JING GAO
Geospatial Data Science
Department of Geography
Data Science Institute
University of Delaware
Email: jinggao@udel.edu

Jing Gao is an Assistant Professor of Geospatial Data Science at the University of Delaware, affiliated with the Department of Geography and the Data Science Institute. Her research investigates large-scale human-environment interactions, especially the relationship between global urban land use, population, and climate change. Trained in Geography and Computer Science, she approaches interdisciplinary scientific inquiries by integrating diverse data, quantitative and computational methods from spatial statistics, machine learning, big data mining, geo-visualization, and remote sensing, with narrative-based scenario analyses of societal development. Gao's research is generating new insights and datasets on global, long-term, spatially-explicit changes in urbanization and population characteristics, extending the SSP-RCP scenario framework used by the IPCC and the research community for understanding global environmental change impacts, spearheading creative data-science practices in long-term spatially-explicit modeling of socioeconomic processes, and developing new methods for evaluating the uncertainty and the success of such practice.

Prior to joining University of Delaware, Gao was a Research Scientist at the U.S. National Center for Atmospheric Research, following a Postdoctoral Fellowship at the U.S. National Center for Supercomputing Applications. She holds a BS in Remote Sensing from Beijing Normal University, an MS in GIS and Cartography and a PhD in Geography with a Minor in Machine Learning from the University of Wisconsin–Madison.

Reflections on Geospatial Data Science Applications in Global Human-Environment Interaction Studies

Global environmental and climatic change is a primary challenge that societies across the world will face over the 21st century. Understanding potential societal impacts of anticipated environmental changes requires high-quality long-term spatial projections of both physical environmental stressors and socioeconomic conditions of communities at risk. Such integrated studies conventionally have more often focused on local to regional scales and short- to mid-term futures, while the big picture of global patterns and long-term trends are needed for national and international assessments and policy making. This convention has roots in the parsimonious availability of large-scale time-series observational data. With recent unprecedented increase in geospatial data availability and development in computational technology, opportunities for new innovations as well as challenges to realize desired advancements have attracted much enthusiasm. As a formally trained geospatial data scientist studying human dimensions of global environmental change, below I briefly share a few reflections from my own experience on how geospatial data science might better serve the needs of large-scale human-environment interaction studies.

The recent advancements in data and computation accentuate the importance of some classic challenges of geospatial data science, with some debates dating back to when the field was more commonly referred to as a part of GIS. For example, scale plays determining roles in the usability of data and analytical methods. As spatially-explicit human-environment interaction studies move towards increasingly larger scales (e.g. global, long-term), data structures and representations often must change along with changes in data coverages and resolutions, which could reduce or diminish the applicability of existing analytical and modeling methods. While new analytical, computational, and modeling methods should be developed, an alternative useful avenue to pursue is the effective and credible fusion of existing domain knowledge and models that have been developed using earlier generations of smaller-scale data and methods. In a way, this means providing ready-to-use tools for multi-scale meta analyses. Meta analyses often attempt to synthesize diverse and patchy knowledge, integrating both quantitative and qualitative information. These operations can introduce much uncertainty into a study's conclusions that are difficult to quantify, which leads to my next point.

Improved uncertainty analysis is much needed for better global human-environment interaction studies. Conventionally, efforts in this area have focused on describing the amount and the characteristics of errors and uncertainties in spatial analysis and modeling. However, from applications' point of view, the more useful types of uncertainty analysis tools are the ones that can inform iterative improvements to the analysis or model of interest. Further, because imperfections are innate to any data, and assumptions and simplifications are necessary for any analysis and modeling, certain amount of error and uncertainty is bound to exist in the results of any study. Therefore, another type of highly useful uncertainty-related tools are the ones that can reveal how errors and uncertainties in analytical results may affect the conclusions of integrated studies. Investigations of uncertainty and error propagation are particularly important in the data era for global change research, as large-scale government-organized observations with well-documented metadata constitute only a small fraction of all geospatial data available today, and new data from non-conventional sources often come with more uncertainties. Without a clear understanding of the implications of these uncertainties, the usability of the new data for supporting important policy and decision making is limited. Ultimately, studies about societal impacts of global, long-term environmental change are studies about uncertainties in future societies and future environments; hence, we must be able to distinguish uncertainties caused by analytical artifacts from uncertainties in fundamental social and environmental processes.

The multidisciplinary nature of global human-environment interaction studies calls for truly interdisciplinary researchers, willing to integrate rather than further partition different fields. The integration applies not only to contents and knowledge but also to people from different professional backgrounds. Although, intuitively speaking, geospatial data driven analysis and modeling are useful for global change studies, geospatial data scientists form a relatively small part of the research community studying global human-environment interactions. Many topics in the field have long been studied by researchers from other backgrounds. To contribute efficiently, it is useful to understand the legacy of existing work and be inclusive of visions and practices from different perspectives. At the Spatial Data Science Symposium, I hope to emerge in a multidisciplinary environment and reflect further with the diverse group of participants on bridging advancements in geospatial methodological research and the needs of large-scale human-environment interaction studies.