PERGAMON

Computers, Environment and Urban Systems
26 (2002) 257–265

Computers, Environment and Urban Systems

www.elsevier.com/locate/compenvurbsys

Editorial

GeoComputation: what is it?

1. Introduction

What is GeoComputation? This is a question that has intrigued us as participants in the GeoComputation Conference series and as local organizers of the Fourth International Conference on GeoComputation, GeoComp 99. There are a number of sources to draw upon—conference keynote addresses, journal editorials, books, and papers—to answer this question. In addition, there are a number of published definitions of geocomputation, ranging from Stan Openshaw's view of geocomputation as the application of high performance computing to solve currently unsolvable or even unknown spatial problems to Paul Longley's statement that geocomputation is simply what its researchers and practitioners do.

We prepared a poster entitled "GeoComputation...defining a discipline" (Fig. 1: downloadable from http://www.geovista.psu.edu/sites/geocomp99/poster.htm) for GeoComputation 99. The poster contained a series of key words and phrases that were extracted from the conference abstracts. The words and phrases were organized hierarchically, assuming that the larger text represented a broader concept and smaller text a more specific concept. Different type styles and colors were used for words representing each of the four basic concepts: space, time, data, and computing. Whereas the poster represented an interesting tool for visualizing the nature of geocomputation, it was clearly biased by the perceptions and experiences of its creators.

The GeoComputation 99 poster served as the inspiration for study of the abstract texts from the five GeoComputation conferences in greater detail, using more advanced tools for text analysis, to try to answer our initial question: What is GeoComputation? Longley (1998) states that at this point we must assume that GeoComputation is "...what its researchers and practitioners do, nothing more, nothing less..." (p. 9) We agree with Longley's statement and therefore used the cumulative body of research as expressed in the abstracts of the papers, posters, and keynote addresses from the five GeoComputation conferences between 1996 and 2000 to characterize GeoComputation (Table 1). In the tradition of exploratory data analysis, this approach provides a glimpse into the GeoComputation literature without introducing the prejudices of the researchers or producing a 'statistically significant' result. If indeed, GeoComputation is simply what its researchers and practitioners do, than this approach should provide some insight into what GeoComputation is.

PII: S0198-9715(01)00047-3

Fig. 1. GeoComputation…defining a discipline.

Table 1
Conference statistics

Conference location	Year	No. abstracts	No. authors/ co-authors	No. countries
University of Leeds, UK	1996	>98	>150	18
University of Otago, Dunedin, NZ	1997	45	100	8
University of Bristol, UK	1998	87	>150	11
Mary Washington College, Fredericksburg, VA USA	1999	83	>160	12
University of Greenwich, Chatham, UK	2000	58	>160	21

To do this, we used text analysis software from the Center for Intelligent Information Retrieval in the Department of Computer Science at the University of Massachusetts. Whereas there is bias in the selection of sources to evaluate, we believed this was the most comprehensive body of literature to date related to GeoComputation, and thus the most likely to successfully answer our question. Furthermore, it represents the range of topics that GeoComputation Conference organizers believed relevant and interesting to the field.

Dr. Judy Ehlen presented the detailed results of the study at GeoComputation 2000 in a keynote address entitled 'The Semantics of GeoComputation' at the University of Greenwich in September 2000. The full paper can be found in the proceedings of the GeoComputation 2000 at http://www.geocomputation.org/2000/GC006/Gc006.htm and in Abrahart and Carlisle (2000). This short editorial will not summarize the larger study, but will review a small selection of previous commentary on GeoComputation; describe the text analysis process; briefly summarize the results; and evaluate the analysis results with respect to the commentary.

2. Commentary on GeoComputation

GeoComputation has been described and defined numerous times in the past. Here, we will examine a series of statements from Stan Openshaw, Helen Couclelis, Bill Macmillan, Paul Longley, and Mark Gahegan, as well as recent commentary in *Geographic Information Systems and Science* (Longley, Goodchild, Maguire, & Rhind, 2001). Stan Openshaw is generally recognized as the father of GeoComputation. In the Preface to *GeoComputation* (2000), Stan Openshaw and Bob Abrahart see GeoComputation as a follow-on revolution to GIS; once the GIS databases are set up and expanded, GeoComputation takes over. They state that "GeoComputation is about using the various different types of geo-data and about developing relevant geo-tools within the overall context of a 'scientific' approach." (p. ix); it is about solving all types of problems, converting computer "toys" into useful tools that can be usefully applied.

In his paper (Openshaw, 2000), Stan identified three aspects that make GeoComputation special. The first is emphasis on "geo" subjects, i.e. GeoComputation is concerned with geographical or spatial information. Second, the intensity of the computation required is distinctive. It allows new or better solutions to be found for existing problems, and also lets us solve problems heretofore insoluble. Finally, GeoComputation requires a unique mind set, because it is based on "...substituting vast amounts of computation as a substitute for missing knowledge or theory and even to augment intelligence." (p. 5) Stan clearly sees GeoComputation as dependent upon high performance computing to solve currently unsolvable or even unknown problems.

Helen Couclelis (1998) adopted the working definition that GeoComputation, is "...the eclectic application of computational methods and techniques 'to portray spatial properties, to explain geographical phenomena, and to solve geographical problems.'" (p. 17) As such, she believes that GeoComputation currently makes little contribution to the scientific community. At present, she sees GeoComputation as merely a tool-based approach using tools derived from artificial intelligence research. GeoComputation must meet the theoretical challenge to "...formulate a model based on the computational notion of machine that justifies the 'geo' prefix." (p. 25)

In the Epilogue of *Geocomputation, A Primer*, the same book in which Couclelis' paper appears, Bill Macmillan (1998) essentially takes issue with her position. He believes that GeoComputation includes the latest forms of computational geography and that it is not an incremental development. He accepts that sound theory is needed, but believes that it has to some extent already been provided by Stan Openshaw, at least as a form of inductivism.

Mark Gahegan (1999), like Helen Couclelis, sees the concern of GeoComputation as "...to enrich geography with a toolbox of methods to model and analyze a range of highly complex, often non-deterministic problems." (p. 204) But he views GeoComputation as an enabling technology, one needed to fill the "...gap in knowledge between the abstract functioning of these tools...and their successful deployment to the complex applications and data sets that are commonplace in geography." (p. 206) Mark's is a practical approach to GeoComputation, but one with promise and vision, different from Helen Couclelis's philosophical, somewhat pessimistic, perspective.

More recently, GeoComputation has been discussed in *Geographic Information Systems and Science* (Longley et al., 2001). Here, the authors state "In some important respects, the term GeoComputation is synonymous with geographic information science...although it has often put greater emphasis upon the use of high-performance computers." (p. 140) This interpretation emphasizes that GeoComputation has more in common with geographic information science and its associated geographic information systems than was previously acknowledged within the GeoComputation community.

3. Text analysis procedures

It is interesting to compare the high level commentary on GeoComputation with the body of literature produced for the five conferences. Text analysis software provides the exploratory data analysis capabilities for evaluating the body of the

GeoComputation Conference Series textual materials. We initially hoped that the text analysis would be a straightforward method of analyzing the conference series abstracts, i.e. combine the abstracts for each conference, run them through the analysis programs, and compare the results. This hands-off approach would limit researcher bias, if not the bias of the text analysis software developers. Unfortunately, because language is sufficiently complicated and ambiguous in a multinational collection of written material, significant editing was required to prepare the files for analysis. In addition, interpretations of the word and phrase analyses were not as straightforward as we had hoped. Despite these limitations, the analyses produced results that were interesting and thought provoking, particularly as an exploratory exercise.

Preparing files for input to the text analysis software was the initial challenge. First, all abstracts for each conference were combined into a single file. These files were then edited to remove all parentheses, numbers, equations, special characters, bolding, etc. All references cited in the text or at the end of an abstract were removed, as were place names and the names of individuals and institutions. All acronyms and abbreviations were written out in full, except for "GIS" and "www". Finally, the English was standardized using the UK English option in the word processor spell checker. This was necessary to correct for differences in spelling such as color (US) and colour (UK) or defense (US) and defence (UK). The files were saved in ASCII format and sent to the Center for Intelligent Information Retrieval in the Department of Computer Science at the University of Massachusetts for word and phrase analysis. The process used to do the analysis is described in Feng and Croft (2000) and in Ehlen, Caldwell, and Harding (2000).

Two files based on these analyses were generated for each conference, one containing words and the other, phrases. Each file consists of the list of words or phrases sorted in decreasing order according to the number of times that word or phrase is used in the abstracts and the number of abstracts in which each word or phrase occurs (Table 2). The phrase "spatial analysis" thus occurs nine times in eight abstracts and the phrase "data models", six times in four abstracts. The two data sets for each conference can be found in Ehlen et al. (2000).

The results were also edited to facilitate the analysis. All words and phrases that occurred in only one abstract were deleted to reduce the files to manageable size.

Table 2
Example of phrase analysis software output

Phrase			Phrase frequency	Abstract frequency
spatial	Analysis		9	8
digital	Elevation	model(s)	7	4
spatial	variation(s)		7	2
functional	Pattern		7	1
data	model(s)		6	4
geographic	Space		6	2
visibility	index(ices)		6	2

Meaningless words such as "versa" in "vice versa", and "priori" as in "a priori"; and phrases such as "paper describes" and "wide variety", were also deleted. In addition, several caveats are required with respect to phrase analysis. First, spot checks of the original show that what the software identified as a phrase is not necessarily so. Examples of such errors include "biochemistry exhibiting reflectance" and "important research remains". This problem is at least partly due to the complexity of the English language: "Remains", for example, can be either a noun or a verb, and in this case the software identified a verb as a noun, producing a meaningless phrase. Second, we found that the software did not identify certain phrases of interest, such as "high performance computing" and "exploratory data analysis". Third, the software was arbitrary in its identification process. For example, we are interested in the phrase "artificial intelligence." Let us say that the software identified this phrase in two abstracts at one conference. But it also identified the words "artificial" and "intelligence" as parts of a different phrase, for example "artificial intelligence technologies", in two abstracts. We cannot combine these and say the phrase "artificial intelligence" occurs in four abstracts because we do not know whether the two abstracts in which "artificial intelligence" occurs are the same as or different from the two in which "artificial intelligence technologies" occur.

Finally, to analyze the results, each word or phrase was normalized by dividing the number of abstracts in which each occurred by the total number of abstracts presented at that conference. This not only permitted the words and phrases from one set of abstracts to be evaluated in a semi-quantitative manner, but also allowed comparison between conferences. Because plural phrase forms were merged with singular forms, some inflation occurred in the phrase frequencies.

Clearly the process was not as straightforward or as objective as initially hoped. The complications caused by multiple forms of spelling and the decisions we made in preparing the files for analysis affected the results. The output from the software did not always produce the expected results. Despite these limitations, the exercise did produce exploratory results that are useful for identifying important aspects of GeoComputation from the perspective of those who consider themselves "GeoComputationalists".

4. GeoComputation in context

Given the text analysis, it is only possible to address a few of the many questions of interest to the GeoComputation community with respect to the meaning of GeoComputation. We consider the following four questions to be most important at the present time with respect to the issues brought forward by Stan Openshaw, Helen Couclelis, Bill Macmillan, Mark Gahegan, and Paul Longley. First, is there a focus on high performance computing? Second, is GeoComputation simply a grab bag of tools? Third, what is the relationship between GIS and GeoComputation? And finally, what are the key concepts of GeoComputation?

Is there a focus on high performance computing? The phrase analysis did not identify the phrase 'high performance computing' in two or more papers at any conference. A quick search for phrases containing the words "performance" or "computing" for one of the early conferences, however, produces six occurrences of five phrases that refer to high performance computing—high performance computing, high performance, performance computing, high performance computing hardware, and performance computing. A comparable search for the word "parallel" as in "parallel processing", produced more than 20 occurrences in more than 10 phrases, including parallel computing, parallel programming, parallel architecture, parallel supercomputing, and parallel hardware. This suggests that although the bulk of the work considered to be GeoComputation by researchers is in fact not related to high performance computing or computers, there is only a small component within the GeoComputation community that is addressing this issue. This is essentially contrary to the ideal set by Stan Openshaw (2000; Openshaw & Abrahart, 2000) and reiterated in Longley et al. (2001).

Is GeoComputation simply a grab bag of tools? Unlike fields within GeoComputation which have developed from theory, such as geostatistics, our text analysis indicates that GeoComputation lacks a single focus. Many different techniques are encompassed in GeoComputation, including "neural networks", "cellular automata", "genetic algorithms", "expert systems", "fuzzy modelling", and "dynamic modelling", but occurrences of these phrases are low, suggesting that a wide variety of approaches and techniques lend themselves to the GeoComputation environment. Although our phrase analysis shows only an increasing interest (see earlier), we believe GeoComputation researchers operate primarily in the realm of Mode 2 science, focusing on applications and results, rather than laws and theories. In fact, the words "law" and "theory" do not occur in the top 25 occurrences at any conference; phrases containing these words, which often do not occur in an entire set of conference abstracts, reference specific theory outside GeoComputation, such as Darcy's Law or graph theory. Whether this is viewed positively, as Gahegan's enriching set of tools, or negatively, as Couclelis' grab bag of tools without theoretical foundation, cannot be answered using text analysis tools.

What is the relationship between GIS and GeoComputation? Longley et al. (2001) state that "In some important respects the term GeoComputation is synonymous with geographic information science." (p. 140) With respect to geographic information science, they further state that "Other terms have much the same meaning: geomatics and geoinformatics, spatial information science, GeoComputation, geoinformation engineering. All suggest a scientific approach to the fundamental issues raised by the use of GIS and related technologies…" (p. 21) Whereas GeoComputation researchers do not often address geographic information science per se, geographic information systems are very much on their minds. GIS was among the top seven words at every conference and was the most commonly used word at the 1997 conference in Dunedin, NZ. However, word analysis does not provide context, a key problem with "GIS". We found the term used in different ways. In some cases, it is mentioned as the antithesis of GeoComputation, while in others it is noted as a tool for addressing GeoComputation applications. The lack of context in word and phrase text analysis is one of its more significant limitations.

What are the key concepts of GeoComputation? The most frequently occurring words resulting from the analysis were data and spatial. Data was the most frequently occurring word at four of the five conferences and second at the fifth. Spatial was the second most frequently occurring word at four of the five conferences and third at the fifth. Clearly the "geo" aspect is well-represented in the literature. The next tier of frequently occurring words includes model, information, analysis, and GIS. Patterns of word usage become more fragmented from conference to conference and less clear further down the lists.

Analysis of phrases over the 5-year period gives a slightly different, and more comprehensive, image of GeoComputation than the analysis of words. Data phrases, such as spatial data, data sets, were the most frequently used phrases at the earlier conferences, whereas phrases relating to the tools used in GeoComputation, e.g. neural networks, genetic algorithms, cellular automata, were the most frequently used in the later conferences. Other frequently used phrases at all five conferences dealt with analysis and modelling (e.g. spatial analysis, hydrologic modelling). Moreover, phrases addressing the more practical aspects of GeoComputation, those dealing with applications (e.g. resource management, drainage basins, urban areas), results (e.g. fractal dimension, correlation coefficients, spatial patterns) and products (e.g. self-organizing maps, operational systems), increased in frequency of use from 1996 to 2000.

5. Reflections

It is always difficult to capture the essence of a discipline like GeoComputation, where discussions and debates on the definition of the term and the nature of the field are on going. Information can be gleaned from the writings of leaders in the field as well as from the body of literature produced by rank and file researchers. Tools, like text analysis, provide a mechanism for exploring this larger body of literature. While imperfect and sometimes difficult to use, text analysis of the literature eliminates some of the biases of the researchers and can reveal unexpected insights. We hope this brief attempt to relate the results of text analysis of abstracts presented in the five GeoComputation conferences between 1996 and 2000 to commentary from leaders in the field has been useful, and will assist in the overall development and evolution of GeoComputation.

References

Abrahart, R. J., & Carlisle, B. H. (Eds.) (2000). *GeoComputation 2000*. Proceedings of the 5th International Conference on GeoComputation, 23–25 August 2000, Chatham, UK. Greenwich, UK: GeoComputation CD-ROM.

Couclelis, H. (1998). Geocomputation in context. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation, a primer*. Chichester: John Wiley and Sons.

Ehlen, J., Caldwell, D. R., & Harding, S. (2000). The Semantics of GeoComputation. In R. J. Abrahart, & B. H. Carlisle (Eds.), *GeoComputation 2000, Proceedings of the 5th International Conference on GeoComputation, 23–25 August 2000, Chatham, UK*. Greenwich, UK: GeoComputation CD-ROM.

Feng, F., & Croft, W. B. (2000). *Probabilistic techniques for phrase extraction* (Technical Report IR-187). Amherst, Massachusetts: University of Massachusetts, Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts.

Gahegan, M. (1999). What is Geocomputation?. *Transactions in GIS*, *3*, 203–206.

Longley, P. A. (1998). Foundations. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation, A Primer*. Chichester: John Wiley and Sons.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic information systems and science*. Chicester: John Wiley and Sons.

Macmillan, B. (1998). Epilogue. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation, a primer*. Chichester: John Wiley and Sons.

Openshaw, S. (2000). GeoComputation. In S. Openshaw, & R. J. Abrahart (Eds.), *GeoComputation*. London: Taylor and Francis.

Openshaw, S., & Abrahart, R. J. (2000). Preface. In S. Openshaw, & R. J. Abrahart (Eds.), *GeoComputation*. London: Taylor and Francis.

Judy Ehlen, Douglas R. Caldwell
U.S. Army Engineer Research and Development Center,
Topographic Engineering Center,
Alexandria, VA 22315-3864,
USA
E-mail address: jehlen@tec.army.mil (J. Ehlen)

Stephen Harding
Center for Intelligent Information Retrieval,
Department of Computer Science,
University of Massachusetts,
Amherst, MA 01003-4610,
USA