ciples explaining how
which are likely to
re. But this promise
nding and controlling
ling it with valid and
phenomena. Yet, as
models become ever
and more difficult to
are doing, and harder
*nipulating* both so as
ecisions come out as

gy has an even more
GIS technology and
l tools because they
straction rather than
encourages the cre-
ding of decision sup-
ng of decisions which
be *explicitly* included
our software code.
ymbol for "church"
res Cathedral, and
annot represent the
Memorial. We need
how much this tech-

olution, and master
have much to reflect

---

# Stepping Over The Line: Technological Constraints And the New Cartography

Michael F. Goodchild

ABSTRACT. Traditional cartography is seen as an optimal response to a highly constrained technology based largely on pen and paper. Although many of the conventions of manual cartography appear to be intelligent choices, they have nevertheless been made in an extremely restricted environment which imposes a limited view of reality. Early digital technology did little to broaden the constraints, and led cartography, map analysis and spatial analysis in different directions. More recent hardware and the results of intensive research have produced a digital cartography which can successfully emulate its analogue parent. However, its true potential lies in less conventional methods of analysis and display and in the degree to which it can escape its traditional constraints.

Traditional, analogue cartography is a highly developed technology, providing an efficient and effective means of storing and communicating geographic information. The printed map vastly outperforms the most sophisticated digital storage devices on any index of data per unit weight or per unit volume. Moreover, the human eye and brain can scan and interpret information on a map more rapidly than the most elegant software can process the equivalent digital data.

Like biological species, the techniques of cartography have evolved over the centuries to optimize performance within constraints, in this case those imposed by available technology. Techniques to draw lines of roughly constant width on relatively stable media have been available since antiquity, and despite steady improvements, particularly in methods of reproduction, remain the basis of conventional cartography today. Thus, linear features such as roads can be shown reasonably accurately. However, road width must often be portrayed by line style or color since the range of possible line widths is constrained by the technology.

Point features can be represented by symbols drawn with the pen. Area features can be shown by drawing bounding lines, and perhaps by crosshatching on the assumption that the symbolized attributes are homogeneous. But the limitations of the technology are quickly apparent in the display of transition zones. A technology based on lines of constant width cannot readily portray fuzziness or uncertainty or continuous change. Similarly, the isoline has evolved as the conventional solution to the portrayal of continuous variation as a compromise between the need to find a method which could be readily understood by the map user, and the constraints imposed by pen technology. Other methods might make the surface easier to per-

---

Michael F. Goodchild is affiliated with the Department of Geography, University of Western Ontario, London, Ontario, Canada N6A 5C2

ceive, and avoid the problems many people experience with understanding isopleth maps, but potential improvements such as continuous density variation, simulated illumination or perspective views, lay beyond the constraints of pen and paper.

The proposition that conventional or analog cartography has evolved under the constraints imposed by the pen, variously supplemented by symbols and fill patterns and recently air brush techniques, provides the starting point for this essay. Subsequent sections follow the development of digital methods from the same perspective of adaptation to technological constraints. Although the approach is primarily historic, the inferences and views expressed are personal, based on an intimate involvement with spatial data handling since the 1960s.

The first part of the essay discusses the early impact of digital technology which because of its limitations had relatively little influence on cartography, while having a much greater effect on other, related fields such as spatial analysis. More rapid development of digital cartography was possible when technological developments fell more into line with traditional needs, and digital methods could be used to emulate conventional products. At the same time, developments in spatial analysis, remote sensing and particularly GIS, all based on digital technology, had begun to explore the possibilities of new methods of spatial data handling and analysis. The final section of the essay suggests that the most significant impact of digital methods on cartography will occur when the field evolves to fit the constraints of the new technology and moves beyond its traditional limitations.

## THE LIMITATIONS OF PEN TECHNOLOGY

The features on the earth's surface most closely approximated by lines of constant width are roads, rivers and shorelines. Of the three, the shoreline, which is a well-defined boundary between land and water with no real width, is arguably the most successful approximation at cartographic scales (but see Mandelbrot, 1967; Maling, 1968). It is probably no more than coincidental that these are among the earliest and most common features to be mapped. The reason is likely found in their uniform importance to transportation. Parcel boundaries are also of zero width, but subject to uncertainty of location since they must be based in many cases on measurement rather than direct observation. But while the degree of approximation may vary, it is clear that all cartographic representations of complex reality as primitive point, line and area objects are abstractions. Furthermore the form of abstraction is to a large degree dictated by the ease with which these primitives can be portrayed using a pen.

It is, of course, debatable whether abstractions and generalizations are to be regarded as necessary consequences of technological constraints or as devices used by the cartographer to improve the effectiveness of the map as a communicator of spatial information. For example, shoreline generalization can be seen as a way of removing confusing clutter in order to emphasize important features. Or shoreline generalization can be viewed as imposed by technical limits to line width and the costs of data collection and manual tracing. In this case the former seems more reasonable. In the case of forest inventory maps, the need for accurate inventory clearly overrides any question of cartographic clarity and ease of perception. However forest inventories continue to be mapped using bounded areas to portray homogeneous forest stands, suggesting in this case that technological constraints, specifically the inability to show transition or heterogeneity, have outweighed any more abstract cartographic principles.

Nowhere is this debate between technical constraint and effective communication clearer than in the discussion over continuous shading.

Tobler (1973) pointed out that digital techniques could remove the need to establish a finite number of levels of shading in making a choropleth map, as it is technically possible to crosshatch each area with a density of lines directly proportional to the value of the mapped attribute for that area. Evans (1977) took the opposing position that while the need for a finite number of levels can certainly be regarded as the consequence of technical constraints in manual cartography, it also has a distinct and legitimate function in communication.

One of the more severe constraints of pen and paper technology operates in the time dimension. Substantial time must be spent in collecting data, drawing, reproducing and distributing conventional analog maps. Moreover, continuous update is impossible because after printing and distribution the copies are physically dispersed. Again, convention has evolved to optimize effectiveness under these technical constraints, in this case by a policy of periodic update and associated obsolescence, the update frequency being determined largely by the rate of change of the mapped information.

It is easy to find examples where the constraints of mapping and distribution affect data collection. Because continuous updating of the distributed products is impossible or at least very expensive, there is no incentive for continuous data collection. Thus it is common to resurvey a forest resource and redraft the inventory maps periodically, rather than undertake continuous monitoring, because of the technical difficulties of updating existing cartographic products.

In this sense, then, the technological constraints of data handling affect not only the product but the data collection process, and ultimately the view of the real world. If transition zones cannot be shown on maps, we may tend to overlook them in reality and to grow so accustomed to cartographic portrayal of homogeneity within zones that we expect it in, or impose it on, reality.

## FALSE STARTS: EARLY DIGITAL DEVELOPMENTS

Although the roots of digital computer technology can be found in the commercial calculating machines of the late 19th century, what we now know as mainframe computers became generally available in the early 1960s. Even in those early days, the market for the new technology was clearly segmented between the numerically intensive scientific applications centered in the universities and the transaction and inventory management common in business applications. In neither area was there any general awareness of potential for graphic applications. I recall struggling in 1965 with the problem of placing a single character at a randomly chosen location on a sheet of printer output using Fortran, and realizing that the solution lay in a simple page buffer, initialized with blanks and printed later with a format of (132A1).

**SYMAP** Perhaps the most serious problem was that in 1965 there were no devices analogous to the pen and paper, and therefore there was no obvious way in which the new technology could be put to cartographic use. Numerical applications were found in geodesy and in calculating projections. SYMAP, arguably the first package for automated cartography, was constrained to the limitations of the line printer as an output device. Nevertheless, it seems in retrospect to have been successful for several reasons. First, it undoubtedly drew attention to the possibility of digital cartography and paved the way for the more useful graphics technology which emerged in the late 1960s. Second, it was, and perhaps continues to be, effective in one particular form of mapping: the rapid production of crude but informative choropleth maps based on constant boundaries. Much of the cost of digital choropleth mapping lies in input and editing of digital boundaries and other map features. Costs are much lower if the same set of boundaries and features can be used for many maps. This situation is common in the preparation of statistical maps based on predefined reporting zones such as census tracts or counties, and SYMAP found its strongest applications in census-based urban atlases.

A third significant aspect of SYMAP serves to introduce a more general point about early digital developments. The nature of manual cartography requires that its procedures be largely subjective. Again this can be regarded either as a consequence of technological limitations or as a desirable aspect of effective communication. SYMAP was perhaps the first software to tackle a process which has become one of the richest areas of research in spatial data handling: the operationalization of subjective cartographic procedures. Its algorithm for spatial interpolation (Shepard, 1984) operationalized a standard but subjective cartographic process and remains one of the most useful of the large number of methods which have been devised for interpolating continuous surfaces from randomly distributed points.

**Spatial Analysis** The widespread availability of computers adapted to numerical and scientific applications was one of the driving forces behind the rapid expansion of interest in statistical methods in the social and earth sciences in the 1960s. From the start, it was clear that the spatial content of geographical data and the nature of spatial processes and forms required the development of specialized techniques of analysis. The term spatial analysis was in widespread use by 1968 (Berry and Marble, 1968). However, three problems seem to account for a lack of interaction between cartography and spatial analysis, and a lack of mutual interest, in this period.

First, the motivation to pursue quantitative, and by implication, scientific methodologies, did not have the same appeal in cartography. On the one hand, cartography has a longstanding tradition of seeing itself as part art, part science; on the other, the more scientific aspects of cartography, including photogrammetry, geodesy and surveying engineering, were already undoubtedly scientific in their methods.

Second, the body of statistical method on which geography attempted to build was primarily aspatial. Although extensive methods of spatial pattern analysis had been developed by the 1960s (Ripley, 1981), the assumption of independence of observations which underlies much of statistics continues to conflict with Tobler's first law of geography, that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970; and see Gould, 1970, pp. 443–4). Inevitably the most immediate returns lay in overlooking the independence assumption, and it is only comparatively recently that spatial analysts have begun to address the problem of adapting standard statistical techniques to spatial situations (Cliff and Ord, 1981). Moreover, it was much easier to apply the available tools, in the form of BMD and later SPSS and SAS with their essentially aspatial views of data, than to develop special codes for more complex spatial data structures.

Finally, while in the late 1960s it was abundantly clear that computers were ideal tools for statistical analysis, their applications to processing graphic images and text were still to come. As a result, spatial analysis has developed almost independently of cartography over the past three decades. Statistical analysis rose rapidly to occupy a central place in the teaching and research programs of geography departments, while cartography remained of merely technical importance. Quantitative analysis was, and continues to be conducted largely aspatially. While digital technology has been regarded as essential for statistical processing, it is likely that any cartographic display of results would have been generated manually. The addition of mapping capabilities to SAS and SPSS and the marketing of

er, the more scientific
including photogram-
ying engineering, were
ientific in their meth-

statistical method on
pted to build was pri-
gh extensive methods
s had been developed
981), the assumption
vations which under-
tinues to conflict with
graphy, that "every-
thing else, but near
than distant things"
Gould, 1970, pp.
ost immediate returns
dependence assump-
ratively recently that
egun to address the
dard statistical tech-
ons (Cliff and Ord,
much easier to apply
form of BMD and
vith their essentially
an to develop special
spatial data struc-

1960s it was abun-
ers were ideal tools
heir applications to
and text were still
patial analysis has
dently of cartogra-
decades. Statistical
cupy a central place
programs of geog-
rtography remained
ance. Quantitative
es to be conducted
digital technology
ntial for statistical
any cartographic
ve been generated
mapping capabili-
the marketing of

mapping software for the PC have done much to change this situation in the past few years, but it has done little to close the gap between the perceptions of the two fields.

**Early GIS** Despite the lack of graphics peripherals in the 1960s, it was nevertheless possible to conceive of applying computer technology to the numerical analysis of map data. While the map is an efficient store of information, it is notoriously ineffective as a basis for numerical analysis, particularly in the measurement of area and the comparison of coverages of the same area. In the first case, conventional analogue methods, specifically planimetry and dot counting (Frolov and Maling, 1969; Goodchild, 1980), are both inaccurate and tedious. In the second case, the conventional response has been to include as many themes as possible on a single sheet. A revival of interest in map analysis in the 1960s (McHarg, 1968) led to the need for more efficient methods. It is significant that while measurement of area and map overlay were the key motivations which led the Canadian government to develop the Canada Geographic Information System (Tomlinson, Calkins and Marble, 1976) in the mid 1960s, the original intention was that the associated cartographic products of the Canada Land Inventory would be generated manually.

Other early developments in the GIS field were similarly aimed at responding to needs for numerical map analysis. Systems such as New York's LUNR (Tomlinson, Calkins and Marble, 1976) were based on coarse rasters and clearly inappropriate for cartographic purposes, although crude line printer graphics could be used for data checking and simple analysis. So while map analysis found ample applications in planning and landscape architecture, it had little influence on the development of spatial analysis in the 1960s and 1970s, and little directly to do with cartography. No significant convergence of interests occurred until the late 1970s, when it finally became clear that all three fields were very strongly related and mutually dependent.

## ENABLING TECHNOLOGIES

**Emulation** The constraints which had limited the development of digital technology in cartography were loosened in stages during the 1970s as a range of new products became available. The pen plotter was perhaps the most important of the early arrivals, since it allowed computer systems to generate something approaching a conventional cartographic product for the first time. With it, software designers were able to concentrate on better and better emulation of the traditional analog document. Early efforts were crude, since the relatively large increments used in early plotters gave lines a jagged appearance, and because such cosmetic features as fonts, neatlines, scales and legends were relatively simple.

The graphics terminals based on storage tube technology which reached the market in the late 1960s and early 1970s allowed interactive editing of graphic information for the first time and led to rapid development of associated software for input of large volumes of spatial data. But cartography has rarely been seen as the primary market for hardware, and cartographic applications have generally lagged behind larger market segments such as CAD (Computer Assisted Design). Cartographic applications also tend to require sophisticated software. The result has been that the timetable of developments in cartography has been dictated more by software than by hardware in the past two decades. The data input and edit workstations which reached the market in the late 1970s and early 1980s were based on technology which had already been available for several years.

One of the issues retarding software development in automated cartography has been the relatively small size of the research community. Successful applications of spatial data handling techniques have required major advances in data structures and algorithms, and these have tended to occur at a rate which has little to do with hardware development. In fact, much of the necessary research agenda of automated cartography remains untackled. We still lack completely successful operationalizations of such basic cartographic processes as label placement and feature generalization.

Nevertheless, by the early 1980s it had become possible, given adequate investment in software and hardware, to replace traditional mapmaking with digital technology. Although the proportion of maps created using digital methods is still relatively small, the technology has advanced to the point where it is impossible to determine by cursory examination which approach has been used to generate the final product. This critical point was probably passed in about 1980. The hardware developments and multi-year investments in software have advanced digital technology to the point where its constraints are no more restrictive than those of pen and paper.

One of the surprising aspects of these developments, and of applications of digital technology in many other areas as well, is the degree to which perfect emulation of the conventional product appears to be regarded as a legitimate objective. Word processing technology has been developed to the point where it is a more cost-effective way to generate the conventional typed text. Surprisingly, it has not been used to explore alternative methods of presentation to the same extent. Yet neither word processing nor digital cartography are constrained by the same limitations that produced the printed page or analogue map.

**New Departures** The key to successful automation of mapmaking was the development of devices which allowed the digital system designer to emulate conventional technology. But among the more significant advances in spatial data handling research in the past 30 years are several which represent clear departures from convention. Perhaps the most significant of these is the raster representation. Although rasters are familiar in the regular scans of written text, traditional cartography is exclusively vector in its approach, using the three vector primitives, point, line and area, rather than pixels or cells. Yet SYMAP was based around raster display, and recently both graphics terminals and hard copy devices have moved to raster technology as memory has become steadily cheaper. Moreover raster has become the increasingly dominant mode of data input since the deployment of remote sensing systems and the development of scanners.

Raster display is radically different from pen and paper. We might debate the likely nature of cartography had the traditional discipline developed as a raster rather than a vector technology. It might not have chosen to represent reality through point, line and area primitives; would likely not have chosen contours as its method of portraying continuous surfaces; and might well have given less importance to assumptions of homogeneity within zones. Unfortunately these concepts are so familiar and they structure our view of spatial reality so completely that such a debate would be very difficult.

A more radical departure is the development in the past 15 years of data structures based on hierarchical subdivision, often loosely referred to as quadtrees (Burrough, 1986). In one sense, these represent a departure from fixed scale in the form of fixed pixel size in rasters or fixed levels of spatial generalization in vectors. They are non-intuitive in that they correspond to no conventional, pictorial view, but have meaning only as digital representations. In quadtree data structures we are beginning to see the emergence of a genuinely new technology in which methods have no obvious conventional analogues. At the same time the constraints imposed by the new technology are radically different.

However, it appears that elements of quadtree structures can be found in conventional methods, and that we have here an example of how a new technology can serve to restructure and refine an old. In both digital and conventional mapmaking it is common to divide a large mapped area, such as an entire nation, into map sheets. Tiles is the corresponding term for this process in digital systems. In conventional mapping this is done because the technology imposes limits on the size of sheet which can be conveniently printed and handled. In digital mapping the corresponding constraints lie in the efficiency of access algorithms and the capacity of storage devices.

Elements of the quadtree structure are present in both tiles and mapsheets, as it is conventional in both to nest smaller tiles within larger ones. For example, the organization of topographic maps typically nests 1:50,000 sheets within 1:250,000 sheets. Moreover, the ordering of sheets and tiles is often designed to assign similar positions in the sequence to sheets or tiles which are close together in space, thus minimizing the time taken to access neighboring information. One of the earliest attempts to place this principle on an objective basis was the tile sequence developed by Morton (1966) for the Canada Geographic Information System.

**Digital Databases** Digital data can be duplicated and distributed comparatively cheaply. Moreover, much spatial data is relatively stable through time. Despite such events as the breaking off of a large part of the Ross Ice Shelf in 1987, spatial data is relatively amenable to archiving. So one of the more pronounced effects of the development of digital spatial data handling has been the impetus provided by the widespread availability of large, relatively cheap data sets, ranging from remote sensing to topography and statistical boundaries.

In an environment of cheap digital data and handling systems, the map begins to lose importance as the final product, and along with it the limitations of a technology geared to the paper medium. Instead, discussion turns to electronic maps, simulated verbal instructions, and responses to plain language queries—elements, in principle, of the fully developed geographic information system.

## THE NEW CARTOGRAPHY

The primary motivation in a new technology is often novelty. It seemed interesting to me that the computers of the 1960s could be used to make maps. The secondary motivation is often efficiency or cost-effectiveness. CGIS provides a clear example of a need which could be satisfied only by the digital option. The edit stations which were installed in the Surveys and Mapping Branch of Energy, Mines and Resources Canada in the early 1970s based on Tektronix storage tubes were clearly an attempt to reduce labour costs.

The first section of this essay argued that while manual cartography is a highly sophisticated discipline, it is, nevertheless, tightly constrained by its dependence on a technology of pen and paper. The consequences of those constraints can be rationalized in many cases as intelligent choices, and are so fundamental that it is difficult to consider alternatives, but they are in actuality severely restricting, and influence not only the way we portray the world but also the way we observe it.

The most straightforward objective of digital technology is the emulation of manual methods, to the point where the two products are indistinguishable. There are many interesting and challenging problems involved in reaching this objective, including the operationalization of many standard but subjective cartographic processes and the development of efficient algorithms and data structures. Indeed the nature of much effort in digital cartography, and the products offered by many vendors, suggest that this objective is the dominant one in much of the profession. But it embodies the view that conventional cartographic methods are somehow optimal irrespective of technological constraints. It is intuitively unreasonable that a technology optimized under the narrow constraints of pen and paper would turn out to be indistinguishable from one optimized under the much broader constraints of digital technology.

We have argued that the divergence which occurred in the 1960s among cartography, map analysis and spatial analysis was due at least in part to the different response of each field to the limited capabilities of the digital technology available at the time. Recent interest in GIS, particularly among cartographers and spatial analysts, indicates that we are now in a period of convergence. The spatial analysis literature is paying increasing attention to the importance of data structures; to the development of effective methods for dealing with spatial dependence; and to the statistical models needed to describe uncertainty in spatial data. At the same time, the cartographic literature appears increasingly concerned with analytical and numerical cartography; with issues of accuracy which arise much more significantly in digital than in manual data handling; and with objective versions of its conventional processes.

Digital technology offers enormous potential to cartography. Raster display can be used to depict continuous surfaces in numerous ways, ranging from contours through simulated illumination to perspective views. The latter two can be combined in the technology of scene generation, which attempts to simulate the actual visual appearance of a landscape using stochastic process models to produce various types of texture, trees, clouds and other finite objects. Here again, we encounter the two conflicting visions of cartography. Would scene generation be an escape from the limits of technology, or an abandonment of the cartographer's art in selecting and abstracting significant elements of the landscape?

Perhaps the greatest potential is in the time dimension. Maps are static views, and require corresponding approaches to data collection and map production. Electronic communication allows instantaneous distribution from one, continuously updated authority. The implications of this technical change extend not only to the way the data are viewed, but also to the way they are collected, and the infrastructure which has grown up around data collection. The decennial census, for example, was a solution to the problems of data collection and dissemination of the mid-19th century. Yet we have only the vaguest notions of how to exploit the new technology in this area; our perceptual systems are so geared to conventional display that we find it difficult, for example, to conceive of color being used to show the time dimension, or of how to structure and store time-dependent data in an efficient manner.

Digital technology does indeed have the potential to stimulate a revolution in cartography, but the revolution has hardly begun. Production of conventional maps by digital means may have been the vision which motivated the development of many of today's automated cartography systems, but it is not revolutionary. Monmonier (1985) has termed the move to digital technology the Electronic Transition. Digital production is not always cost-effective, and it is even difficult to use the familiar avoidance-of-drudgery argument so common in other automation debates. The real revolution will be in a redefinition of cartography's role, away from the limitations of pen and paper technology, based on a vision which has emerged much more recently and which sees the central role played by spatial data handling technology in fields as diverse as automated mapping, remote sensing, AM/FM, urban planning, forestry and natural resource management. The challenge to cartography will be to maintain its traditional balance between art and science and its emphasis on the communication of spatial information, in a period in which the science becomes increasingly complex and in which the art must be continually adapted to new opportunities.

## REFE

Berry, E
*ysis*
Ha.
Burroug
*mat*
Mo
12.
Cliff, A.
*Mo*
Evans, I
*Insti*
98–
Frolov,
of ar
*Carr*
Goodchil
accu
*Geo*
Gould, P
nam
439
Maling, I
*Cart*

## REFERENCES

Berry, B.J.L. and D.F. Marble, 1968. *Spatial Analysis: A Reader in Statistical Geography*. Prentice-Hall, Englewood Cliffs, N.J.

Burrough, P.A., 1986. *Principles of Geographic Information Systems for Land Resources Assessment.* Monographs on Soil and Resources Survey No. 12. Oxford University Press, London.

Cliff, A.D. and J.K. Ord, 1981. *Spatial Processes: Models and Applications*. Pion, London.

Evans, I.S., 1977. The selection of class intervals. *Institute of British Geographers, Transactions 2(NS)* 98–114.

Frolov, Y.S. and D.H. Maling, 1969. The accuracy of area measurements by point counting techniques. *Cartographic Journal 6* 21–35.

Goodchild, M.F., 1980. A fractal approach to the accuracy of geographical measures. *Mathematical Geology 12* 85–98.

Gould, P.R., 1970. Is *statistix inferens* the geographical name for a wild goose? *Economic Geography 46* 439–48.

Maling, D.H., 1968. How long is a piece of string? *Cartographic Journal 5* 147–56.

Mandelbrot, B.B., 1967. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science 156* 636–8.

McHarg, I.L., 1969. *Design with Nature*. Doubleday-Natural History Press, New York.

Monmonier, M.S., 1985. *Technological Transition in Cartography*. University of Wisconsin Press, Madison.

Morton, G.M., 1966. A computer oriented geodetic data base, and a new technique in file sequencing. IBM Canada Ltd., mimeo.

Ripley, B.D., 1981. *Spatial Statistics*. Wiley, New York.

Shepard, D.S., 1984. Computer mapping: the SYMAP interpolation algorithm. In G.L. Gaile and C.J. Wilmott, editors, *Spatial Statistics and Models*. Reidel, Dordrecht, 133–45.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography 46* 234–40.

Tobler, W.R., 1973. Choropleth maps without class intervals? *Geographical Analysis 5* 262–5.

Tomlinson, R.F., H.W. Calkins and D.F. Marble, 1976. *Computer Handling of Geographical Data.* UNESCO Press, Paris.