COMPUTATIONAL HUMAN GEOGRAPHY: TOWARDS A RESEARCH AGENDA

Stan Openshaw

WORKING PAPER 93/24

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

WORKING PAPER 93/24

COMPUTATIONAL HUMAN GEOGRAPHY: TOWARDS A RESEARCH AGENDA

Stan Openshaw

School of Geography
University of Leeds
Leeds LS2 9JT

Seemingly over the last decade a major computing revolution has started in many of the 'hard' sciences. Slowly at first, but now at an ever increasing pace, the principal methods used at the research frontiers of science have become increasingly computation dependent. New ways of doing science have been developed in physics, chemistry, biology, engineering, geology and environmental modelling; indeed major grand challenge areas have been recognised in many sciences which now possess the relatively new adjective "computational" (Catlow, 1991). It is important to realise that these changes reach far beyond the hard sciences and have the most profound implications for geography and most social sciences too. The purpose of this commentary is to investigate some of the scientific and technological opportunities in geography that are to be created by the new computational era likely soon to exist. Maybe we are witnessing the beginnings of a new technological and intellectual revolution in science with consequences far beyond what we can currently imagine. Certainly these developments constitute an argument for the development of a computational human geography (CHG). Indeed, there is now a large gap between what most geographers see computers as offering and what is now, or will soon be, computationally feasible. The computing world has started to change in big ways and these changes provide the basis for exciting new computationally dependent ways of doing better geography. So what are these developments?

Massively Parallel Processing

Many computer scientists are today very excited about the impending massively parallel computing era that is dawning. As Hillis (1992) puts it "... there is a significant technology transition that is taking place in computers, and since it radically changes the costs and capabilities of information processing, it is likely to change our lives. The new technology is called massively parallel computing ... the implication of these ideas is that the computer revolution has not yet begun to reach its limits ... It is always easiest to believe in a future that is a minor extrapolation of current-day trends. Such an extrapolated present is unlikely to happen in a time of rapid technological change". Expressed another way, a new era in computation is beginning with far reaching consequences for both science and geography. After a long gestation period, distributed memory massively parallel processor (MPP) hardware now seems to be on the verge of offering a realistic prospect of teraflop computing with terabyte memory spaces by the late 1990s. If the MPP route to teraflop computing falters, it is noted that the more conventional vector-parallel route is only a few years behind. It is likely that compute speeds will increase from a peak of about 2 gigaflops on a Cray YMP8 to about 10,000 gigaflops within the next 10 years, and memory sizes from about 1 gigabyte towards terabytes. We are at the beginning of a new era of computing, an era likely to be characterised by the most dramatic changes in computing capabilities ever witnessed.

When these developments in computing technology are translated into a quantitative geographical perspective, the magnitude of the changes are perhaps more readily appreciated. The next 10 years will probably see a speed-up in number crunching ability by about 10^9 times since the first quantitative revolution of the early 1960s; a 10^8 times speed-up since the mathematical modelling revolution of the early 1970s, a colossal 10^6 speed-up since the beginning of the GIS revolution in the mid-1980s, and a 10^3 improvement over today's supercomputers. Computer memory sizes alone will have increased by about 10^5 times since the quantitative revolution and are continuing to rapidly expand to accommodate even the largest databases. This is fortunate since over this 30 year period there has also been a very large increase in the size and numbers of geography relevant databases and it is only now that the computing hardware is showing signs of being able to cope. Many of these changes are exponential and it seems that it is only now we are about to reach the steep part of the curve. The challenge for geography is to discover how best to exploit the vast explosion in computational power and information resource to develop a new and more appropriate form of human geography, suitable for the 21st Century. Perhaps, we can make progrss by becoming more computational. It can be observed that the nature of geography has always reflected the technology and the needs of the societies in which it was performed. It seems that geography maybe today approaching a key turning point in its long history.

A scientific human geography

The increasing severity of many of the world's social, economic, political and environmental problems together with great uncertainty about global change of an environmental and climatic sort, places a growing emphasis (or it should be) on the science of developing new technologies targeted at analysing, understanding, modelling, and controlling all the "key" human systems. At a time when most of the world's "silly" global climatic change models essentially treats human systems as fixed unmodellable exogenous inputs, how can geographers continue to ignore the modelling and analysis challenges being presented to them by contemporary society? It seems absurd that we are developing massively sophisticated models of oceanic circulation and weather systems on a global scale but there is not a single useful model of the physical, environmental, and socioecononic systems operating in even a single city, at a time when most of the world's population live in cities. I wonder why; and who is to blame for this neglect? Is the notion of a new scientific human geography still so philosophically absurd or irrelevant as some undoubtedly either consider or once considered it to be? In my view there is much good basic science yet to be done in human geography and its time we started doing much more of it. The awesome changes in computing resource, the emergence of the IT state, the spatial data riches being generated by the GIS revolution in the context of a computer saturated bureaucracy, create many, many, new opportunities for doing computational human geography and for re-assessing what science can now be done in a vast range of different human geographic contexts.

A definition

So what is computational human geography (CHG)? It is the adoption of a computational approach and everything that means for doing human geography. It is not the use of computers that distinguishes the forthcoming revolution but the development of a new computationally intensive and totally computer dependent paradigm in geography. Some of the basic principles can be listed as follows:

(i) You seek to solve problems involving numeric and symbol (non numeric) data by adopting a computational solution; including those previously tackled using analytical methods.

(ii) You seek to try to work in an explicitly and massively data parallel and vector/parallel fashion in order to exploit the superfast hardware.

(iii) You embrace the basics of an implicit new scientific faith, which encourages you to believe that many previously impossible or poorly handled analysis and modelling tasks can now be solved properly and optimally once sufficient compute power is available. This is very relevant because Geography is a computable subject for two reasons: first, very little of what geographers do is not computable; second, most of the raw material of geography is already in computer forms.

Ok, you think (maybe), it sounds good but what problems in human geography need such a heavy computational intense approach? Ah, its here where your prejudices may well either tunnel-vision or obliterate your imagination.

That famous, dead, physicist and nobel prize winner Richard Feynman once had this to say about the social sciences in a 1981 interview that was repeated in a 1993 TV documentary (Feynman, 1993). He said:

> "Because of the success of science there is a kind of pseudo-science. Social science is an example of a science which is not a science. They don't do scientific things. They follow the forms, they gather data, you do so and so, and so forth, but they don't get any laws; they haven't found anything; they haven't got anywhere, yet. Maybe some day they will, but it's not very well developed. But what happens is, at an even more mundane level, we get experts on everything that sounds like they are sort of scientific experts, but they're not scientific. They sit at a typewriter and they make up something like, 'Food grown with fertiliser that's organic is better for you than food grown with fertiliser that's inorganic.' May be true, may not be true, but it hasn't been demonstrated one way or the other. But they'll sit there on the typewriter and make up all this stuff as if it's science and then become an

expert on foods, organic foods. There are all kinds of myths and pseudo-science all over the place."

It is true Feynman was not explicitly referring to human geography, but he might well have been; it certainly fits large parts of it! There is nothing intrinsically 'bad' about a 'soft' pseudo-science path to describing and maybe even understanding how this or that system appears to operate. Openshaw (1991) termed this the Catherine Cookson approach to geographical explanation, and it does work to some degree. Its not all rubbish. It offers a simplified, essentially qualitative picture of seemingly high complex and hitherto unmodellable human systems and processes, not previously thought susceptible to any more rigorous form of scientific understanding. Faced with such complexity of hard to model human systems, the frailness and conceptual poverty of the original 1960's scientific quantitative geographical paradigm soon became apparent and long ago a counter-revolution began. Now see Figure 1; for how much longer will such key human processes remain unmodellable? A computational human geography will provide the tools for a much more rigorous scientific treatment of this type of problem. Feynman's criticism of social science was the lack of testing of qualitative assertions ("maybe true, maybe false, but it hasn't been demonstrated one way or the other") and the consequential lack of any scientifically acceptable theory (and science credibility). An obvious analogy here is between qualitative human geography and medieval astronomy; in the latter case the descriptions were plausible, declared ideologically sound, but were later shown to be scientific rubbish.

Re-Modelling people based systems

Recently, I re-read most of the early books of Alan Wilson (1970, 1974, 1981); my predecessor at Leeds. What stands out so clearly is his concern with both science and the rigorous development of soundly based mathematical models of people related systems. His works show that considerable progress was made but that there were also very severe limits as to what could be achieved 20 years ago. For example, the mathematical complexity of using entropy-maximising methods as a general purpose model building tool kit have now been greatly reduced by the development of software packages able to solve in a symbolic form the relevant equations. This combined with the availability of a billion times more compute power opens up the propsect of realising some of Wilson's 1970 dreams.

However, not all old problems can be so easily solved. The mathematical model building tools turned out to be best suited to Weaver type I and II systems. Wilson demonstrated that very large systems containing large numbers of weakly interacting components (systems of disorganised complexity) often possess simple macrolaws that could be used as the basis for mathematical modelling. However, many human systems in geography are much more complex with large numbers of components some of which are strongly interacting. There are no longer any simple macrolaws and basically these Weaver type III systems of organised complexity were beyond reach. New approaches are needed and these are now becoming available. It is useful to remember Wilson's original work was performed at a time when computers were about 10^8 times slower than today, small in memory, and there was virtually no relevant data. Additionally, there was also a resolution and fundamental representational difficulty - the models could only deal with zones rather than people, whilst those microanalytical simulation models that focused on people suffered such severe data constraints and computational bottlenecks as to be considered almost infeasible.

Some Computational Geographical Grand Challenges

Today, none of these previously insoluble problems appear quite so insuperable. An improved understanding of the science of complexity is yielding new approaches that in principle seem powerful enough to offer some hope of being able to cope with the Weaver type III problems of human geography. We could in principle, if we now wished to do so, seek to model entire populations at the people level as well as at zonal levels of resolution. Indeed, lets be quite clear here, there is probably no longer any human system that could not be modelled in all its spatial and aspatial complexity if we had a good reason so to do, over the next 10-20 years. If we prefer not to contemplate this challenge, then we could at least start by improving the inadequate subsystem models still in use, extending their scope, and optimising their usefulness. Maybe this task is more than enough for the rest of the 1990s. CHG is not a process of rediscovering old methods or even that of re-engineering old models to meet old problems in new contexts, but is predicated on a belief in the importance of the need to developing new models of previously unmodellable systems to meet new problems and needs in new contexts.

In other areas too computational geography provides the basis for potentially useful scientific advances. Smart explorations of the tri-space (space-time-attribute) data domains that characterise GIS databases for evidence of pattern and relationships by suggesting WHERE, WHEN AND WHAT to look for should be feasible; Openshaw (1992) outlines one way of doing this, but there are doubtlessly many others. How wonderful it would be, if we could release some of the 'trapped' patterns, relationships, and latent knowledge that are held in the world's spatial data resources. Some of the findings might represent major new theoretical advances, some might yield vital patterns that are commercially valuable or save lives, others might be complete rubbish. The point is, we can now think about looking without having to know answers to what/where/when questions before we can even start, provided the search and analysis processes we use are capable of massively parallel computation. This is urgent and important because currently we are essentially blind when faced with the immense multidimensional complexity of the cyberspaces that are our own creations. In the 18 and 19th Centuries geographers explored new continents (a three dimensional map based exploration); in the 21st Century we will have to become explorers again but this time of the geocyberspaces which now abound and which are infinitely more dimensionally complex.

It is also most important not to neglect theory. Sadly, we have not got very far, we do not have much at present. What exists is typically old (Von Thunen, Christaller, Hoyt etc) or the sort of pseudo scientific myth based stuff that Feynman would have found so objectionable. At the very least we should be able to properly test, extend, and improve on the simpler map related theories of our ancestors. Pattern recognition and robotic vision technologies might be most useful here. For instance, is the residential social structure pattern of Leeds similar to that of Tyneside? The traditional statistical approach is to derive a null hypothesis that can be tested but this is unlikely to contain more than a minute fraction of the content of the original geographical question. The focus in the original geographical question was on "pattern" as a spatial object. Traditional statistical methods have always demanded that geographers be more precise than it is safe to be and in consequence have generally failed the first test of geographical usefulness. If you take the same question and adopt a more general computer vision perspective, then you would now rephrase the question and ask whether there is any scale and rotationally invariant generalised spatial pattern that is common to both Leeds and Newcastle. The answer might well be "yes", or at least one that allows the creation of a small library of different but recurrent residential social structure spatial pattern types. We can just about do this now and thus stumble towards the beginnings of a new concepts and theory rich form of geographical analysis. The secret is again to automate the concepts discovery process and apply it universally for gaining insight and creating knowledge at the cost of massively parallel machine time.

In CHG we aspire to greater intelligence by first becoming less sophisticated in our analysing and modelling technology and then more than compensating by computational intensity. In ten or so years time many of these developments will have been taken for granted. The challenge for the present is how to bridge the gap between the current situation of virtually no significant use of high performance computing in geography to a situation where there can be no significant geographic research without it. There is now much less than 10 years left.

A Teracomputing research manifesto for geography

There are a number of obvious and important geographical themes that are waiting for teraflop computing, indeed some potential applications are so compute intensive it just was not simply worthwhile even considering them previously for supercomputing. In many senses the "super" was simply not "super" enough (neither big enough nor fast enough to make their use worthwhile). This situation still exists but now is the time to start planning for the new era of parallel computation. It is an appropriate moment to move on from the data dominated GIS era towards a computational geography. A number of important needs exist, including that of raising awareness of what is now possible; training oneself in new programming skills, conversion of existing code to provide demonstrators and a corpus of basic supercomputer dependent applications, and to write new code to exploit the massively parallel machines by seeking to perform tasks that were previously impossible. As the GIS revolution showed so clearly, all this takes time; planning needs to start now to significantly exploit the potential offered by teraflop hardware in both geography and the social sciences more generally. New research initiatives are needed and skeletal research agendas drawn up.

What follows is my attempt to define some of the perceived methodological themes and opportunities in the areas relevant to Environment and Planning. They are listed in no particular order:

(1) Human systems modelling involving at one extreme microlevel representation of the key components and, at the other, directly modelling the nonlinear dynamics, chaotic and emergent behaviour, and error propagation at fine spatial levels of resolution of complex adaptive systems.

(2) Location optimisation problems include optimal siting, utility maximisation, and zone design can now be re-visited with both many more zones, greater realism in embedded models, and improved heuristics.

(3) Exploratory spatial analysis and modelling of very, very, large GIS databases using smart search and AI based data mining procedures, in a whole range of relevant

applications areas which are now data rich, theory poor, and result deficient.

(4) Natural language processing of text information relevant to geography held in an electronic form; for example, place and map feature names, fore-and surnames, and street address databases. Also related is spatial cognition and studies of the usage spatial language as contained in speech, in texts of political debate, in interviews, in wayfinding, in music and literature. Digitised sound and digital images can also be processed for geographical content. The complexity and size of some of these non-numeric applications should not be underestimated;

(5) A re-working of basic spatial geographic theories and concepts via robotic vision and pattern recognition techniques, and other automated scientific discovery systems.

(6) Use of neurocomputing and other AI based methods to model soft information, putting science back into qualitative research, where ironically it might be argued a computational approach is both highly relevant and offers considerable potential early benefits.

(7) Certain specialist digital map and GIS operations that are massively compute intensive; for example, automated context sensitive digital map generalisation; and also involving massive amounts of data.

It should go without saying but often needs to be said, "you do not compute or model just for the hell of it". The success or failure of a CHG is not sensibly measured in terms of terabytes or gigaflops but rests on the substantive results that are obtained.

These are exciting times to be a human geographer provided you are neither blinded by past prejudices about science, nor scared of the words computational human geography, nor too hyped-up by an over enthusiasm for AI or infected by the neural net virus. It needs to be recognised that there are new paths to scholarship which are entirely computer based that will significantly extend our capabilities. This is true in CHG as it is in the humanities as a whole (British Library, 1993). At the research frontiers, it now seems that the faster the computer technology becomes the greater the range of new opportunities and the better the scholarship. With a little luck, it should be just about possible to re-establish human geography as a key science in the most important grand challenge area of them all; understanding people and their behaviour in space. Certainly, this CHG is technology led but we would be really daft if we don't seek to exploit the new opportunities to the maximum extent; wouldn't we?

References

British Library, 1992, **Information Technology in Humanities Scholarship: British Achievements, Prospects and Barriers** The British Library Research and Development Department and British Academy, Report 6097, HMSO, London

Feynman, R., 1993 Horizon: No Ordinary Genius: Part 2, Transcript p6-7, BBC Enterprises, London

Hillis, W.D. 1992, 'What is massively parallel computing, and why is it important?! **Daedalus Journal of the American Academy of Arts and Sciences**, Winter 1992 1-16

Openshaw, S., 1991, 'A view on the GIS crisis in geography or using GIS to put Humpty-Dumpty back together again; **Environment and Planning** A 23, 621-628

Openshaw, S., 1992, 'Some suggestions concerning the development of artificial intelligence tools for spatial modelling and analysis in GIS; **Annals of Regional Science** 26, 35-51

Wilson, A.G., 1970, **Entropy in Urban and Regional Modelling**, Pion, London

Wilson, A.G., 1974, **Urban and regional models in geography and planning**, Wiley, London

Wilson, A.G., 1981, **Geography and the Environment**, Wiley, London