Title
A Data Science Framework for Movement

Journal
Geographical Analysis, 53(1)

ISSN
0016-7363

Author
Dodge, Somayeh

Publication Date
2021

DOI
10.1111/gean.12212

Peer reviewed

GEOGRAPHICAL ANALYSIS
50th Anniversary Special Issue

A Data Science Framework for Movement

Somayeh Dodge

Correspondence
Department of Geography, University of California Santa Barbara
Email: sdodge@ucsb.edu

Funding information

Movement is the driving force behind the form and function of many ecological and human systems. Identification and analysis of movement patterns that may relate to behavior of individuals and their interactions is a fundamental first step in understanding these systems. With advances in IoT and the ubiquity of smart connected sensors to collect movement and contextual data, we now have access to a wealth of geo-enriched high-resolution tracking data. These data promise new forms of knowledge and insight into movement of humans, animals, and goods, and hence can increase our understanding of complex spatiotemporal processes such as disease outbreak, urban mobility, migration, or effects of human activity on behavior of competing species. To take advantage of the evolution in our data, we need a revolution in how we visualize, model, and analyze movement as a multidimensional process that involves space, time, and context. This paper introduces a data science paradigm with the aim of advancing research on movement.

KEYWORDS
Spatial Data Science, Movement Analytics, Movement Model, Prediction, Trajectory Data, Patterns, Process

Abbreviations: GPS, Global Positioning Systems; LAT, Location-Aware Technologies; RFID, radio-frequency identification; GIScience, Geographic Information Science.

Corresponding author.

1 | INTRODUCTION

Motion is an integral characteristic of almost all organisms and spatiotemporal phenomena (Nathan et al., 2008). As such, it is an important driving mechanism behind the operation of many ecological and human systems, shaping individuals trajectories and their interactions across multiple spatial and temporal scales. Today, with the ubiquity of location-aware technologies (LATs) and the Internet of Things (IoT), we now have an unprecedented ability to track almost any mobile entity anywhere on the planet. High frequency tracking of individuals and things is becoming cheaper and more accurate through the increasing usage of smart sensors, social media, and activity loggers. With the major breakthrough in location-aware and multimodal sensors, ecologists have gained access to new forms of information about organisms. Although continuous and long-term tracking of wildlife still remains challenging, especially in remote areas and for endangered species, we are able to follow migratory species remotely along their annual migrations and monitor locomotion of animals (Kays et al., 2015).

Like many other fields (e.g. business, behavioral sciences, intelligence), the study of movement has significantly benefited from the ubiquitous collection of tracking data at large volumes and fine spatial and temporal granularities. Before the advent of satellite positioning and radio telemetry technologies in the twentieth century, our ability to track and study movement was limited to direct observations of moving individuals, travel surveys, sporadic presence-absence logs, cellphone usage signals, and close range video-tracking. Since then, movement data has evolved from coarse sporadic observations (i.e. presence-absence information in space and time) to fine resolution geo-enriched trajectories (i.e. multidimensional time series of locations representing footprints of individuals and their ambient context). Today using tracking, we know *where* things and individuals are exactly and *how* they move in space and time (Miller et al., 2019a). More importantly, we have the ability to record new forms of information about the condition of the environment, geographic context, and even the type of activity and the behavioral mode of individuals using multimodal sensors, thermometers, accelerometers, gyroscopes, heart rate sensors and so on. Contextual information and other ancillary data can be linked and analyzed with raw movement observations to increase our understanding of *why* individuals move the way they do and *what* do they do along their movement paths (Lee and Kwan, 2018; Wu et al., 2014; Dodge et al., 2013; Wilson et al., 2012). While widespread tracking, especially in the case of human movement, raises serious privacy concerns and security issues (Valentino-DeVries et al., 2018), it provides a tremendous opportunity to generate new insights into the behavior of moving individuals (Miller et al., 2019a). The question is how these new forms and large arrays of data can be exploited for a deeper understanding of complex dynamic processes such as immigration, migration of animals, spread of diseases, natural disasters, autonomous navigation, smart and sustainable mobility.

With the rise of 'Big Data' (Cukier and Mayer-Schoenberger, 2013), the Geographic Information Science (GIScience) community has played a key role in advancing methodologies and computational techniques for the collection, processing, analysis, visualization, and sense-making of large sets of geographic data. As a result, GIScience has intersected with data science to a great extent, specifically where the data has a spatiotemporal nature. On one hand, GIScience leverages data science approaches such as parallel computing, machine learning, artificial intelligence, and data mining techniques to analyze and handle large volumes of data sets for geographic knowledge discovery and predictions (Bowlick and Wright, 2018). On the other hand, GIScience continues to make significant contributions to data science by introducing more meaningful and spatially aware analytics (involving spatial indexing, spatial autocorrelation, spatial heterogeneity, geographic visualization, geovisual analytics and spatialization) (Singleton and Arribas-Bel, 2019). Spatial thinking has been infused in data-driven computing and cyberinfrastructures to analyze and map large arrays of geospatial data sets which are streamed in a variety of formats, granularity, and dimensions (UCGIS, 2018). Geographic visualization has become an integral part of data science as it aids human understanding of spatial patterns and analytical reasoning of geographic processes (Andrienko et al., 2017). In particular, visualization, processing, and analysis of multivariate and high frequency movement data have gained significant attention in GIScience, because of its ubiquity and importance for knowledge discovery in a broad application domain (Long et al., 2018; Holloway and Miller, 2018; Dodge, 2016).

Movement is a complex multidimensional process, and its study requires an interdisciplinary approach. While GIScience has significantly advanced computational movement analytics (Long et al., 2018), with the emergence of new forms of movement data, new interdisciplinary approaches are needed to revolutionize the way we analyze and model movement. Examples of these new forms of movement data include trajectory data streams (Coelho et al., 2016), or data obtained from wearable motion sensors equipped with magnetometers and accelerometers (Akhavian and Behzadan, 2015; Shoaib et al., 2014; Nathan et al., 2012), social network activities (Mckenzie et al., 2013; Azmandian et al., 2013), contextual data from remote sensing (Dodge et al., 2013), camera-traps (Karanth and Nichols, 1998) and surveillance cameras (Choi et al., 2013), checkpoint data (Tao et al., 2018), observations through IoT and Information Communications Technology (ICT) (Batty et al., 2012) and geosensor networks (Laube et al., 2011; Both et al., 2013). As movement data become increasingly available at large volumes, high dimensions, and multiple granularities, development of new computational data analytics and modeling approaches have gained momentum to advance movement research (Thums et al., 2018; Kays et al., 2015; Demšar et al., 2015; Dodge et al., 2013). There is a growing need to understand how these unique forms of linked and multidimensional data can best be exploited to make new discoveries about the behavior of moving individuals and their interactions in dynamic systems. This necessitates a new research paradigm to better study and understand movement as a multidimensional process involving space, time, and context.

This paper intends to promote leveraging the data science paradigm in advancing research on movement. Data science builds on a combination of expertise from computer science, applied sciences (for domain knowledge), math and statistics (Figure 1). It takes an interdisciplinary approach to convert large amounts of data into insights and useful predictions often using statistical research, data analytics, and machine learning (Blei and Smyth, 2017). In this paper, I introduce a data science framework for movement by integrating the main components of the data science paradigm (Figure 1) into the continuum of movement research introduced previously in Dodge (2016). The framework brings together **data-driven analytics** and **theory-driven modeling** of movement, through an interdisciplinary approach supported by **data**, **domain knowledge** and **visualization**. The aim is to blend data and knowledge on both movement and its embedding context for discovery of new insights about the behavior of moving individuals, interaction of mobile agents, and prediction of changes in dynamic systems. The premise is that the study of movement in various applications (e.g. human mobility, public health, movement ecology) can benefit from high-performance computing, intelligent algorithms and machine learning approaches, especially when movement data are multidimensional and represent the collective behavior of a large number of moving individuals.

2 | A DATA SCIENCE FRAMEWORK FOR MOVEMENT

This section introduces a framework as a marriage of the data science paradigm (Figure 1) and the continuum of movement research developed in (Dodge, 2016). The framework enables the study of movement through an iterative process (as shown in Figure 2), in which raw movement observations are transformed to meaningful information about the behavior of movement processes through data-driven analysis and knowledge discovery approaches. This knowledge is then used to enhance modeling and prediction of movement with the aid of theories, and mathematical and physical principles. These models can increase our knowledge of movement, which can lead to a better understanding

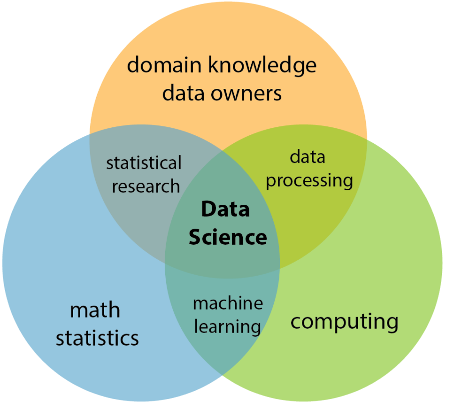

FIGURE 1 Data science paradigm, modified from Conway (2010).

of how movement patterns are formed and how mobile individuals interact with their environment. This process is heavily reliant on movement observations (i.e. tracking data sets), computing, math, statistics, spatial knowledge, and domain knowledge.

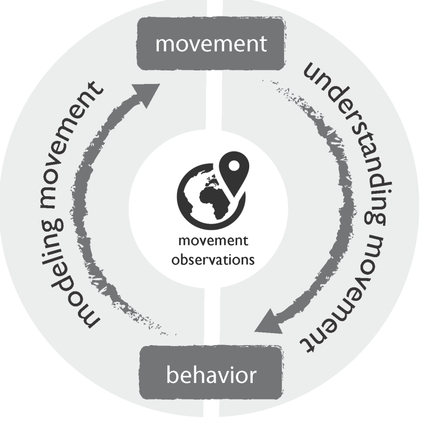

FIGURE 2 Process model for the study of movement

Figure 3 visualizes a the proposed framework. This is a multi-layered system, in which all components are tightly intertwined and contribute to one another through an iterative research process. The main components of this framework consist of **data**, **visualization**, **data-driven analytics**, **theory-driven modeling**, and **domain knowledge**. As in any spatial data science system, analytics and modeling of movement is facilitated by computing approaches and theories from Computer Science, GIScience, Mathematics, and Statistics. The right side of the framework supports a bottom-up approach in which raw observations are analyzed for finding patterns and associations in the data through data-driven methods. The aim is to infer the process from observed patterns in order to advance our understanding of dynamic processes in social and ecological systems. While the left side of the framework is assisted by a top-down approach to enable a deductive science of movement. The aim is to combine insights obtained from raw data, domain knowledge, and theories to enhance theory-driven modeling and prediction of movement and behavioral changes given a set of conditions.

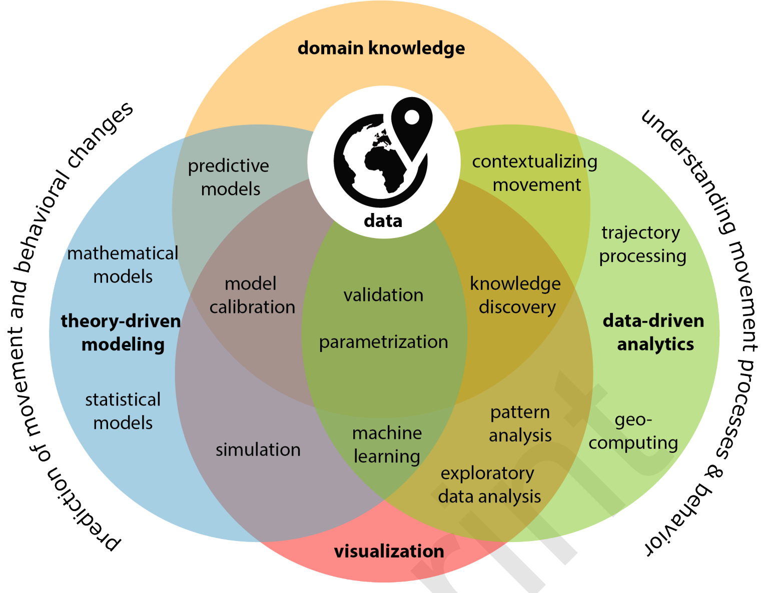

FIGURE 3 The data science framework for movement

At the heart of this framework are movement and contextual **data**, which are raw observations collected in a great volume, velocity, and veracity every day. Using sensor fusion or modern tracking sensors, location data can be recorded at high accuracies (i.e. centimeter and millimeter) for example for mobile robots, drones, and automated guided vehicles to help them navigate more autonomously (Sukkarieh et al., 1999; Yudanto and Petre, 2015; Whitton, 2019). High-resolution and long-term movement observations enable **data-driven analytics** to generate insights on the behavior of individuals and increase our knowledge of movement processes through geo-computing, pattern mining, machine learning, and knowledge discovery techniques. The generated knowledge from the analytics can be used in conjunction with **theory-driven modeling** approaches to facilitate predictions and reconstruction of movement paths of individuals. In this framework, the incorporation of **domain knowledge** and collaboration with domain scientists is vital to assure meaningful outcomes, reliable interpretations, and verification of observed patterns and predictions. **Visualization** plays a key role to streamline such interdisciplinary work and communication of the results and patterns (Dodge, 2018). It also facilitates the validation of analytical algorithms and monitoring the work progress of methodologies. Validation of results, parametrization of models and calibration of computational techniques are also enhanced through the use of actual observations and meaningful visualization techniques. The following sections describe different layers of the framework and their connections in more details.

2.1 | Data

Movement observations convey important information about the behavior and interactions of individuals along their space-time lifeline. These observations are collected in different forms and granularity. Depending on the tracking approach, raw observations can describe movement of individuals in two different perspectives, Lagrangian or Eulerian as shown in Figure 4. In the Lagrangian approach, individuals are observed and followed along their movement path over time mainly using LAT sensors such as GPS or RFID tags. This approach directly constructs a discrete *trajectory* (i.e. a time-ordered sequence of locations) of the entity. In the Eulerian approach, movement of individuals is observed at fixed locations using local sensors such as cameras (e.g. camera traps, security cameras), wifi stations, cell towers, transit gates, or through social media check-ins. The Eulerian approach specifies the presence or absence of a moving entity in various locations at certain times. Focusing on specific locations in the space through which the entities move as time passes, the Eulerian perspective can describe the patterns of space use by moving entities (e.g. home range of animals, activity space of humans). A time-ordered sequence of observations (i.e. presence) of an entity obtained from the Eulerian approach can be used to construct the Lagrangian specification of the entity's movement (or the so-called the trajectory). Trajectory data can be collected at regular or irregular time intervals depending on the type of sensors used for tracking or the tracking aims. Laube (2017, 2014) describes these two movement perspectives and different forms of movement data collection in a greater detail.

Regardless of the data collection approach, movement observations can be geo-enriched with environmental and behavioral information. That is, tracking data can be annotated with additional context variables at each location and timestamp (Mandel et al., 2011). Context variables can be related to behavioral parameters that are internal to the state of the mobile individual (e.g. age, gender, behavior, movement purpose), or they may represent external factors which influence the individual's movement (e.g. transport mode, weather condition, presence of other individuals, etc.). Nowadays, geo-enrichment can be done automatically at the same time of tracking, using new generation of LAT sensors capable of recording motion, direction, orientation, sound, video, weather condition simultaneously. Geo-enrichment of movement data can also be done after tracking, by applying trajectory annotation algorithms using remote sensing data of the environment or travel logs and surveys (Dodge et al., 2013). Long-term and high-frequency tracking of individuals and geo-enrichment of trajectory data can result in large volumes of multidimensional data of movement, and hence can provide promising opportunities for data science in this domain. These data sets capture movement of individuals either in a Euclidean space (represented as a continuous 2D or 3D space, regular grid, irregular tessellation) or a network space (represented as a graph, vector network). For more information on movement spaces, the readers are referred to Laube (2014). Miller et al. (2019a) compare two different types of movement data in the domains of animal movement in the Euclidean space and human mobility in the network space. They conclude that although there are significant differences in the collection and structure of tracking data in human and animal contexts, there is a great synergy in the types of research questions and analytical methods applied to animal movement and human mobility domains. Hence, a universal data science framework to study movement can potentially facilitate cross-collaboration between the two domains by promoting data-driven methods that can apply across the board — e.g. for measuring activity patterns, measuring interaction among moving entities, analyzing and visualizing movement in relation to its geographic and environmental context.

2.2 | Visualization

Visualization empowers data science. Visualization and visual analytics enable direct interaction with the data and hence can guide analysis. In data science, data visualization is used for effective understanding of data, revealing unknowns, reasoning and interpreting the results to support decision making (Van Der Aalst, 2016). Figure 5 highlights how visualization facilitates various processes involved in movement data science. The first common step in most data science research is to visualize raw observations, and speculate about what they represent and how to strategize data processing and analytics. In Geographic Information Science, information visualization enables researchers for

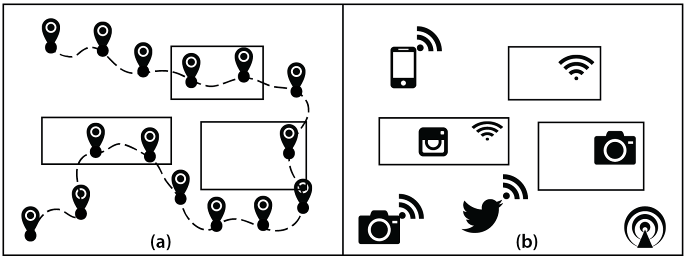

FIGURE 4 Movement observations in (a) Lagrangian and (b) Eulerian perspectives.

visual inspection of data records, summarizing data, and capturing geographic patterns. For example, Andrienko and Andrienko (2013) applied a set of visual exploration tools to superimpose temporal diagrams, mosaic diagrams, and bar diagrams over geographic maps to aggregate and summarize movement patterns of taxis in Milan and white storks in Africa. Using large tracking data sets, these visual data summaries can help to depict the patterns of movement at an aggregate level by hiding complex details of the underlying trajectories. Interaction with the data using visual displays can also help researchers to discover unknowns in their movement data sets, explore hidden relationships, and formulate hypotheses about the behavioral patterns represented in the data (Andrienko et al., 2008). This is especially important in interdisciplinary data science efforts. Andrienko and Andrienko (2012) review existing exploratory data analysis techniques for visualizing trajectories (e.g. using 2D static maps or space-time cubes), summarizing movement parameters along the trajectories (e.g. using trajectory segmentation and time bars), aggregating movement patterns in space and time (e.g. using clustering and treemaps), and exploring the relationship between movement and context (e.g. using multivariate representations).

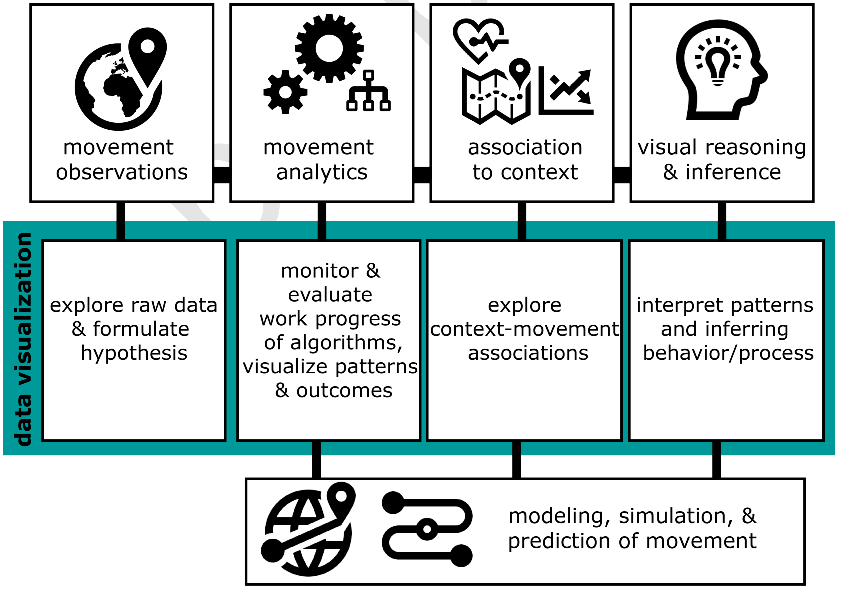

FIGURE 5 Visualization facilitates movement data science. Figure is modified from Dodge (2018).

In comparison to traditional static maps, dynamic and interactive representations are better suited to support the study of movement from the Lagrangian perspective as they can depict how entities move in space and time (Dodge, 2018; Miller et al., 2019a). In addition, multivariate and hybrid representations can help us explore complex relationships between movement and its environmental and geographic context (Andrienko and Andrienko, 2012). For example, Figure 6 uses a multivariate representation to map how a Galapagos albatross uses wind during its regular foraging flights between the Galapagos Island to the coastal areas of Peru. The visualization (Figure 6) illustrates albatross's strain into the crosswind (the wind blowing across the direction of movement) on the way to Peru. This forces the bird to fly slower (shown as thin red line in Figure 6b) along its outbound flight to the coast. In contrast, the albatross flight back to the Galapagos is assisted by a more favorable tailwind resulting in a higher movement speed (shown as thick blue line).

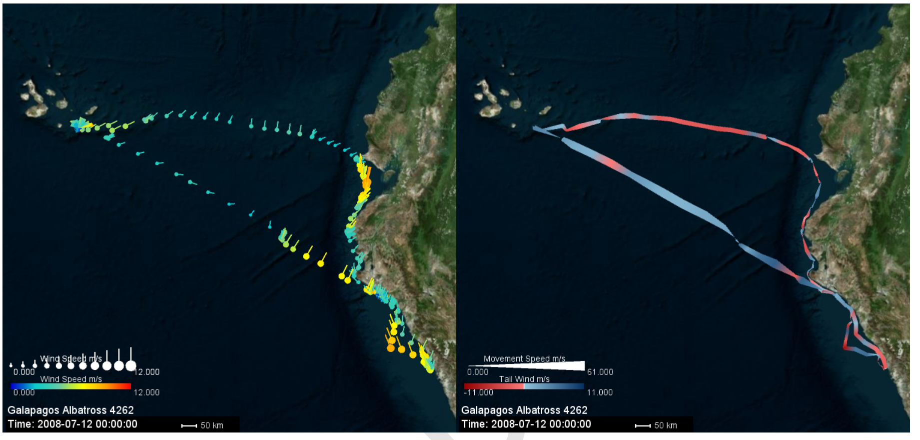

FIGURE 6 Map of an albatross movement track and its association with the wind: (a-left) using directional vectors for wind and color for tailwind speed, and (b-right) using thickness for movement speed and color for tailwind speed, created using DYNAMOvis (Dodge et al., 2018)

In modeling movement, visualization can be used to monitor the work process of the algorithms and simulations, and hence aids the validation process of methods and results. Finally, visualization can be used as a platform to communicate and interpret analytical results, and to promote visual reasoning to understand the observed patterns, extract meaningful information, and derive relevant knowledge.

2.3 | From data analytics to prediction

The study of movement in GIScience has had tremendous progress with the rapid progress in tracking technologies (Holloway and Miller, 2018). Early work in late 1990s and early 2000s involved quantification of trajectories, formalization of movement metrics, and analysis of movement patterns (Imfeld, 2000; Laube, 2005; Dodge et al., 2008; Nathan et al., 2008; Holyoak et al., 2008). The focus of these studies was more on animal movement. More recent work embarked on developing computational techniques for semantic knowledge discovery from geo-enriched movement data sets (Dodge et al., 2013; Parent et al., 2013; Long et al., 2018; Brum-Bastos et al., 2018; Miller et al., 2019b). These studies aimed to connect movement to its embedding environment and geographic context to investigate the underlying mechanisms of movement.

2.3.1 | Data-driven analytics

Highlighted with a green circle in Figure 3, data-driven analytics involve a set of computational techniques to analyze and mine trajectory data. Laube (2014, p.6) categorizes these techniques as 'Computational Movement Analysis (CMA)', defined as the use of "computational techniques for capturing, processing, managing, structuring, and ultimately analyzing data describing movement phenomena, both in geographic and abstract spaces, aiming for a better understanding of the processes governing that movement". Assisted by geocomputing and visualization, the aim of data-driven analytics is to extract patterns from observation and relate them to processes by connecting movement to its context. The ultimate goal is to learn about the behavior of mobile entities and the mechanisms behind movement processes at multiple scales. Guided by data science approaches such as deep learning and knowledge discovery techniques and often through interdisciplinary studies, a common workflow is as follows:

Exploratory data analysis The CMA process may start with an explanatory data visualization to inspect data, and formulate hypotheses and research questions. It then proceeds with outlier detection and elimination of erroneous data. Here visualization can also help to summarize patterns in data and facilitate communication between developers of computational methods and domain scientists (Andrienko and Andrienko, 2012).

Trajectory processing encompasses quantitative techniques for computing movement metrics and if needed decomposing trajectories using segmentation. Movement metrics are measurable quantities and derivatives of trajectories. They describe geometric, spatial and temporal characteristics of trajectories such as basic movement parameters (e.g. speed, turn angle, path tortuosity, etc.), trajectory's fractal dimension, movement rates, home range, first-passage time, return rate, etc. (Seidel et al., 2018). In this phase, often a data base management system and a spatial indexing technique (e.g. quad-tree, KD-tree, R-tree) are established for fast data access and efficient retrieval and query of trajectory data (Pfoser, 2002). Trajectory segmentation is the process of decomposing a long and convoluted trajectory into parts of similar characteristics (Ahearn and Dodge, 2018). It is used to simplify trajectory data for further analyses and pattern detection. It can also be used to distinguish various behavioral modes and different activities represented in the data.

Contextualizing movement involves using geo-enriched or annotated trajectory data and other contextual data sets (e.g. remote sensing data of the environment, demographic data, road networks, etc.) to explore relationships between movement and its embedding context. This step is usually done prior or parallel to movement pattern analysis and knowledge discovery phases. For data which lack contextual variables, first a trajectory annotation process is employed to attribute tracking data with environmental and behavioral variables and generate geo-enriched trajectories (Dodge et al., 2013). For heterogeneous data sets obtained from multiple sensors in different structures and granularities, data fusion techniques can help to blend the data together to facilitate linking movement to its context. Once geo-enriched trajectories are generated, dependencies between movement and environmental parameters can be explored through hypothesis-driven statistical analysis or machine learning approaches. This area has recently attracted a number of studies to explore movement-context dependencies using data integration in a variety of applications such as movement ecology (Dodge et al., 2013; Nathan et al., 2012), human mobility (Lee and Kwan, 2018; Brum-Bastos et al., 2018), and engineering (Akhavian and Behzadan, 2015).

Pattern analysis entails computational approaches to formalize and analyze movement patterns (Dodge et al., 2008; Laube, 2009). Movement patterns are regularities, arrangements, and structures that are existed in tracking data. Examples of movement patterns include convergence, divergence, single files, moving clusters, trend-setting, avoidance, attraction, flocking, etc. (Dodge et al., 2008). These movement patterns can be seen as footprints of processes or behavioral signatures in dynamic systems (O'Sullivan and Perry, 2013). For instance, the similarity in the daily trajectory of people can inform us about their regular commuting behavior, or the similarity of albatross movement patterns can be indicative of commonalities in their foraging behaviors. In order to quantify how similar movement of individuals are in space and time, various similarity analysis methods have been proposed using distance measures such as Euclidian distance, edit distance, Fréchet distance, longest common subsequence, dynamic time warping, to name a few (Dodge et al., 2012; Miller et al., 2019b). Next, computed similarities can be used together with a clustering technique in mining similarity patterns and clusters in the data.

Knowledge discovery & data mining (KDD) involve applying data mining and machine learning algorithms to detect and classify patterns and extract rules and associations in the data, often using large and multidimensional tracking data sets (Miller and Han, 2001; Mennis and Guo, 2009). Data science approaches are used to detect similarities, clusters, anomalies, interactions, and collective patterns in the movement of multiple individuals (Lee and Kwan, 2018; Soleymani et al., 2014; Zhong et al., 2015; Li et al., 2007). The aim is to transfer raw observations to new knowledge about forms and functions of dynamic processes and to develop useful insight into the behavior of individuals and their interactions. Movement research can benefit from high-performance geocomputing approaches to train and apply machine learning algorithms in extracting patterns and dependencies in large geo-enriched trajectory data sets. For instance, Miller and Han (2001); Cao and Philip (2012); Giannotti and Pedreschi (2008); Holzinger and Jurisica (2014); Jiang and Shekhar (2017); Evans et al. (2019) document recent advancements and challenges of KDD using large data sets in different domains such as public health, transportation, climate change, and disaster management. Provost and Fawcett (2013) argue that a data science perspective to knowledge discovery goes beyond merely data mining using geo-computing, and incorporation of domain knowledge of the processes is necessary. Therefore in this framework, 'knowledge discovery' is situated at the intersection of 'domain knowledge' and 'data-driven analytics'.

2.3.2 | Modeling and prediction of movement

As highlighted in a blue circle in Figure 3, the study of movement relies on theories and domain knowledge for computational modeling and simulation of movement with the aim of gaining a deeper understanding and more realistic predictions of movement and behavioral changes in dynamic systems. Models capture a simple representation of movement and its patterns (O'Sullivan and Perry, 2013) and hence can enable exploratory simulations of complex systems (Manson and O'Sullivan, 2006). This is especially important in cases for which observational data is limited, the system is not accessible or unsafe for empirical experiments, or ethical concerns over privacy or security is evident (e.g. modeling virus spread, migration of endangered species, human exposure to risk).

Generally speaking, given a set of conditions or for a given behavior, the aim of modeling is to: (1) predict the path or activity range of moving individuals, (2) better understand behaviors and dynamic processes, (3) predict missing points and gap filling in tracking data, (4) reconstruct movement patterns, (5) estimate population distribution and diffusion of mobile agents, and (6) predict behavioral responses to a changing environment and vice versa. Depending on the purpose of the study, motion can be modeled at a range of scales (local to global) and at the individual or collective levels (Kokko, 2006). At local scales, the goal is to predict movement choices of individuals and their step-by-step locomotions. While at global scales, the activity space and general flows of individuals are modeled. In animal movement ecology, most studies focus on modeling animals' local choices and resource selection, for instance for modeling foraging strategies (Schick et al., 2008). While global models of movement are mainly used to estimate the home ranges of animals. In human mobility, the main emphasis is more on global patterns of movement such as origin-destination flows and activity spaces, rather than the exact route of individuals and their local movement decisions (Miller et al., 2019a). Modeling locomotions and local choices of movement become important for studying navigation strategies in human mobility and designing algorithms for robot navigation (Metz, 2019) and autonomous vehicles (Whitton, 2019). This section provides a more general overview and a brief summary of relevant modeling approaches. The readers are encouraged to see Schick et al. (2008) for a review of models in movement ecology, and Barbosa et al. (2018) for modeling human mobility.

Existing movement models can broadly be classified into two groups: (a) theory-driven techniques based on deterministic mathematical functions or likelihood-based stochastic models, (b) data-driven and machine learning-based approaches using statistical inferential models. Computational data analytics (described in Section 2.3.1) provides the foundation of data-driven models. In addition, they can contribute useful knowledge (i.e. more realistic movement metrics computed from actual observations) for better validation, calibration and parametrization of theory-driven models (Holloway, 2018; Downs et al., 2018). The most common theory-driven approaches to model locomotions of individuals include: random walks and their variations (including correlated random walks and biased correlated random walks) (Codling et al., 2008), Lévy flights (Rhee et al., 2011), and step-selection functions (Thurfjell et al., 2014). These models are often referred to as 'process models' which can predict movement at space and time for a given state based on the Lagrangian perspective (Schick et al., 2008). These process models can be integrated in more advanced models such as state-space models, hidden morkov models, or agent-based simulations to capture multiple behavioral states or the heterogenous context of trajectories. Furthermore, they can be enriched with incorporation of physiological and environmental parameters and domain knowledge to encapsulate movement-context dependencies for more realistic modeling (Patterson et al., 2008; Schick et al., 2008; Ahearn et al., 2017; Holloway, 2018). At global scales, models such as deterministic and probabilistic time-geography (Miller, 2005; Winter and Yin, 2010) and Brownian bridge models (Horne et al., 2007; Kranstauber et al., 2012) can be used to estimate the activity space or home ranges of mobile individuals. These models compute the probability of a mobile entity being present at a given location and time between a specific origin and a destination. With these models the Eulerian perspective of movement can be captured in terms of the presence-absence probability of individuals in space and time.

Movement models are usually incorporated in agent-based simulation systems and Monte Carlo approaches in order to study and understand complex dynamic systems, test hypotheses and scenarios, make inferences, and predict patterns. Data science approaches can streamline high-performance simulation and computation of complex models which often involve a large number of interacting agents. For instance, simulation models can be strengthened by integrating enhanced data models through pattern recognition from large data sets using machine learning, data mining, and artificial intelligence approaches to make more reliable and accurate predictions and for actionable knowledge creation (Dhar, 2013; Kim et al., 2017). Simulation modeling is based on our knowledge of causal relationships between elements in a given system, while data-driven modeling is based on correlations observed in the data. Data science advances simulation modeling through high-performance data modeling and computing. The key is to incorporate knowledge and new insights obtained from actual observations in the calibration and parametrization of simulations (Ahearn et al., 2017). For example, Akhavian and Behzadan (2015) applied process-level knowledge obtained from construction labor activity recognition using machine learning on activity data (captured with GPS, accelerometer and gyroscope using smartphone sensors) to verify and update the input parameters of simulation models for construction engineering and management. Besides providing means to enhance input parameters, data science has also helped to optimize simulations and prediction models through cloud computing and high-performance computing by enabling "multiple runs of large-scale simulation models to search for optimal or near-optimal solutions" (Xu et al., 2015, p.1550019-2).

2.4 | Domain Knowledge and Validation

An important success factor in data science is the incorporation of domain knowledge in computational algorithms and analytical reasoning (Provost and Fawcett, 2013). Even the most efficient and powerful computational models and analytical techniques might result in misleading or useless information, without having a full understanding of what the actual data represent or what types of information and patterns can be extracted from the data. In early computational geometry approaches analyze movement patterns, movement trajectories were treated as a set of isolated points over time, often as moving point objects. While purely geometric algorithms advanced our ability to find structures and patterns in movement data sets, the extracted patterns were lacking semantics and a clear connection to the processes creating these patterns (Laube, 2017; Dodge, 2016). Incorporation of context information and domain knowledge is essential to the validation and interpretation of patterns. Having meaningful research questions and hypotheses about the underlying processes captured in the data can increase the validity and applicability of developed computational methods. This becomes even more important when our data increasingly grow in size and dimension. Therefore, it is essential to include data owners and domain experts in computational research to ensure meaningful and useful products. Because of the ubiquity of movement and its broad applications, the science of movement is interdisciplinary by nature. Developers of computational techniques and quantitative experts (e.g. GIScientists, computer scientists) can significantly benefit from working closely with domain scientists (e.g. ecologists, social scientists, transportation planers, health scientists) for (a) formulation of sound hypotheses and research questions, (b) incorporation of valid parameters and domain knowledge, especially in rule-based machine learning algorithms, and (c) more accurate reasoning, verification, and interpretation of observed patterns and associations in the data. As mentioned earlier in Section 2.2, these collaborations can be facilitated with visualization as a common visual language.

While domain knowledge is essential in data science, it should not be treated as the main source for verification of the results and parametrization of computational models. Nowadays with our increasing access to actual tracking data, parametrization and calibration of algorithms are enhanced through deep learning and using knowledge derived from actual observations (Ahearn et al., 2017). Furthermore, data science relies on validation processes to ensure stable analytics and models. That is, machine learning algorithms and models developed based on a specific data set should be generalizable to other independent data. This can be achieved through rigorous cross-validation processes and sensitivity analyses of input parameters. Proper null models, which are often used as a base for validation in movement ecology, can enable the evaluation of the outcomes and observed patterns (O'Sullivan and Perry, 2013; Miller, 2015). "A null model approach involves generating a pattern based on randomization that lacks some process or mechanism of interest in order to test its influence" (Miller, 2015, p.349). As an example, Ahearn et al. (2017) applied a correlated random walk as a null model to evaluate a context-sensitive random walk for modeling movement of tigers through a Monte Carlo simulation. High-performance computing can facilitate applications of the null model approach for large-scale systems.

3 | CONCLUSION: WHAT IS NEXT?

This paper introduced a general data science framework to advance the study of movement. The framework leverages domain knowledge, data-driven analytics, theory-driven computational models, and visualization to transfer raw movement observations to meaningful knowledge discovery and predictions of movement. What has changed from the previous generations of movement analytics is that movement data sets are becoming available in larger volumes and in more heterogeneous and multifaceted forms. For example, using multimodal GPS sensors or smart phones equipped with magnetometer and accelerometer we can collect activity information of movement entities along with their trajectories. Using Internet of Things, geosensor networks, or data fusion, these trajectories can be linked to weather information, traffic information, images of the surrounding environment of movement, etc. The premise is that high-performance computing, powerful machine learning and visual analytics approaches are well suited to support recognition of patterns and discovery of dependencies in such large and linked movement observations to support the study and modeling of complex dynamic systems such as human dynamics (González et al., 2008; Candia et al., 2008; Shoaib et al., 2014; Thums et al., 2018), social networks (Azmandian et al., 2013; Wu et al., 2014; Jurdak et al., 2015), animal behaviors and ecosystems (Boettiger et al., 2011; Polansky et al., 2010; Soleymani et al., 2014), to name but a few.

Recently, a number of papers reviewed the state of computational movement research in Geographic Information Sciences and related disciplines (Miller et al., 2019a; Holloway and Miller, 2018; Long et al., 2018; Yuan, 2018; Thums et al., 2018). These articles provide a list of recommendations for future studies and call for development of novel and more efficient methodologies to address potential research challenges stemming from the emergences of new forms of geo-enriched movement data sets. To conclude this paper, below I highlight some of these challenges and open opportunities for future data science research in advancing the science of movement.

Having access to fine-resolution tracking data, we now have the means to develop new approaches for modeling movement across spatial and temporal scales. This will open up new opportunities for multi-scale modeling of movement by integrating a range of patterns from locomotions to more global scale movement patterns. Similarly, there is a need for further research into movement interaction and modeling the contribution of individuals to the larger population level patterns or the collective behavior of dynamic systems. New developments in movement studies highlight the promising potential of deep-learning, artificial intelligence and machine learning techniques for leveraging massive and multidimensional movement data sets in understanding human dynamics and ecological systems (Wu et al., 2014; Jurdak et al., 2015; Thums et al., 2018; Lee and Kwan, 2018; Yuan, 2018). Therefore, there is a growing need in leveraging data science and our unprecedented access to large volumes of high-resolution data of movement and context, for more realistic modeling of patterns and reliable predictions of behavioral changes in dynamic systems.

The need for effective data fusion techniques is growing for integrating a variety of data obtained from different types of sensors and collected in multiple forms and granularity. These data may include GPS observations, ancillary biological, behavioral and activity information, environmental data, social media posts and check-ins, and transaction traces. Geo-enriched tracking of individuals and goods through the Internet of Things (IoTs) and sensor networks are on the rise. Data fusion and sensor integration can advance the capacities of future movement analytics and deepen our understanding of social and natural systems. In the era of 'Big Data', we increasingly record Eulerian-type observations at fixed locations over time (e.g. time series of sales, transactions, weather information, pollution, traffic counts, disease cases, etc.). Accordingly, we have a higher access to Eulerian observations of moving individuals through ICT, Closed-circuit television (CCTV) and surveillance cameras, check points, transit nodes, traffic count sensors, and social media posts. While our ability to observe movement with the Lagrangian perspective have improved through modern trackers, the access to these types of trajectories remains limited due to the higher cost of data collection and concerns over the privacy of observed individuals. Data science and big data analytics provide promising tools for understanding and prediction of dynamic processes in human and natural systems using large arrays of Eulerian-type data sets (Tao et al., 2018; Sowmya and Suneetha, 2017; Miller, 2010). Computational movement studies have yet to fully embrace and extend the potential of data science approaches for movement data analytics based on the Lagrangian approach. New movement models and fusion techniques are required to integrate the Lagrangians and Eulerian specifications of observations to gain a better understanding of these processes from different perspectives. For example, data science approaches on indexing large heterogeneous data structures and flow simulations can be applied in the integration of Lagrangians and Eulerian perspectives in the study of movement (Sauer et al., 2017; Xia et al., 2017).

While context-aware methodologies are on the rise, as Laube (2017) suggests the semantic gap between patterns and processes still exits. Data science approaches for movement should consider the fact that movement is influenced by its embedding context and the observed patterns may be indicative of a behavior or other external processes (i.e. environmental and social processes). Early geo-computation algorithms for moving point pattern analysis were mainly based on the geometric representation of movement. Trajectories were treated as lines or a set of isolated points in space and time, ignoring the context of movement and the spatial and temporal dependencies of trajectory data. However, movement is a continuous dynamic process, and as such it requires careful consideration of autocorrelations and interdependencies in spatial, temporal, and context dimensions. In addition, data science can provide a great opportunity for incorporation of deep knowledge obtained from data-driven analytics into more effective predictive models. Inclusion of more contextual information on abiotic and biotic factors influencing movement is important for enriching movement models to fulfill their potential in analysis and prediction of movement. For this, collaboration with domain scientists who have a better understanding of the subject social and ecological systems is vital to the success of movement research.

It is important to emphasize the role of domain scientists and experts in movement analytics in developing any new method or computing approaches for movement. While developers of computational methods have good understanding of data fusion and analytics, their knowledge of the phenomena might be limited and this can compromise the accuracy of the outcomes (Roos, 2001). For example, it is common in data science to use parallel computing and data indexing approaches to optimize large data analytics by partitioning the data into smaller pieces, with often little attention to the semantic of artificially created data segments (Farmer and Keßler, 2016; Papadias et al., 2002). Further research should explore the potential of these methods on trajectory data sets, for which the data continuity in space and time is key to the integrity of the information.

Visualization provides a common language for communication in interdisciplinary research and facilitates the collaboration between domain experts, data owners, and developers of methods. An unresolved technical challenge here is how to create effective visualizations capable of representing multiple dimensions of movement (space, time, context) from a large number of entities in a single display (Andrienko et al., 2008). New forms of data require new data visualization techniques for proper display of movement and its patterns. Currently, we are lacking cartographic theories and principles for mapping movement and its patterns in cognitively plausible ways. Future research should evaluate how humans perceive movement patterns using dynamic, interactive, and multidimensional displays. A broader question is how to best capture movement and interaction of dynamic entities in space and time in a more effective and meaningful way.

Finally, there are other data-related methodological questions that require careful consideration in future movement research. As mobile data collection has become prevalent, the question is how data science can help meaningful knowledge discovery while protecting the privacy and security of tracked individuals in the context of both humans or endangered species? Analytical and visualization approaches for movement should be enhanced with proper geomasking, generalization and aggregation techniques to avoid inferential disclosure of private information of individuals and ensure safe delivery of information (Andrienko and Andrienko, 2011; Kounadi and Resch, 2018). Another challenge is to account for uncertainty, bias, and error in observations, especially when using volunteered or participatory sensing of moving entities. Movement research may benefit from advanced statistical and data science approaches applied in other disciplines to alleviate this problem through data linkage and cross-validation using multiple sources of data (Ruiz-Gutierrez et al., 2016; Chen et al., 2017; Lavigne et al., 2018). Finally, data fusion and knowledge discovery should be informed and guided by the scale of observations, especially when integrating heterogeneous movement and context data sets from multiple sources. This is essential in movement data science in order to make relevant inferences from observed patterns at proper granularity and frequency levels in space and time (Ahearn and Dodge, 2018).

This paper intended to provoke thoughts on leveraging data science to advance knowledge of movement. Movement research is rapidly progressing owing to the tremendous advances in data collection and computing technologies. Therefore, future studies can thrive on the fusion of data science and movement science. The framework presented here encourages using actual movement observations and domain knowledge, to bring together bottom-up data-driven analytics and top-down theory-driven predictive models for a more holistic study of movement processes in natural and human systems. As such, the knowledge derived from analytics is used to advance theoretical models for a deeper understanding and reliable predictions of movement.

references

Ahearn, S. C. and Dodge, S. (2018) Recursive multi-frequency segmentation of movement trajectories (ReMuS). *Methods in Ecology and Evolution*, **9**, 1075–1087.

Ahearn, S. C., Dodge, S., Simcharoen, A., Xavier, G. and Smith, J. L. D. (2017) A context-sensitive correlated random walk: a new simulation model for movement. *International Journal of Geographical Information Science*, **31**, 867–883. URL: http://dx.doi.org/10.1080/13658816.2016.1224887.

Akhavian, R. and Behzadan, A. H. (2015) Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. *Advanced Engineering Informatics*, **29**, 867–877. URL: http://dx.doi.org/10.1016/j.aei.2015.03.001.

Andrienko, G. and Andrienko, N. (2013) A General Framework for Using Aggregation in Visual Exploration of Movement Data A General Framework for Using Aggregation in Visual Exploration of Movement Data. *The Cartographic Journal*, **47**, 22–40.

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I. and Wachowicz, M. (2008) Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. *Information Visualization*, **7**, 173–180.

Andrienko, G., Andrienko, N. and Weibel, R. (2017) Geographic data science. *IEEE Computer Graphics and Applications*, **37**, 15–17.

Andrienko, N. and Andrienko, G. (2011) Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, **17**, 205–219.

— (2012) Visual analytics of movement : An overview of methods , tools and procedures. *Information Visualization*, **12**, 3–24.

Azmandian, M., Singh, K. and Gelsey, B. (2013) Following Human Mobility Using Tweets. *International Workshop on Agents and Data Mining Interaction, Lecture Notes in Computer Science*, **7607**, 139–149.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F. and Tomasini, M. (2018) Human mobility: Models and applications. *Physics Reports*, **734**, 1–74. URL: https://doi.org/10.1016/j.physrep.2018.01.001.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A. and Bazzani, A. (2012) Smart cities of the future. *The European Physical Journal Special Topics*, **518**, 481–518.

Blei, D. M. and Smyth, P. (2017) Science and data science. *Proceedings of the National Academy of Sciences*, **114**, 8689–8692. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1702076114.

Boettiger, A. N., Wittemyer, G., Starfield, R., Volrath, F., Douglas-Hamilton, I. and Getz, W. M. (2011) Inferring ecological and behavioral drivers of African elephant movement using a linear filtering approach. *Ecology*, **92**, 1648–57. URL: http://www.ncbi.nlm.nih.gov/pubmed/21905431.

Both, A., Duckham, M., Laube, P., Wark, T. and Yeoman, J. (2013) Decentralized Monitoring of Moving Objects in a Transportation Network Augmented with Checkpoints. *The Computer Journal*, **56**, 1432–1449.

Bowlick, F. J. and Wright, D. J. (2018) Digital Data-Centric Geography: Implications for Geography's Frontier. *Professional Geographer*, **70**, 687–694.

Brum-Bastos, V. S., Long, J. A. and Demšar, U. (2018) Computers , Environment and Urban Systems Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*, **71**, 131–152.

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G. and Barabási, A.-L. (2008) Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, **41**, 224015. URL: http://stacks.iop.org/1751-8121/41/i=22/a=224015?key=crossref.97d23b44de724a7398482cd45c7fe01a.

Cao, L. and Philip, S. Y. (2012) *Behavior computing: modeling, analysis, mining and decision*. Springer.

Chen, D.-G., Jin, Z., Li, G., Li, Y., Liu, A. and Zhao, Y. (2017) *New Advances in Statistics and Data Science.* Springer.

Choi, W., Pantofaru, C. and Savarese, S. (2013) A General Framework for Tracking Multiple People from a Moving Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 1577–1591.

Codling, E. A., Plank, M. J. and Benhamou, S. (2008) Random walk models in biology. *Journal of The Royal Society Interface*, **5**, 813–834.

Coelho, T. L., Zeitouni, K. and Mac?, A. F. D. (2016) A Framework for Online Mobility Pattern Discovery from Trajectory Data Streams. In *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, 365–368. IEEE.

Conway, D. (2010) Data science venn diagram by Drew Conway. URL: https://s3.amazonaws.com/aws.drewconway.com/viz/venn_diagram/data_science.html.

Cukier, K. and Mayer-Schoenberger, V. (2013) The Rise of Big Data : How It's Changing the Way We Think About the World. *Foreign Affairs*, **92**, 28–32.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R. (2015) Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, **3**, 5. URL: http://www.movementecologyjournal.com/content/3/1/5.

Dhar, V. (2013) Data Science and Prediction. *Communications of the ACM*, **56**, 64–73.

Dodge, S. (2016) From Observation to Prediction: The Trajectory of Movement Research in GIScience. In *Advancing Geographic Information Science: The Past and Next Twenty Years* (eds. H. Onsrud and W. Kuhn), chap. 9, 123 – 136. GSDI Association Press.

— (2018) Embracing visualization as a key element in computational movement analytics. In *AutoCarto/UCGIS 2018. The 22nd International Research Symposium on Computer-based Cartography and GIScience* (eds. S. Freundschuh and D. Sinton), 52–57. Cartography and Geographic Information Society (CaGIS).

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

Dodge, S., Laube, P. and Weibel, R. (2012) Movement Similarity Assessment Using Symbolic Representation of Trajectories. *International Journal of Geographic Information Science*, **26**, 1563–1588.

Dodge, S., Weibel, R. and Lautenschütz, A.-K. (2008) Towards a taxonomy of movement patterns. *Information Visualization*, **7**, 240–252. URL: http://www.palgrave-journals.com/doifinder/10.1057/palgrave.ivs.9500182.

Dodge, S., Xavier, G. and W.Y., . W. (2018) DynamoVis - Dynamic Visualization of Animal Movement Data.

Downs, J., Horner, M., Lamb, D., Loraamm, R. W., Anderson, J. and Wood, B. (2018) Testing time-geographic density estimation for home range analysis using an agent-based model of animal movement. *International Journal of Geographical Information Science*, **32**, 1505–1522.

Evans, M. R., Oliver, D., Yang, K., Zhou, X., Ali, R. Y. and Shekhar, S. (2019) Enabling spatial big data via cybergis: Challenges and opportunities. In *CyberGIS for Geospatial Discovery and Innovation*, 143–170. Springer.

Farmer, C. J. and Keßler, C. (2016) Hierarchical prism trees for scalable time geographic analysis. In *The Annual International Conference on Geographic Information Science*, 34–47. Springer.

Giannotti, F. and Pedreschi, D. (2008) *Mobility, data mining and privacy: Geographic knowledge discovery*. Springer Science & Business Media.

González, M. C., Hidalgo, C. a. and Barabási, A.-L. (2008) Understanding individual human mobility patterns. *Nature*, **453**, 779–82. URL: `http://www.nature.com/doifinder/10.1038/nature07850http://www.ncbi.nlm.nih.gov/pubmed/18528393`.

Holloway, P. (2018) Simulating Movement-Related Resource Dynamics to Improve Species Distribution Models: A Case Study with Oilbirds in Northern South America. *Professional Geographer*, **70**, 528–540. URL: `https://doi.org/10.1080/00330124.2018.1479972`.

Holloway, P. and Miller, J. A. (2018) Analysis and Modeling of Movement. In *Comprehensive Geographic Information Systems*, vol. 1, 162–180. Elsevier.

Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E. and Spiegel, O. (2008) Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19060–19065.

Holzinger, A. and Jurisica, I. (2014) Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. In *Interactive knowledge discovery and data mining in biomedical informatics*, 1–18. Springer.

Horne, J. S., Garton, E. O., Krone, S. M. and Lewis, J. S. (2007) Analyzing animal movements using Brownian bridges. *Ecology*, **88**, 2354–2363.

Imfeld, S. (2000) *Time, Point and Space- Towards a Better Analysis of Wildlife Data in GIS*. Ph.D. thesis, Universitaet Zuerich.

Jiang, Z. and Shekhar, S. (2017) *Spatial Big Data Science: Classification Techniques for Earth Observation Imagery*.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M. and Newth, D. (2015) Understanding human mobility from Twitter. *PLoS ONE*, **10**, 1–16.

Karanth, K. U. and Nichols, J. D. (1998) Estimation of Tiger Densities in India Using Photographic Captures and Recaptures. *Ecology*, **79**, 2852–2862.

Kays, R., Crofoot, M. C., Jetz, W. and Wikelski, M. (2015) Terrestrial animal tracking as an eye on life and planet. *Science*, **348**, aaa2478.

Kim, B. S., Kang, B. G., Choi, S. H. and Kim, T. G. (2017) Data modeling versus simulation modeling in the big data era: Case study of a greenhouse control system. *Simulation*, **93**, 579–594.

Kokko, H. (2006) From Individual Dispersal to Species Ranges: Perspectives for a Changing World. *Science*, **313**, 789–791. URL: `http://www.sciencemag.org/cgi/doi/10.1126/science.1128566`.

Kounadi, O. and Resch, B. (2018) A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics*, **13**, 203–222.

Kranstauber, B., Kays, R., Lapoint, S. D., Wikelski, M. and Safi, K. (2012) A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, **81**, 738–746.

Laube, P. (2005) *Analysing Point Motion Spatio-Temporal Data Mining of Geospatial Lifelines*. Ph.D. thesis, Universitaet Zuerich.

— (2009) Progress in Movement Pattern Analysis. In *Behaviour Monitoring and Interpretation - Ambient Assisted Living* (eds. B. Gottfried and H. K. Aghajan), vol. 3 of *Behaviour Monitoring and Interpretation ? BMI - Smart Environments*, 43–71. IOS Press BV. URL: http://www.booksonline.iospress.nl/Content/View.aspx?piid=13879.

— (2014) *Computational Movement Analysis*. SpringerBriefs in Computer Science. Springer International Publishing, 1 edn.

— (2017) Representation: Trajectories. In *International Encyclopedia of Geography: People, the Earth, Environment and Technology* (eds. D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu and R. A. Marston), 1–11. Wiley Online Library.

Laube, P., Duckham, M. and Palaniswami, M. (2011) Deferred decentralized movement pattern mining for geosensor. *International Journal ofGeographical Information Science*, **25**, 273–292.

Lavigne, M., Goulden, R., Habib, B., Joseph, L., Girard, N., Buckeridge, D. and Tamblyn, R. (2018) A bayesian way to correct for measurement error in drug risk estimates from ehr data. *International Journal of Population Data Science*, **3**.

Lee, K. and Kwan, M.-P. (2018) Automatic physical activity and in-vehicle status classification based on GPS and accelerometer data: A hierarchical classification approach using machine learning techniques. *Transactions in GIS*, 1–28. URL: http://doi.wiley.com/10.1111/tgis.12485.

Li, X., Han, J., Kim, S. and Gonzalez, H. (2007) Roam: Rule-and motif-based anomaly detection in massive moving object data sets. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, 273–284. SIAM.

Long, J. A., Weibel, R., Dodge, S. and Laube, P. (2018) Moving ahead with computational movement analysis. *International Journal of Geographical Information Science*, **32**, 1275–1281. URL: https://doi.org/10.1080/13658816.2018.1442974.

Mandel, J. T., Bohrer, G., Winkler, D. W., Barber, D. R., Houston, C. S. and Bildstein, K. L. (2011) Migration path annotation: cross-continental study of migration-flight response to environmental conditions. *Ecological Applications*, **21**, 2258–2268.

Manson, S. and O'Sullivan, D. (2006) Complexity theory in the study of space and place. *Environment and Planning A*, **38**, 677–692.

Mckenzie, G., Deutsch, K. and Raubal, M. (2013) What, When and Where: The real-world activities that contribute to online social networking posts. In *International workshop on action and interaction in volunteered geographic information* (eds. P. Mooney and K. Rehrl), 25–30.

Mennis, J. and Guo, D. (2009) Spatial data mining and geographic knowledge discovery?an introduction. *Computers, Environment and Urban Systems*, **33**, 403–408.

Metz, C. (2019) Deepmind can now beat us at multiplayer games, too. *New York Times*. Https://nyti.ms/2MgKTUj.

Miller, H., Dodge, S., Miller, J. and Bohrer, G. (2019a) Towards an integrated science of movement: converging research on animal movement ecology and human mobility. *International Journal of Geographical Information Science*. Accepted on 22-Dec-2018, Published online: 15 Jan 2019.

Miller, H. J. (2005) A Measurement Theory for Time Geography. *Geographical Analysis*, **37**, 17–45.

— (2010) The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, **50**, 181–201.

Miller, H. J. and Han, J. (2001) *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, second edn.

Miller, H. J., Jaegal, Y. and Raubal, M. (2019b) Measuring the Geometric and Semantic Similarity of Space?Time Prisms Using Temporal Signatures. *Annals of the American Association of Geographers*, **109**, 730–753. URL: https://doi.org/10.1080/24694452.2018.1484686.

Miller, J. A. (2015) Towards a Better Understanding of Dynamic Interaction Metrics for Wildlife: a Null Model Approach. *Transactions in GIS*, **19**, 342–361.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. and Smouse, P. E. (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19052–19059.

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M. and Getz, W. M. (2012) Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *The Journal of experimental biology*, **215**, 986–96.

O'Sullivan, D. and Perry, G. L. W. (2013) *Spatial Simulation: Exploring Pattern and Process*. Wiley-Blackwell. URL: http://onlinelibrary.wiley.com/book/10.1002/9781118527085.

Papadias, D., Tao, Y., Kalnis, P., Zhang, J., Bay, C. W. and Kong, H. (2002) Indexing Spatio-Temporal Data Warehouses. In *Proceedings 18th International Conference on Data Engineering*, 166–175. IEEE.

Parent, C., Pelekis, N., Theodoridis, Y., Yan, Z., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A. and Macedo, J. (2013) Semantic trajectories modeling and analysis. *ACM Computing Surveys*, **45**, 1–32. URL: http://dl.acm.org/citation.cfm?doid=2501654.2501656.

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O. and Matthiopoulos, J. (2008) State?space models of individual animal movement. *Trends in Ecology & Evolution*, **23**, 87–94. URL: http://linkinghub.elsevier.com/retrieve/pii/S0169534707003588.

Pfoser, D. (2002) Indexing the trajectories of moving objects. *IEEE Bulletin of the Technical Committee on Data Engineering*, **25**, 3–9.

Polansky, L., Wittemyer, G., Cross, P. C., Tambling, C. J. and Getz, W. M. (2010) From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology*, **91**, 1506–1518.

Provost, F. and Fawcett, T. (2013) Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, **1**, 51–59. URL: http://www.liebertpub.com/doi/10.1089/big.2013.1508.

Rhee, I., Shin, M., Hong, S. and Lee, K. (2011) On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking (TON)*, **19**, 630–643. URL: http://dl.acm.org/citation.cfm?id=2042974.

Roos, D. S. (2001) Bioinformatics–Trying to Swim in a Sea of Data. *Science*, **291**, 1260–1261.

Ruiz-Gutierrez, V., Hooten, M. B. and Grant, E. H. C. (2016) Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, **7**, 900–909.

Sauer, F., Xie, J. and Ma, K. L. (2017) A Combined Eulerian-Lagrangian Data Representation for Large-Scale Applications. *IEEE Transactions on Visualization and Computer Graphics*, **23**, 2248–2261.

Schick, R. S., Loarie, S. R., Colchero, F., Best, B. D., Boustany, A., Conde, D. a., Halpin, P. N., Joppa, L. N., McClellan, C. M. and Clark, J. S. (2008) Understanding movement data and movement processes: Current and emerging directions. *Ecology letters*, **11**, 1338–50. URL: http://www.ncbi.nlm.nih.gov/pubmed/19046362.

Seidel, D. P., Dougherty, E., Carlson, C. and Getz, W. M. (2018) Ecological metrics and methods for GPS movement data. *International Journal of Geographical Information Science*, **32**, 2272–2293. URL: https://doi.org/10.1080/13658816.2018.1498097.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H. and Havinga, P. J. M. (2014) Fusion of Smartphone Motion Sensors for Physical Activity Recognition. *Sensors*, **14**, 10146–10176.

Singleton, A. and Arribas-Bel, D. (2019) Geographic Data Science. *Geographical Analysis*, **0**, 1–15. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gean.12194.

Soleymani, A., Cachat, J., Robinson, K., Dodge, S., Kalueff, A. and Weibel, R. (2014) Integrating cross-scale analysis in the spatial and temporal domains for classification of behavioral movement. *Journal of Spatial Information Science*, **8**, 1–25. URL: http://josis.org/index.php/josis/article/view/162.

Sowmya, R. and Suneetha, K. R. (2017) Data Mining with Big Data. *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*, **26**, 246–250.

Sukkarieh, S., Nebot, E. M. and Durrant-Whyte, H. F. (1999) A high integrity imu/gps navigation loop for autonomous land vehicle applications. *IEEE transactions on robotics and automation*, **15**, 572–578.

Tao, Y., Both, A. and Duckham, M. (2018) Analytics of movement through checkpoints. *International Journal of Geographical Information Science*, **32**, 1282–1303.

Thums, M., Fernández-Gracia, J., Sequeira, A. M. M., Eguíluz, V. M., Duarte, C. M. and Meekan, M. G. (2018) How Big Data Fast Tracked Human Mobility Research and the Lessons for Animal Movement Ecology. *Frontiers in Marine Science*, **5**, 1–12. URL: http://journal.frontiersin.org/article/10.3389/fmars.2018.00021/full.

Thurfjell, H., Ciuti, S. and Boyce, M. S. (2014) Applications of step-selection functions in ecology and conservation. *Movement Ecology*, **2**, 4. URL: http://www.movementecologyjournal.com/content/2/1/4.

UCGIS (2018) A UCGIS Call to Action: Bringing the Geospatial Perspective to Data Science Degrees and Curricula. *Tech. Rep. July*, University Consortium for Geographic Information Science.

Valentino-DeVries, J., Singer, N., Keller, M. and Krolik, A. (2018) Your apps know where you were last night, and they're not keeping it secret. *New York Times*. Https://nyti.ms/2G6l9XC.

Van Der Aalst, W. (2016) Data science in action. In *Process Mining*, 3–23. Springer.

Whitton, R. (2019) Microlocation ushers In the next advancement in mapping and navigation of robots. *Tech. rep.*, ABI Research for Visionaries.

Wilson, R. P., Quintana, F. and Hobson, V. J. (2012) Construction of energy landscapes can clarify the movement and distribution of foraging animals. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 975–80.

Winter, S. and Yin, Z.-C. (2010) The elements of probabilistic time geography. *GeoInformatica*, **15**, 417–434. URL: http://link.springer.com/10.1007/s10707-010-0108-1.

Wu, L., Zhi, Y., Sui, Z. and Liu, Y. (2014) Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, **9**.

Xia, C., Cochrane, C., DeGuire, J., Fan, G., Holmes, E., McGuirl, M., Murphy, P., Palmer, J., Carter, P., Slivinski, L. and Sandstede, B. (2017) Assimilating Eulerian and Lagrangian data in traffic-flow models. *Physica D: Nonlinear Phenomena*, **346**, 59–72. URL: http://dx.doi.org/10.1016/j.physd.2017.02.004.

Xu, J., Huang, E., Chen, C.-H. and Lee, L. H. (2015) Simulation Optimization: A Review and Exploration in the New Era of Cloud Computing and Big Data. *Asia-Pacific Journal of Operational Research*, **32**, 1550019.

Yuan, M. (2018) Human dynamics in space and time: A brief history and a view forward. *Transactions in GIS*, **22**, 900–912.

Yudanto, R. G. and Petre, F. (2015) Sensor fusion for indoor navigation and tracking of automated guided vehicles. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 1–8. IEEE.

Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q., Chen, X. and Zhang, T. (2015) A big data approach for logistics trajectory discovery from rfid-enabled production data. *International Journal of Production Economics*, **165**, 260–272.