Chapter 36

Integrating CyberGIS and Urban Sensing for Reproducible Streaming Analytics

Shaowen Wang, Fangzheng Lyu, Shaohua Wang, Charles E. Catlett, Anand Padmanabhan, and Kiumars Soltani

Abstract Increasingly pervasive location-aware sensors interconnected with rapidly advancing wireless network services are motivating the development of near-real-time urban analytics. This development has revealed both tremendous challenges and opportunities for scientific innovation and discovery. However, state-of-the-art urban discovery and innovation are not well equipped to resolve the challenges of such analytics, which in turn limits new research questions from being asked and answered. Specifically, commonly used urban analytics capabilities are typically designed to handle, process, and analyze static datasets that can be treated as map layers and are consequently ill-equipped in (a) resolving the volume and velocity of urban big data; (b) meeting the computing requirements for processing, analyzing, and visualizing these datasets; and (c) providing concurrent online access to such analytics. To tackle these challenges, we have developed a novel cyberGIS framework that includes computationally reproducible approaches to streaming urban analytics. This framework is based on CyberGIS-Jupyter, through integration of cyberGIS and real-time urban sensing, for achieving capabilities that have previously been unavailable toward helping cities solve challenging urban informatics problems.

S. Wang (✉) · F. Lyu · S. Wang · A. Padmanabhan
Department of Geography and Geographic Information Science, and CyberGIS Center for
Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Urbana, USA
e-mail: shaowen@illinois.edu

S. Wang
e-mail: shaohua@illinois.edu

C. E. Catlett
Argonne National Laboratory, and the University of Chicago, Chicago, USA
e-mail: catlett@anl.gov

K. Soltani
Zillow Inc, Seattle, USA
e-mail: soltani2@illinois.edu

36.1 Introduction and Background

Harnessing urban big data to support scientific investigations into the impacts, challenges, and opportunities associated with increasing urbanization promises to enable the combination of analysis, observation, and modeling capabilities and to set and evaluate urban development policies and goals. Urban areas account for 70% of greenhouse gas emissions and energy use while contributing nearly 80% of total gross national product (GNP) (UN-Habitat 2011). They are consequently important levers to address environmental sustainability. For example, in the Chicago urban area, over 120 cities, towns, and villages have formally adopted a joint sustainability plan called the “Greenest Region Compact” (Marka 2019), understanding that challenges such as the reduction of greenhouse gas emissions or the improvement of air quality are regional in nature, requiring holistic approaches. Setting and tracking progress toward meeting those goals requires harnessing urban big data from not only traditional sources but from new sensor networks, high-bandwidth instruments such as light detection and ranging (LiDAR) and camera systems, and new sources such as those related to remote imaging or mobility. This will require a new approach to urban spatial analytics to support scientific investigations into the impacts, challenges, and opportunities associated with increasing urbanization. These investigations will require applying analysis, observation, and modeling capabilities to set and evaluate urban development policies and goals.

In this context, complex and massive urban data are increasingly collected for understanding and tackling such grand challenges, motivating many urban observatories that could play essential roles in resolving these challenges through science, engineering, and policy innovations (Miller et al. 2019). However, such observatories require innovative approaches to integrating dynamic and voluminous urban data with associated analytics for a variety of scientific problem-solving and decision-making purposes. Therefore, the overarching objective of this research is to develop an innovative cyberGIS (i.e. geographic information science and systems or GIS, based on advanced cyberinfrastructure: Wang 2010) framework for integrating urban sensing and analytics in a computationally reproducible way.

36.1.1 *Urban Sensing Data*

With recent rapid advances in and widespread adoption of location-aware devices and sensors, researchers in many fields now have an overwhelming wealth of dynamic urban data to investigate pressing scientific questions (Armstrong et al. 2019). These data streams from fixed as well as mobile platforms pose significant challenges to urban analytics. The past decade of open-data initiatives has similarly resulted in diverse new datasets related to urban infrastructure, operations, and activities (Huijboom and Van den Broek 2011). Anonymized open data is also available for many US cities such as the City of Chicago, with detailed records of over a decade

of crimes, 311 service calls, permits, inspections, traffic flow, and other operational data. Integrating and analyzing these varied data sources will not only enable new questions about, and insights into the interdependencies of urban phenomena, but also new approaches to understanding complex environmental and urban systems (Xu et al. 2017). For example, a science question may be posed to explore the relationships between social factors such as crime or school performance and the environmental characteristics of urban neighborhoods (e.g. with or without green spaces, weak or strong local economy, etc.).

For many questions, data such as those related to air quality or urban heat lack the spatial and temporal resolutions that are needed to better understand neighborhoods. The National Science Foundation (NSF)-funded Array of Things (AoT), a partnership of the University of Chicago, Argonne National Laboratory, and the City of Chicago, set out to use new sensor technologies and embedded (or “edge”) computation to create an experimental “instrument” comprising hundreds of intelligent sensing devices. The “nodes” were designed to measure Chicago’s urban environment, air quality, and activity such as traffic or pedestrian flow at neighborhood resolution. The project integrates established and emerging sensor technologies to measure several dozen urban environmental conditions, with remotely programmable machine learning capabilities to measure factors for which no sensors are available, such as the flow of pedestrians through a park or of bicycles through an intersection (Catlett et al. 2017). AoT has deployed more than 130 nodes in Chicago. Test deployments are under way in over a dozen cities around the globe.

To illustrate the nature of data from such measurement instruments, a single month of AoT data is in the range of 2 GB compressed, or about 10 GB uncompressed. This is several times larger than the entire Chicago crimes database from 2001 to present (18 years) comprising 7 million rows of crime records.

36.1.2 *CyberGIS*

During the past decade, cyberGIS has emerged as a new generation of GIS, comprising a seamless integration of advanced cyberinfrastructure, GIS, and spatial analysis and modeling capabilities while leading to widespread research advances and broad societal impacts (Anselin and Rey 2012; Wang and Goodchild 2019). CyberGIS has provided a solid foundation for breakthroughs in diverse science, technology, and application domains, and contributed to the innovation of cyberinfrastructure overall (Wright and Wang 2011). During the past several years, cyberGIS has grown as a vibrant interdisciplinary field while the cyberGIS community has achieved significant advances in tackling challenging environmental and geospatial problems (e.g. Hu et al. 2017, Liu et al. 2018).

36.1.3 Spatial Data Synthesis

Substantial progress has been made through a data science project funded by NSF to establish core spatial data synthesis capabilities (e.g. integrating geotagged data streams from social media, census data, and urban infrastructure registry data; Wang 2016). The core capabilities were developed and deployed using cyberGIS supercomputing and cloud architecture to support spatial big data analytics. These capabilities include: (a) vector-data processing; (b) raster processing; (c) integration of heterogeneous spatial data streams; (d) spatial data visualization; and (e) spatial data retrieval and storage.

Developing synthesis capabilities for varied data from a multitude of sources poses new challenges due to the dynamic nature of the data sources and the user-driven nature of data synthesis, which requires the process to be always-on and highly available, demanding innovative computational capabilities. The NSF project has demonstrated powerful synthesis capabilities for spatial data that were developed to overcome the challenge of handling urban big data by researchers who may not be fully trained to employ advanced cyberinfrastructure (Soliman et al. 2017). The developed capabilities benefit from integrated high-performance and cloud computing to overcome some key challenges such as providing on-demand access to virtual distributed processing clusters with elastic resource provision. The cyberGIS framework described in this chapter integrates these capabilities to enable urban discovery and innovation based on streaming data and related urban analytics (Fig. 36.1).

36.1.4 Cyberinfrastructure

The varied types of urban data and associated analytics introduce critical requirements for innovating cyberinfrastructure and cyberGIS. The varied types, sizes, and formats of data pose a need for varied modalities of computing. For example, fast-streaming data from numerous AoT nodes will need an elastic and integrated high-performance computing (HPC) and cloud infrastructure to manage and process the data in near-real time, while historical datasets like census and topographic datasets can be processed in an HPC batch environment.

Resourcing Open Geospatial Education and Research (ROGER) has been established using experiences gained from an NSF Major Research Instrumentation project for computation- and data-intensive processing and analysis of geospatial data. It provides hybrid computing modalities, including high-performance computing (HPC) batch, data-intensive computing based on Hadoop and Spark, and cloud computing, backed by a petascale common data store (Wang 2017). Moreover, ROGER offers a wide variety of geospatial software packages, forming the core computational environment of the cyberGIS framework.

Fig. 36.1 A cyberGIS framework for streaming analytics to enable urban discovery and innovation

36.2 Framework

36.2.1 Architecture

The framework is designed to integrate cyberGIS with urban sensing data for (1) facilitating user interactions with streaming urban analytics through an online environment; (2) providing cyberGIS capabilities to achieve scalable urban analytics; and (3) managing the execution of analytics and their interactions with measurements. These functions are accomplished by: (a) the speed layer; (b) the batch layer; and (c) the serving layer, which are coupled with scalable computing capabilities including a workload-aware data and computation management capability (Fig. 36.2).

Fig. 36.2 Architecture

The framework takes a holistic system approach to: a) varying workloads including low-latency read, fast update, and ad-hoc queries; and b) linear scalability (Yang et al. 2014). When data arrive (e.g. via Apache Kafka; Kreps et al. 2011), they are ingested separately by the speed layer and batch layer. The speed layer is required to specifically make data immediately available for both real-time queries and analysis that are critical for some application scenarios (e.g. emergency management). Hence, the speed layer focuses on the most recent data and streaming analytics and is built on event-processing frameworks (e.g. Apache Storm 2020). On the other hand, the batch layer is designed to handle the integration with large historical datasets, with computationally intensive tasks performed on it. Therefore, the speed layer is designed to sustain high-frequency writes and provide a real-time view into the data while the batch layer is developed for read intensive and analytical workloads. Both batch and speed layers are connected to end users by the serving layer, which accesses the results of previous operations through a diverse range of data stores, including in-memory databases (e.g. REDIS 2020), NoSQL databases (e.g. Cassandra; Apache Cassandra 2020) and big data storage systems (e.g. HDFS; Shvachko et al. 2010). The serving layer provides the interactive user interfaces described in the following section.

36.2.2 *User Environment*

The user environment is established by enhancing CyberGIS-Jupyter to achieve reproducible and scalable computational tasks (Yin et al. 2019). Through this online environment, a user may invoke a CyberGIS-Jupyter notebook with a suite of analysis tasks, perform the tasks that can be executed on cyberinfrastructure resources, and customize the notebook for specific reproducible investigations that can be shared with other users. The user may also be interested to access automated workflows using cyberGIS visual analytics with a particular focus on specifying workflow parameters, interpreting workflow results, assessing visualizations, and sharing results and visualizations with pertinent collaborators and communities. The user environment is designed for a large number of users to simultaneously conduct streaming analytics.

36.2.3 *Analytics*

Spatial references and spatiotemporal resolutions are fundamental characteristics of urban data. Conflating urban data for both analytics and visualization purposes necessitates transforming the data into common projection systems and spatiotemporal units. For example, map reprojection achieves this transformation by applying common map operations such as coordinate translation, framing, forward- and inverse-mapping, and interpolation or resampling. Our earlier work has developed techniques to do reprojection using HPC resources (Finn et al. 2019). Another core capability aims to provide friendly interfaces through which users can interact with urban sensing data and related analyses based on map layers, charts, and tables. We have developed a Web-based and Open Geospatial Consortium (OGC) compliant solution capable of providing interoperable access to heterogeneous spatiotemporal data through the support of several Web services such as WMS, WFS, WCS, and WPS, and state-of-the-art mapping libraries (e.g. leaflet, d3.js) to enhance the visual representation of urban data.

36.3 *Case Study*

36.3.1 *Study Area*

The Chicago Metropolitan Area (CMA) provides an ideal test case for the framework. The CMA covers approximately 28,000 km² with a population of over 10 million people and is the third largest economy in the USA. It is at the crossroads of the rail, road, and air transportation infrastructures in North America. Extreme heat has already had detrimental effects on the Chicago urban population and by extension on the regional and US economy (Karl and Knight 1997). Elevated night temperatures

over multiple days, exacerbated by urban heat-island (UHI) effects, are implicated in human health impacts (Semenza et al. 1996) as is neighborhood economic vitality (Browning et al. 2012). Given that the average summer time temperatures in the midwest are expected to increase by 3–6°F in the next 25–50 years (Wuebbles and Hayhoe 2004), the framework is crucially important for examining the urban microclimate at finer spatial and temporal granularities and directly coupling data with urban heat-related analytics to enhance our understanding of related issues in urban environments.

36.3.2 *AoT Data*

This case study uses data from AoT, coupled with computationally intensive spatial analyses, to explore a “smart city” vision that can make urban planning and policy adjustment possible on time scales of days or weeks rather than more traditional multi-year time windows. AoT nodes include both sensors (including cameras and a microphone) and embedded (“edge”) computing resources, enabling remotely programmed machine learning to analyze data in situ. Currently, AoT nodes measure temperature, relative humidity, barometric pressure, light, vibration, carbon monoxide, nitrogen dioxide, sulfur dioxide, ozone, ambient sound pressure, and particulate matter. Nodes analyze images at 30 s intervals to count pedestrians and vehicles, transmitting these numbers along with readings from the sensors every 30 s to a central data repository. A map for the locations and types of sensors of AoT in Chicago is available at the project website (Catlett 2020).

Data are open and free, available for bulk download and through a real-time API. With respect to climate, AoT data have been used as part of a project funded by the Department of Energy’s Exascale Computing Program for calibration and parameterization of fine-resolution weather models (Jain et al. 2018). Figure 36.3 shows a general workflow of how AoT measurement data can be translated into useful smart city applications.

Initiated with experimental nodes deployed in 2016, the project is implemented using Argonne’s Waggle hardware/software platform (Beckman et al. 2016). As of late 2019, the 130 nodes in Chicago and over 60 nodes being deployed in partner cities represent the fourth generation of the platform (Fig. 36.4). Recent funding from NSF for the SAGE (Beckman et al. 2019) project aims to move to the fifth generation with substantially increased edge computing power, new sensors, and with experimental deployments in multiple observatories including the NSF’s National Ecological Observation Network (NEON; Keller et al. 2008) and High-Performance Wireless Research and Education Network (HPWREN; Hansen et al. 2002).

The spatial distribution of nodes is illustrated in Fig. 36.5 showing the municipality of Chicago (589 km²). The density of deployment varies from every block along several streets in the downtown area to more sparse distribution in residential areas. Locations are selected in cooperation with science teams, city officials, and community groups. An analysis by the University of Chicago’s Center for Spatial Data

Fig. 36.3 A workflow from AoT sensor data to smart city applications

Science showed that 80% of Chicago’s population lives within 2 km of an AoT node and 42% live within 1 km. While traditional sources for measurements such as air quality are available, for instance, there are fewer than 10 Environmental Protection Agency sites in the Chicago municipality, and most only measure 1 or 2 pollutants. AoT is an experimental instrument with respect to the technologies, and similarly, the density of nodes is aimed at optimal placement for various research or policy questions and their associated measurement requirements.

Another issue worth noting is that different generations (or “models”) of AoT nodes (three models are in operation as of late 2019) vary with respect to sensors and capabilities. Only a few of the early nodes measured particulate matter, but all of the fourth-generation nodes are equipped with particulate-matter sensors. Similarly, the microphone in early nodes measured aggregate sound pressure, while new nodes provide measurements for ten octaves. As shown in Fig. 36.5, several nodes may be not working at a specific time, and during software updates and experimental software deployments, many nodes may be unavailable for periods of time. Figure 36.5 indicates that orange nodes are active while blue nodes denote inactive nodes. In reality, the number of nodes that are available may not equal the total number of AoT nodes deployed. The Waggle platform provides resiliency to communication outages, caching all measurements until the data have been transmitted to the central servers and acknowledged as received. Thus, in periods where nodes appear unavailable, the data for that period of time may become available later. Such factors are less visible in the bulk downloads than in using the real-time API.

Fig. 36.4 The deployment information for AoT nodes in Chicago

36.3.3 *CyberGIS-Jupyter*

CyberGIS-Jupyter serves as the foundational engine for capturing and analyzing real-time streaming AoT data. CyberGIS-Jupyter is equipped with cyberGIS libraries scaling to both high-performance computing and cloud resources (Padmanabhan et al. 2019) and hence can support computationally intensive spatial analysis for users not only to capture the real-time, high-frequency data, but also to conduct urban analytics with AoT data. In this case study, real-time location-based AoT data can be used for understanding Chicago’s heat environment. For example, temperature patterns can be derived based on AoT data as shown in Fig. 36.6. For all AoT nodes with temperature sensors, the temporal trend on September 30, 2019, is visualized

Fig. 36.5 The spatial distribution of AoT nodes in Chicago

Fig. 36.6 The temperature curve derived from AoT data

on CyberGIS-Jupyter with different colors indicating different nodes. The AoT data have high frequency with the temperature data recorded every 26 s on average.

Due to the huge amount of data stored, with 2–3 GB of data captured from AoT every week, AoT's API cache only keeps 3–4 weeks of fresh data. In order to get the data back in 2017, for example, we need to download the whole dataset from the AoT bulk download website (or a subset of months of interest) and start our data processing from there.

Using the AoT streaming API as our data access option, spatial analysis of the temperature data and the geolocation of the AoT nodes can be conducted based on CyberGIS-Jupyter. Considering the need for identifying dense concentrations of high-temperature areas, Fig. 36.7 shows temperature patterns within one week

Fig. 36.7 Temperature maps in Chicago based on AoT sensors using a spatial interpolation algorithm. Temperature measurements are in degrees Celsius. From the top left to last map in the last row, each map represents the temperature distribution captured at 6am on September 30th, October 1st, October 2nd, October 3rd, October 4th, October 5th, and October 6th, respectively

in 2019. One can distinguish some hot spots from these heat maps. A workflow has been developed to capture the temperature data of the Chicago area based on CyberGIS-Jupyter from all of the available temperature sensors in Chicago using AoT's API at 6am in the morning from September 30th to October 6th on a daily basis. Combining with the geolocation of the sensors, the dynamic maps shown in Fig. 36.7 were generated using an inverse-distance weighted algorithm for spatial interpolation (Wang and Armstrong 2003). As shown in Fig. 36.7, throughout the week, the temperature in northwest Chicago, near Jefferson Park and North Park, and oftentimes in southeast and downtown Chicago, was higher than the average temperature in other areas. It is straightforward to understand that the temperature in downtown and southeast Chicago was higher due to human activities, since those areas have high population density. We investigated the sensor located in northwest Chicago (latitude 41.97 N, longitude 87.76 W, Fig. 36.8) and found it is installed near an underground transformer and some external air conditioners, which seem to be the heat sources. In addition, the density of sensors in northwest Chicago is lower than in other urban areas as shown in Fig. 36.5, leading to the skewed spatial interpolation result near Jefferson Park. The workflow for this analysis and associated data is represented as a CyberGIS-Jupyter notebook that can be shared with other users for reproducing the same results. The notebook can be adapted to accommodate data from different AoT nodes and time ranges and support different parameter values of the analysis (e.g. the number of the nearest neighbors in the spatial interpolation algorithm).

Fig. 36.8 A Google Steetview image of the AoT node located at latitude 41.97 N, longitude 87.76 W near Jefferson Park on Chicago's Northwest side

Similar to the example for analyzing temperature patterns demonstrated above, CyberGIS-Jupyter allows users to select other measurements from specific AoT nodes and specify temporal ranges to retrieve corresponding data streams for conducting computationally intensive analytics based on advanced cyberinfrastructure. Each workflow for combining AoT and other related data with specific analytics can be represented as CyberGIS-Jupyter notebooks that can record the provenance of computational steps in the workflow. Many users can simultaneously compose and run their notebooks on CyberGIS-Jupyter without noticing that their notebooks are executed on advanced cyberinfrastructure. While it is often challenging to “freeze” dynamic data streams to experiment with various analytical scenarios, CyberGIS-Jupyter notebooks can be shared among users to enable collaborative development and computational reproducibility of urban analytics with dynamic data (<https://go.illinois.edu/CyberGIS-UrbanInformatics>).

36.4 Concluding Discussion

Large cities like Chicago increasingly engage data-driven methods for urban planning and management, including for example land-use and transportation modeling, economic forecasts, and environmental monitoring. However, the ability to continuously monitor and alter policies of urban planning and management in a responsive manner is hampered by the difficulty of harnessing high-quality, spatially explicit, and temporally continuous data. In the USA, for example, large-scale land-use planning requires fine-resolution land cover data that is only available every five years from the National Land Cover Database. Similarly, socioeconomic models depend heavily on a census that is conducted on a ten-year interval. Due to these difficulties, though cities incorporate data-driven approaches in their planning processes, it is still challenging to implement the “smart city” vision based on fast data streams. A key barrier is the inability to make timely interventions and management decisions when environmental, social, or economic processes take place dynamically.

To address these challenges, this research has demonstrated that users can conduct computationally intensive streaming analytics using CyberGIS-Jupyter and AoT data without having to possess in-depth technical knowledge of cyberGIS or cyberinfrastructure. AoT data can be harnessed through CyberGIS-Jupyter to help users to monitor urban heat and other key indicators of urban dynamics. The cyberGIS framework described in this chapter is able to resolve the volume and velocity of urban big data through the support of advanced cyberinfrastructure; meet the computing requirements for processing, analyzing, and visualizing these datasets; and support concurrent online access to CyberGIS-Jupyter notebooks for collaborative development and computational reproducibility of urban analytics.

Regarding future research in urban informatics involving fast data streams, it is both important and challenging to achieve reproducible urban analytics. Without computationally reproducible urban analytics, it would be difficult, if not entirely impossible, to convince decision makers and practitioners to adopt such analytics in any real-world settings. Fast data streams produce data continuously and pose significant challenges that must be addressed through novel algorithms that treat spatial and temporal characteristics synergistically. Furthermore, exciting and important cyberGIS research is urgently needed to better understand and support computational reproducibility of urban analytics, which requires holistic approaches to optimizing access and management of cyberinfrastructure resources, trading off performance and uncertainty of spatial and spatiotemporal algorithms, and generalizing standards and specifications for the building blocks of urban analytics.

Acknowledgements This chapter and associated materials are based in part upon work supported by the National Science Foundation (NSF) under grant numbers 1443080, 1532133, 1743184, 1833225 and 1935984. The work used the ROGER supercomputer, which was supported by NSF under grant number 1429699. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

- Anselin L, Rey SJ (2012) Spatial econometrics in an age of CyberGIScience. *Int J Geogr Inform Sci* 26(12):2211–2226
- Apache Cassandra (2020) <https://cassandra.apache.org/>. Accessed 30 Jan 2020
- Apache Storm (2020) <https://storm.apache.org/>. Accessed 30 Jan 2020
- Armstrong MP, Wang S, Zhang Z (2019) The Internet of Things and fast data streams: prospects for geospatial data science in emerging information ecosystems. *Cartography Geogr Inform Sci* 46(1):39–56
- Beckman P, Catlett C, Altintas I, Collis S, Kelly E (2019) Mid-Scale RI-1: SAGE: a software-defined sensor network. Award 1935984, national science foundation
- Beckman P, Sankaran R, Catlett C, Ferrier N, Jacob R, Papka M (2016) Waggle: an open sensor platform for edge computing. *IEEE SENSORS*
- Browning CR, Cagney KA, Iveniuk J (2012) Neighborhood stressors and cardiovascular health: crime and C-reactive protein in Dallas, USA. *Soc Sci Med* 75(7):1271–1279
- Catlett CE (2020) <https://arrayofthings.github.io/>. Accessed January 30, 2020
- Catlett CE, Beckman PH, Sankaran R, Galvin KK (2017) Array of things: a scientific research instrument in the public way: platform design and early lessons learned. In: *Proceedings of the 2nd international workshop on science of smart city operations and platforms engineering*. ACM
- Finn MP, Liu Y, Mattli DM, Behzad B, Yamamoto KH, Shook E, Padmanabhan A, Stramel M, Wang S (2019) High-performance small-scale raster map projection empowered by cyberinfrastructure. In: Wang S, Goodchild MF (eds) *CyberGIS for geospatial discovery and innovation*. Springer, Dordrecht, pp 171–188

- Hansen T, Yalamanchili P, Braun HW (2002) Wireless measurement and analysis on HPWREN. In: Proceedings of passive and active measurement workshop, fort collins, Co
- Hu H, Lin T, Wang S, Rodriguez LF (2017) A cyberGIS approach to uncertainty and sensitivity analysis in biomass supply chain optimization. *Appl Energy* 203:26–40
- Huijboom N, Van den Broek T (2011) Open data: an international comparison of strategies. *Eur J EPractice* 12(1):4–16
- Jain R, Luo X, Sever G, Hong T, Catlett C (2018) Representation and evolution of urban weather boundary conditions in downtown Chicago. *J Build Perform Simul* 1–14
- Karl TR, Knight RW (1997) The 1995 Chicago heat wave: how likely is a recurrence? *Bull Am Meteor Soc* 78(6):1107–1120
- Keller M, Schimel DS, Hargrove WW, Hoffman FM (2008) A continental strategy for the national ecological observatory network. *Front Ecol Environ* 6(5):282–284
- Kreps J, Narkhede N, Rao J (2011) Kafka: a distributed messaging system for log processing. In: Proceedings of NetDB
- Liu YY, Maidment DR, Tarboton DG, Zheng X, Wang S (2018) A CyberGIS integration and computation framework for high-resolution continental-scale flood inundation mapping. *J Am Water Resour Assoc* 54(4):770–784
- Marka E (2019) Greenest region compact. Metropolitan mayors caucus. <https://mayorscaucus.org/initiatives/environment/rec/>. Accessed 29 Jan 2020
- Miller H, Akar G, Clifton K, Gopalakrishnan S, Tufte K (2019) Exploring a research network of urban sustainability observatories via data-enabled university-community partnership. NSF workshop report
- Padmanabhan A, Yin D, Lyu F, Wang S (2019) Bridging local cyberinfrastructure and XSEDE with CyberGIS-Jupyter. In: Proceedings of the practice and experience in advanced research computing on rise of the machines (learning). ACM
- REDIS (2020) <https://redis.io>. Accessed 30 Jan 2020
- Semenza JC, Rubin CH, Falter KH, Selanikio JD, Flanders WD, Howe HL, Wilhelm JL (1996) Heat-related deaths during the July 1995 heat wave in Chicago. *N Engl J Med* 335(2):84–90
- Shvachko K, Kuang H, Radia S, Chansler R (2010) The hadoop distributed file system. In: 26th IEEE symposium on mass storage systems and technologies (MSST). IEEE
- Soliman A, Soltani K, Yin J, Padmanabhan A, Wang S (2017) Social sensing of urban land use based on analysis of Twitter users' mobility patterns. *PLoS ONE* 12(7):e0181657
- UN-Habitat (2011) Cities and climate change: global report on human settlements 2011. https://unhabitat.org/?mbt_book=cities-and-climate-change-global-report-on-human-settlements-2011. Accessed 29 Jan 2020
- Wang S (2010) A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Ann Assoc Am Geogr* 100(3):535–557
- Wang S (2016) CyberGIS and spatial data science. *GeoJournal* 81(6):965–968
- Wang S (2017) CyberGIS. In: Richardson D, Castree N, Goodchild MF et al (eds) The international encyclopedia of geography. Wiley
- Wang S, Armstrong MP (2003) A quadtree approach to domain decomposition for spatial interpolation in grid computing environments. *Parallel Comput* 29(10):1481–1504
- Wang S, Goodchild MF (eds) (2019) CyberGIS for geospatial discovery and innovation. Springer
- Wright DJ, Wang S (2011) The emergence of spatial cyberinfrastructure. *Proc Natl Acad Sci* 108(14):5488–5491
- Wuebbles DJ, Hayhoe K (2004) Climate change projections for the United States Midwest. *Mitig Adapt Strat Glob Change* 9(4):335–363
- Xu L, Kwan MP, McLafferty S, Wang S (2017) Predicting demand for 311 non-emergency municipal services: an adaptive space-time kernel approach. *Appl Geogr* 89:133–141

Yang F, Tschetter E, Léauté X, Ray N, Merlino G, Ganguli D (2014) Druid: a real-time analytical data store. In: Proceedings of the 2014 ACM SIGMOD international conference on management of data. ACM

Yin D, Liu Y, Hu H, Terstriep J, Hong X, Padmanabhan A, Wang S (2019) CyberGIS-Jupyter for reproducible and scalable geospatial analytics. *Concurrency Comput: Pract Exp* 31(11):e5040

Shaowen Wang is a Professor and Head of the Department of Geography and Geographic Information Science, and founding director of the CyberGIS Center for Advanced Digital and Spatial Studies at the University of Illinois at Urbana-Champaign. His research interests primarily include geographic information science and systems (GIS) and cyberGIS.

Fangzheng Lyu is a Ph.D. student in the Department of Geography and Geographic Information Science at the University of Illinois at Urbana-Champaign. He received his BE from the University of Hong Kong in computing engineering. His research interests focus on GIS, urban multi-sensing, remote sensing, cyberGIS, and big data analysis.

Shaohua Wang is a postdoctoral fellow in the Department of Geography and Geographic Information Science, and the CyberGIS Center for Advanced Digital and Spatial Studies at the University of Illinois at Urbana-Champaign. His major research interests include GIS, cyberGIS, spatial optimization, and spatiotemporal big data analytics.

Charlie E. Catlett is a Senior Research Scientist at the University of Illinois Discovery Partners Institute. He is also a Senior Computer Scientist at Argonne National Laboratory and Visiting Senior Scientist at the University of Chicago. His research focuses on intelligent sensing systems using edge computation to understand urban dynamics.

Anand Padmanabhan is a Research Associate Professor in the Department of Geography and Geographic Information Science at the University of Illinois at Urbana-Champaign. His research interests focus on middleware for advanced cyberinfrastructure; cyberGIS and geospatial problem solving; geospatial big data analysis; and interdisciplinary education.

Kiumars Soltani is a Software Engineer in Zillow Group. He received his Ph.D. from the University of Illinois at Urbana-Champaign. His research is focused on spatially aware distributed systems and algorithms for ingesting, processing, and visualizing large and high-throughput geosensor data.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.