Frank, A. U. "Geographic Information Science: New Methods and Technology." *Journal of Geographical Systems, Special Issue: Spatial Analysis and GIS* 2, no. 1 (2000): 99-105.

Part 4 Technology and the future of GIS and spatial analysis

Geographic Information Science: New methods and technology

Andrew U. Frank

Department of Geoinformation, Technical University Vienna, Gusshausstrasse 27-29/127, A-1040 Vienna, Austria (e-mail: frank@geoinfo.tuwien.ac.at)

Abstract. Geographical Information Science is essentially computational geography and has its own research program, namely all aspects of formal models for spatial natural processes and the interaction of humans with the environment in space and time. This is not a question of technology and technology-related research; but technology influences what questions can be researched effectively. Collection of data in the field and the simulation of field experience through Virtual Reality are just two questions of how spatial reality and human experience are linked. The focus on human spatial cognition is similarly found in software engineering for interoperable Geographic Information Systems.

Key words: Multi-agent systems, computational geography, field-computing

1 Introduction

In a time of rapid technology change a reflection on the interaction between technology and methods used in GIScience is necessary. During the second half of the 20th century a new 'data processing' technology developed, the effects of which are only comparable to Gutenberg's invention of the printing press 500 years ago (Negroponte 1995). What we see is only the beginning: the fundamental limits to faster and smaller computers are still orders of magnitudes away and the trend to faster, smaller and less expensive computers will continue.

Changes in technology also change the research environment: some questions we battled with are solved by sheer computational speed. Other questions come into scope as we acquire tools to successfully tackle them. While these changes are going on we must guide university research to concentrate on fundamental questions of lasting importance, not on the 'problem of the day'. This implies that some topics which were appropriate questions yesterday are not worth our effort today. University research needs several years from start to results and is most successful if it addresses questions that will be very important in five or more years. Issues which are basically further development of known solutions to make them more efficient should be left to the R&D efforts of industry.

Technology is not the issue for GIScience. Technology-driven research is in danger of obsolescence, often before results are published. The computing power for most of the questions in geography and spatial analysis is available today, often directly on the desktop of the researcher. As an example, the PC-Raster program shows simulated geological processes with details and speed inconceivable even last year (Burrough – this issue). Academic research must concentrate on the fundamental questions of GIScience, even if the devil is tempting us with attractive R&D questions funded by industry, or well-funded research projects in other sciences, for which we can provide GIS technology.

There is a GIScience research agenda in its own right and I will elaborate here on what I think are the fundamental questions posed today. There is first the need for computational geography, which produces formal results which can be incorporated in computer programs and then widely used. Pressing, but difficult to solve problems are mostly related to an understanding of meaning, not only in geography but generally in science (Dreyfuss 1998), and are discussed in 'national bestsellers' (Pinker 1997). I select three viewpoints, namely computational geography, field computing, and the connection between software engineering and science, as three topics to point to the challenges posed and to the new methods to attack them.

2 Computational geography

In one philosophical tradition, understanding means the construction of models (Newell 1997). Models are useful if they predict the outcome of actions. Much of human learning can be seen as construction of mental models – even in the first months of life (Gopnik and Meltzoff 1997).

Geography is not about collecting facts, expressed as proposition in logic, but about understanding the causes – the processes in space and time – which created these facts. "Information itself is nothing special; it is found wherever causes leave effects" (Pinker 1997). Actual research is limited by our ability to describe natural processes in a formal framework.

Logic is not the appropriate formalism for a description of change and process – despite efforts (McCarthy 1980; Reiter in preparation).

In the past, differential equations were the only well known formalism to describe processes which affect change in time and space. The quantitative revolution in geography (Gould 1999) used this tool (Tobler 1976), but two limitations emerged: (1) most problems which can be solved in closed form with differential equation are not interesting, and most interesting problems cannot be solved in closed form, usually due to problems with complex boundary conditions; (2) differential equation forces a quantitative description of a problem, which is adequate for problems of physical geography, but not suitable for topics from human-geography.

The computing power available now overcomes the first limitation. Instead of closed solutions, numerical approximations are used. Physical geography uses these tools with success: PC-Raster (Peter Burrough – this issue) is perhaps the most impressive result. Numerical solutions allow a visual control of the result and convince of the 'correctness' of the model with respect to processes we find in the real world.

Topics in human geography resist quantification, which is necessary for numerical solution. Statistical aggregation is necessary to use methods based on differential equation. Spatial analysis research has developed an impressive array of methods. What is missing is the linkage between the individual behavior, which is the cause of what we observe, and the analyzable statistical values.

A large array of computational methods are used in disciplines closely related to geography: dynamic models are successfully used in biology and economy to simulate the interaction of complex agents for similar problems, in general not including effects of spatial distribution (Zeigler 1976; Swarm Development Group 1996; Ruth and Hannon 1997). The new research discipline of **multi-agent systems**, emerging in computer science (O'Hare and Jennings 1996; Weiss 1999), produces models of multiple agents, e.g. human beings, agencies, or companies that act and interact in and with a common environment. Multi-agent models can locate agents and other resources of the environment in space and thus include the effects of space on the behavior of the agents and the effects of the agents on the environment. Geographers are asked to introduce the spatial dimension in multi-agent systems. Geographical research, especially in traffic, wayfinding and population movement (Benenson and Portugali 1995) can be tackled with such methods. A few years ago, a model of 100 agents in a 'landscape' of 200 by 200 cells, where resources are distributed irregularly (Epstein and Axtell 1996), was at the limit of computational power but larger systems can be tackled today.

In such multi-agent systems – a further development from cellular automata already used in geography (Takeyama and Couclelis 1997) – global properties of the system as a whole emerge from the individual behavior of the agents based on local information. It appears that category theory (Barr and Wells 1990; Asperti and Longo 1991) could contribute to understanding how these global properties emerge from local rules. For early applications of category theory to GIS see the contributions from Herring (1989; Herring, Egenhofer et al. 1990).

3 Field computing

Modern computers are smaller and portable – we can bring computing power into the field. At the same time, Virtual Reality systems make it possible to bring realistic representations of the field to the office. Field computing investigates the effects of these two enabling technologies. The collection of data in the field is very expensive and will remain expensive even with new methods. Can data quality be improved? More data can be collected at the same time, especially data that can be sensed and recorded automatically. Checking for error and completeness, and comparison with already available data in the field may allow data collection by non-specialists. Exciting, practical problems are posed, but we must not lose sight of the fundamental research questions:

1. How do observations of properties in the field relate to the recorded observation values? Data quality descriptions focusing on data and collection methods (Morrison 1988; Goodchild and Gopal 1990) can be extended to take the use of the data into account (Timpf, Raubal et al. 1996).
2. It has been suggested (A. Buckley position paper for Varenius meeting) that field trips – an important, time consuming and expensive part of geo-

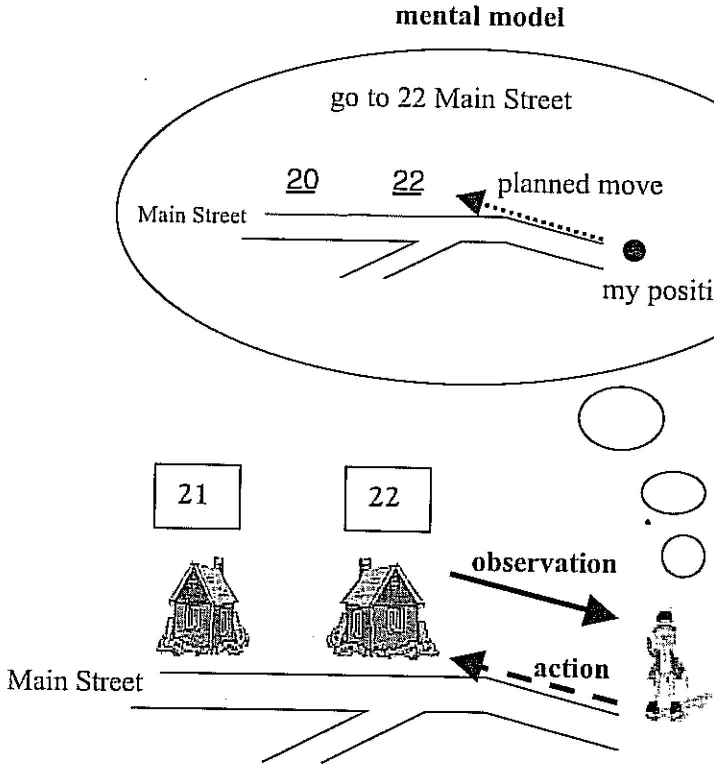

Fig. 1. A person imperfectly observing the world and using the observation to decide an activity

graphic education – can be replaced by Virtual Reality experiences. What are the cognitive, experiential or emotional differences between a field trip with 'bus and boots' compared to a Virtual Reality experience?

These two questions are inverse to each other: The first investigates how experiences in the real world relate to data values, the second how data values can be translated into experiences. The results can be compared with the original experience in the real environment. This creates a new paradigm to discuss data quality in a wide sense, including human dimensions like emotions (Picard 1997). Generally, it shows how to construct links from the experience of reality to the information system and then back to reality (Fig. 1). This is comparable to a topographic mapping agency observing reality and producing a map that is then used to navigate (Frank to appear). In this 'closed loop', data quality and semantics of data in general can be formalized with algebraic methods (up to isomorphism) tying the data collection process to the process of using the data for decisions about actions which are executed in the world (Fig. 2).

Field computing is a much larger field than scientific geography: all humans are permanently on a field trip. Often we are in an unknown environment and lack vital information (Kottman 1998): Where is a good restaurant? Where is an open gas station? In which century was this building built? Field computing could answer such questions – using a combination of existing technology GPS, GSM and personal assistant (like Palm Pilot). Several research groups

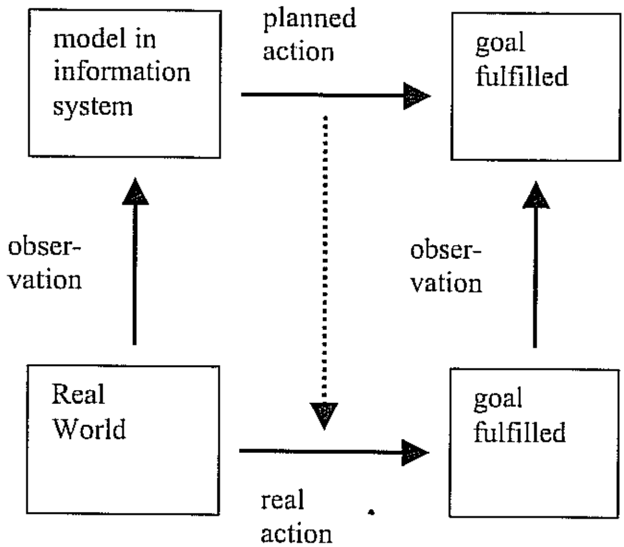

Fig. 2. Categorical diagram for isomorphism in Information system

investigate such systems; the first application is in tourism, the most important growth sector in today's economy.

4 Software engineering: Object-orientation, component ware and open GIS

As computer systems become more powerful, the focus of software engineering shifts from the production of efficient computer programs to the efficient production of computer programs. Writing computer software is not unlike writing computational models in GIScience. The problems the scientist faces are similar to the problems the analyst and programmer must overcome. Most advanced software engineering methods can be useful in computational research (Frank 1999).

Combining pieces of knowledge that seem to be well understood and described in some fashion is difficult. The complexity of the potential interactions surprises – with the result that all too often the resulting program does not work. Software engineering studies methods to formalize the interaction of formalized subsystems, forming clusters where interaction within the cluster is maximized and interaction with the outside minimal. Performance of the code becomes a secondary issue behind the desire to build software that works. Of what use is a very fast system that produces the wrong results? Geography integrates results from different sciences with respect to their spatial effects and therefore faces much the same problems.

The tendency in software engineering is to move from a technology concern towards a cognitive orientation and to better mathematical foundations: object-orientation is the current focus. The commercial languages (especially C^{++}) are efficiency-dominated (Stroustrup 1986) and are fundamentally flawed by a theoretical point of view (contravariance of functions, Abadi and Cardelli 1996). Cleaner languages, closer to mathematics emerge, directly linking to denotational semantics (Stoy 1977). For example, functional programming has overcome the usability problems, which made Backus (1978) describe them as theoretically interesting but not practical, and have integrated classes (Bird 1998) to give a mathematically clean form of object-orientation in the form of parameterized polymorphism (Wadler 1989). To make best use of such tools is crucial for computational geography. Object-orientation based on algebra (Birkhoff 1945) and category theory (Barr and Wells 1990; Asperti and Longo 1991) opens the door for a computational approach which is not forced to quantification.

5 Conclusions in 6 theses

1. Geography is about how people interact with their environment. GIScience is computational geography, modeling processes in space. Research must focus on advancing formalization of our understanding of processes in space and human interactions with the environment.
2. Technology is not the issue: More computing power is available than we know how to use; research to improve computational efficiency is nearly always obsolete after a short while.
3. Technology development influences what are researchable topics. New problems come into focus as they can be tackled. Academic research must focus on issues of lasting importance, not the 'problem of the day' suggested by available technology and its current limitations.
4. The formal tools available to describe processes in space and time limit geographic research. Standard logic is not suited to describe change and differential equations force quantification. Multi-agent systems – a further development of finite element methods, cellular automata and dynamic models – provide a general framework to describe and simulate the behavior of autonomous agents in an environment. Geographers are invited to contribute the spatial dimension to this research.
5. Technology affects the way we collect data and how we communicate it to the user. We can bring computing power to the field to assist with the data collection. Virtual Reality, on the other hand, allows recreating the field in the lab, asking questions of how data values are translated into human experience. Linking the process of data collection and data use in decisions affecting the world allows a comprehensive description of data quality.
6. The methods we use to conduct research, especially the methods used to formalize results in a computational science, affect the research. Logic to describe discrete situations does not lend itself ideally to the description of processes. Differential equations limit descriptions to quantitative problems. Computational human geography needs better tools: higher order formal languages, algebra and category theory.

References

Abadi M, Cardelli L (1996) *A theory of objects.* Springer-Verlag, New York

Asperti A, Longo G (1991) *Categories, types and structures – An introduction to category theory for the working computer scientist.* MIT Press, Cambridge, MA

Backus J (1978) Can programming be liberated from the von Neumann style? A functional style and its algebra of programs. *Communications of the ACM* 21:613–641

Barr M, Wells C (1990) *Category theory for computing science*. Prentice Hall

Benenson I, Portugali J (1995) *Internal vs. external spatial information and cultural emergence in a self-organising city*. In: Frank AU, Kuhn W (eds) Spatial information theory – A theoretical basis for GIS. Int. Conference COSIT'95. LNCS 988:431–441, Springer-Verlag, Berlin, Heidelberg New York

Bird R (1998) *Introduction to functional programming using Haskell*. Prentice Hall Europe, Hemel Hempstead, UK

Birkhoff G (1945) *Universal algebra*. First Canadian Mathematical Congress, Toronto University Press

Dreyfuss (1998) *What computers still cannot do*. MIT Press, Cambridge, MA

Epstein JM, Axtell R (1996) *Growing artificial societies*. Brookings Institution Press, Washington, DC

Frank AU (1999) *One step up the abstraction ladder: Combining algebras – from functional pieces to a whole*. In: Freksa C, Mark DM (eds) Spatial Information Theory – Cognitive and Computational Foundations of Geographic Information Science. Int. Conference COSIT'99. LNCS 1661:95–107, Springer, Berlin Heidelberg New York

Frank AU (to appear) *Communication with maps: A formalized model*. International Workshop on Maps and Diagrammatical Representations of the Environment, Hamburg, Germany

Goodchild MF, Gopal S (1990) *The accuracy of spatial databases*. Taylor & Francis, London

Gopnik A, Meltzoff AN (1997) *Words, thoughts, and theories*. MIT Press, Cambridge, MA

Gould P (1999) *Becoming a geographer*. University of Syracuse Press, Syracuse, NY

Herring JR, Egenhofer MJ et al. (1990) *Using category theory to model GIS applications*. 4th Int. Symposium on Spatial Data Handling, SDH'90. IGU Commission on GIS, Zurich, Switzerland

Herring JR (1989) *The category model of spatial paradigms*. Languages of Spatial Relations: NCGIA Initiative 2 Specialist Meeting Report. In: Mark DM, Frank AU, Egenhofer MJ et al. (eds) National Center for Geographic Information and Analysis, Goleta, CA, pp. 47–51

Kottman C (1998) *Progress toward interoperability and a geospatial infrastructure at the open GIS consortium*. 4th EC-GIS Workshop, Budapest. European Commission, DG III

McCarthy J (1980) Circumscription – A form of non-monotonic reasoning. *Artificial Intelligence* 13:27–39

Morrison J (1988) The proposed standard for digital cartographic data. *American Cartographer* 15(1):9–140

Negroponte N (1995) *Being digital*. Knopf, New York

Newell A (1997) *Unified theories of cognition*. Harvard University Press, Cambridge, MA

O'Hare GMP, Jennings NR (eds.) (1996) *Foundations of distributed artificial intelligence*, John Wiley, New York

Picard R (1997) *Affective computing*. MIT Press, Cambridge, MA

Pinker S (1997) *How the mind works*. W.W. Norton, New York

Reiter R (in preparation) *Knowledge in action: Logical foundations for describing and implementing dynamical systems*

Ruth M, Hannon BM (eds.) (1997) *Modeling dynamic economic systems*. Springer-Verlag, New York

Stoy J (1977). *Denotational semantics*. MIT Press, Cambridge, MA

Stroustrup B (1986) *The C++ programming language*. Addison-Wesley, Reading, MA

Swarm Development Group (1999) *Reference guide for Swarm 2.0.1*. Santa Fé Institute, New Mexico. Available from: http://www.santafe.edu/projects/swarm/swarmdocs/refbook/refbook.html

Takeyama M, Couclelis H (1997) Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science* 11(1):73–91

Timpf S, Raubal M et al. (1996) *Experiences with metadata*. 7th Int. Symposium on Spatial Data Handling, SDH'96. IGU Commission on GIS, Delft, The Netherlands

Tobler WR (1976) Analytical cartography. *The American Cartographer* 3(1):21–31

Wadler P (1989) *Theorems for free*! In: Functional programming languages and computer architecture. 4th Int. Conference FPCA, London. ACM, pp 347–359

Weiss G (1999) *Multi-agent systems: A modern approach to distributed artificial intelligence*. MIT Press, Cambridge, MA

Zeigler BP (1976) *Theory of modelling and simulation*. John Wiley, New York