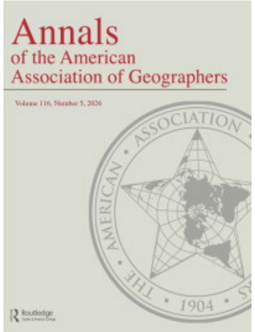

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: www.tandfonline.com/journals/raag21

Movement Research in the Era of Big Data, AI, and Open Science

Somayeh Dodge, Robert Weibel, Christophe Claramunt, Gavin McArdle, Zhiyong Zhou, Yanan Xin, Anita Graser & Tumasch Reichenbacher

To cite this article: Somayeh Dodge, Robert Weibel, Christophe Claramunt, Gavin McArdle, Zhiyong Zhou, Yanan Xin, Anita Graser & Tumasch Reichenbacher (08 Jun 2026): Movement Research in the Era of Big Data, AI, and Open Science, Annals of the American Association of Geographers, DOI: 10.1080/24694452.2026.2671994

To link to this article: https://doi.org/10.1080/24694452.2026.2671994

Published online: 08 Jun 2026.

ANNALS OF THE AMERICAN ASSOCIATION OF GEOGRAPHERS
https://doi.org/10.1080/24694452.2026.2671994

Movement Research in the Era of Big Data, AI, and Open Science

Somayeh Dodge,[a] Robert Weibel,[b] Christophe Claramunt,[c] Gavin McArdle,[d] Zhiyong Zhou,[e] Yanan Xin,[f] Anita Graser,[g] and Tumasch Reichenbacher[h]

[a]*Department of Geography, University of California, Santa Barbara, USA;* [b]*Professor Emeritus of Geographic Information Science, University of Zurich, Switzerland;* [c]*Computer Science, Naval Academy Research Institute Lanvéoc, France;* [d]*School of Computer Science, University College Dublin, Ireland;* [e]*Department of Geography, University of Wisconsin–Madison, USA;* [f]*Department of Transport and Planning, Delft University of Technology, The Netherlands;* [g]*Austrian Institute of Technology, Vienna, Austria;* [h]*Geographic Information Science Center, University of Zurich, Switzerland*

The increasing availability of movement data and rapid advancements in artificial intelligence have opened new frontiers in movement pattern analysis with implications for a wide range of domains such as urban planning, transportation management, public health, pandemic management, disaster response, and wildlife conservation. The sensitive nature and heterogeneity of novel movement data, however, create technical and methodological challenges around data governance and open movement analytics. This article explores these challenges and identifies key opportunities for the geography discipline to drive analytical advances in human mobility research and open science. In this vision, the geographic perspective is emphasized in future developments of artificial intelligence for movement research for its ability to enrich mobility analyses with spatial and temporal context. Strategies for effective governance of large-scale mobility data are discussed, with special attention to fostering open science and reproducible research practices to ensure transparency and transferability in this rapidly evolving field. *Key Words: big data, GeoAI, location privacy, movement analytics, open science.*

Movement research[1] is a multidisciplinary field bridging geography, computer science, data science, human mobility, and movement ecology (Miller et al. 2019; Mokbel et al. 2024). It aims to understand how individuals interact with their environments and others (Dodge 2021), through five core processes (National Academies of Sciences, Engineering, and Medicine 2019): (1) developing methods to represent, investigate, and describe movement behaviors and patterns; (2) employing reasoning to contextualize patterns; (3) using theory, observation, and evidence to model and predict behavioral responses; (4) disseminating research outcomes; and (5) informing application domains such as urban planning, emergency and pandemic preparedness, and conservation.

In pursuit of these goals, researchers have advanced spatiotemporal methodologies[2] to analyze individual trajectories and aggregate movement data. Implementation of these analytics has accelerated across diverse domains, including health care science (Oliver et al. 2020; Linardon et al. 2025), healthy aging (Röcke et al. 2023), sports and fitness sciences (Rana and Mittal 2021), urban resilience (Haraguchi et al. 2022), sustainable transportation (Miller 2020), urban analytics (Su, Newsham, and Dodge 2024; W. Xu 2022), maritime research (Del Mondo et al. 2021), and wildlife ecology (Nathan et al. 2022). These approaches facilitate the extraction of meaningful patterns and the identification of drivers shaping movement across spatial and temporal dimensions.

Two advancements have driven movement research forward (Mokbel et al. 2024): First, advanced sensing technologies provide unprecedented access to high-resolution, large-scale movement data. These data sets offer detailed information about movement of individuals in human,

ARTICLE HISTORY
Initial submission, September 2025; revised submissions, January and March 2026; final acceptance, March 2026

CORRESPONDING AUTHOR Somayeh Dodge sdodge@ucsb.edu

technological, and ecological systems, recorded in diverse formats, scales, and modalities, generated at volumes that far exceed what traditional mobility studies could accommodate. Second, rapid developments in geographic artificial intelligence (GeoAI) are redefining pattern recognition and predictive modeling within movement research. GeoAI refers to AI-driven approaches—including machine learning (ML), deep learning (DL), large language models (LLMs), and generative AI (GenAI) techniques—that integrate geospatial data and geographic principles about space and place to solve data-intensive geographic problems (W. Li et al. 2024). GeoAI along with the emerging and foundational models have substantially augmented the analytical power of movement research for a broad spectrum of tasks, including extracting knowledge from vast unstructured digital information, movement pattern discovery, behavior analysis, and movement response predictions (Mai et al. 2024).

Given the impact of big data, AI, and the growing urgency for open and reproducible science, this article presents a forward-looking vision for the discipline of geography, particularly GIScience, to guide future advancements in movement research. We establish this vision by examining contemporary challenges and opportunities at the intersection of movement data, AI innovations, and open science. Following a conceptual overview, we trace the field's progress through the data, open science, and AI revolutions, framing contemporary challenges as catalysts for methodological innovation. We then outline a future research agenda. Although we acknowledge the broad relevance to movement ecology (Miller et al. 2019; Demšar et al. 2021), our primary focus remains on frameworks specific to human movement.

Movement: A Multidimensional and Multiscale Process

Movement is a multidimensional process unfolding across a space–time–context continuum, and is observed through two different data collection perspectives (Laube 2014; Dodge 2021). The Eulerian perspective is a place- or space-based measure recording movement at fixed locations or aggregated within spatial units, such as public transit smartcard logs, call detail records (CDRs), or visit counts at points of interest (POIs) or census tracts. In contrast, the Lagrangian perspective follows an entity-based principle, capturing detailed individual trajectories over time via Global Positioning System (GPS) tracking or location-based apps. Although data generated from these two perspectives can occasionally be transformed (e.g., aggregating trajectories into flow counts or reconstructing paths from CDRs or check-in records), they require distinct storage formats and analytical methods.

Movement patterns emerge at the intersection of three primary scales: spatial (ranging from micro-local to macro-global), temporal (short- to long-term), and social (from individual to collective, where "collective" could mean a household, a community, or an entire population). As illustrated in Figure 1, movement observations capture a spectrum of processes: Fine-grained, individual routing decisions at the microscale are typically captured via the Lagrangian perspective, whereas coarser meso- and macrolevel behaviors (e.g., global migration patterns) emerge through aggregate data analyzed via the Eulerian lens. Ultimately, the selected measurement perspective and data representation models inherently shape the understanding of these processes.

At the spatial scale, movement is represented through trajectories, networks, or thematic maps (Dodge and Noi 2021). Trajectories capture fine-scale, individual-level paths as continuous routes or semantic stop-move sequences, where stops denote activity places and moves describe the transitions between them (Parent et al. 2013). Trajectories provide precise information about where, when, and how movement occurs. In contrast, networks (e.g., mobility, transportation, or disease dispersion networks) model movement between locations as origin–destination (OD) flows (Nanni et al. 2020; Chin et al. 2024). Networks provide information of where individuals are and when and what locations they visit, without depicting the exact routes. Thematic maps use polygons or tessellations to map movement intensity (e.g., visitation counts, in-flows, out-flows) within spatial units (Zhong and Bian 2023). These aggregate representations reveal population-level regularities and where movement is concentrated but lack the individual-scale how found in trajectory data.

At the temporal scale, movement ranges from high-frequency, short-term microsteps (e.g., routing decisions) to long-term processes (e.g., daily

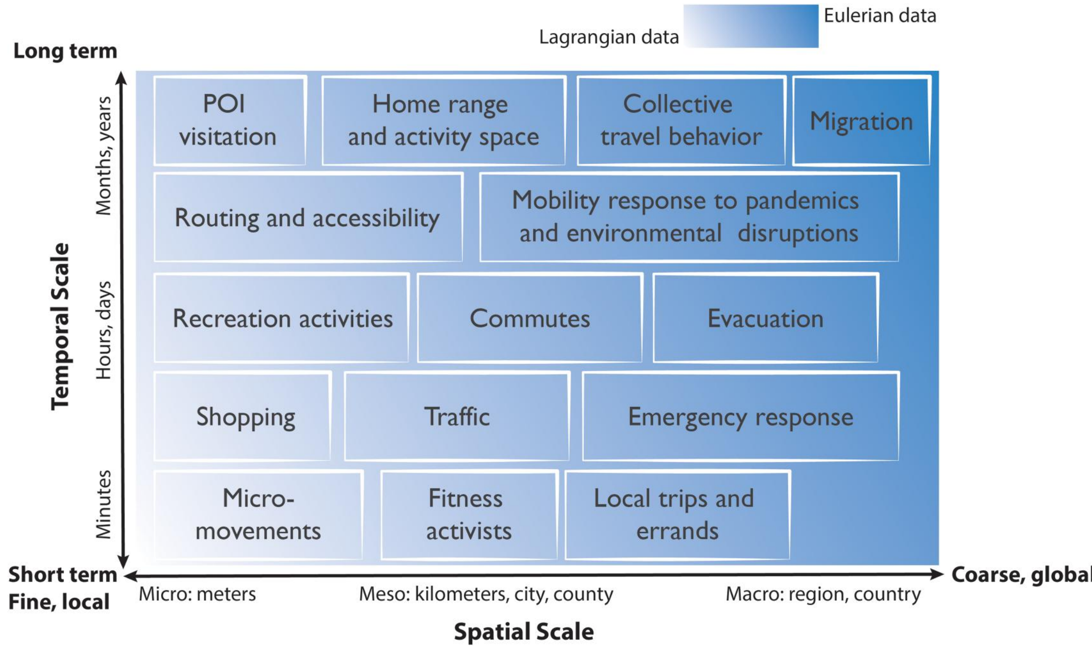

Figure 1. Movement processes occur across spatial and temporal scales through movement of individuals or collectives (e.g., dyads, groups, population). These processes can be observed through Lagrangian and Eulerian perspectives. *Note:* POI = point of interest.

commutes, migration). Movement can be seen as a sequence of local choices made at specific times, such as destination choice, mode of transport, departure time, and specific routes. Fine-grained timestamps capture immediate behavioral shifts, whereas coarser intervals reveal stable trends in macro-urban mobility.

At the social scale individual microsteps aggregate into interactions between individuals (e.g., dyads, social groups), to collective patterns of households and populations in cities or regions. Whereas microsteps occur at high frequency, population-level movement patterns typically evolve over longer durations. These patterns are further shaped by environmental and structural factors, as well as interpersonal interactions (Dodge and Nelson 2023). These factors are often integrated as contextual variables (attributes or thematic layers) to investigate the underlying drivers or motivations behind movement patterns (the *why questions*).

Current Progress and Pressing Challenges

Until the advent of affordable GPS trackers in the early 2000s, and the subsequent proliferation of smartphones, movement observations were primarily collected via travel surveys and radio telemetry, resulting in trajectory data sets that were relatively small, context-light, and characterized by coarse temporal resolution. Since the late 2000s, technological and methodological advances, occurring alongside the broader data and AI revolutions, have facilitated new insights and expanded the utility of movement analytics. This section reviews existing progress and challenges as they relate to data and methods, with a central focus on the AI developments and promoting open and reproducible analytics.

Dimensions of Heterogeneity in Movement Data

Although expanding access to movement data creates new research opportunities (e.g., study of detailed movement patterns at the population level), it introduces significant challenges in managing multidimensional data sets collected across diverse modalities (e.g., through cell phone apps, CDRs, or in-situ counters) and scales. Contemporary data sets frequently integrate Lagrangian (object-based) and Eulerian (field-based) perspectives. For example, by annotating individual trajectories with environmental remote sensing or population density data, researchers can enrich high-volume movement data with complex geographic and environmental contexts. This allows for an analysis that considers both the path of the individual and the characteristics of the space through which they move. Such heterogeneity, however, complicates data representation, integration, and privacy preservation. For example, movement data could be represented as discrete points at locations of interest, or as aggregated flows within tessellations, networks, or polygonal maps. Geographers have long addressed heterogeneity as it pertains to spatial data structure (vector, raster, network), scale, and granularity (Goodchild 2004). Movement data, however, introduce unique dimensions of heterogeneity, including different collection modes and perspectives (Lagrangian vs. Eulerian), structural forms (e.g., OD networks vs. trajectories and time series), and social granularity (individual vs. collectives). Consequently, traditional geographic frameworks must evolve to accommodate these unique dimensions of movement-specific heterogeneity. Additionally, individual-level data pose particular challenges to privacy preservation. For example, data that are captured in trajectory forms must be preprocessed to mask identifiable information (e.g., significant home work locations) or be aggregated in different forms and granularity.

Moreover, heterogeneity in data scale creates additional challenges when passing from one scale to another, whether moving from fine-grained local observations of short-term events at microscale (e.g., stop-moves) to broader regional and long-term patterns (e.g., commuting, migration) at meso- and macroscales, or vice versa. These transitions are rarely straightforward, as the spatial relationships and contextual drivers observed at one resolution might dissipate or transform at another due to scale-dependent processes and aggregation effects. Consequently, methodological complexities arise when performing cross-scale analysis on data sets containing nested patterns (Ahearn and Dodge 2018) across multiple spatiotemporal resolutions (e.g., from micromovements and short-term trips to daily commutes) or when integrating data captured from disparate Eulerian and Lagrangian perspectives. These challenges fundamentally limit the generalizability of movement analytics when results are extrapolated across varying spatial or temporal scales.

Geographic research has managed data heterogeneity through developing established ontologies, which have facilitated the development of robust geographic information systems (GIS) techniques for map generalization, data integration, and cross-scale analysis (Brassel and Weibel 1988; Fonseca et al. 2002). Conversely, movement research has remained less guided by ontology-driven frameworks, resulting in fragmented solutions for addressing data granularity. Hornsby and Egenhofer (2002) proposed the "geospatial lifeline" following Hägerstrand's (1970) time geography as a framework to model individual-level trajectory data across different spatial and temporal scales. Yan et al. (2008) provided a foundational trajectory ontology by combining spatial and temporal ontology components (stops, moves, time, and geospace). Parent et al. (2013) extended this to the concept of semantic trajectories by adding contextual information and activity types to the location of stops. Although these developments have been valuable in spatiotemporal query and moving object analysis, there remains a critical void in universal frameworks capable of integrating Lagrangian and Eulerian perspectives across different aggregation structures and scales.

Uncertainty in Contemporary Movement Data

Following the COVID-19 pandemic, providers such as Cuebiq, Mapbox, and SafeGraph (now Advan) expanded access to aggregated mobility data for research through "Data for Good" initiatives. These data sets, however, originate from opaque, fluctuating sets of smartphone applications, complicating transparency regarding the underlying user base, tracking change over time. Although these resources enabled rapid responses to crises, they present significant challenges for long-term and valid inquiry. At the same time, many mobile-driven data sets capture on average about 7.5 percent of the U.S. population (Z. Li et al. 2024), with coverage varying significantly across space, time, demographics, environment, transport mode, and activity type (Heidger, Nelson, and Willet 2025). For example, Figure 2 illustrates how signal persistence is often confined to high-traffic corridors, leaving secondary roads, pedestrians, and remote areas with unreliable sparse observations. Moreover, not all individuals carry smartphones, and contributing apps are not used uniformly across the population (Sourbati and Behrendt 2021; Fischer, Nelson, and Winters 2022). Similarly, although user-generated social media traces have been instrumental in mapping regional mobility (Hawelka et al. 2014; Q. Huang and Wong 2015), inherent temporal gaps

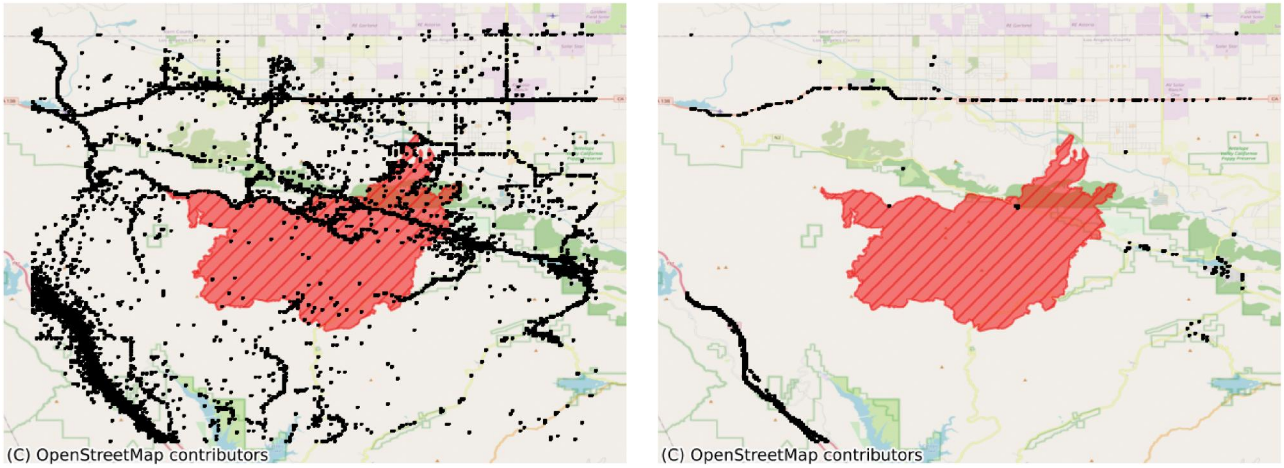

Figure 2. Mobile location data around Lake Fire (2025) in Santa Barbara County, California. The left map illustrates mobile data coverage aggregated over one month. The right map highlights areas where activity signal persisted for at least seven days.

and sample biases constrain their utility for granular behavioral discovery. Therefore, these contemporary movement data require rigorous preprocessing and explicit considerations during analytical modeling to avoid skewed scientific conclusions (Mokbel et al. 2024).

Researchers often shift to high-resolution individual-level GPS traces to bypass aggregate data limitations, yet these remain inherently noisy. Spatial and contextual attributes can contain substantial gaps and errors caused by technical challenges of tracking, signal loss, or environmental interference (Park and Kwan 2025). Spatial errors fluctuate based on environmental conditions, such as indoor versus outdoor, weather patterns, and the built environment. Moreover, temporal gaps emerge when devices are inactive or located in remote areas with sparse signal coverage. Furthermore, population representativeness remains a fundamental challenge: Because data are anonymized, the demographic profiles of individuals are often indeterminate.

Limitations of Movement Data for Open and Reproducible Science

Open science relies on data that can be shared, reused, and critically examined. Although repositories such as Movebank (Kays et al. 2022), Microsoft Geolife (Zheng et al. 2011), Vessel Traffic AIS portals (Tritsarolis, Kontoulis, and Theodoridis 2025), and various micromobility portals have catalyzed innovation, fully open human mobility data remain scarce. Most high-resolution data sets are proprietary, costly, or restricted by terms of use. Even when accessible, a lack of shared standards and metadata often limits interoperability and the comparability of results.

The primary barriers to open repositories are locational privacy, commercial interests, and data heterogeneity. High-resolution trajectories inherently reveal sensitive information, including habits, lifestyle, home and work addresses, and even health or religious affiliations (Monreale and Pellungrini 2023). Consequently, researchers are often unable to share raw trajectory data without obfuscating identifiable information, such as precise coordinates or demographic attributes. Although necessary, these privacy-preserving practices can introduce spatial gaps and inconsistencies that hinder reproducibility.

Because robust anonymization is difficult to guarantee, most open human mobility data sets are provided mainly in aggregate form. Although aggregation reduces privacy risks, it constrains the types of questions that can be addressed and creates a "data gap" for academic researchers who are disadvantaged by the high costs of data license fees or even inaccessibility of private-sector data. The recent transition of mobile location data sets behind paywalls reflects both the economic interests of data providers and the liability of handling sensitive information. As a result, high-resolution research is typically restricted to small groups of collaborators governed by costly contracts. This limits external replication unless the research generates secondary, nonidentifiable data, creating a significant barrier to open science.

To safeguard personal data, regulations such as the General Data Protection Regulation (GDPR; European Union 2016) and other data protection laws (Lim and Oh 2025) impose strict rules on the collection, storage, processing, and transfer of movement and personal data. These include limited lawful bases for collection of personal data, restrictions on data retention and purpose, and adherence to data minimization principles. Because GDPR does not apply to anonymous information that is not identifiable to a natural person (Recital 26, GDPR), researchers frequently rely on anonymization, pseudonymization, or aggregated data products (Monreale and Pellungrini 2023; Yao et al. 2024). Although these frameworks introduce constraints on data access, processing, and innovation, they are essential for safeguarding personal privacy and must be central to any future vision for open and reproducible human movement research.

Advancements in Open Movement Analytics

Geographers have long been leading efforts in developing open and reproducible approaches to spatial data science and GIScience research (Brunsdon and Comber 2021; Kedron et al. 2021). These efforts have resulted in the development of widely adopted open geospatial tools such as PySAL (Rey and Anselin 2010) and QGIS (Graser, Sutton, and Bernasocchi 2025), as well as crowdsourced maps such as OpenStreetMap and guidelines for open GIScience processes (Shannon and Walker 2018). In movement research, these practices have become more prevalent in recent years, particularly in movement ecology (Joo et al. 2020) Human mobility science has begun to adopt open data and code practices. Examples of trajectory analysis Python libraries include MovingPandas (Graser 2019), Trackintel (Martin et al. 2023), ORTEGA (Su, Liu, and Dodge 2024), and scikit-mobility (Pappalardo et al. 2022).

A review of open public movement analytics libraries indicates a lack of standardization in tool development (Graser and Dragaschnig 2020). These tools, often developed using diverse programming languages, tend to be application-specific, making it challenging to assess their cross-applicability with various data sets. There is also the question of best-of-breed algorithms available in such software packages. For instance, most of the preceding packages will offer some functionality for stop detection from trajectory data, and many further algorithms have been published in the literature for that purpose. It is difficult to assess, however, which of these algorithms actually best serves the purpose, despite full availability and transparency of source code. Moreover, the lack of accessibility to raw data sets (see the previous section) and original analytical code presents a major challenge to the reproducibility of research using mobile location data, making large-scale human movement studies nearly impossible for the broader scientific community and creating obstacles for global research initiatives.

Progress in Data Fusion and Data Integration to Contextualize Movement and Enhance Data Quality

Movement is driven (constrained or facilitated) by internal and external contextual factors such as behavior and the surrounding environment, biological markers, policy, and infrastructure (Dodge and Nelson 2023). Movement data are frequently collected in isolation from contextual variables, however. Furthermore, as mentioned earlier, the inherent complexities of tracking often result in incomplete data sets, with gaps and missing information. As a result, advanced data integration and fusion methods have become essential for contextualizing movement patterns and linking them to environmental and behavioral information (Nathan et al. 2022), and to combine multiple sources of movement data.

Contextual data integration, linking trajectories with geographic, environmental, and behavioral variables, supports more comprehensive and meaningful knowledge discovery (Röcke et al. 2023). For instance, spatial interpolation can be used to align trajectory data with environmental and socioeconomic data sets (e.g., population or the built environment) across space and time, creating enriched observations that reveal movement–context associations (Dodge et al. 2013). In contrast, data fusion is employed to synthesize multiple sources and scales of observations to enhance data quality and mitigate sparsity. For example, tensor decomposition approaches have been used to fuse multimodal movement data (e.g., POI visits and activity indexes) to enhance the continuity and fidelity of movement observations for longitudinal studies (Noi and Dodge 2023).

Several studies have systematically reviewed data fusion and integration techniques, across a range of applications, including human activity analysis, intelligent transportation systems, and smart cities (Lau et al. 2019; Ounoughi and Ben Yahia 2023). Despite significant progress in data fusion and integration, persistent challenges remain. Current methods often struggle with imputation of gaps in long trajectories, addressing missingness in aggregate mobility data, and inconsistent sampling rates and data collection across large populations or complex environments. A few studies have applied statistical approaches and deep learning techniques in filling short gaps for GPS trajectory data and computing movement parameters (Barnett and Onnela 2020), and removing erroneous trajectory points using enhanced map matching (Wan and Dodge 2025).

Emerging GeoAI Paradigms for Movement Analytics

Graser, Jalali et al. (2025) highlighted the transformative role of GeoAI in movement analysis. The applications include trajectory segmentation and behavioral mode classification (Dabiri et al. 2020), anomaly detection (Belhadi et al. 2021), map matching (Wan and Dodge 2024), trajectory prediction (H. Li, Jiao, and Yang 2023), and trajectory synthesis and generation (Rao, Gao, and Zhu 2023; Kapp, Hansmeyer, and Mihaljević 2024). DL models, such as recurrent neural networks, attention-based architectures, convolutional neural networks, and generative models including generative adversarial networks, have been used to predict future movement paths and traffic flows (Altche and De La Fortelle 2017; Feng et al. 2018), and forecast next activity locations based on historical trajectory data (Q. Liu et al. 2016). These models excel at capturing temporal dependencies and complex, nonlinear movement patterns, making them well-suited for trajectory data sets. More recently, LLMs have emerged for trajectory representation and reasoning (Nie, Sun, and Ma 2025). By treating movement as semantic sequences of stops and moves, LLMs can infer trip purposes and demographic attributes (Gong et al. 2024; Luo et al. 2024), reconstruct paths from text (L. Liu et al. 2025), predict mobility (Liang et al. 2024), and generate synthetic mobility data sets for model validation and scenario testing (S. Li et al. 2024; Van de Weghe et al. 2025).

Despite these advantages, emerging GeoAI and LLM-based approaches face notable limitations in movement analytics. Although LLMs are effective at capturing high-level patterns, they often struggle with microlevel details such as irregular trajectories, varying spatial resolutions, and the influence of external or rapidly changing contextual factors. To address these shortcomings, recent work has explored hybrid frameworks that combine LLMs with retrieval-augmented generation (RAG) frameworks (Yuan, Zhang, et al. 2025). By retrieving relevant external knowledge or historical trajectory data, RAG-enhanced models can provide richer contextual grounding and improve performance in complex spatiotemporal settings, particularly when data are incomplete or noisy (H. Xu et al. 2025). Despite this, LLMs remain prone to hallucinations, where they generate outputs that might appear plausible but are factually incorrect. The risk of hallucinations can lead to misleading predictions or flawed insights. As a result, although LLMs can offer valuable high-level insights, their application in precision-driven fields remains limited without further refinement in handling fine-grained, context-specific information (Nie, Sun, and Ma 2025).

Broader challenges further constrain AI adoption. These models require large volumes of high-quality, representative data for proper training, yet movement data sets are often sparse, biased, or incomplete as previously discussed. These issues can result in systematic underrepresentation of certain populations or regions and introduce missingness that limits model generalizability (B. Hong et al. 2021). Consequently, training DL models inevitably becomes difficult, and the resulting outputs might not capture the full complexity of real-world movement patterns (Yuan, Ding, et al. 2025). Additionally, the high computational costs of training AI models, coupled with the fact that many data sets reside on closed, paywalled platforms, create significant technical and financial barriers for the development of advanced AI-driven solutions in movement analytics.

Future Outlook

Data Governance: A Call for an Academically Run Movement Data Consortium

Future advances in human movement research require open tracking data supported by transparent, reproducible analytical methods. In animal ecology, Movebank provides a leading model for secure data management through tiered privacy protection and metadata disclosure (Kays et al. 2022). Building on this, Malekzadeh et al. (2025) proposed the OpenGPS platform, which emphasizes encryption and access control for human movement. Although the Open Geospatial Consortium (OGC) and the Open Mobility Foundation (OMF) provide essential standards for spatial interoperability and shared-mobility formats (e.g., MDS), their frameworks are primarily optimized for industrial utility and urban governance. A gap remains not in the consideration of standards, but in their implementation and application within a researcher-centric ecosystem. Existing consortia focus on interoperability for shared transportation modes (e.g., e-scooters, taxi); however, they lack the specific models required to host the high-resolution, longitudinal high-resolution traces necessary for scientific discovery.

An academically run consortium would bridge this gap by acting as a "trusted third party" to manage the legal and ethical frameworks that OGC and OMF are not designed to navigate for the global research community. By prioritizing reproducibility over commercial application, this entity would implement OGC standards to ensure high-resolution trajectory data are not only standardized but also accessible for rigorous peer-reviewed study. Given the sensitive nature of human trajectories, such a trusted entity is essential to navigate the surveillance implications of granular, real-time monitoring. By establishing transparent governance structures, the consortium could build public trust and foster international collaboration through privacy-preserving computation and robust consent management systems. Key strategies to be established by the consortium include real-time anonymization, privacy-preserving computation methods, metadata standards, robust consent management systems, sustainable operating models, and the development of shared ethical frameworks that ensures responsible data use across jurisdictions (Anthony 2025).

The consortium could establish standardized metadata schemas to enhance the reproducibility, replicability, and generalizability of movement research. Comprehensive metadata should delineate the context and scales of the processes captured, alongside the sensor specifications, spatial resolution, and spatiotemporal coverage of the data set. Furthermore, it should provide demographic details regarding the subject population, specifically defining inclusion and exclusion criteria, and the data collection methodology, including a full inventory of contributing applications and whether passive (background) data collection occurred. Critical technical metadata should address location uncertainty, systematic missingness, and privacy safeguards, such as consent provenance and anonymization protocols. Finally, documentation should include the underlying algorithms used to derive mobility metrics, such as transportation mode detection, distance calculations, visitation counts, and dwell times, as well as any baseline data sets incorporated during analysis.

To promote open science, academia and industry should collaborate on developing shared secure repositories, benchmark data, and standardized protocols for creating anonymized human trajectories and aggregate indexes. By implementing strict deidentification policies, these resources could be made freely available for research while balancing scientific progress with personal privacy (Malekzadeh et al. 2025). Private intelligence and mobility analytics companies could also adopt these standards and contribute to the consortium by offering reduced-cost or open-license access for academic institutions, thereby furthering "Data for Good" initiatives beyond pandemic-related research. Such partnerships would unlock the potential of movement analytics for public health, urban resilience, and emergency planning while maintaining high ethical standards.

Data Quality: Addressing Bias, Representation, and Privacy

To advance the field, there is a pressing need for innovative approaches that can intelligently interpolate missing data, integrate multiple sources of mobility data and auxiliary contextual information, and quantify uncertainty in both individual and aggregate mobility analyses. The development of robust and flexible and cross-domain data fusion frameworks capable of combining various forms of individual- and aggregate-level data will be critical for generating more complete, reliable, and meaningful behavioral insights from movement data, because movement itself is only the manifestation of behavior, whereas the reasons that actually trigger the movement, and thus behavior, lie in contextual factors and the environment. Additionally, open data must be accompanied by robust preprocessing strategies and thorough exploratory data analysis to ensure transparency and readiness for movement analytics. Movement research efforts should seriously take sampling biases into account and develop strategies to enhance the representation and address data bias in analysis, for example through spatial poststratification methods (Anganuzzi and Buckland 1993) or generating synthetic, more representative population (Embury et al. 2024).

Moreover, addressing and acknowledging data representation (considered population) and data quality issues prior to data sharing and providing comprehensive documentation and data collection protocols will enhance the ethical usage and sharing of location data (Nelson et al. 2022) and support open movement research. These documentations should describe the scope and limitations of the data collection approach, the operational scales of the data (spatial, temporal, and social), movement and contextual variables, known data quality issues, and recommendations for handling those issues. For an example of a data protocol that could be adopted in GPS tracking studies to enhance future use and reproducibility, see Röcke et al. (2023).

Benchmarking: The Necessity for Synthetic and Benchmark Movement Data Sets

There is a fundamental distinction between data volume and its actual utility. Although "big data" offers very large, high-resolution trajectories, and long-term observations at a global scale, it often remains "sparse" regarding spatial coverage, demographic representation, and temporal continuity, limiting its utility of the investigation of processes at different scales (Figure 1). This creates a paradox in modern movement research and for GeoAI developments: We have an abundance of unlabeled, noisy, and biased opportunistic data (e.g., mobile location data), yet we lack the high-quality, labeled benchmarks required to train and validate precise models.

Future efforts must focus on creating benchmarks, defined as standardized, task-specific data sets with fixed evaluation protocols that serve as shared "ground truth" for comparing algorithms on clearly defined problems at specified scales and contexts (Figure 1). Examples include labeled data for transportation modes, routing preferences, activity types (e.g., commuting vs. recreation), and evacuation strategies. These data sets provide the verified labels necessary to represent complex processes accurately, serving as an essential baseline for evaluating and comparing algorithmic performance in movement classification and prediction. Typically generated through controlled studies, longitudinal surveys, or high-fidelity simulations, benchmarks provide the verified labels required to represent complex processes accurately, while safeguarding privacy via synthetic alternatives.

Following the success of benchmarks in remote sensing and computer vision benchmarks (X. X. Zhu et al. 2017), the community can develop "ground truth" trajectory data sets through data challenges, hackathons, and crowdsourcing initiatives (Russakovsky et al. 2015). Recent advancements include WorldTrace, a global GPS trajectory data set with road-level semantics (Y. Zhu et al. 2024), and WorldMove, a synthetic privacy-preserving global mobility data set generated via diffusion models integrating population, POI maps, and flow data (Yuan, Zhang, et al. 2025). Developing such resources for movement analytics remains a significant challenge, but it is a pressing need for the field's advancement.

Anonymized trajectory data sets, enriched by detailed behavioral and contextual data, are vital for training models in classification, segmentation, anomaly detection, and interaction analysis. Achieving this would require large-scale survey studies such as the California Household Travel Survey (NuStats 2013), which have supported mobility research (Su, McBride, and Goulias 2020). The high cost and privacy concerns of human subject research at large scale, however, often necessitate government management. Although these aggregate-level efforts enhance population-scale research, we note that nuanced, individual-level behavioral discovery remains dependent on access to high-resolution trajectory data and customized, domain-specific analytical frameworks.

Simulation offers a promising alternative for generating large, synthetic trajectory data sets (Technitis et al. 2015; Yuan, Zhang, et al. 2025). Universal models that satisfy both diverse utility and privacy protection are still lacking, however. There is in fact a need for clearer utility definitions, transparent evaluation protocols, and real-world test cases to guide model development and ensure practical relevance (Kapp, Hansmeyer, and Mihaljević 2024). The research community should leverage AI and LLMs to develop advanced movement simulation techniques and GenAI models that can produce representative synthetic trajectories reflecting a range of contexts and behaviors. Ontologies, discussed in detail in the next section, or other preexisting, structured data about geographical facts (e.g., historical trajectories), could also be used in RAG frameworks (Van de Weghe et al. 2025; Yuan, Zhang, et al. 2025), allowing LLMs to be augmented by a system component that retrieves given facts about the real world and thus improves the accuracy and trustworthiness of simulated trajectories (and reduces the likelihood of LLM hallucinations). Future research should also leverage AI models that can learn emergent patterns from smaller samples of high-quality data sets (Hollmann et al. 2025), to generate realistic movement patterns that better represent larger populations. Synthetic data generated by these models must be subject to rigorous validation frameworks, such as checking for physical plausibility (e.g., speed and connectivity constraints) and statistical alignment with real-world distributions.

Finally, synthetic data generation serves as a dual-purpose strategy: It addresses data sparsity due to spatial and temporal data gaps as well as lack of continuity, while acting as a privacy-preserving mechanism through the decoupling of individual identities from movement patterns. Future research should focus on exploring how AI-driven simulation can enhance the diversity and representativeness of movement data sets, improve the scalability of data generation, and support the development of more generalizable movement analytics models.

Standardization: Developing Movement Ontologies and Standardized Mobility Metrics

Beyond benchmark and synthetic data sets, and responding to the challenge mentioned previously, we advocate the development of domain-specific movement ontologies grounded in geographic science (Fonseca et al. 2002; Agarwal 2005). Formalizing the concepts, entities, and relationships inherent in movement processes (e.g., disease dispersion, mobility disruption, accessibility, or environmental exposure) is critical for situating trajectories and aggregate movement data within their broader geographic context across the three spatial, temporal, and social scales. These ontologies provide the semantic structure necessary to bridge the gap between AI developers and domain experts, facilitating more robust collaborations. By establishing a common logical language across disciplines such as geography, urban planning, public health, and transportation, ontologies ensure that patterns are identified using consistent definitions. Furthermore, this semantic clarity enhances the capacity for AI-driven logical reasoning, allowing models to infer behavioral patterns with greater contextual accuracy and improving the interoperability of disparate data sources. Integrating semantic and topological ontologies into the AI training process might reduce the data hungriness of these algorithms (discussed earlier) by providing the model with structural priors, thereby enhancing both the interpretability and the knowledge discovery capabilities of the resulting models.

In parallel, the field requires a standardized library of movement metrics and open-source algorithms. Establishing a unified suite of aggregate measures (e.g., visitation frequencies, movement rates, and activity classifications) is vital for cross-data-set comparability. Similarly, suites of metrics should also be developed and agreed on to describe daily mobility on the individual level. For example indicator sets for use in health and aging studies, see Perchoux et al. (2014) and Fillekes et al. (2019). When these metrics are coupled with standardized computational procedures and transparent parameterization, the research community can generate more scientifically grounded and reproducible measures of movement. Such standardization facilitates a transition from isolated, site-specific case studies toward comparative analyses that generalize across diverse geographical and social contexts. Ultimately, achieving this rigor requires extending taxonomical foundations (Dodge, Weibel, and Lautenschütz 2008) to provide a comprehensive characterization of movement descriptors and processes.

Computational Advancements: Expanding Next-Generation Analytics and GeoAI

Addressing data complexity challenges requires next-generation movement analytics that provide privacy-preserving, cross-scale solutions capable of supporting a wide range of movement and contextual data across different scales and formats. Future systems must enable scalable, real-time integration of multisource data, tackling the complex task of synchronizing heterogeneous data streams, ranging from legacy systems to high-velocity, real-time feeds, while harmonizing spatial and temporal resolutions. This entails overcoming challenges such as managing computational scalability, ensuring interoperability across diverse platforms, and maintaining data consistency. Promising solutions include adopting stream-processing architectures, leveraging distributed edge–cloud computing, and implementing standardized application programming interfaces and reference frameworks to enable scalable seamless data fusion (Mokbel et al. 2024).

Because movement data inherently integrate spatial, temporal, and auxiliary context, effective representation requires creative visualization techniques that extract patterns without compromising individual location privacy. Methods optimized for specific data sets often fail to generalize across domains; therefore, it is imperative that new analytics remain open, transparent, and extensible. To ensure broad adoption, developers should provide comprehensive documentation, standardized data schemas, and open-source implementations that illustrate tool application through published case studies. For example, DynamoVis uses the standard Movebank data format to generate dynamic visualization of movement trajectories integrated with environmental factors (Dodge, Toka, and Bae 2021). The standard structure has enabled its adoption by movement ecologists. Other examples of open libraries with documentation and usage guidelines include OSMnx for visualizing complex street networks (Boeing 2017), ORTEGA for movement interaction analysis (Su, Liu, and Dodge 2024), and the MovingPandas library for general-purpose movement data handling and mapping (Graser 2019).

GeoAI has already proven transformative for movement analysis, offering sophisticated tools for extracting insights from high-dimensional spatiotemporal data (Luca et al. 2023; Pappalardo et al. 2023). Although the majority of these developments have occurred in the computer science and machine learning community, it is necessary for researchers to bring a geographic perspective to developing more spatially explicit techniques that are grounded in spatiotemporal theories and that are better suited for movement pattern analysis at various spatial and temporal granularities. For instance, incorporating the first law of geography regarding the spatial autocorrelation can enable a DL model to generate more realistic mobility flows across geographic spaces, even with scarce training data (Simini et al. 2021). Moreover, the movement contexts of geographic spaces are increasingly being captured and modeled using multisource geospatial data, such as points of interest and street-view imagery, and spatial representation learning techniques, significantly improving the accuracy of mobility predictions (Y. Hong et al. 2023; F. Huang et al. 2025).

Significant opportunities remain to deepen the integration of geographic theory within AI-based movement analytics (Van de Weghe et al. 2025). Following the revolution in computational map generalization (Brassel and Weibel 1988; Burghardt, Duchêne, and Mackaness 2014), GeoAI models must be enhanced by incorporating contextual information and knowledge of the movement processes involved for plausible pattern recognition, knowledge discovery, and abstraction. Future research should leverage ontologies to improve the generalizability of AI and DL models, particularly when extracting representative patterns from biased or imperfect tracking data. Additionally, developing scalable generative methods to synthesize domain-specific benchmark data sets will be crucial for model validation. As these technologies evolve, the research community must prioritize ethical frameworks and robust privacy-preserving measures to ensure the responsible use of location intelligence (Mokbel et al. 2024). For example, federated learning enables multiple mobility data providers to collaboratively train models on movement data without sharing raw, sensitive records (Gecer and Garbinato 2024; Graser, Lorencio Abril 2025). By keeping data local and only exchanging model updates, it preserves privacy while still capturing patterns across providers.

Finally, critical research is needed to mitigate hallucinations in LLMs and generative AI during trajectory analysis and prediction. To ensure physical and logical consistency, as discussed earlier, AI models must be grounded in movement ontologies and formal knowledge graphs that serve as domain-specific "fact-checkers" (S. Li et al. 2024; Liang et al. 2024). By anchoring LLMs within an ontological framework, models can be constrained to adhere to valid entity behaviors and transitions. Furthermore, given the strong power and flexibility of graphs in modeling spatial (e.g., street network topologic graph and POI accessibility graphs) and nonspatial relationships of geographic environments and agents, another promising direction is to integrate graph learning with LLMs to enhance the reasonability and reduce spatial hallucinations in trajectory analysis and prediction (Jin et al. 2024; Chai et al. 2026). A straightforward example is that a LLM can be given a street network description beyond the generation prompts to generate a more plausible trajectory (Wu et al. 2024).

Methodological Rigor: Leveraging Causal Inference for Movement Research

Because movement research is largely data-driven, with few controlled experiments and often unknown data generation processes, reasoning causality is fundamentally challenging. Correlations discovered in observational movement data analysis can be misinterpreted as causal, leading to misleading conclusions or even misguided policy interventions. To move beyond descriptive analytics, future research must leverage the conceptual and methodological foundations of causal inference to rigorously quantify the relationships between movement patterns and their drivers. In recent years, there has been a growing interest in using causal discovery and causal effect estimation methods for movement research, with applications including, but not limited to, understanding the causality between urban form and travel behavior, identifying causal relationships in traffic states, and evaluating causal effects of transportation policy interventions (M. Xin et al. 2021; Dong, Zhang, and Zhang 2025).

Future research agendas should prioritize the refinement of spatiotemporal causal inference to account for the unique interference effects inherent to mobility, such as spatial spillover and temporal carryover. Although progress has been made in modeling spatiotemporal dimensions (Kolak and Anselin 2020; Akbari, Winter, and Tomko 2023) and mitigating dependence effects (Papadogeorgou and Samanta 2023; Runge et al. 2023), there is a critical need for frameworks specifically engineered for the high-velocity data streams and the continuous trajectory data. Establishing robust benchmarks for spatiotemporal causal models (Credit and Lehnert 2024) will be essential for validating these methods across different geographical contexts.

Furthermore, the integration of causal inference and machine learning and AI combines complementary strengths for movement research. AI's strong representation power and scalability accelerate causal discovery and effect estimation from large-scale mobility data. In turn, causal inference also strengthens the interpretability and robustness of AI models in movement analysis by focusing on causation rather than merely association (Y. Xin et al. 2022). Recent advances in world models, which explicitly capture latent environmental structure and the consequences of actions, offer a complementary pathway toward more reliable, causally consistent mobility prediction (Le Cun 2022). Therefore, we envision a move toward causally augmented AI, where models are grounded in causal reasoning to enhance the analysis, prediction, simulation, and interpretation of movement data. Recent studies have already yielded promising results. For example, augmenting trajectory forecasting with causal features improves generalizability and reliability in unseen environments. (Y. Liu et al. 2022). Furthermore, counterfactual explanations increase the transparency of black-box models (Wang et al. 2025), and causal interventions enable the generation of controlled synthetic data sets to test "what-if" scenarios in mobility prediction (Y. Hong et al. 2025). Given the rapid development of LLMs and its widespread adoption, causal inference has also been used to enhance LLMs' reasoning ability and logical deduction, mitigate hallucinations, and improve robustness against adversarial attacks (X. Liu et al. 2025). For example, LLMs can be trained or fine-tuned using data annotated by known cause-and-effect relationships or guided by causal graph structures during reasoning (e.g., causal chain-of-thought). By integrating causal knowledge into LLMs, the research community can develop "responsible AI" frameworks that move beyond pattern matching to understand the why behind movement. This transition from "correlation-based" to "causation-based" movement analysis is vital for creating reliable, actionable insights for urban planning, public health, and social policy.

Open Science: Moving toward a Culture of Open Movement Research

Reproducibility, replicability, and generalizability are fundamental to open science, ensuring the reliability and trustworthiness of scientific inquiry (National Academies of Sciences 2019). Achieving computational reproducibility necessitates that the movement data and code used to develop methods and insights be made publicly available, a goal that could be realized through the establishment of an open movement data consortium. Replicability, or the ability to obtain consistent results across independent studies, must be evaluated in relation to the specific study context, the underlying generative processes, and the spatial and temporal scales of movement patterns. Furthermore, generalizability aims to extend insights derived from a specific subset of moving entities or locations to broader populations and diverse geographical settings.

The core components of movement research are movement data and analytical methods, with both originating from, and applying to, a range of application domains, including human mobility and animal ecology. To foster open science and reproducible movement research, these components must adhere to the FAIR (Findable, Accessible, Interoperable, and Reusable) data sharing principles, which were introduced in 2016 to improve data-driven science and discovery (Wilkinson et al. 2016). Enhancing reproducibility requires that movement observations be accompanied by comprehensive metadata as described earlier (Kinkade and Shepherd 2022; Malekzadeh et al. 2025). Similarly, analytical code must be published with rigorous documentation regarding the spatiotemporal constraints of the algorithms, enabling researchers to independently verify findings and transfer knowledge across disciplines. Beyond individual data sets, standardized and openly available methods facilitate cross-setting replication, and collaborative platforms allow diverse research groups to test and extend localized findings to a global context.

Looking ahead, the evolution of open movement research lies in the transition from static and isolated data sharing to dynamic, interoperable knowledge ecosystems through a trusted data consortium. Future research should prioritize the development of comprehensive repositories where code, data, and ontologies are linked, allowing for replication across varying context, scales and resolutions. By leveraging decentralized frameworks and privacy-preserving federated learning, the research community can transcend the limitations of siloed data sets, enabling the discovery of universal laws of movement that are robust to geographic and social variations. Ultimately, the goal is to establish a global "research commons" for movement research, where the synthesis of geographic theory and open-source AI empowers researchers to address urgent societal challenges, such as pandemic and disaster response and urban climate adaptation.

Roadmap: Open Research Questions

Building on the challenges already identified, we propose the following set of research questions to guide future work. Addressing these challenges by integrating intellectual contributions from geography, GIScience, AI, and computer science provides a pathway for movement research to evolve into a more cohesive and interdisciplinary science. In this collaborative future, GeoAI will move beyond predictive modeling to provide deeper insights into the underlying drivers of movement patterns, while upholding rigorous privacy standards through secure, ethical systems. These research frontiers represent a strategic roadmap toward a unified, transparent, and impactful ecosystem for open movement research.

- *Data, privacy, and ethics:* How can an academically run consortium implement a privacy-preserving and ethical federated query system that enables the analysis of high-resolution mobility without raw data ever leaving its secure origin?
- *Data integration and data fusion:* What are the characteristics of contextual data integration and fusion frameworks to effectively combine sparse GPS trajectories with environmental and social auxiliary data to correct for systemic sampling biases and data gaps in digitally underrepresented populations?
- *Generative AI:* To what extent can generative AI models leverage small-sample, high-resolution seed data (e.g., travel surveys with GPS tracking of a subset of a population) to synthesize representative, large-scale populations that accurately reflect the behaviors of the overall population?
- *Benchmarking:* Can a unified library of movement metrics serve as a gold standard to benchmark GeoAI performance across both Lagrangian and Eulerian representations, and diverse scales from micropedestrian maneuvers to macro-urban flows?
- *Ontologies and LLMs:*
    - How can GeoAI frameworks integrate spatiotemporal ontologies with LLMs to improve logical consistency and mitigate spatial hallucinations in mobility predictions?
    - To what extent can formal movement ontologies be leveraged to enable the generation of physically plausible trajectories in LLMs compared to purely data-driven approaches?
- *Causal inference:* How can causal modeling be integrated into GeoAI frameworks to distinguish between coincidental correlations and true behavioral drivers in multisource mobility data?

- *Interoperability and open science:* What are the defining characteristics of open-standard code libraries and data standards required to ensure cross-disciplinary interoperability and reproducibility in movement research across diverse data scales and formats?

Conclusions

This article reviewed the transformative developments and emerging opportunities within the field of movement research. We examined the critical challenges to building an open, reproducible, and ethical research ecosystem—obstacles that are amplified by the large-scale, multidimensional, heterogeneous, and privacy-sensitive nature of modern movement data. To address these barriers, we propose, centered on the establishment of ethical governance through academically run consortia, the development of scalable infrastructure for multimodal, multisource data integration, and the creation of standardized benchmarks and movement-specific ontologies. Furthermore, we advocate for the advancement of GeoAI founded on formal movement ontologies to capture emergent behavioral patterns, alongside the integration of causal inference to move beyond correlation toward meaningful knowledge discovery and robust behavioral modeling.

Acknowledgments

The authors acknowledge that this article emerged from discussions held during the Movement Analytics in the Era of Big Data and Open Science workshop (19-20 June 2024, Irchel Campus, University of Zurich), sponsored by the ETH/UZH Collegium Helveticum. We thank Alexan-dra-Ioana Georgescu and Changyu Han for their notes, as well as the other workshop participants for their valuable contributions.

Author Contributions

CRediT: **Somayeh Dodge**: Conceptualization, Funding acquisition, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing; **Robert Weibel**: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing; **Christophe Claramunt**: Writing – review & editing; **Gavin McArdle**: Writing – review & editing; **Zhiyong Zhou**: Writing – review & editing; **Yanan Xin**: Writing – review & editing; **Anita Graser**: Writing – review & editing; **Tumasch Reichenbacher**: Conceptualization.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Somayeh Dodge gratefully acknowledges the support of the National Science Foundation (Award #BCS-2043202). Furthermore, this work was supported by fellowships from the ETH/UZH Collegium Helveticum and the University of Zurich's Digital Society Initiative during a sabbatical, which provided the foundational research and inspiration for this article.

ORCID

Somayeh Dodge http://orcid.org/0000-0003-0335-3576
Robert Weibel http://orcid.org/0000-0002-2425-0077
Christophe Claramunt http://orcid.org/0009-0008-5208-1208
Gavin McArdle http://orcid.org/0000-0003-0613-546X
Zhiyong Zhou http://orcid.org/0000-0002-6117-2071
Yanan Xin http://orcid.org/0000-0003-3866-821X
Anita Graser http://orcid.org/0000-0001-5361-2885
Tumasch Reichenbacher http://orcid.org/0000-0002-7833-185X

Notes

1. Movement research is also known as mobility research, movement data science, data-driven movement analysis, or computational movement analysis. In this article we use the term *movement research* to refer to the field, and *movement analysis* or *movement analytics* to refer to the computational and data-driven methods used to study movement patterns.

2. Collectively referred to as movement analytics, movement analysis, or mobility analytics.

References

Agarwal, P. 2005. Ontological considerations in GIScience. *International Journal of Geographical Information Science* 19 (5):501–36. doi: 10.1080/13658810500032321.

Ahearn, S. C., and S. Dodge. 2018. Recursive multi-frequency segmentation of movement trajectories (ReMuS). *Methods in Ecology and Evolution* 9 (4):1075–87. doi: 10.1111/2041-210X.12958.

Akbari, K., S. Winter, and M. Tomko. 2023. Spatial causality: A systematic review on spatial causal inference. *Geographical Analysis* 55 (1):56–89. doi: 10.1111/gean.12312.

Altche, F., and A. De La Fortelle. 2017. An LSTM network for highway trajectory prediction. In *IEEE Conference on Intelligent Transportation Systems (ITSC)*, ed. Institute of Electrical and Electronics Engineers, 353–59. Yokohama, Japan: IEEE. doi: 10.1109/ITSC.2017.8317913.

Anganuzzi, A. A., and S. T. Buckland. 1993. Post-stratification as a bias reduction technique. *The Journal of Wildlife Management* 57 (4):827. doi: 10.2307/3809085.

Anthony, B. 2025. Data governance policies for human-centric urban mobility services in smart cities. Urban Human Mobility 2025, 62–73. doi: 10.1201/9781003503262-9.

Barnett, I., and J. P. Onnela. 2020. Inferring mobility measures from GPS traces with missing data. *Biostatistics* 21 (2):e98–e112. doi: 10.1093/BIOSTATISTICS/KXY059.

Belhadi, A., Y. Djenouri, G. Srivastava, D. Djenouri, J. C. W. Lin, and G. Fortino. 2021. Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection. *Information Fusion* 65:13–20. doi: 10.1016/j.inffus.2020.08.003.

Boeing, G. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65:126–39. doi: 10.1016/j.compenvurbsys.2017.05.004.

Brassel, K. E., and R. Weibel. 1988. A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems* 2 (3):229–44. doi: 10.1080/02693798808927898.

Brunsdon, C., and A. Comber. 2021. Opening practice: Supporting reproducibility and critical spatial data science. *Journal of Geographical Systems* 23 (4):477–96. doi: 10.1007/S10109-020-00334-2/TABLES/3.

Burghardt, D., C. Duchêne, and W. Mackaness. 2014. *Abstracting geographic information in a data rich world.* Lecture Notes in Geoinformation and Cartography. Cham, Heidelberg, New York, Dordrecht, and London: Springer. doi: 10.1007/978-3-319-00203-3.

Chai, Z., T. Zhang, L. Wu, K. Han, X. Hu, X. Huang, and Y. Yang. 2026. GraphLLM: Boosting graph reasoning ability of large language model. *IEEE Transactions on Big Data* 12 (2):475–83. doi: 10.1109/TBDATA.2025.3627488.

Chin, W. C. B., C. C. Feng, C. H. Leong, H. E. Clapham, J. Pang, and Y. C. Wang. 2024. The networked community of urban mobility during the pandemic. *Annals of the American Association of Geographers* 114 (8):1704–17. doi: 10.1080/24694452.2023.2278689.

Credit, K., and M. Lehnert. 2024. A structured comparison of causal machine learning methods to assess heterogeneous treatment effects in spatial data: A structured comparison of causal machine learning methods to assess heterogeneous treatment. *Journal of Geographical Systems* 26 (4):483–510. doi: 10.1007/S10109-023-00413-0/TABLES/6.

Dabiri, S., C. T. Lu, K. Heaslip, and C. K. Reddy. 2020. Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data. *IEEE Transactions on Knowledge and Data Engineering* 32 (5):1010–23. doi: 10.1109/TKDE.2019.2896985.

Del Mondo, G., P. Peng, J. Gensel, C. Claramunt, and F. Lu. 2021. Leveraging spatio-temporal graphs and knowledge graphs: Perspectives in the field of maritime transportation. *ISPRS International Journal of Geo-Information* 10 (8):541. doi: 10.3390/ijgi10080541.

Demšar, U., J. A. Long, F. Benitez-Paez, V. Brum Bastos, S. Marion, G. Martin, S. Sekulić, K. Smolak, B. Zein, and K. Siła-Nowicka. 2021. Establishing the integrated science of movement: Bringing together concepts and methods from animal and human movement analysis. *International Journal of Geographical Information Science* 35 (7):1273–1308. doi: 10.1080/13658816.2021.1880589.

Dodge, S. 2021. A data science framework for movement. *Geographical Analysis* 53 (1):92–112. doi: 10.1111/gean.12212.

Dodge, S., G. Bohrer, R. Weinzierl, S. C. Davidson, R. Kays, D. Douglas, S. Cruz, J. Han, D. Brandes, and M. Wikelski. 2013. The environmental-data automated track annotation (Env-DATA) system: Linking animal tracks with environmental data. *Movement Ecology* 1 (1):3. doi: 10.1186/2051-3933-1-3.

Dodge, S., and T. A. Nelson. 2023. A framework for modern time geography: Emphasizing diverse constraints on accessibility. *Journal of Geographical Systems* 25 (3):1–19. doi: 10.1007/s10109-023-00404-1.

Dodge, S., and E. Noi. 2021. Mapping trajectories and flows: Facilitating a human-centered approach to movement data analytics. *Cartography and Geographic Information Science* 48 (4):353–75. doi: 10.1080/15230406.2021.1913763.

Dodge, S., M. Toka, and C. J. Bae. 2021. DynamoVis 1.0: An exploratory data visualization software for mapping movement in relation to internal and external factors. *Movement Ecology* 9 (1):55. doi: 10.1186/s40462-021-00291-5.

Dodge, S., R. Weibel, and A. K. Lautenschütz. 2008. Towards a taxonomy of movement patterns. *Information Visualization* 7 (3–4):240–52. doi: 10.1057/PALGRAVE.IVS.9500182.
Dong, P., X. Zhang, and X. Zhang. 2025. A data-driven approach for spatio-temporal causal analysis in large-scale urban traffic networks. *Transportation Research Part E: Logistics and Transportation Review* 202:104244. doi: 10.1016/j.tre.2025.104244.
Embury, J., A. Nara, S. Rey, M. H. Tsou, and S. Ghanipoor Machiani. 2024. Detecting synthetic population bias using a spatially-oriented framework and independent validation data. *International Journal of Geographical Information Science* 38 (9):1912–38. doi: 10.1080/13658816.2024.2358399.
European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation). *Official Journal of the European Union* L (119):1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj.
Feng, J., Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin. 2018. DeepMove: Predicting human mobility with attentional recurrent networks. In *The Web Conference 2018—Proceedings of the World Wide Web Conference, WWW 2018*, 1459–68. Geneva: Association for Computing Machinery. doi: 10.1145/3178876.3186058.
Fillekes, M. P., E. Giannouli, E. K. Kim, W. Zijlstra, and R. Weibel. 2019. Towards a comprehensive set of GPS-based indicators reflecting the multidimensional nature of daily mobility for applications in health and aging research. *International Journal of Health Geographics* 18 (1):17. doi: 10.1186/S12942-019-0181-0.
Fischer, J., T. Nelson, and M. Winters. 2022. Changes in the representativeness of Strava bicycling data during COVID-19. *Findings* 2022:80. doi: 10.32866/001c.33280.
Fonseca, F. T., M. J. Egenhofer, P. Agouris, and G. Câmara. 2002. Using ontologies for integrated geographic information systems. *Transactions in GIS* 6 (3):231–57. doi: 10.1111/1467-9671.00109.
Gecer, M., and B. Garbinato. 2024. Federated learning for mobility applications. *ACM Computing Surveys* 56 (5):1–28. doi: 10.1145/3637868.
Gong, L., Y. Lin, X. Zhang, Y. Lu, X. Han, Y. Liu, S. Guo, Y. Lin, and H. Wan. 2024. Mobility-LLM: Learning visiting intentions and travel preference from human mobility data with large language models. *Advances in Neural Information Processing Systems 37*:36185–217. doi: 10.52202/079017-1141.
Goodchild, M. F. 2004. GIScience, geography, form, and process. *Annals of the Association of American Geographers* 94 (4):709–14. doi: 10.1111/j.1467-8306.2004.00424.x.
Graser, A. 2019. MovingPandas: Efficient structures for movement data in Python. *GI_Forum* 1 (1):54–68. doi: 10.1553/giscience2019_01_s54.
Graser, A., and M. Dragaschnig. 2020. Open geospatial tools for movement data exploration. *KN – Journal of Cartography and Geographic Information* 70 (1):3–10. doi: 10.1007/s42489-020-00039-y.
Graser, A., A. Jalali, J. Lampert, A. Weißenfeld, and K. Janowicz. 2025. MobilityDL: A review of deep learning from trajectory data. *GeoInformatica* 29 (1):115–47. doi: 10.1007/S10707-024-00518-8/FIGURES/6.
Graser, A., J. A. Lorencio Abril, A. Weißenfeld, and A. Wachsenegger. 2025. ST-SplitVFL: Spatio-temporal split vertical federated learning. Paper presented at the 6th Spatial Data Science Symposium (SDSS), December 4-5, 2025. doi: 10.5281/ZENODO.17661670.
Graser, A., T. Sutton, and M. Bernasocchi. 2025. The QGIS Project: Spatial without compromise. *Patterns* 6 (7):101265. doi: 10.1016/j.patter.2025.101265.
Hägerstrand, T. 1970. What about people in regional science? *Papers of the Regional Science Association* 24 (1):6–21. doi: 10.1007/BF01936872.
Haraguchi, M., A. Nishino, A. Kodaka, M. Allaire, U. Lall, L. Kuei-Hsien, K. Onda, K. Tsubouchi, and N. Kohtake. 2022. Human mobility data and analysis for urban resilience: A systematic review. *Environment and Planning B: Urban Analytics and City Science* 49 (5):1507–35. doi: 10.1177/23998083221075634.
Hawelka, B., I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* 41 (3):260–71. doi: 10.1080/15230406.2014.890072.
Heidger, L., T. Nelson, and D. Willet. 2025. Uncertainty in cell-phone generated bike and pedestrian volumes. *Findings* 2025 (October):234. doi: 10.32866/001c.145234.
Hollmann, N., S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S. B. Hoo, R. T. Schirrmeister, and F. Hutter. 2025. Accurate predictions on small data with a tabular foundation model. *Nature* 637 (8045):319–26. doi: 10.1038/s41586-024-08328-6.
Hong, B., B. J. Bonczak, A. Gupta, and C. E. Kontokosta. 2021. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nature Communications* 12 (1):1870. doi: 10.1038/s41467-021-22160-w.
Hong, Y., Y. Xin, S. Dirmeier, F. Perez-Cruz, and M. Raubal. 2025. A causal intervention framework for synthesizing mobility data and evaluating predictive neural networks. *Transportation Research Interdisciplinary Perspectives* 31:101398. doi: 10.1016/j.trip.2025.101398.
Hong, Y., Y. Zhang, K. Schindler, and M. Raubal. 2023. Context-aware multi-head self-attentional neural network model for next location prediction. *Transportation Research Part C: Emerging Technologies* 156:104315. doi: 10.1016/j.trc.2023.104315.
Hornsby, K., and M. J. Egenhofer. 2002. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence* 36 (1–2):177–94. doi: 10.1023/A:1015812206586.

Huang, F., J. Rong Lv, G. L. Li, and Y. Yue. 2025. SPOK: Tokenizing geographic space for enhanced spatial reasoning in GeoAI. *International Journal of Geographical Information Science* 39 (12):2768–2808. doi: 10.1080/13658816.2025.2497810.

Huang, Q., and D. W. S. Wong. 2015. Modeling and visualizing regular human mobility patterns with uncertainty: An example using Twitter data. *Annals of the Association of American Geographers* 105 (6):1179–97. doi: 10.1080/00045608.2015.1081120.

Jin, B., G. Liu, C. Han, M. Jiang, H. Ji, and J. Han. 2024. Large language models on graphs: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering* 36 (12):8622–42. doi: 10.1109/TKDE.2024.3469578.

Joo, R., M. E. Boone, T. A. Clay, S. C. Patrick, S. Clusella-Trullas, and M. Basille. 2020. Navigating through the R packages for movement. *The Journal of Animal Ecology* 89 (1):248–67. doi: 10.1111/1365-2656.13116.

Kapp, A., J. Hansmeyer, and H. Mihaljević. 2024. Generative models for synthetic urban mobility data: A systematic literature review. *ACM Computing Surveys* 56 (4):1–37. doi: 10.1145/3610224.

Kays, R., S. C. Davidson, M. Berger, G. Bohrer, W. Fiedler, A. Flack, J. Hirt, C. Hahn, D. Gauggel, B. Russell, et al. 2022. The Movebank system for studying global animal movement and demography. *Methods in Ecology and Evolution* 13 (2):419–31. doi: 10.1111/2041-210X.13767.

Kedron, P., W. Li, S. Fotheringham, and M. Goodchild. 2021. Reproducibility and replicability: Opportunities and challenges for geospatial research. *International Journal of Geographical Information Science* 35 (3):427–45. doi: 10.1080/13658816.2020.1802032.

Kinkade, D., and A. Shepherd. 2022. Geoscience data publication: Practices and perspectives on enabling the FAIR guiding principles. *Geoscience Data Journal* 9 (1):177–86. doi: 10.1002/gdj3.120.

Kolak, M., and L. Anselin. 2020. A spatial perspective on the econometrics of program evaluation. *International Regional Science Review* 43 (1–2):128–53. doi: 10.1177/0160017619869781.

Lau, B. P. L., S. H. Marakkalage, Y. Zhou, N. U. Hassan, C. Yuen, M. Zhang, and U. X. Tan. 2019. A survey of data fusion in smart city applications. *Information Fusion* 52 (December):357–74. doi: 10.1016/j.inffus.2019.05.004.

Laube, P. 2014. *Computational movement analysis*. Cham, Heidelberg, New York, Dordrecht, and London: Springer International. doi: 10.1007/978-3-319-10268-9.

Le Cun, Y. 2022. A path towards autonomous machine intelligence. *Open Review* 62 (1):1–62.

Li, H., H. Jiao, and Z. Yang. 2023. Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis. *Engineering Applications of Artificial Intelligence* 126:107062. doi: 10.1016/j.engappai.2023.107062.

Li, S., T. Tran, H. Lin, J. Krumm, C. Shahabi, L. Zhao, K. Shafique, and L. Xiong. 2024. Geo-Llama: Leveraging LLMs for human mobility trajectory generation with spatiotemporal constraints. https://arxiv.org/pdf/2408.13918.

Li, W., S. T. Arundel, S. Gao, M. F. Goodchild, Y. Hu, S. Wang, and A. Zipf. 2024. GeoAI for science and the science of GeoAI. *Journal of Spatial Information Science* 29 (29):1–17. doi: 10.5311/JOSIS.2024.29.349.

Li, Z., H. Ning, F. Jing, and M. Naser Lessani. 2024. Understanding the bias of mobile location data across spatial scales and over time: A comprehensive analysis of SafeGraph data in the United States. *PLoS ONE* 19 (1):e0294430. doi: 10.1371/journal.pone.0294430.

Liang, Y., Y. Liu, X. Wang, and Z. Zhao. 2024. Exploring large language models for human mobility prediction under public events. *Computers, Environment and Urban Systems* 112 (September):102153. doi: 10.1016/j.compenvurbsys.2024.102153.

Lim, S., and J. Oh. 2025. Navigating privacy: A global comparative analysis of data protection laws. *IET Information Security* 2025 (1):5536763. doi: 10.1049/ise2/5536763.

Linardon, J., K. Chen, S. Gajjar, A. Eadara, S. Wang, M. Flathers, J. Burns, and J. Torous. 2025. Smartphone digital phenotyping in mental health disorders: A review of raw sensors utilized, machine learning processing pipelines, and derived behavioral features. *Psychiatry Research* 348 (June):116483. doi: 10.1016/J.PSYCHRES.2025.116483.

Liu, L., T. Pei, Z. Fang, X. Yan, C. Zheng, X. Wang, C. Song, W. Luan, and J. Chen. 2025. Extracting individual trajectories from text by fusing large language models with diverse knowledge. *International Journal of Applied Earth Observation and Geoinformation* 141 (July):104654. doi: 10.1016/j.jag.2025.104654.

Liu, Q., S. Wu, L. Wang, and T. Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. *Proceedings of the AAAI Conference on Artificial Intelligence* 30 (1):194–200. doi: 10.1609/aaai.v30i1.9971.

Liu, X., P. Xu, J. Wu, Y. Yang, F. Liu, T. Guan, H. Wang, T. Yu, J. McAuley, W. Ai, et al. 2025. Large language models and causal inference in collaboration: A comprehensive survey. In *Proceedings of the 63rd annual meeting of the Association for Computational Linguistics, vol. 1: Long papers*, ed. W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar. Ithaca, NY: Cornell University Library. doi: 10.18653/V1/2025.

Liu, Y., R. Cadei, J. Schweizer, S. B. Alexandre, and A. Alahí. 2022. Towards robust and adaptive motion forecasting: A causal representation perspective. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17081–17092. New Orleans: IEEE. doi: 10.1109/CVPR52688.2022.01657.

Luca, M., G. Barlacchi, B. Lepri, and L. Pappalardo. 2023. A survey on deep learning for human mobility. *ACM Computing Surveys* 55 (1):1–44. doi: 10.1145/3485125.

Luo, Y., Z. Cao, X. Jin, K. Liu, and L. Yin. 2024. Deciphering human mobility: Inferring semantics of trajectories with large language models. *Proceedings – IEEE International Conference on Mobile Data Management*, 289–94. Brussels: Institute of Electrical and Electronics Engineers. doi: 10.1109/MDM61037.2024.00060.

Mai, G., W. Huang, J. Sun, S. Song, D. Mishra, N. Liu, S. Gao, T. Liu, G. Cong, Y. Hu, et al. 2024. On the opportunities and challenges of foundation models for GeoAI. *ACM Transactions on Spatial Algorithms and Systems* 10 (2):1–46. doi: 10.1145/3653070.

Malekzadeh, M., H. J. Ha, K. Sila-Nowicka, V. Brum-Bastos, J. Lee, U. Demšar, and J. A. Long. 2025. How can we make GPS tracking studies more open, reproducible, and collaborative? A vision for the OpenGPS platform. *Data in Brief* 60 (June):111603. doi: 10.1016/j.dib.2025.111603.

Martin, H., Y. Hong, N. Wiedemann, D. Bucher, and M. Raubal. 2023. Trackintel: An open-source Python library for human mobility analysis. *Computers, Environment and Urban Systems* 101 (April):101938. doi: 10.1016/j.compenvurbsys.2023.101938.

Miller, H. J. 2020. Movement analytics for sustainable mobility. *Journal of Spatial Information Science* 20 (20):115–23. doi: 10.5311/JOSIS.2020.20.663.

Miller, H. J., S. Dodge, J. Miller, and G. Bohrer. 2019. Towards an integrated science of movement: Converging research on animal movement ecology and human mobility science. *International Journal of Geographical Information Science* 33 (5):855–76. doi: 10.1080/13658816.2018.1564317.

Mokbel, M., M. Sakr, L. Xiong, A. Züfle, J. Almeida, T. Anderson, W. Aref, G. Andrienko, N. Andrienko, Y. Cao, et al. 2024. Mobility data science: Perspectives and challenges. *ACM Transactions on Spatial Algorithms and Systems* 10 (2):1–35. doi: 10.1145/3652158.

Monreale, A., and R. Pellungrini. 2023. A survey on privacy in human mobility. *Transactions on Data Privacy* 16 (1):51–82. https://arpi.unipi.it/handle/11568/1176246.

Nanni, M., L. Tortosa, J. F. Vicent, and G. Yeghikyan. 2020. Ranking places in attributed temporal urban mobility networks. *PLoS ONE* 15 (10):e0239319. doi: 10.1371/JOURNAL.PONE.0239319.

Nathan, R., C. T. Monk, R. Arlinghaus, T. Adam, J. Alós, M. Assaf, H. Baktoft, C. E. Beardsworth, M. G. Bertram, A. I. Bijleveld, et al. 2022. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* 375 (6582):eabg1780. doi: 10.1126/science.abg1780.

National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and replicability in science*. Washington, DC: National Academies Press. doi: 10.17226/25303.

Nelson, T. A., M. F. Goodchild, D. J. Wright, S. Elwood, D. Liverman, and N. Schuurmann. 2022. Accelerating ethics, empathy, and equity in geographic information science. *Proceedings of the National Academy of Sciences of the United States of America* 119 (19):e2119967119. doi: 10.1073/PNAS.2119967119.

Nie, T., J. Sun, and W. Ma. 2025. Exploring the roles of large language models in reshaping transportation systems: A survey, framework, and roadmap. *Artificial Intelligence for Transportation* 1 (July):100003. doi: 10.1016/j.ait.2025.100003.

Noi, E., and S. Dodge. 2023. A data fusion framework for exploring mobility around disruptive events. In *12th International Conference on Geographic Information Science (GIScience 2023)*. Leibniz International Proceedings in Informatics (LIPIcs), Vol. 277, 57:1–57:6. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi: 10.4230/LIPIcs.GIScience.2023.57.

NuStats. 2013. *2010–2012 California household travel survey final report*. NuStats Research Solution, LLC. https://www.nlr.gov/transportation/secure-transportation-data/tsdc-california-travel-survey

Oliver, N., B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M. De Nadai, E. Letouzé, A. A. Salah, R. Benjamins, C. Cattuto, et al. 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* 6 (23):eabc0764. doi: 10.1126/sciadv.abc0764.

Ounoughi, C., and S. Ben Yahia. 2023. Data fusion for ITS: A systematic literature review. *Information Fusion* 89 (January):267–91. doi: 10.1016/j.inffus.2022.08.016.

Papadogeorgou, G., and S. Samanta. 2023. Spatial causal inference in the presence of unmeasured confounding and interference. https://arxiv.org/pdf/2303.08218.

Pappalardo, L., E. Manley, V. Sekara, and L. Alessandretti. 2023. Future directions in human mobility science. *Nature Computational Science* 3 (7):588–600. doi: 10.1038/S43588-023-00469-4.

Pappalardo, L., F. Simini, G. Barlacchi, and R. Pellungrini. 2022. Scikit-mobility: A Python library for the analysis, generation, and risk assessment of mobility data. *Journal of Statistical Software* 103 (4):1–38. doi: 10.18637/jss.v103.i04.

Parent, C., N. Pelekis, Y. Theodoridis, Z. Yan, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, et al. 2013. Semantic trajectories modeling and analysis. *ACM Computing Surveys* 45 (4):1–32. doi: 10.1145/2501654.2501656.

Park, Y. M., and M.-P. Kwan. 2025. Revisiting the uncertain geographic context problem: Expanding its scope to include indoor geographic contexts and dynamics in environmental health and social science research. *Annals of the American Association of Geographers* 115 (5):1055–70. doi: 10.1080/24694452.2025.2472974.

Perchoux, C., Y. Kestens, F. Thomas, A. Van Hulst, B. Thierry, and B. Chaix. 2014. Assessing patterns of spatial behavior in health studies: Their socio-demographic determinants and associations with transportation modes (the RECORD Cohort Study). *Social Science & Medicine* 119 (October):64–73. doi: 10.1016/J.SOCSCIMED.2014.07.026.

Rana, M., and V. Mittal. 2021. Wearable sensors for real-time kinematics analysis in sports: A review. *IEEE Sensors Journal* 21 (2):1187–1207. doi: 10.1109/JSEN.2020.3019016.

Rao, J., S. Gao, and S. Zhu. 2023. CATS: Conditional adversarial trajectory synthesis for privacy-preserving trajectory data publication using deep learning approaches. *International Journal of Geographical Information Science* 37 (12):2538–74. doi: 10.1080/13658816.2023.2262550.

Rey, S. J., and L. Anselin. 2010. PySAL: A Python library of spatial analytical methods. In *Handbook of applied spatial analysis*, ed. M. M. Fischer and A. Getis, 175–93. Berlin: Springer. doi: 10.1007/978-3-642-03647-7_11.

Röcke, C., M. Luo, P. Bereuter, M. Katana, M. Fillekes, V. Gehriger, A. Sofios, M. Martin, and R. Weibel. 2023. Charting everyday activities in later life: Study protocol of the mobility, activity, and social interactions study (MOASIS). *Frontiers in Psychology* 13 (January):1011177. doi: 10.3389/FPSYG.2022.1011177.

Runge, J., A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls. 2023. Causal inference for time series. *Nature Reviews Earth & Environment* 4 (7):487–505. doi: 10.1038/s43017-023-00431-y.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 115 (3):211–52. doi: 10.1007/S11263-015-0816-Y/FIGURES/16.

Shannon, J., and K. Walker. 2018. Opening GIScience: A process-based approach. *International Journal of Geographical Information Science* 32 (10):1911–26. doi: 10.1080/13658816.2018.1464167.

Simini, F., G. Barlacchi, M. Luca, and L. Pappalardo. 2021. A deep gravity model for mobility flows generation. *Nature Communications* 12 (1):6576. doi: 10.1038/S41467-021-26752-4.

Sourbati, M., and F. Behrendt. 2021. Smart mobility, age and data justice. *New Media & Society* 23 (6):1398–1414. doi: 10.1177/1461444820902682.

Su, R., Y. Liu, and S. Dodge. 2024. ORTEGA v1.0: An open-source Python package for context-aware interaction analysis using movement data. *Movement Ecology* 12 (1):20. doi: 10.1186/s40462-024-00460-2.

Su, R., E. C. McBride, and K. G. Goulias. 2020. Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *Transportation Research Part C: Emerging Technologies* 120 (April):102796. doi: 10.1016/j.trc.2020.102796.

Su, R., N. Newsham, and S. Dodge. 2024. Spatiotemporal dynamics of ethnoracial diversity and segregation in Los Angeles County: Insights from mobile phone data. *Computers, Environment and Urban Systems* 114 (December):102203. doi: 10.1016/j.compenvurbsys.2024.102203.

Technitis, G., W. Othman, K. Safi, and R. Weibel. 2015. From A to B, randomly: An endpoint-to-endpoint random trajectory generator for animal movement. *International Journal of Geographical Information Science* 29 (6):912–34. doi: 10.1080/13658816.2014.999682.

Tritsarolis, A., Y. Kontoulis, and Y. Theodoridis. 2025. The Piraeus AIS dataset for large-scale maritime data analytics. Accessed August 6, 2025. doi: 10.5281/ZENODO.6323416.

Van de Weghe, N., L. De Sloover, A. G. Cohn, H. Huang, S. Scheider, R. Sieber, S. Timpf, and C. Claramunt. 2025. Opportunities and challenges of integrating geographic information science and large language models. *Journal of Spatial Information Science* 30 (30):93–116. doi: 10.5311/JOSIS.2025.30.389.

Wan, Z., and S. Dodge. 2025. Medark: A map-matching error detection and rectification framework for vehicle trajectories. *International Journal of Geographical Information Science* 39 (4):872–99. doi: 10.1080/13658816.2024.2436482.

Wang, R., Y. Xin, Y. Zhang, F. Perez-Cruz, and M. Raubal. 2025. Counterfactual explanations for deep learning-based traffic forecasting. *Communications in Transportation Research* 5 (December):100176. doi: 10.1016/j.commtr.2025.100176.

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3 (1):160018. doi: 10.1038/sdata.2016.18.

Wu, X., Y. Shen, C. Shan, K. Song, S. Wang, B. Zhang, J. Feng, H. Cheng, W. Chen, Y. Xiong, et al. 2024. Can graph learning improve planning in LLM-based agents? Advances in Neural Information Processing Systems 37:5338–83. doi: 10.52202/079017-0173.

Xin, M., A. Shalaby, S. Feng, and H. Zhao. 2021. Impacts of COVID-19 on urban rail transit ridership using the synthetic control method. *Transport Policy* 111:1–16. doi: 10.1016/J.TRANPOL.2021.07.006.

Xin, Y., N. Tagasovska, F. Perez-Cruz, and M. Raubal. 2022, November. Vision paper: Causal inference for interpretable and robust machine learning in mobility analysis. In *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, ed. M. Renz and M. Sarwat. New York: Association for Computing Machinery. doi: 10.1145/3557915.3561473.

Xu, H., J. Yuan, A. Zhou, G. Xu, W. Li, X. J. Ban, and X. Ye. 2025. Generative artificial intelligence–powered multi-agent paradigm for smart urban mobility: Opportunities and challenges for integrating large language models and retrieval augmented generation with intelligent transportation systems. In *Urban human mobility: Practices, analytics, and strategies for smart cities*, 123–37. CRC Press. doi: 10.1201/9781003503262-14.

Xu, W. 2022. The contingency of neighbourhood diversity: Variation of social context using mobile phone application data. *Urban Studies* 59 (4):851–69. doi: 10.1177/00420980211019637.

Yan, Z., J. Macedo, C. Parent, and S. Spaccapietra. 2008. Trajectory ontologies and queries. *Transactions in GIS* 12 (Suppl. 1):75–91. doi: 10.1111/j.1467-9671.2008.01137.x.

Yao, A., S. Pal, X. Li, Z. Zhang, C. Dong, F. Jiang, and X. Liu. 2024. A privacy-preserving location data collection framework for intelligent systems in edge computing. *Ad Hoc Networks*.161 (August):103532. doi: 10.1016/j.adhoc.2024.103532.

Yu, D., R. Bao, R. Ning, J. Peng, G. Mai, and L. Zhao. 2025. Spatial-RAG: Spatial retrieval augmented generation for real-world geospatial reasoning questions. https://arxiv.org/pdf/2502.18470.

Yuan, Y., J. Ding, D. Jin, and Y. Li. 2025. Learning the complexity of urban mobility with deep generative network. *PNAS Nexus* 4 (5):pgaf081. doi: 10.1093/PNASNEXUS/PGAF081.

Yuan, Y., Y. Zhang, J. Ding, and Y. Li. 2025. World move, a global open data for human mobility. https://arxiv.org/pdf/2504.10506.

Zheng, Y., H. Fu, X. Xie, W.-Y. Ma, and Q. Li. 2011. *Geolife GPS Trajectory Dataset – User Guide*. ACM Press. https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/.

Zhong, S., and L. Bian. 2023. How regularly do people visit service places? *Computers, Environment and Urban Systems* 99 (January):101896. doi: 10.1016/J.COMPENVURBSYS.2022.101896.

Zhu, X. X., D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* 5 (4):8–36. doi: 10.1109/MGRS.2017.2762307.

Zhu, Y., J. J. Yu, X. Zhao, X. Wei, and Y. Liang. 2024. UniTraj: Learning a universal trajectory foundation model from billion-scale worldwide traces. https://arxiv.org/pdf/2411.03859.

SOMAYEH DODGE is an Associate Professor of Spatial Data Science in the Department of Geography at the University of California, Santa Barbara, Santa Barbara, CA 93106. E-mail: sdodge@ucsb.edu. Her research develops spatiotemporal data analytics, machine learning, modeling, and visualization techniques to study movement in human and ecological systems.

ROBERT WEIBEL is a Professor Emeritus of Geographic Information Science at the University of Zurich (UZH), 8006 Zurich, Switzerland, now affiliated with the UZH Healthy Longevity Center. E-mail: robert.weibel@geo.uzh.ch. He is interested in mobility analytics with applications in health and animal ecology, spatial analysis for linguistic applications, and computational cartography.

CHRISTOPHE CLARAMUNT is a Professor of Computer Science at the Naval Academy Research Institute Lanvéoc, 29160, France. E-mail: christophe.clarammunt@ecole-navale.fr. His research focuses on the conceptualization and development of spatiotemporal and computational models, as well as their application to environmental and urban sciences.

GAVIN MCARDLE is an Associate Professor in the School of Computer Science at University College Dublin, Belfield, Dublin 4, Ireland. E-mail: gavin.mcardle@ucd.ie. His research focuses on data analytics, machine learning, and spatial data analysis, particularly in environmental and urban science.

ZHIYONG ZHOU is a Postdoctoral Researcher in the Department of Geography at the University of Zurich, 8006 Zurich, Switzerland, and the Department of Geography at the University of Wisconsin–Madison, Madison, WI 53706. E-mail: zhiyong.zhou@wisc.edu. His research focuses on spatial data generalization, smart mobility, and sustainable built environments.

YANAN XIN is an Assistant Professor at the Department of Transport and Planning at Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands. E-mail: yanan.xin@tudelft.nl. Her research focuses on developing causality-inspired AI methods that are interpretable and robust to support mobility modeling, prediction, and policy intervention.

ANITA GRASER is a Senior Scientist at the Austrian Institute of Technology, 1210 Vienna, Austria. E-mail: anita.graser@ait.ac.at. Her research focuses on developing spatiotemporal analytics methods and open-source research software with a focus on mobility applications.

TUMASCH REICHENBACHER is a Group Leader at the Geographic Information Science Center, University of Zurich, 8006 Zurich, Switzerland. E-mail: tumasch.reichenbacher@uzh.ch. His research interests focus on mobility analytics, mobile cartography, and location-based services.