular machine learning and data mining have been applied to modeling civil unrest and other forms of collective action [71, 95], and agent-based modeling has found success in multiple applications, including modeling crime [75], incarceration [54], and technology adoption [107].

On the whole, however, large-scale agent-based simulations have not been applied to computational social science at the same scale as for disaster modeling and epidemiology. This is not surprising because computational social science, and social science more broadly, deals with more abstract and nebulous concepts like collective identity [31], misinformation [21], and social change [86]. The relevant sociological theories can be hard to operationalize into computational models, and correspondingly hard to validate, which has largely limited the application of agent-based models to explanation rather than forecasting or response.

Why live simulations would be transformative: We envision a scenario in which a live simulation platform would continually integrate information about events around the world in combination with sentiment and opinion from social media and other open source indicators, with a model of the population and their activity patterns. This would allow a highly spatiotemporally-resolved analysis of unfolding events and offer feedback and insight into policy implementation and its effects on population welfare in real-time.

Such a platform would undoubtedly be very hard to build, but we argue that its benefits would be unprecedented through making democracy and governance more data-driven and transparent, thereby making it more robust and resilient. We are living in times of very rapid physical, social, and technological change. This kind of platform would help us to better prepare for and rapidly adapt to these changes by providing an ability to assess consequences of policies in real-time and to optimize allocation of resources for social good.

Path to achievement: Multiple relevant data sources already exist, e.g., projects like GDELT [50], ICEWS [12], and EventRegistry [48] provide up to date news from around the world, coded in machine-understandable forms. Methodology is also progressing for understanding the spread of information in various online social media [58], for how these social networks grow [97], and for relating online and offline events [74]. Social media analysis is a big research area, far beyond what we can summarize here, and will evolve as information and communication technologies themselves change, but these data sets and insights are important components of a live simulation research program.

New methods are needed, though, for integration of data from anonymous platforms such as social media into agent-based simulations. Typically, data-driven agent-based models rely on demographic matching to integrate data from multiple sources [52, e.g.]. While some demographic attributes can be inferred on some platforms [100], there are multiple other obstacles to creating a consistent representation for a simulation. For example, users can have multiple accounts on the same platform, can assume different roles on different platforms, and can exhibit significantly different opinions and behavior online and offline.

It has also been argued that multi-scale modeling, integrating cognitive science and social science, is essential to the proper simulation of human behavior [64]. This is an active area of research from a scientific perspective, so we are quite far from having well-accepted models of how such integration is to be achieved. An ongoing challenge in both cognitive science and sociology is how to operationalize theories in way that allows computational modeling. Operationalization broadly refers to making a theoretical variable measurable. A computational simulation, however, requires specifying the cognitive/sociological process algorithmically so that it can be implemented as an agent, even if only some aspects of this process are directly measurable. This kind of computational operationalization is a step beyond what is typically done in cognitive science and sociology and requires methodological advancement at the intersection of those fields and the field of multi-agent modeling.

From an engineering perspective, live simulation in computational social science is challenging because of the range of scales. It can span multiple timescales because of the range of phenomena. For example, people can tweet in seconds, but opinions percolate on Twitter over hours and days; they can lead to social movements that span weeks or months, or possibly years in the case of sustained social and political efforts. It can also span multiple spatial scales, from small groups to cities, countries, and the entire world. Finally, it can span multiple scales of complexity, from simple voter model-like simulations to very complex reasoning agents. Creating a live simulation platform that can function at all these scales will require especially robust, HPC-based designs, sophisticated data and information management architectures, and real-time analytics capabilities to complement the live simulations.

2 CHALLENGES

The discussion of three domains in the previous section has highlighted some common, general challenges. Before we address those, there is one very important, overarching challenge: can live simulations be done in a value-sensitive way?

By "value-sensitivity" [33], we mean designing tools in a manner that is ethical, moral, and respectful of a broad range of human values, including security, privacy, dignity, fairness, accountability, and transparency. Values are embedded into every stage of the design and use of computational tools, whether we are explicitly aware of them or not. There is a range of potential biases in big data [41] and numerous "ethical tensions" in their use [18]. When biased data are uncritically incorporated into response procedures, outcomes can exacerbate inequality [53]. Live simulations, incorporating streaming data, are going to require a novel methodology for removing or compensating for biases. Similarly, great care must be taken with respect to their use. Over-reliance on any one tool can lead to a reduction in critical thinking. Privacy is another important issue, since it is well known that deidentification is not enough [61]. Synthetic data, combined with differential privacy, may offer a solution in this regard [7], but much work remains to be done on this front.

Technical challenges. Live simulations face several challenges, one broad class being efficient, scalable *data collection* operating close to *real-time*. Additionally, sensors providing data feeds may operate under adverse conditions, and may have inherent uncertainties due to engineering limitations. A computational platform integrating data collection and simulation models must handle data volume, variety, velocity, and veracity (the four Vs) in a robust manner. Naturally, the simulation models residing in this platform must be designed to flexibly adapt to the scale or resolution of each V-dimension in a manner that is meaningful for simulations to be considered live. Designing such an ecosystem of models, data, and analytical tools in a way that supports, for example, realistic policy formation is no small task, in particular when one adds the need for complete provenance tracking of data. The latter may be possible in constrained environments (e.g., within Python or R) but is a serious undertaking for more flexible combinations of computational tools. Naturally, it is desirable that new data sources and suitably designed simulation models and analytic tools can be integrated with relative ease in an HCI-sensible manner.

3 RELEVANT RESEARCH

There are number of disparate streams of research that need to be brought together to create live simulation technology. We discuss these briefly below, moving from data collection to integration, modeling, system engineering, use, and privacy.

Data collection, curation, annotation, feature extraction:

There are a number of efforts underway to collect data from multiple sources during disasters [9, 77, 99], epidemics [16, 46, 65], and more broadly about world events [12, 48, 50]. The sources include social media, news, satellite imagery, GPS traces, call detail records, and more. These complement data collection efforts related to population estimation, such as WorldPop [104] and LandScan [10], and regular large-scale survey efforts such as national censuses in many countries, national and multi-national surveys of activities [17, 98], health conditions [62], and attitudes, beliefs, and behaviors [28]. These are just a few examples, there are many more.

Data synthesis: Despite this extensive data collection, there are some kinds of data which are not available, yet highly relevant to the application domains we have described. An example is, detailed models of the population where every individual is represented. In this case, data are collected, e.g., by national censuses, but are disclosed only at some level of aggregation. Another example of this is networks of physical collocation of people, which are needed for simulating disease transmission at high resolution. In this case, data are not possible to collect through survey methods because most people do not know all the people they are collocated with during a typical day. For these scenarios, synthetic population methods have been developed to create estimates [1, 3, 19, 35, 36, 60, 87]. However, integrating synthetic population data with live data streams such as above needs both methodological and engineering innovation. We believe this is an open area of research that can provide immediate benefits, such as improving disease forecasting, while also creating a stepping stone to the broader challenge of creating live simulations.

Data assimilation: New methods are being developed for assimilating data from observations into multi-agent simulations [13, 23, 51, 101]. These methods include filtering-based approaches for state estimation and calibration of agent-based models, as well as visual environments that allow humans to participate in the situation assessment process. Scalability and real-time performance are ongoing challenges for these systems, where parallel and distributed simulation platforms should find immediate application.

Active learning and state estimation: A live simulation could, in principle, actively trigger information collection efforts where more information is needed to improve the accuracy of state estimates and forward projections. In disaster situations, this might require interfacing with sensing hardware, such as drones, in order to gather data in areas where communication infrastructure may be damaged. Research in low-power drone technology for extended sensing applications is progressing [80].

Prediction and simulation of mobility: In multiple applications in the domains discussed, the agent model for the simulation requires predicting human mobility (e.g., disaster evacuation scenarios, population mixing scenarios for epidemics, and migration scenarios in social science.) There is much work in this area, both in the context of disasters [84, 88] and more generally [43]. These methods rely on digital traces, such as GPS or cellphone data, though possibly anonymized and aggregated. Thus methods are needed for integrating these with a population model where not everyone has a device generating such a signal, or where signal detection is limited due to lack of access to data or damaged infrastructure.

Behavior modeling and inference: Depending on the context, the kinds of behaviors that people engage in can vary greatly. For example, behaviors in disasters, such as evacuation, looking for family members, and aiding and assisting others, are quite different from behaviors during epidemics, which might include staying at home, getting vaccinated, and avoiding crowded places. Simulating behaviors requires both modeling these behaviors and inferring them from the data streams in order to accurately estimate and predict state. There have been several efforts at behavior modeling in multi-agent simulations [66, 70, 81, 96]. Evacuation behaviors have long been studied in the transportation literature [59]. More recently, deep learning methods and inverse reinforcement learning are being applied to learning behavior models [49, 83]. However, there is little work that combines behavior modeling in a simulation with behavior inference from data. This is a promising direction for methodological advancement [72].

Platforms for scalable simulation: Platforms for simulation can be designed with various conceptions of scalability, e.g., computational scalability [6, 11], scalability in the scope of data and models that can be integrated [39], and scalability in terms of rapid development and model composability [4], to name just a few. All these properties are relevant to a platform for live simulations, and there are lessons to be learned from each perspective on scalability in the design of the proposed kind of platform.

There are several computational frameworks and workflow systems that address parts of the challenges faced by live simulations, notable examples being Taverna [93], Airavata [2, 57, 67], Swift [103], Tapis [25, 92], and SciDuct [42]. Examples of factors limiting these frameworks include specialization to particular scientific domains, omission of rigorous provenance tracking, or simply having had design goals that do not include all the facets needed for live simulation. Other approaches include Notebook environments such as Jupyter Notebook [44], Apache Zeppelin [106], and PyCharm [69], however these have inherent limitations regarding the required scale. Data integration and data fusion represent a serious challenge with respect to management and automation. Approaches such as Frictionless Data [32] provide a clean way to standardize data declarations, in particular for tabular data. ORM technologies (e.g., SQLAlchemy [85]) should naturally also be considered.

Simulation analytics: Simulations can produce more data than they consume. Sense-making with complex simulations can be hard. New methods for analytics are needed that can exploit the repeatability and completeness of the data generated by a simulation to generate insights into, e.g., causality [91]. Tools for doing analytics can also be integrated into the simulation platform to enable end-to-end analysis in real time [94]. Applications of machine learning to create surrogate models [45] or response surfaces [15, 63] can also be integrated into simulation platforms.

4 CONCLUSION

Though the challenges are significant, the opportunities in these domains are just beginning to be realized [14, 34, 38]. In all these domains, policies and human factors drive outcomes through dynamical interactions. Thus it is important to have a means of "putting the data into motion" and answering what-if questions [40]. At the same time, we are not advocating building a simulation of everything. A notion of adequacy for use is very important [89]. We believe that live simulations will greatly expand the scope and value of application of MABS technology in the world.

ACKNOWLEDGMENTS

This work was partially supported by DTRA Grant HDTRA1-17-F-0118, NASA grant 80NSSC18K1594, NSF Grant No. SMA-1520359 and by DARPA, via AFRL Contract No. FA8650-19-C-7923.

REFERENCES

[1] Abhijin Adiga, Aditya Agashe, Shaikh Arifuzzaman, Christopher L. Barrett, Richard J. Beckman, Keith R. Bisset, Jiangzhuo Chen, Youngyun Chungbaek, Stephen G. Eubank, Sandeep Gupta, Maleq Khan, Christopher J. Kuhlman, Eric Lofgren, Bryan L. Lewis, Achla Marathe, Madhav V. Marathe, Henning S. Mortveit, Eric Nordberg, Caitlin Rivers, Paula Stretz, Samarth Swarup, Amanda Wilson, and Dawen Xie. 2015. *Generating a Synthetic Population of the United States.* Technical Report NDSSL 15-009. Network Dynamics and Simulation Science Laboratory.

[2] Airavata [n. d.]. Apache Airavata. ([n. d.]). https://airavata.apache.org/index.html Last accessed: 14 November 2019.

[3] Robert L. Axtell. 2016. 120 Million Agents Self-Organize into 6 Million Firms: A Model of the U.S. Private Sector. In *Proc. AAMAS*.

[4] Christopher Barrett, Keith Bisset, Shridhar Chandan, Jiangzhuo Chen, Youngyun Chungbaek, Stephen Eubank, Yaman Evrenosoğlu, Bryan Lewis, Kristian Lum, Achla Marathe, Madhav Marathe, Henning Mortveit, Nidhi Parikh, Arun Phadke, Jeffrey Reed, Caitlin Rivers, Sudip Saha, Paula Stretz, Samarth Swarup, James Thorp, Anil Vullikanti, and Dawen Xie. 2013. Planning and Response in the Aftermath of a Large Crisis: An Agent-based Informatics Framework. In *Proceedings of the 2013 Winter Simulation Conference*, R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl (Eds.). IEEE Press, Piscataway, NJ, USA, 1515–1526.

[5] Christopher Barrett, Keith Bisset, Jonathan Leidig, Achla Marathe, and Madhav Marathe. 2010. An Integrated Modeling Environment to Study the Co-evolution of Networks, Individual Behavior, and Epidemics. *AI Magazine* 31, 1 (2010), 75–87.

[6] Christopher L. Barrett, Keith R. Bisset, Stephen G. Eubank, Xizhou Feng, and Madhav V. Marathe. 2008. EpiSimdemics: an Efficient Algorithm for Simulating the Spread of Infectious Disease over Large Realistic Social Networks. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing (SC '08)*. 37:1–37:12.

[7] Steven M. Bellovin, Preetam K. Dutta, and Nathan Reitinger. 2019. Privacy and Synthetic Datasets. *Stanford Technology Law Review* 22, 1 (2019).

[8] Linus Bengtsson, Xin Lu, Anna Thorson, Richard Garfield, and Johan von Schreeb. 2011. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Med* 8, 8 (2011), e1001083.

[9] Kelly J. Bennett, Jennifer M. Olsen, Sara Harris, Sumiko Mekaru, Alicia A. Livinski, and John S. Brownstein. 2013. The Perfect Storm of Information: Combining Traditional and Non-traditional Data Sources for Public Health Situational Awareness During Hurricane Response. *PLoS Currents Disasters* 1 (Dec 16 2013).

[10] Budhendra L Bhaduri, Eddie A Bright, and Jerome E Dobson. 2002. LandScan: Locating People is What Matters. *Geoinfomatics* 5, 2 (1 2002).

[11] Parantapa Bhattacharya, Saliya Ekanayake, Chris Kuhlman, Christian Lebiere, Don Morrison, Samarth Swarup, Mandy Wilson, and Mark Orr. 2019. The Matrix: An Agent-Based Modeling Framework for Data Intensive Simulations. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. Montreal, Canada.

[12] Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. 2015. ICEWS Coded Event Data. (2015). https://doi.org/10.7910/DVN/28075

[13] S. John Brownstein, Shuyu Chu, Achla Marathe, V. Madhav Marathe, T. Andre Nguyen, Daniela Paolotti, Nicola Perra, Daniela Perrotta, Mauricio Santillana, Samarth Swarup, Michele Tizzoni, Alessandro Vespignani, S. Anil Kumar Vullikanti, L. Mandy Wilson, and Qian Zhang. 2017. Combining Participatory Influenza Surveillance with Modeling and Forecasting: Three Alternative Approaches. *JMIR Public Health Surveill* 3, 4 (01 Nov 2017), e83. https://doi.org/10.2196/publichealth.7344

[14] Annetta Burger, Talha Oz, William G. Kennedy, and Andrew T. Crooks. 2019. Computational Social Science of Disasters: Opportunities and Challenges. *Future Internet* 11, 103 (2019).

[15] Kathleen M. Carley, Natalia Y. Kamneva, and Jeff Reminga. 2004. *Response Surface Methodology.* CASOS Technical Report CMU-ISRI-04-136. Carnegie Mellon University.

[16] Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boëlle, Antoine Flahault, and Neil M. Ferguson. 2008. Estimating the Impact of School Closure on Influenza Transmission from Sentinel Data. *Nature* 452 (2008), 750–755.

[17] Center for Time Use Research. [n. d.]. The Multinational Time Use Study. ([n. d.]). https://www.timeuse.org/mtus

[18] Stevie Chancellor, Michael L. Birnbaum, Eric D. Caine, Vincent M. B. Silenzio, and Munmun De Choudhury. 2019. A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media. In *Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3287560.3287587

[19] Kevin Chapuis, Patrick Taillandier, Misslin Renaud, and Alexis Drogoul. 2018. Gen*: a generic toolkit to generate spatially explicit synthetic populations. *International Journal of Geographical Information Science* 32, 6 (2018), 1194–1210.

[20] X. Chen and F. B. Zhan. 2008. Agent-based modelling and simulation of urban evacuation: Relative effectiveness of simultaneous and staged evacuation strategies. *Journal of the Operational Research Society* 59, 1 (2008), 25–33.

[21] Giovanni Luca Ciampaglia. 2018. Fighting fake news: A role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science* 1, 1 (01 Jan 2018), 147–153. https://doi.org/10.1007/s42001-017-0005-6

[22] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J.-P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel, and D. Helbing. 2012. Manifesto of Computational Social Science. *Eur Phys J Special Topics* 214 (2012), 325–346.

[23] Yulin Ding, Qing Zhu, and Hui Lin. 2014. An Integrated Virtual Geographic Environmental Simulation Framework: A Case study of Flood Disaster Simulation. *Geo-spatial Information Science* 17, 4 (2014), 190–200.

[24] David S. Dixon, Pallab Mozumder, William F. Vásquez, and Hugh Gladwin. 2017. Heterogeneity Within and Across Households in Hurricane Evacuation Response. *Networks and Spatial Economics* 17, 2 (01 Jun 2017), 645–680. https://doi.org/10.1007/s11067-017-9339-0

[25] Rion Dooley, Steven R. Brandt, and John Fonner. 2018. The Agave Platform: An Open, Science-as-a-Service Platform for Digital Science. In *Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18)*. ACM, New York, NY, USA, Article 28, 8 pages. https://doi.org/10.1145/3219104.3219129 URL: https://www.tacc.utexas.edu/tapis.

[26] Lila R. Elveback, John P. Fox, Eugene Ackerman, Alice Langworthy, Mary Boyd, and Lael Gatewood. 1976. An Influenza Simulation Model for Immunization Studies. *Am. J. Epidemiology* 103, 2 (1976), 152–165.

[27] S. Eubank, H. Guclu, V. S. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature* 429 (May 2004), 180–184.

[28] European Social Survey. [n. d.]. ([n. d.]). http://www.europeansocialsurvey.org/

[29] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL ©2016 DigitalGlobe. (2016). https://data.humdata.org/dataset/highresolutionpopulationdensitymaps Accessed on Nov 8, 2019.

[30] Neil Ferguson. 2007. Capturing Human Behavior. *Nature* 446 (2007), 733.

[31] Cristina Flesher Fominaya. 2010. Collective Identity in Social Movements: Central Concepts and Debates. *Sociology Compass* 4, 6 (2010), 393–404.

[32] Frictionless [n. d.]. Frictionless Data. ([n. d.]). https://frictionlessdata.io/ Last accessed: 14 November 2019.

[33] Batya Friedman, Peter H. Kahn, Jr., and Alan Borning. 2006. Value Sensitive Design and Information Systems. In *Human-Computer Interaction in Management Information Systems: Foundations*, P. Zhang and D. Galleta (Eds.). M. E. Sharpe, Inc, NY, 348–372.

[34] Richard M. Fujimoto, Christopher Carothers, Alois Ferscha, David Jefferson, Margaret Loper, Madhav V. Marathe, and Simon J. E. Taylor. 2017. Computational Challenges in Modeling and Simulation of Complex Systems. In *Proceedings of the 2017 Winter Simulation Conference*, W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page (Eds.).

[35] Shannon Gallagher, Lee F. Richardson, Samuel L. Ventura, and William F. Eddy. 2018. SPEW: Synthetic Populations and Ecosystems of the World. *Journal of Computational and Graphical Statistics* 27, 4 (2018), 773–784.

[36] Andrea E. Gaughan, Forrest R. Stevens, Zhuojie Huang, Jeremiah J. Nieves, Alessandro Sorichetta, Shengjie Lai, Xinyue Ye, Catherine Linard, Graeme M. Hornby, Simon I. Hay, Hongjie Yu, and Andrew J. Tatem. 2016. Spatiotemporal patterns of population in mainland China, 1990 to 2010. *Scientific Data* 3 (2016), 160005.
[37] Andres Gomez-Lievano, Oscar Patterson-Lomba, and Ricardo Hausmann. 2016. Explaining the prevalence, scaling, and variance of urban phenomena. *Nature Human Behavior* 1 (2016), 0012.
[38] Peter J. Haas. 2014. Model-data Ecosystems: Challenges, Tools, and Trends. In *Proc. PODS*. Snowbird, UT, USA.
[39] Peter J. Haas, Nicole C. Berberis, Piyaphol Phoungphol, Ignacio G. Terrizzano, Wang-Chiew Tan, Patricial G. Selinger, and Paul P. Maglio. 2012. Splash: Simulation Optimization in Complex Systems of Systems. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control and Computing*.
[40] Peter J. Haas, Paul P. Maglio, Patricia G. Selinger, and Wang-Chiew Tan. 2011. Data is Dead... Without What-If Models. In *Proceedings of the VLDB Endowment*, Vol. 4. 1486–1489.
[41] Eszter Hargittai. 2015. Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science* 659, 1 (2015), 63–76. https://doi.org/10.1177/0002716215570866
[42] Stefan Hoops, Brian Klahn, Chris J. Kuhlman, Dustin Machi, Madhav V. Marathe, Henning S. Mortveit, Amanda Wilson, and Dawen Xie. 2019. *SciDuct: a Cyberinfrastructure for Reproducible Science*. Technical Report. NSSAC, University of Virginia. NSSAC Technical Report Series: No. 2019–002.
[43] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C. González. 2016. The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences* 113, 37 (2016), E5370–E5378. https://doi.org/10.1073/pnas.1524261113
[44] Jupyter [n. d.]. Jupyter Notebook. ([n. d.]). https://jupyter.org/ Last accessed: 14 November 2019.
[45] Francesco Lamperti, Andrea Roventini, and Amir Sani. 2018. Agent-based Model Calibration Using Machine Learning Surrogates. *Journal of Economic Dynamic & Control* 90 (2018), 366–389.
[46] Vasileios Lampos and Nello Cristianini. 2012. Nowcasting Events from the Social Web with Statistical Learning. *ACM Trans. Intell. Syst. Technol.* 3, 4, Article 72 (Sept. 2012), 22 pages. https://doi.org/10.1145/2337542.2337557
[47] Vasileios Lampos, Tijl De Bie, and Nello Cristianini. 2010. Flu Detector – Tracking Epidemics on Twitter. In *Proceedings of ECML PKDD (LNAI)*, J. L. Balcázar and et al. (Eds.), Vol. 6323, Part III. Springer Verlag, Berlin Heidelberg, 599–602.
[48] Gregor Leban, Blaz Fortuna, Janez Brank, and Marko Grobelnik. 2014. Event Registry: Learning About World Events from News. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. ACM, New York, NY, USA, 107–110. https://doi.org/10.1145/2567948.2577024
[49] Kamwoo Lee, Mark Rucker, William T. Scherer, Peter A. Beling, Matthew S. Gerber, and Hyojung Kang. 2017. Agent-based Model Construction Using Inverse Reinforcement Learning. In *Proceedings of the 2017 Winter Simulation Conference*, W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page (Eds.). 1264–1275.
[50] Kalev Leetaru and Philip A. Schrodt. 2013. GDELT: Global Data on Events, Location and Tone, 1979-2012.. In *International Studies Association meeting*. San Francisco, CA, USA.
[51] Jordan Lueck, Jason H. Rife, Samarth Swarup, and Nasim Uddin. 2019. Who Goes There? Using an Agent-based Simulation for Tracking Population Movement. In *Proceedings of the Winter Simulation Conference*, N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son (Eds.). National Harbor, MD, USA.
[52] Kristian Lum, Youngyun Chungbaek, Stephen G. Eubank, and Madhav V. Marathe. 2016. A two-stage, fitted values approach to activity matching. *International Journal of Transportation* 4, 1 (2016), 41–56.
[53] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x
[54] Kristian Lum, Samarth Swarup, Stephen Eubank, and James Hawdon. 2014. The contagious nature of imprisonment: An agent-based model to explain racial disparities in incarceration rates. *J. R. Soc. Interface* 11 (2014), 20140409.
[55] Ian S. Lustick and Dan Miodownik. 2009. Abstractions, Ensembles, and Virtualizations: Simplicity and Complexity in Agent-Based Modeling. *Comparative Politics* 41, 2 (2009), 223–244.
[56] Angellar Manguvo and Benford Mafuvadze. 2015. The impact of traditional and religious practices on the spread of Ebola in West Africa: Time for a strategic shift. *Pan Afr Med J.* 22, Suppl 1 (2015), 9.
[57] Suresh Marru, Lahiru Gunathilake, Chathura Herath, Patanachai Tangchaisin, Marlon Pierce, Chris Mattmann, Raminderjeet Singh, Thilina Gunarathne, Eran Chinthaka, Ross Gardler, Aleksander Slominski, Ate Douma, Srinath Perera, and Sanjiva Weerawarana. 2011. Apache Airavata: A framework for distributed applications and computational workflows. *GCE'11 - Proceedings of the 2011 ACM Workshop on Gateway Computing Environments, Co-located with SC'11*, 21–28. https://doi.org/10.1145/2110486.2110490
[58] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. 2017. Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter Bots. *PLoS ONE* 12, 9 (2017), e0184148.
[59] Pamela Murray-Tuite and Brian Wolshon. 2013. Evacuation transportation modeling: An overview of research, development, and practice. *Transportation Research Part C: Emerging Technologies* 27, Supplement C (2013), 25 – 45. https://doi.org/10.1016/j.trc.2012.11.005 Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII).
[60] Mohammad-Reza Namazi-Rad, Payam Mokhtarian, and Pascal Perez. 2014. Generating a Dynamic Synthetic Population - Using an Age-Structured Two-Sex Model for Household Dynamics. *PLoS ONE* 9, 4 (2014), e94761.
[61] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust De-anonymization of Large Sparse Datasets. In *Proceedings of 29th IEEE Symposium on Security and Privacy*. IEEE Computer Society, Oakland, CA, 111–125.
[62] National Center for Health Statistics (U.S.). Division of Health Interview Statistics. 1987. *National Health Interview Survey*. U.S. Public Health Service, National Center for Health Statistics, Division of Health Interview Statistics. https://books.google.com/books?id=f9BxqSorsoIC
[63] H. Gonda Neddermeijer, Gerrit J. van Oortmarssen, Nanda Piersma, and Rommert Dekker. 2000. A Framework for Response Surface Methodology for Simulation Optimization. In *Proceedings of the 32nd Conference on Winter Simulation (WSC '00)*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick (Eds.). Society for Computer Simulation International, 129–136.
[64] Mark G. Orr, Christian Lebiere, Andrea Stocco, Peter Pirolli, Bianica Pires, and William G. Kennedy. 2019. Multi-scale resolution of neural, cognitive and social systems. *Computational and Mathematical Organization Theory* 25, 1 (01 Mar 2019), 4–23. https://doi.org/10.1007/s10588-018-09291-0
[65] D. Paolotti, A. Carnahan, V. Colizza, K. Eames, J. Edmunds, G. Gomes, C. Koppeschaar, M. Rehn, R. Smallenburg, C. Turbelin, S. Van Noort, and A. Vespignani. 2014. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. *Clinical Microbiology and Infection* 20, 1 (2014), 17–21.
[66] Nidhi Parikh, Harshal G. Hayatnagarkar, Richard J. Beckman, Madhav V. Marathe, and Samarth Swarup. 2016. A Comparison of Multiple Behavior Models in a Simulation of the Aftermath of an Improvised Nuclear Detonation. *Autonomous Agents and Multi-Agent Systems, Special Issue on Autonomous Agents for Agent-Based Modeling* 30, 6 (2016), 1148–1174.
[67] Marlon Pierce, Suresh Marru, Lahiru Gunathilake, Thejaka Amila Kanewala, Raminder Singh, Saminda Wijeratne, Chathuri Wimalasena, Chathura Herath, Eran Chinthaka, Chris Mattmann, Aleksander Slominski, and Patanachai Tangchaisin. 2014. Apache Airavata: Design and Directions of a Science Gateway Framework. In *Procceedings of the 2014 6th International Workshop on Science Gateways*. IEEE, 48–54. https://doi.org/10.1109/IWSG.2014.15
[68] Kiesha Prem, Alex R. Cook, and Mark Jit. 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology* 13, 9 (09 2017), 1–21. https://doi.org/10.1371/journal.pcbi.1005697
[69] PyCharm [n. d.]. PyCharm. ([n. d.]). https://www.jetbrains.com/pycharm/ Last accessed: 14 November 2019.
[70] David V. Pynadath, Heather Rosoff, and Richard S. John. 2016. Semi-Automated Construction of Decision-Theoretic Models of Human Behavior. In *Proc. AAMAS*.
[71] Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, Chris Kuhlman, Achla Marathe, Liang Zhao, Ting Hua, Feng Chen, Chang Tien Lu, Bert Huang, Aravind Srinivasan, Khoa Trinh, Lise Getoor, Graham Katz, Andy Doyle, Chris Ackermann, Ilya Zavorin, Jim Ford, Kristen Summers, Youssef Fayed, Jaime Arredondo, Dipak Gupta, and David Mares. 2014. 'Beating the News' with EMBERS: Forecasting Civil Unrest Using Open Source Indicators. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. ACM, New York, NY, USA, 1799–1808. https://doi.org/10.1145/2623330.2623373
[72] Jason Rife, Samarth Swarup, and Nasim Uddin. 2019. A Behavior-Based Population Tracker Can Parse Aggregate Measurements to Differentiate Agents. In *Proceedings of the IEEE International Symposium on Technologies for Homeland Security (IEEE HST)*. Woburn, MA, USA.
[73] Caitlin M. Rivers, Eric T. Lofgren, Madhav Marathe, Stephen Eubank, and Bryan L. Lewis. 2014. Modeling the Impact of Interventions on an Epidemic of Ebola in Sierra Leone and Liberia. *PLOS Currents Outbreaks* Nov 6, Edition 2 (2014).
[74] Daniel M. Romero, Brian Uzzi, and Jon Kleinberg. 2019. Social Networks under Stress: Specialized Team Roles and Their Communication Structure. *ACM Transactions on the Web* 13, 1 (2019), Article 6.
[75] Raquel Rosés, Cristina Kadar, Charlotte Gerritsen, and Chris Rouly. 2018. Agent-based Simulation of Offender Mobility: Integrating Activity Nodes from Location-based Social Networks. In *Proc. AAMAS*. Stockholm, Sweden.
[76] Allen G.P. Ross, Suzanne M. Crowe, and Mark W. Tyndall. 2015. Planning for the Next Global Pandemic. *International Journal of Infectious Diseases* 38 (2015), 89 – 94. https://doi.org/10.1016/j.ijid.2015.07.016
[77] Naina Said, Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Laiq Hassan, Nasir Ahmad, and Nicola Conci. 2019. Natural disasters detection in social media

and satellite imagery: A survey. *Multimedia Tools and Applications* (17 Jul 2019). https://doi.org/10.1007/s11042-019-07942-1

[78] Marcel Salathé, Linus Bengtsson, Todd J. Bodnar, Devon D. Brewer, John S. Brownstein, Caroline Buckee, Ellsworth M. Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L. Mabry, and Alessandro Vespignani. 2012. Digital Epidemiology. *PLoS Comp. Biol.* 8, 7 (2012), e1002616.

[79] Thomas C. Schelling. 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* 1, 2 (1971), 143–186.

[80] Vishal Sharma, Ilsun You, Giovanni Pau, Mario Collotta, Jae Deok Lim, and Jeong Nyeo Kim. 2018. LoRaWAN-Based Energy-Efficient Surveillance by Drones for Intelligent Transportation Systems. *Energies* 11, 3 (2018). https://doi.org/10.3390/en11030573

[81] Dhirendra Singh, Lin Padgham, and Brian Logan. 2016. Integrating BDI Agents with Agent-Based Simulation Platforms. *Auton Agent Multi-Agent Syst* 30, 6 (2016), 1050–1071. https://doi.org/10.1007/s10458-016-9332-x

[82] Meghendra Singh, Achla Marathe, Madhav V. Marathe, and Samarth Swarup. 2018. Behavior Model Calibration for Epidemic Simulations. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. Stockholm, Sweden.

[83] Xuan Song, Ryosuki Shibasaki, Nicholas Jing Yuan, Xing Xie, Tao Li, and Ryutaro Adachi. 2017. DeepMob: Learning Deep Knowledge of Human Emergency Behaviorand Mobility from Big and Heterogeneous Data. *ACM Transactions on Information Systems* 35, 4 (2017), Article 41.

[84] Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Ryosuke Shibasaki, Nicholas Jing Yuan, and Xing Xie. 2016. Prediction and Simulation of Human Mobility Following Natural Disasters. *ACM Transactions on Intelligent Systems and Technology* 8, 2 (2016), Article 29.

[85] SQLAlchemy [n. d.]. SQL Alchemy. ([n. d.]). https://www.sqlalchemy.org/ Last accessed: 14 November 2019.

[86] Flaminio Squazzoni. 2008. A (computational) social science perspective on societal transitions. *Computational and Mathematical Organization Theory* 14, 4 (01 Dec 2008), 266–282. https://doi.org/10.1007/s10588-008-9038-y

[87] Forrest R. Stevens, Andrea E. Gaughan, Catherine Linard, and Andrew J. Tatem. 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* 10, 2 (2015), e0107042.

[88] Akihito Sudo, Takehiro Kashiyama, Takahiro Yabe, Hiroshi Kanasugi, Xuan Song, Tomoyuki Higuchi, Shin'ya Nakano, Masaya Saito, and Yoshihide Sekimoto. 2016. Particle Filter for Real-time Human Mobility Prediction Following Unprecedented Disaster. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

[89] Samarth Swarup. 2019. Adequacy: What Makes a Simulation Good Enough?. In *Proceedings of the Spring Simulation Conference (SpringSim)*. Tucson, AZ.

[90] Samarth Swarup, Stephen Eubank, and Madhav Marathe. 2014. Computational Epidemiology as a Challenge Domain for Multiagent Systems. In *Proceedings of the Thirteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Paris, France.

[91] Samarth Swarup, Achla Marathe, Madhav V. Marathe, and Christopher L. Barrett. 2019. Simulation Analytics for Social and Behavioral Modeling. In *Social-Behavioral Modeling for Complex Systems*, Paul K. Davis, Angela O'Mahony, and Jonathan Pfautz (Eds.). Wiley, 617–632.

[92] Tapis [n. d.]. TACC CLOUD API SYSTEM. ([n. d.]). https://www.tacc.utexas.edu/tapis Last accessed: 14 November 2019.

[93] Taverna [n. d.]. Taverna Workflow System. ([n. d.]). https://taverna.incubator.apache.org/ Last accessed: 14 November 2019.

[94] William J. Tolone and Mark Armstrong. 2011. Integrated Analytics: Understanding Critical Infrastructure Behaviors for Resilience Analysis. *The Homeland Security Review* 5, 3 (2011), 241–258.

[95] Mark Tremayne. 2014. Anatomy of Protest in the Digital Era: A Network Analysis of *Twitter* and Occupy Wall Street. *Social Movement Studies* 13, 1 (2014), 110–126.

[96] Jason Tsai, Natalie Fridman, Emma Bowring, Matthew Brown, Shira Epstein, Gal Kaminka, Stacy Marsella, Andrew Ogden, Inbal Rika, Ankur Sheel, Matthew Taylor, Xuezhi Wang, Avishay Zilka, and Milind Tambe. 2011. ESCAPES - Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social comparison. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[97] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural Diversity in Social Contagion. *PNAS* 109, 16 (April 2012), 5962–5966.

[98] United States Federal Highway Administration. 2017. The National Household Travel Survey. (2017). https://nhts.ornl.gov/.

[99] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1079–1088. https://doi.org/10.1145/1753326.1753486

[100] Svitlana Volkova, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. 2015. Inferring Latent User Properties from Texts Published in Social Media. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[101] Jonathan A. Ward, Andrew J. Evans, and Nicolas S. Malleson. 2016. Dynamic calibration of agent-based models using data assimilation. *Royal Society Open Science* 3, 4 (2016), 150703. https://doi.org/10.1098/rsos.150703

[102] Lawrence M. Wein, Youngsoo Choi, and Sylvie Denuit. 2010. Analyzing Evacuation Versus Shelter-in-Place Strategies After a Terrorist Nuclear Detonation. *Risk Analysis* 30, 9 (2010), 1315–1327.

[103] Michael Wilde, Mihael Hategan, Justin M. Wozniak, Ben Clifford, Daniel S. Katz, and Ian Foster. 2011. Swift: A language for distributed parallel scripting. *Parallel Comput.* 37, 9 (2011), 633–652. https://doi.org/10.1016/j.parco.2011.05.005 http://swift-lang.org/main/.

[104] WorldPop. [n. d.]. WorldPop. ([n. d.]). https://www.worldpop.org Last accessed: 14 November 2019.

[105] Jingchao Yang, Manzhu Yu, Han Qin, Mingyue Lu, and Chaowei Yang. 2019. A Twitter Data Credibility Framework—Hurricane Harvey as a Use Case. *ISPRS Int. J. Geo-Inf.* 8 (2019), Article 111.

[106] Zeppelin [n. d.]. Apache Zeppelin. ([n. d.]). https://zeppelin.apache.org/ Last accessed: 14 November 2019.

[107] Haifeng Zhang, Yevgeniy Vorobeychik, Joshua Letchford, and Kiran Lakkaraju. 2016. Data-driven agent-based modeling, with application to rooftop solar adoption. *Auton Agent Multi-Agent Syst* 30, 6 (2016), 1023–1049. https://doi.org/10.1007/s10458-016-9326-8