



## Editorial

## Advances in geocomputation (1996–2011)

## 1. What is geocomputation?

The first international conference on 'Geocomputation', hosted by the School of Geography at the University of Leeds in 1996, launched a new research agenda in geographical analysis and modelling (Openshaw & Abrahart, 1996): "The art and science of solving complex spatial problems with computers" (GeoComputation, 2012). The interest generated in this field was established as a yearly conference in its early stages (Otago, New Zealand, 1997; Bristol, UK, 1998; Virginia, USA, 1999; Greenwich, UK, 2000; Queensland, Australia, 2001), later maturing into an event every 2 years (Southampton, United Kingdom, 2003; Michigan, USA, 2005; Maynooth, Ireland, 2007; Sydney, Australia, 2009; London, UK, 2011). The next conference will be held at Wuhan University (China) in 2013, the first time in Asia.

Alongside the conferences, and in seeking to advance the field conceptually, debate around a definition of 'What is Geocomputation' has continued. An early definition offered by Couclelis (1998, p. 18) simply states that 'geocomputation just means the universe of computational techniques applicable to spatial problems'. However, Openshaw argues that such a statement is limited in extent and that geocomputation further presents a "new paradigm for applying science in a geographical context" (Openshaw, 2000, p. 5). In this sense, geocomputation is not simply about applying computational methods to explore geographical concepts; it offers an extensive toolkit for the examination and identification of new perspectives on spatial processes.

Longley describes geocomputation in terms of the unexplored geographical processes that it allows one to examine – "The hallmarks of geocomputation are those of research-led applications which emphasise process over form, dynamics over statics, and interaction over passive response" (Longley, 1998, p. 3). Longley sees geocomputation as providing the framework for the execution of geocomputational science – in both advancing the state-of-the-art in the computation of geography as well as extending our understanding of geographical phenomena. In taking this view, it is said that geocomputation extends geographical information systems (GISs), with the latter offering a toolkit that allows the practice of geocomputational science. Openshaw (2000, p. 9), in agreement, suggests that the relationship between GIS and geocomputation is important, "yet may be just as important; for example, with computer science or numerical methods or statistics". Geocomputation, then, represents a broad framework, and despite significant methodological advances since these definitions were initially offered (which have been showcased at the geocomputation conference series), the principles of geocomputation continue to follow the vein offered by Openshaw and Longley.

No matter how the science of geocomputation is defined, it is clear that a number of fundamental methods and technologies encapsulate the majority of research within the area. Gahegan (1999) describes this framework as consisting of four 'enabling technologies' (ETs), described as core to the execution of geocomputational research, which are: (1) computer architecture and design; (2) search, classification, prediction and modelling; (3) knowledge discovery (spatial data mining); and (4) visualisation. These ETs represent the key tasks of geocomputation. They are not mutually exclusive and in fact are highly interrelated, and one could even argue that ET2 and ET3 differ only in the algorithms that are used. For instance, knowledge discovery, or spatial data mining (SDM), is defined as "the process of discovering interesting and previously unknown, but potentially useful, patterns from large spatial datasets" (Shekhar, Zhang, & Huang, 2010, p. 837). The tasks listed under ET2 could easily be subsumed under, or seen as logical extensions to, this definition. The progress made in spatial data mining could be traced in two editions of the book edited by Miller and Han (2001, 2009).

Visualisation (ET4) remains an important challenge and active research area for the geocomputation community, for whom complex, heterogeneous datasets, varying across space and time, are the norm. Visualisation, and in particular visual analytics, enables the communication of information to both experts and non-experts and can be useful in breaking down conceptual and technical barriers in the interpretation of data analyses (Thomas & Cook, 2005). However, effective 3D visualisation of space–time data still represents a major bottleneck in our understanding and conceptualisation of space–time processes. Despite much research, the static 2D map remains the predominant medium for displaying spatial information. Significant challenges still exist regarding scalability, interoperability, visualisation of complex space–time structures, and the linkages between exploration and validation (Andrienko et al., 2007). In the coming years, geovisual analytics will need to fully embrace space–time in order to produce output that is meaningful and interpretable.

In the following review, the development of the field of Geocomputation over the past 15 years is detailed, with respect to the ETs outlined by Gahegan (1999). We begin with a review of modern computational algorithms and models in Section 2 (ET2 and ET3). The trends described here reflect the continuing ability of geocomputation to work with and between other scientific disciplines, while focusing on the specific and unique issues introduced in space and time. We focus on three key areas; statistical approaches, machine learning approaches, and simulation, whose applications span ET2 and ET3. It is clear that ETs 2–4 are facilitated by advancements in ET1. In Section 3, we proceed to review the progress in high performance computing (ET1), and its

consequences for geocomputational research. We do not focus specifically on ET4 in this review, but direct readers to recent reviews (Andrienko, Andrienko, Dykes, Fabrikant, & Wachowicz, 2008; Andrienko et al., 2007; Cheng, Haworth, Anbaroglu, Tanaksarand, & Wang, in press). In Section 4, as introduction to this special issue, we summarise the contributions of each of the accepted papers. These papers provide a snapshot of current research in Geocomputation. Finally, we conclude in offering some thoughts as to the direction of geocomputation over the next 10 years in Section 5.

## 2. Modern computational algorithms and models

The field of Geocomputation exists at the boundary of geography and computational science. It has facilitated the mixture of established theoretical/statistical models with modern computational techniques, as well as bringing a plethora of new computational methods to take advantage of a data rich environment in order to greatly advance the field of geography. In this section, some of the methods that have either benefitted from, or been a result of Geocomputation are reviewed.

### 2.1. Statistical (parametric and numerical) approaches

Statistical models are the backbone of spatial analysis. Spatial regression models allow researchers to explore the dependency relationships between spatially referenced variables, leading to the creation of new knowledge about spatial processes. Geostatistical models provide a framework for spatial interpolation and forecasting. The main attraction of such methods is that they are theoretically well founded and interpretable. The theory behind them is not new, but the era of Geocomputation has led to great advancements in these models in three key areas, which are described below.

The first key advancement is in the way statistical models are estimated and their assumptions tested. Turton and Openshaw (1998, p. 1840) noted that “much of the quantitative technology presently used in geography dates from the 1960s and that many assumptions were made in an effort to minimise computation”. Ordinary least squares (OLSs) regression does not provide an unbiased estimator in the presence of spatial and/or temporal autocorrelation, and methods involving maximum likelihood are sometimes intractable or inappropriate (Griffith & Lagona, 1998). Advances in technology have allowed researchers to use computationally intensive numerical methods to approximate model specifications. For instance, Monte Carlo simulation methods are now widely used both to estimate spatial regression models (Gu & Zhu, 2001; Huffer & Wu, 1998) and to test and scrutinise their assumptions (Farber, Páez, & Volz, 2009; Fingleton, 1999; Mizruchi & Neuman, 2008).

The second key advancement in statistical models has come in the way we look at spatial processes. Traditional statistical models were necessarily global in order to satisfy their assumptions and make them computationally feasible. Advances in geocomputation allow us to consider heterogeneity in spatial processes. Geographically weighted regression (GWR) is the prime example of this (Brunsdon, Fotheringham, & Charlton, 1996; Fotheringham, Brunsdon, & Charlton, 2002). It takes ideas from traditional statistics (OLS), fuses them with non-parametric methods (kernel regression), and uses computational methods (Monte Carlo simulation) to test hypotheses and parameter stability. Heterogeneity can now also be accounted for in spatial panel data models (Elhorst, 2003) and has been studied in the context of geostatistical models (Atkinson & Lloyd, 2007), leading to the development of locally calibrated (Kleiber, Raftery, & Gneiting, 2011) and multiple-point (Caers & Zhang, 2004) geostatistical models, amongst others. In a

different vein, Eigenvector spatial filtering (ESF) allows spatial processes to be viewed in a new way. In ESF, Moran's I is decomposed into map patterns using the eigenvectors of the spatial weights matrix. It is conceptually similar to principal components analysis (PCA), the difference being that it uses the eigenvectors themselves as synthetic variables rather than as coefficients (Griffith, 2003). Crucially, OLS, a well understood method, can be used to model the filtered data, opening up the range of standard diagnostic tests. This is an excellent example of how geocomputation not only allows us to look at space from different perspectives but also to use modern techniques to make our analyses easier.

The third advancement is in scale of the problems that we are able to study, and in particular the exploration of space–time. Spatio-temporal statistical models such as space–time autoregressive integrated moving average (STARIMA) models and variants have been around since the 1970s (Martin & Oeppen, 1975). However, it is only recently that we have had access to either the data or the computing power to make full use of them. Recent years have seen a sharp increase in development of theory and methods related to space–time. For instance, Huang, Wu, and Barry (2010) extended GWR to space–time in their geographically and temporally weighted regression (GTWR) model. They incorporate the spatial and temporal dimensions into the weight matrix to account for spatial and temporal nonstationarity, applying the method to the modelling of residential housing sales in Calgary, Canada. Moreover, space–time geostatistics is an increasingly developed field with much research concerned with deriving space–time covariance structures and semivariograms for the purpose of space–time interpolation and forecasting (Gneiting, 2002; Heuvelink & Griffith, 2010). The aim is to build a process that mimics some patterns of the observed spatiotemporal variability, without necessarily following the underlying governing equations (Kyriakidis & Journel, 1999). Since the correlation in data cannot be adequately described by parameters that are globally fixed spatially and/or temporally (such as STARIMA), a localized STARIMA (LSTARIMA) is developed based upon a dynamic spatial weight matrix that captures the autocorrelation locally (heterogeneity) and dynamically (unstationarity) (Cheng, Wang, Haworth, Heydecker & Chow, 2011; 2012).

Spatio-temporal scan statistics (STSSs) is a clustering technique that automatically detects regions of space that are “anomalous,” “unexpected,” or otherwise “interesting” (Brunsdon & Charlton, 2011; Neill, 2008). Spatial and temporal proximities are exploited by scanning the entire study area via overlapping space–time regions (STRs). The dimensions of the STR are allowed to vary in order to detect outbreaks of varying sizes. STSS was developed based upon the Spatial Scan Statistic (Kulldorff, 1997), which was proposed to overcome the multiple hypothesis testing problem encountered in the Geographical Analysis Machine (GAM) developed by Openshaw, Charlton, Wymer, and Craft (1987). STSS has been widely used in crime (Nakaya & Yano, 2010) and health (Kulldorff, 2001) studies, but efforts are still needed to improve computational efficiency and to reduce the false alarm rate, in particular for emergence detection.

### 2.2. Machine learning (non-parametric) approaches

As well as enhancing and advancing the theory and application of statistical models, one of the defining features of geocomputation has been a willingness to explore non-traditional computational methods from the fields of artificial intelligence and machine learning. Fischer (1998, p. 1874) argued that traditional spatial analysis techniques “overemphasize linear statistical model designs when nonlinearities prevail in reality, they tend to neglect rather than take into account the special nature of spatial data”. Machine learning methods, such as Artificial Neural Networks

(ANNs) and Support Vector Machines (SVMs) are inherently suited to the challenges of spatial and spatio-temporal data due to their innate ability to model complex nonlinear relationships. As well as being applied to the classic regression tasks of spatial analysis, machine learning methods can be applied to a wide range of tasks such as classification, clustering and novelty detection. ANNs are able to learn from empirical data and can be used in cases where the modelled phenomena are hidden, non-evident or not very well described. This makes them particularly useful in modelling the complex dependency structures present in spatial and spatio-temporal data that cannot be described theoretically. Kanevski, Timonin, and Pozdnoukhov (2009) and Hsieh (2009) have applied various types of ANN to spatial and environmental modelling problems.

Kernel methods are another set of machine learning algorithms that are growing in popularity in the geocomputation community. The strength of kernel methods lies in their use of the kernel trick, whereby observations are mapped to a higher (possibly infinite) dimensional feature space without the need to explicitly compute the mapping. Linear algorithms can be applied in the feature space to find solutions that are nonlinear in the input space. The most widely used kernel method is the Support Vector Machine (SVM for classification, SVR in the regression case) (Boser, Guyon, & Vapnik, 1992). The key to its strong performance is that the learning task is formulated as a convex optimisation problem meaning that, for a given set of parameters, the solution is globally optimal provided one can be found. Therefore, SVMs avoid the problems of local minima which are traditionally associated with ANNs. SVMs have been applied to reservoir porosity and Caesium contamination prediction (Kanevski et al., 2009) and, more recently, to spatio-temporal avalanche forecasting (Pozdnoukhov et al., 2011). The approach involves incorporating the outputs of simple physics based and statistical approaches to interpolate meteorological and snowpack related data over a digital elevation model of the region. This study highlights another feature of geocomputation, which is the hybrid combination of statistical and machine learning methods.

### 2.3. Simulation

While statistical and machine learning methods are enabling an important development of our insight into the nature and function of existing geographical systems, there remains a need to explore how spatial systems develop and evolve over time, and how future influences upon an environment may subsequently affect any underlying systems. Over recent years, the simulation approaches of Cellular Automata and, later, Agent-based Modelling have emerged as key methods in achieving this. While the two approaches differ in a number of respects, they are united in their treatment of the individual entity, exploring how the actions and interactions of constituent elements result in the formation of complex geographic phenomena. This section will describe both approaches as well as their application in describing geographical systems.

Cellular Automata (CA) models are simplified representations of spatial systems, where space is discretised into an environment of grid cells, and time represented across time steps. Within a CA simulation, at each time step, a cell *state* is updated according to the set of predefined rules governing all behaviour (known as *transition rules*) that make reference to the status of other cells around it (known as the *cell neighbourhood*). Through the interaction of neighbouring cells according to a specified set of rules, this simple arrangement – consisting of only four design elements – is capable of generating complex macroscopic patterns. The nature of the approach means that considerably complex results, unintelligible at the system scale, may emerge from an initial system state of a single root cell (Batty, 2007).

These principles have been extended widely to demonstrate emergence in a range of real world systems. The discretisation of space in CA lends itself well to exploring wide scale geographic development, and so has been extensively used in the development of land-use and urban growth models (Almeida, Gleriani, Castejon, & Soares-Filho, 2008; Batty, 2007; Clarke, Hoppen, & Gaydos, 1997; White, Engelen, & Uljee, 1997). Social processes may be explored similarly at a higher level of abstraction, as demonstrated by work on dynamic social processes in cities (Schelling, 1978; Yeh & Li, 2002). The approach has been extended to spatio-ecological processes, where predator–prey networks (Resnick, 1994) and forest dynamics (Grimm et al., 2005) have been simulated.

Agent-based Modelling (ABM) represents a conceptual development that extends CA by introducing greater flexibility and individual autonomy to the simulation environment. Rather than constraining behaviour and interactions to a grid-like structure, individual entities (known as *agents*) are able to move throughout the environment and interact with others (according to a given set of rules). ABM enables a more sophisticated description of individual behaviour than is possible using CA, as well as the incorporation of different types of agent, defined by distinct rule sets. In their most advanced form, agent behaviour may be represented in considerable detail, describing cross-population heterogeneity in goals, preference, knowledge and perspective. Agent rules may describe both physical and mental processes. Like CA, the nature of the simulation conditions allows global phenomena to emerge through individual interactions and in response to new conditions, yet according to a more sophisticated set of rules governing agent behaviour. The platform represents a natural method of simulation for many spatial phenomena (Bonabeau, 2002).

ABM has been applied in the simulation of a wide range of geographic phenomena. As seen with CA, these have included land-use, housing and regional models (Bretagnolle & Pumain, 2010; Xie, Batty, & Zhao, 2007), yet have been extended to other urban phenomena, most notably transportation systems. ABM has been shown to represent a strong paradigm for the simulation of pedestrian movement patterns, including the formation of patterns in crowd activity (Batty, 2007; Helbing & Molnar, 1995; Torrens, 2012). The approach has also been employed in the replication of traffic flows in urban systems (Chen & Cheng, 2010). Recent work has demonstrated a link between individual choice and behaviour in the formation of urban road congestion (Manley & Cheng, 2011).

## 3. High performance computing

### 3.1. HPC

One of the defining features of Geocomputation has been an eagerness to exploit the potential of high performance computing (HPC). Prior to the 1990s, a lack of computational resources constrained much of the progress being made in the geographic field to theoretical and experimental research. The advent of HPC opened up new ways of conducting science in areas too complex to be handled by other means or where the data volume was too large to be handled in a traditional computing environment (Turton & Openshaw, 1998). HPC has facilitated a shift from deductive science to inductive, data driven science, and has gone hand in hand with the willingness to explore techniques and methods from computational science (Armstrong, 2000).

Within HPC there are two main paradigms. The first is grid (distributed) computing, whereby computing resources are distributed widely and opportunistically used whenever available. Since the turn of the century, there has been considerable public investment in grid computing, termed e-Science in the UK (NeSC, 2012) and Cyberinfrastructure in the US (Office of Cyberinfrastructure, 2012). This has enabled researchers to begin to tackle hard

problems in physics, biology, engineering and geosciences, amongst others, facilitating collaboration across distance and disciplines and enabling scientific breakthroughs that would not have been possible otherwise (Wright & Wang, 2011). A closely related service is cloud computing, which is the delivery of computing services over the internet. This area has seen growth recently, with online services such as AmazonEC2 providing businesses and researchers with access to powerful computing facilities without the need for hardware investment.

The second paradigm is cluster computing, whereby a large number of processors are located in close proximity, the classic form of the supercomputer. This type of architecture is now being replicated in single processors with multiple cores, such as multi-core central processing units (CPUs) and graphical processing units (GPUs), the latter of which gives access to many hundreds of cores inside a desktop workstation. The goal of both is the same; massive parallelisation of computationally intensive tasks.

### 3.2. Parallel geocomputation

Geography as a discipline stands to benefit greatly from HPC for two reasons. Firstly, it encompasses a diverse range of complex processes and interactions, both human and physical. Secondly, the sheer amount of data being collected with spatial and temporal dimensions exceeds the limits of traditional computing techniques. Consequently, a select few geographers were forward thinking in adopting HPC methods. In the early 1990s, parallel GIS was identified as a technology whose time would come, and two research avenues were opened; the parallelisation of existing algorithms and the exploration of new analytical methods made possible through HPC (Clematis, Mineter, & Marciano, 2003). It was the work of Stan Openshaw and colleagues at the University of Leeds in particular that pioneered this second direction, believing in the potential to simulate whole societies and model the previously unmodellable (Turton & Openshaw, 1998).

These days, HPC is more accessible to geographers than ever, and this is due in no small part to a general trend within science and research towards parallelisation. Parallel implementations of key mathematical operations such as matrix multiplications and inversions, as well as interfaces to parallel computing architectures, are now available in many of the software that geographers use. For example, R statistical package (R) now has support for multicore, GPU and distributed computing (McCallum & Weston, 2011). Additionally, many universities and research institutes now have access to e-Science infrastructure as well as their own, dedicated supercomputers which can be used with very little training.

The extent to and ease with which an algorithm can be parallelised depends on the nature of the problem. Conveniently, many problems in geographical research have the “embarrassingly parallel” property that a single large problem involving  $n$  locations can be divided into  $n$  small problems, each of which can be solved independently of the  $n - 1$  others. Harris, Singleton, Grose, Brunson, and Longley (2010) showed how this type of problem could be tackled in the context of geographically weighted regression (GWR) using a combination of R and the UK's National Grid Service (National e-Infrastructure Service, 2012). This type of problem requires little modification to be parallelised and can be run on a grid or computer cluster. In other problems, the challenge lies in identifying specific parts of an algorithm that can be parallelised in this way. For example, Guan, Kyriakidis, and Goodchild (2011) developed a parallel approach to geostatistical interpolation by recognising the fact that the computation of each entry of the covariance matrix is independent of the other entries.

A more difficult set of problems arises when the division of the geographic space is nontrivial. Parallel algorithms must balance

the amount of data sent to each processor (data parallelism) with the partitioning of tasks across the processors (task parallelism). In irregular and heterogeneous data, regular partitioning of geographic space is inappropriate (Guan & Clarke, 2010). Furthermore, when the inputs/outputs of each of the problems being executed in parallel are not independent of each other a message passing interface (MPI) must be used. Tomlin's (1990) classification of map algebra divides the scope of such problems into local, neighbourhood, regional and global. In global scope problems, each spatial location requires information from the whole of the geographic space at each evaluation. In iterative algorithms, this leads to challenges in developing communications systems between different processors. These types of problems require what Turton and Openshaw (1998, p. 1842) termed “thinking in parallel”, which is important but hard.

It is a major challenge in geocomputation to get everyone thinking in parallel and to make HPC the norm rather than the exception. Computing technology in general is moving towards parallelism and it is vital that this shift can be harnessed. Arguably the current work with the most wide reaching significance is that which makes use of GPU technology. GPUs with multiple processor cores are included as standard in most modern desktop workstations, making HPC possible without grids or supercomputers. Even the latest CPU processor cannot match the potential speedup provided by a modest GPU with a few hundred cores. However, despite significant advances, parallelism is far from being common practice in geography and falls short of the lofty ambitions of Openshaw and others.

### 4. Geocomputation 2011 and this special issue

The 11th International Conference on Geocomputation 2011 was held at University College London (UCL) on 20th–22nd July 2011. Scholars from 19 countries and regions joined the conference – from Australia, Austria, Canada, China, Finland, Iran, Ireland, Israel, Japan, Malta, Portugal, The Netherlands, Spain, Slovenia, Sweden, Switzerland, Turkey, UK and USA. The Local Organising Committee received 104 submissions which were each reviewed by a minimum of two reviewers. There were 78 papers scheduled for full paper presentation at the conference, following revisions made in the light of reviewer comments.

The strength and breadth of geocomputation is reflected in the 16 paper sessions of the 2011 Conference. Eight sessions present advances in algorithms and models: “Agent-based Modelling” (two sessions), “Genetic Algorithms & Cellular Automata Modelling”, “Geographically Weighted Regression”, “Geostatistics”, “Machine Learning”, and “Space-Time Modelling and Analysis” (two sessions). Four sessions contribute to important domain specific applications of geocomputation: “Geodemographics”, “Network Complexity”, “Location-Based Services”, and “GeoVisual and Terrain Analysis”, with two further sessions with more general focus on environmental and urban studies. The broader environment to geocomputation provides the focus for two other sessions – one on “Uncertainty and Accuracy”, and the other on cloud computation – “VGI and Computational Infrastructure”. One poster session consisting of 14 papers was also included.

Five world-renowned scholars kindly gave keynotes at the Conference. They addressed ‘Digital Environments and Real World Geographies’ (Peter Nijkamp, Free University Amsterdam); ‘Does Visualization with Geocomputation Offer Anything We Didn't Know Already?’ (Jo Wood, City University London); ‘Geographically Weighted Regression and Geocomputation: an Overview of Recent Developments’ (Stewart Fotheringham, National University of Ireland at Maynooth); and ‘The Future of Geocomputation’ (Keith Clarke, University of California, Santa Barbara, USA). Another keynote presentation – ‘Visualising Space-Time Dynamics: Graphs and

*Maps, Plots and Clocks*' was contributed by Mike Batty, UCL, as a joint event with the ISPRS-sponsored International Symposium on Spatio-Temporal Analysis and Data Mining, which ran immediately before the Conference at the same location. This session was intended as a forum to foster closer dialogue between these two groups.

This special issue consists of seven papers that were accepted from selected papers presented in the Geocomputation 2011 conference. They are substantially revised and extended for resubmission to this special issue; each paper was reviewed by a minimum of three blind reviewers. We hope this process is fully reflected in the quality of this special issue. These seven papers cover most themes of Geocomputation 2011, which also reflects the broad applications of Geocomputation, in *location-based services* (Jacob, Winstanley, Togher, Roche, & Mooney, *in press*), *traffic survey* (Bolbol, Cheng, Tsapakis, & Haworth, *in press*; Horner, Zook, & Downs, *in press*), *traffic prediction* (Haworth & Cheng, *in press*), *crime* (Malleon & Birkin, *in press*), and *community resource management* (Guise and Crooks, *in press*). They also show the breadth of computational methods: *statistical analysis* (Horner, Zook, Downs, *in press*; Mack, Malizia, & Rey, *in press*), *machine learning* (SVM and Kernel-based) based reasoning (Bolbol et al., *in press*; Haworth & Cheng, *in press*), and *agent-based simulation* (Malleon & Birkin, *in press*; Guise and Crooks, *in press*). All these papers deal with data with spatio-temporal dimensions, which requires seamless and simultaneous integration of space and time. Specifically, this special issue includes the following papers:

"Where were you? Development of a time-geographic approach for activity destination re-construction" (Horner, Zook, & Downs, *in press*) presents a framework and basic computational approach for exploring unlocateable activity locations inherent to travel surveys, based upon a network-based probabilistic time geography. "Population Shift Bias in Tests of Space-Time Interaction" (Mack et al., *in press*) quantifies and compares the population shift bias present in the results of the Knox, Mantel, and Jacquez tests of space-time interaction. It illustrates that population shift bias can be a serious problem for short study periods whenever spatially heterogeneous population change is identified within a study area. In "Pedestrian navigation using the sense of touch" (Jacob et al., *in press*), haptics is implemented in four distinct prototypes for mobile pedestrian navigation. The prototypes are classified based on the user's navigation guidance requirements, the user type (based on spatial skills), and overall system complexity.

"Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification" (Bolbol et al., *in press*), solves a classification problem of GPS data into hybrid transportation modes (i.e. car, walk, cycle, underground, train and bus) based on Support Vector Machines (SVMs) classification. In "Non-parametric regression for space-time forecasting under missing data" (Haworth & Cheng, *in press*), a non-parametric spatio-temporal kernel regression model is developed to forecast the future unit journey time values of road links in central London, UK, under the assumption of sensor malfunction.

In "Analysis of crime patterns through the integration of an agent-based model and a population microsimulation" (Malleon & Birkin, *in press*), microsimulation is combined with agent-based simulation to understand crime variations at progressively finer spatial scales, right down to individual streets or even houses. In "Agent-based modeling for community resource management: Acequia-based agriculture" (Wise & Crooks, *in press*), an agent-based model is constructed to investigate the significant interaction and cumulative impact of the physical water system, local social and institutional structures which regulate water use, and the real estate market on the sustainability of traditional farming as a lifestyle in the northern New Mexico area.

## 5. Summary and outlook

This review, along with the breadth and quality of contributions to this special issue, has demonstrated that great progress has been made in the development of algorithms and methods along the lines of Longley's vision for "applications which emphasize process over form, dynamics over statics, and interaction over passive response" (Longley, 1998, p. 3). The field of Geocomputation stemmed from dissatisfaction with the status quo in geographical research in the early 1990s, where restrictive statistical methods that originated from the data poor, computation poor 1960s and 1970s were still the norm. Since that time, the era of geocomputation has heralded new ways of analysing and interacting with geospatial data that have advanced geography into the data rich 21st century. Nowadays, along with autocorrelation, heterogeneity is fundamental to our observation (or our empirical test) of reality, and is accounted for in an increasing range of models. Furthermore, as we begin to analyse massive spatio-temporal datasets, the challenge now is how to calibrate, explain and validate space-time models. Quantifying spatio-temporal autocorrelation and heterogeneity presents greater challenges than examining the spatial and temporal dimensions in isolation, and the field is still in its infancy. Although models have been developed to represent higher order spatial autocorrelation, but their pitfalls are also obvious in terms of complexity of computation and calibration (LeSage & Pace, 2011). We also think that the statistical distributions widely used in most spatial analysis are hardly approvable which reduce our confidence in using these models, though we have deep trust rooted on statistical analysis in the past (Cheng et al., *in press*).

Due to their ability to deal with multi-dimensional, nonlinear data, machine learning methods have always been at the forefront of the Geocomputation movement. Among others, ANNs, SVMs and other kernel-based approaches have been widely used with excellent results. However, they do not represent a magic wand and still often fail to provide the explanatory power and interpretability that make traditional statistical methods so appealing. Indeed, statistical modelling remains the mainstay of GeoComputation. However, as we increasingly begin to analyse massive spatio-temporal datasets, accurate forecasting will supersede interpretability in time-critical applications. Often, it is found that hybrid frameworks combining both statistical and machine learning approaches can offer the best compromise (Cheng, Wang, & Li, 2011).

A relatively new phenomenon that is providing fresh avenues for Geocomputational research is the rapid growth in volunteered geographic information (VGI). Goodchild (2007, p. 212) describes VGI as "the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information". VGI mapping sources such as OpenStreetMap provide alternatives to those provided by national mapping agencies, and are often timelier and of very high quality (Haklay, 2010). These data look to provide significant scope for research in the coming years (Sui, Elwood, & Goodchild, 2012). In the context of Geocomputation, the challenge lies in extracting meaningful patterns and developing new theories based on VGI. In particular, social media data from services such as Twitter and Facebook allow us to look at how social networks are constructed and evolve in time. We can also look at how people's interactions on social networks lead to manifest effects in geographic space during times of crisis, such as the Arab spring and London riots of 2011.

However, VGI can be noisy, biased, and incomplete. New methods and theory are needed to deal with missing and incomplete data in ways that are applicable in real time (Haworth & Cheng, *in press*; Marks, 2012), and to evaluate the quality of the data (Brunsdon & Comber, 2012; Haklay, 2010). These issues are closely

linked with the notion of trust in VGI, which will be fostered by the development of standards for its collection, validation and maintenance. Furthermore, theory and methods need to be developed to extract meaningful patterns from VGI and put them under the framework of networks such as transport and social-networks that made up of those individuals. More importantly, the dynamics of interactions and movements among individuals should be considered in mining spatio-temporal patterns. This is challenging that we are facing in modelling network complexity, especially in detecting emergence and non-recurrent patterns.

STDM will play an ever increasing role as we look to develop new methods for analysing data of massive spatio-temporal resolution in order to elicit new knowledge on spatio-temporal processes, mobility patterns and spatial networks. Technically, grid computation and cloud computation, which allows geocomputation to be implemented at multiple computer sources, will facilitate research in this area. Even so, when the data volume is increased, the capacity of software and hardware is still limited. How to scale algorithms to larger networks will always be a challenge for Geocomputation given the increase of data volume is far quicker than the improvement in the performance of data processors. This is a challenge that goes hand in hand with a need to shift thinking towards “thinking in parallel”.

Looking ahead, we should be confident that Geocomputation will be as active as it has been in the past 15 years. Its on-going integration with network and complexity theory, spatio-temporal data mining, and parallel computation is needed to promote further prosperity the Geocomputation community.

## Acknowledgements

Tao Cheng would like to thank all the contributors to this special issue for their interest. Special thanks to all the reviewers of this special issue and editor-in-chief of CEUS, Prof. Jean-Claude Thill. All the authors acknowledge the support of the STANDARD project funded by UK EPSRC (EP/G023212/1).

## References

- Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographic Information Science*, 22(9), 943–963.
- Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M.-J., MacEachren, A., et al (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, 21(8), 839–857.
- Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I., & Wachowicz, M. (2008). Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. *Information Visualization*, 7(3–4), 173–180.
- Armstrong, M. P. (2000). Geography and computational science. *Annals of the Association of American Geographers*, 90(1), 146–156.
- Atkinson, P. M., & Lloyd, C. D. (2007). Non-stationary variogram models for geostatistical sampling optimisation: An empirical investigation using elevation data. *Computers & Geosciences*, 33(10), 1285–1300.
- Batty, M. (2007). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: MIT Press.
- Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Supplement 3), 7280–7287.
- Bolbol, A., Cheng, T., Tsapakis, I., & Haworth, J. (in press). Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems* 36(6).
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th annual ACM workshop on computational learning theory* (pp. 144–152). ACM Press.
- Bretagnolle, A., & Pumain, D. (2010). Simulating urban networks through multiscalar space–time dynamics: Europe and the United States, 17–20th centuries. *Urban Studies*, 13, 2819–2939.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281–298.
- Brunsdon, C., & Charlton, M. (2011). An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environment and Planning B: Planning and Design*, 38(2), 216–230.
- Brunsdon, C., & Comber, A. (2012). Assessing the changing flowering date of the common Lilac in North America: A random coefficient model approach. *Geoinformatica*. <http://dx.doi.org/10.1007/s10707-012-0159-6>.
- Caers, J., & Zhang, T. (2004). Multiple-point geostatistics: A quantitative vehicle for integration geologic analogs into multiple reservoir models. In: *Integration of outcrop and modern analog data in reservoir models* AAPG memoir 80 (pp. 383–394).
- Chen, B., & Cheng, H. H. (2010). A review of the applications of agent technology in traffic and transportation systems. *EEE Transactions on Intelligent Transportation Systems*, 11, 485–497.
- Cheng, T., Wang, J., & Li, X. (2011). A hybrid framework for space–time modeling of environmental data. *Geographical Analysis*, 43(2), 188–210.
- Cheng, T., Wang, J., Haworth, J., Heydecker, B.J. and Chow, A.H.F. (2011). Modelling dynamic space–time autocorrelations of urban transport network, *Proceedings of the 11th International Conference on Geocomputation 2011*, July 20th–22nd, p. 215–210.
- Cheng, T., Wang, J., Haworth, J., Heydecker, B.J. and Chow, A.H.F. (2012). Dynamic spatial weight matrix and localised STARIMA for Network Modelling, under review of *Geographical Analysis*.
- Cheng, T., Haworth, J., Anbaroglu, B., Tanaksaranond, G., & Wang, J. (in press). Spatio-temporal data mining. In P. Nijkamp, & M. Fischer (Eds.), *Handbook of regional science*. Springer.
- Clematis, A., Mineter, M., & Marciano, R. (2003). Guest editorial: High performance computing with geographical data. *Parallel Computing*, 29(10), 1275–1279.
- Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, 24, 247–261.
- Couclelis, H. (1998). Geocomputation in context. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation: A primer* (pp. 17–29). Chichester: Wiley.
- Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3), 244–268.
- Farber, S., Páez, A., & Volz, E. (2009). Topology and dependency tests in spatial and network autoregressive models. *Geographical Analysis*, 41(2), 158–180.
- Fingleton, B. (1999). Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration. *Journal of Regional Science*, 39(1), 1–19.
- Fischer, M. M. (1998). Computational neural networks: A new paradigm for spatial analysis. *Environment and Planning A*, 30(10), 1873–1891.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.
- Gahegan, M. (1999). Guest editorial: What is geocomputation? *Transactions in GIS*, 3, 203–206. <http://dx.doi.org/10.1111/1467-9671.00017>.
- GeoComputation (2012). *What is geocomputation? A history and outline*. <<http://www.geocomputation.org/what.html>> Accessed 27.09.12.
- Gneiting, T. (2002). Nonseparable, Stationary Covariance Functions for Space–Time Data. *Journal of the American Statistical Association*, 97, 590–600.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221.
- Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. New York: Springer.
- Griffith, D. A., & Lagona, F. (1998). On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. *Journal of Statistical Planning and Inference*, 69(1), 153–174.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., et al (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310, 987–991.
- Gu, M. G., & Zhu, H.-T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 339–355.
- Guan, Q., & Clarke, K. C. (2010). A general-purpose parallel raster processing programming library test application using a geographic cellular automata model. *International Journal of Geographical Information Science*, 24(5), 695–722.
- Guan, Q., Kyriakidis, P. C., & Goodchild, M. F. (2011). A parallel computing approach to fast geostatistical areal interpolation. *International Journal of Geographical Information Science*, 25(8), 1241–1267.
- Guise, S. & Crooks, A. (in press). Agent-based modeling for community resource management: Acequia-based agriculture, *Computers, Environment and Urban Systems* 36(6).
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B*, 37(4), 682–703.
- Harris, R., Singleton, A., Grose, D., Brunsdon, C., & Longley, P. (2010). Grid-enabling geographically weighted regression: A case study of participation in higher education in England. *Transactions in GIS*, 14(1), 43–61.
- Haworth, J., & Cheng, T. (in press). Non-parametric regression for space–time forecasting under missing data. *Computers, Environment and Urban Systems* 36(6).
- Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51, 4282–4286.
- Heuvelink, G. B. M., & Griffith, D. A. (2010). Space–time geostatistics for geography: A case study of radiation monitoring across parts of Germany. *Geographical Analysis*, 42, 161–179.

- Horner, M., Zook, B., & Downs, J. (in press). Where were you? Development of a time-geographic approach for activity destination re-construction. *Computers, Environment and Urban Systems* 36(6).
- Huffer, F. W., & Wu, H. (1998). Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, 54(2), 509–524.
- Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383–401.
- Hsieh, W. W. (2009). *Machine learning methods in the environmental sciences: Neural networks and kernels* (1st ed.). Cambridge University Press.
- Jacob, R., Winstanley, A. C., Togher, N., Roche, R., & Mooney, P. (in press). Pedestrian navigation using the sense of touch. *Computers, Environment and Urban Systems* 36(6).
- Kanevski, M., Timonin, V., & Pozdnoukhov, A. (2009). *Machine learning for spatial environmental data: Theory, applications, and software* Har/Cdr. EFPL Press.
- Kleiber, W., Raftery, A. E., & Gneiting, T. (2011). Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *Journal of the American Statistical Association*, 106(496), 1291–1303.
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics – Theory and Methods*, 26(6), 1481.
- Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1), 61–72.
- Kyriakidis, P. C., & Journel, A. G. (1999). Geostatistical space–time models: A review. *Mathematical Geology*, 31(6), 651–684.
- LeSage, J. P., & Pace, R. K. (2011). Pitfalls in higher order model extensions of basic spatial regression methodology. *The Review of Regional Studies*, 41(1), 13–26.
- Longley, P. A. (1998). Foundations. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation: A primer* (pp. 1–15). Chichester: Wiley.
- Mack, E. A., Malizia, N., & Rey, S. (in press). Population shift bias in tests of space–time interaction. *Computers, Environment and Urban Systems* 36(6).
- Malleon, N., & Birkin, M. (in press). Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* 36(6).
- Manley, E. J., & Cheng, T. (2011). Multi-agent simulation of drivers reactions to unexpected incidents on urban road networks. In *Proceedings of GIS research UK conference*. Portsmouth, April 2011.
- Martin, R. L., & Oepfen, J. E. (1975). The identification of regional forecasting models using space: Time correlation functions. *Transactions of the Institute of British Geographers*, 66, 95–118.
- McCallum, Q. E., & Weston, S. (2011). *Parallel R* (1st ed.). Sebastopol, CA: O'Reilly Media.
- Miller, H., & Han, J. (2001). *Geographic data mining and knowledge discovery*. London: Taylor and Francis.
- Miller, H., & Han, J. (2009). *Geographic data mining and knowledge discovery* (2nd ed.). London: CRC Press.
- Mizuruchi, M. S., & Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, 30(3), 190–200.
- Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space–time cube: An exploratory data-analysis approach using space–time kernel density estimation and scan statistics. *Transactions in GIS*, 14(3), 223–239.
- National e-Infrastructure Service (2012). *The NES: Who we are and what we do*. <<http://www.ngs.ac.uk/>> Accessed 27.09.12.
- Neill, D. B. (2008). Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting*, 25(3), 498–517.
- NeSC (2012). *National e-science centre*. <<http://www.nesc.ac.uk/>> Accessed 27.09.12.
- Office of Cyberinfrastructure (2012). *Cyberinfrastructure – stimulating advances in 21st century science and engineering*. <<http://www.nsf.gov/dir/index.jsp?org=OCI>> Accessed 27.09.12.
- Openshaw, S., Charlton, M., Wymer, C., & Craft, A. (1987). A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1(4), 335–358.
- Openshaw, S., & Abrahart, R. J. (1996). Geocomputation. In R. J. Abrahart (Ed.), *Proc 1st international conference on geocomputation* (pp. 665–666). UK: University of Leeds.
- Openshaw, S. (2000). GeoComputation. In S. Openshaw & R. J. Abrahart (Eds.), *GeoComputation* (pp. 1–31). London: Taylor and Francis.
- Pozdnoukhov, A., Matasci, G., Kanevski, M., & Purves, R. S. (2011). Spatio-temporal avalanche forecasting with support vector machines. *Natural Hazards and Earth System Sciences*, 11, 367–382.
- Resnick, M. (1994). *Turtles, termites and traffic jams*. MIT Press.
- Schelling, T. (1978). *Micromotives and macrobehavior*. New York: W.W. Norton & Company.
- Shekhar, S., Zhang, P., & Huang, Y. (2010). Spatial data mining. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 837–854). US: Springer.
- Sui, D. Z., Elwood, S., & Goodchild, M. F. (Eds.). (2012). *Volunteered geographic information, public participation, and crowdsourcing production of geographic knowledge*. Berlin: Springer.
- Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path: The research and development agenda for visual analytics*. IEEE: Los Alamitos.
- Tomlin, C. D. (1990). *Geographic information systems and cartographic modelling*. Englewood Cliffs, NJ: Prentice-Hall.
- Torrens, P. M. (2012). Moving agent-pedestrians through space and time. *Annals of the Association of American Geographers*, 102, 1–33.
- Turton, I., & Openshaw, S. (1998). High-performance computing and geography: Developments, issues, and case studies. *Environment and Planning A*, 30(10), 1839–1856.
- White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B*, 24(1997), 323–343.
- Wise, S., & Crooks, A. (in press). Agent-based modeling for community resource management: Acequia-based agriculture. *Computers, Environment and Urban Systems* 36(6).
- Wright, D. J., & Wang, S. (2011). The emergence of spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences*, 108(14), 5488–5491.
- Xie, Y. C., Batty, M., & Zhao, K. (2007). Simulating emergent urban form using agent-based modeling: Desakota in the Suzhou–Wuxian region in China. *Annals of the Association of American Geographers*, 97(3), 477–495.
- Yeh, A. G. O., & Li, X. (2002). A cellular automata model to simulate development density for urban planning. *Environment and Planning B*, 29, 431–450.

Tao Cheng\*

James Haworth

Ed Manley

SpaceTimeLab, Department of Civil,  
Environmental and Geomatic Engineering,  
University College London, Gower Street,  
WC1E 6BT London, United Kingdom

\* Corresponding author.

E-mail addresses: tao.cheng@ucl.ac.uk (T. Cheng),  
j.haworth@ucl.ac.uk (J. Haworth),  
edward.manley.09@ucl.ac.uk (E. Manley)

Available online 2 November 2012